Paper 3 - Causal Roadmap

Lorenzo Fabbri

Table of contents

List Of Acronyms

ANT Attention Network Test
CBCL Child Behavior Checklist
CPM Raven's Coloured Progressive Matrices
EDC endocrine disruptor
EDCs endocrine disruptors
MR Mendelian Randomization
MTP modified treatment policy
MTPs modified treatment policies
NPSEM non-parametric Structural Equation Model
PRS polygenic risk score
SL Super Learner
TMLE Targeted Minimum Loss-Based Estimator

Most of the following was **copy-pasted** from published papers.

Formulate the research question(s)

The **aim** of the present study is to research the short-term effects of postnatal exposure to non-persistent endocrine disruptors (EDCs) on neurodevelopment and neurobehavior in childhood, and how the metabolome and the proteome might mediate these effects. We will strengthen the estimated effects by making use of the principles and criteria of triangulation (Lawlor, Tilling, and Davey Smith 2016). To estimate these effects, we will primarily rely on the use of modified treatment policies (MTPs) (Muñoz and van der Laan 2012; Haneuse and Rotnitzky 2013; Díaz et al. 2021) in combination with Targeted Minimum Loss-Based Estimator (TMLE) (van der Laan, Benkeser, and Sofrygin 2018), to avoid relying on arbitrary parametric assumptions. We will consider both additive and multiplicative shifts of the exposures' distribution.

The primary study **population** is based on the HELIX sub-cohort, consisting of N = 1200 mother-child pairs from six existing European birth cohorts (Vrijheid et al. 2014). We will replicate the obtained results in the HELIX Child Panel, consisting of N = 150 children followed twice for one week. **Exposures** consisted of non-persistent EDCs (phenols, phthalates, and organophosphate compounds), measured in childhood in a pool of two urine samples. A single-spot blood sample, collected during the visit, was used for serum metabolomics and plasma proteomics. **Outcomes** related to childhood neurodevelopment and neurobehavior, included:

- Raven's Coloured Progressive Matrices (CPM), for assessing non-verbal intelligence.
- Computerised n-back test, for assessing working memory.
- Attention Network Test (ANT).
- Child Behavior Checklist (CBCL), for assessing behavioural and emotional problems.

Descriptive analyses

TODO

Define a realistic statistical model

For time index t, let W_t denote the set of potential confounders, A_t the observed exposures, and Y_t the clinical outcome of interest. For each subject i and each time index t, we assume its observed data $O_{ti} = (W_{ti}, A_{ti}, Y_{ti})$ were generated by sampling from a distribution $\mathbb{P}_{0,t}$ compatible with the causal model specified in Equation 1.

Specify a causal model and causal quantity of interest

We specify the following non-parametric Structural Equation Model (NPSEM) to present the data generating process, including its measured confounders W_t, exposures A_t, and outcome Y_t, for each time index t:

$$\begin{aligned} W_t &= f_{W_t}(U_{W_t}) \\ A_t &= f_{A_t}(W_t, U_{A_t}) \\ Y_t &= f_{Y_t}(W_t, A_t, U_{Y_t}), \end{aligned} \tag{1}$$

where $(f_{W_t}, f_{A_t}, f_{Y_t})$ are the non-parametric structural equations, and $(U_{W_t}, U_{A_t}, U_{Y_t})$ are the unmeasured factors contributing the confounders, exposures, and outcome, respectively. The corresponding causal graph (without indexes for ease of notation) is given in Figure 1.

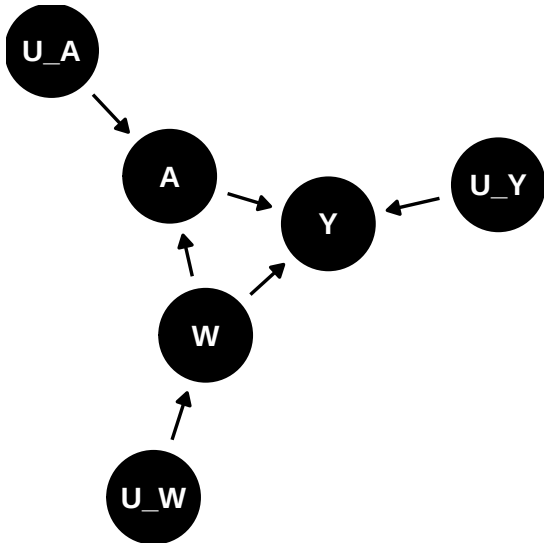

Figure 1: Causal graph corresponding to structural equations.

In order to test the NPSEM's implications (conditional independencies), we will perform simulations from the DAGs, checking assumptions and data consistency, using the `dagitty` R package.

Following the modified treatment policy (MTP) framework, we will generate counterfactual outcomes $Y^d = f_Y(W, A^d, U_Y)$ by intervening on the causal model to shift the observed exposures by some user-specified function $A^d = d(\cdot)$. We will focus on simple interventions to shift the observed exposure A by an additive constant c or multiplicative constant k: $d(A, c) = c + A$ and $d(A, k) = k \times A$, respectively.

We specify our causal parameter ϕ as the difference in the expected counterfactual outcome under the shifted exposures and the expected outcome under the observed exposures:

$$\phi^{\Delta} = \mathbb{E}[Y^d] - \mathbb{E}[Y],$$

where the expectation is over the individuals of our target population.

As **primary analyses**, we will estimate the short-term effects of childhood exposure to individual non-persistent EDCs in relation to one or more neurodevelopmental outcomes of interest. We will further explore the potential mediating role of the serum metabolome and the plasma proteome.

In the following sections, we present a brief summary of each research question.

Research question 1: What are the short-term effects of childhood exposure to EDCs on the outcomes?

We will estimate the short-term effects of childhood exposure to non-persistent EDCs on the outcomes of interest measured in childhood. We will perform an ExWAS kind of analysis with the dependent variable y being the outcome, and the independent variable x_i being the levels of endocrine disruptor (EDC) i. We will assume that all associations are confounded by the same set of confounders, independently of the EDC considered.

Research question 2: What are the short-term effects of childhood exposure to EDCs on metabolites and proteins?

We will estimate the short-term effects of childhood exposure to non-persistent EDCs on metabolites and proteins measured in childhood. We will perform an MWAS analysis with the dependent variable m_i being the metabolite or protein i, and the independent variable x_j being the levels of EDC j. We will assume that all associations are confounded by the same set of confounders, independently of the EDC considered.

Research question 3: What are the short-term effects of childhood *exposure* to metabolites and proteins on the outcomes?

We will estimate the short-term effects of childhood *exposure* to metabolites and proteins on the outcomes of interest measured in childhood. We will perform an MWAS analysis with the dependent variable y being the outcome, and the independent variable m_i being the metabolite or protein i. We will assume that all associations are confounded by the same set of confounders, independently of the EDC considered.

Identification and the statistical estimand

For the causal parameter ϕ^{Δ}, which involves a summary measure of the distribution of counterfactuals, to be identified in terms of the observed data distribution, several assumptions would be required:

- **No unmeasured confounding**: There are no unmeasured common causes of the exposure and the subsequent outcome, which we can formalize as $Y^d \perp\!\!\!\perp A|W$. In our context, this assumption would be violated if, for example, an unmeasured variable influences both the exposure levels and the clinical outcome. We cannot guarantee that this assumption holds, and therefore limit our interpretations to statistical associations rather than causal effects.
- **Positivity**: If (a, w) is within the support of A, W, then $(d(a, c), w)$ for additive shifts and $(d(a, k), w)$ for multiplicative shifts must also be within the support of A, W. In practice, this means that for any given time index and set of adjustment covariates, there is a positive probability of finding a subject with the same covariate values and a exposure level matching the shifted value. We will attempt to improve plausibility of this assumption by considering small shifts, while recognizing that this approach makes our causal effect data-adaptive.
- **Independence** of subjects. This assumption also implies **no interference**: the exposure level a for a given subject does not affect the outcomes of the other subjects.
- **Consistency**: If $A = a$ for any subject, then $Y(a) = Y$, and hence the full observed set of outcomes when $A^d = A$ is simply $Y(A^d) = Y$. This means that the counterfactual outcome for a subject with its observed exposure level is the observed outcome.
- **Time-ordering**: The confounders W precede the exposure A, which also precedes the outcome Y.

If these identifiability assumptions held, we could specify and focus our estimation efforts on a statistical estimand that equals the wished-for causal effect. In the likely case that they are not satisfied, we could still specify and focus our estimation efforts on a statistical estimand that is as close as possible to the causal parameter. Factoring the joint distribution of the observed data $\mathbb{P}_0$ into $\mathbb{P}_0(O) = \mathbb{P}_0(Y|A, W)\mathbb{P}_0(A|W)\mathbb{P}_0(W)$, it can be shown that the statistical estimand corresponding to expected counterfactual outcome under shift d, $\mathbb{E}[Y^d]$, is given by

$$\psi_0(A^d) = \int \mathbb{E}(Y|A = a^d, W = w) dF_{A,W}(a, w),$$

with $dF_{A,W}(a, w)$ as the joint density of received exposures A and covariate levels W being integrated over. We refer to $\psi_0(A^d)$ as the *shift parameter*. Under no shift, the expected outcome was identified as $\psi_0(A^d) = \mathbb{E}(Y)$. Therefore, our statistical estimand of interest, corresponding to the expected difference in the outcome under shifted and observed exposure levels, is

$$\psi_0^{\Delta} = \psi_0(A^d) - \psi_0(A) = \int \mathbb{E}(Y|A = a^d, W = w) dF_{A,W}(a, w) - \mathbb{E}(Y).$$

Estimation from data and statistical inference

Triangulation: Improving causal inference in aetiological epidemiology

From (Lawlor, Tilling, and Davey Smith 2016):

> *Triangulation is the practice of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other. With respect to causal questions in aetiological epidemiology, if the results of different approaches all point to the same conclusion, this strengthens confidence in the finding. This is particularly the case when the key sources of bias of some of the approaches would predict that findings would point in opposite directions if they were due to such biases.*

Table 1 summarizes the approaches that we will use to perform triangulation, based on (Lawlor, Tilling, and Davey Smith 2016).

Table 1: Triangulation for Paper 3

Effect	Approach	Description	Key sources of bias
Chemical → outcome	Multivariate analysis in observational data	Application of multivariable regression to observational data.	Residual confounding. Reverse causality. Misclassification of exposure is related to the outcome, or vice versa.
	Outcome negative control study	Aims to reproduce the same conditions as the real study, but using a different (negative control) outcome that is not plausibly causally related to the exposure.	There are differences in the sources of bias between the real and negative control outcome: attempts to explore this (e.g., exploring the association of observed confounders with the negative control outcome) should be made. There is a real (but unknown) causal effect of the negative control outcome on the exposure.
	Literature search	Aims to validate the obtained results with data from the published scientific literature.	Positive-results bias, a type of publication bias. Use of different nomenclatures for, e.g., metabolites.
	Qualitative polygenic risk score	TODO	TODO
Chemical → omic	Multivariate analysis in observational data	Application of multivariable regression to observational data.	Residual confounding. Reverse causality. Misclassification of exposure is related to the outcome, or vice versa.
	Cross-cohort comparison	Compares results between two or more populations in different contexts that result in different confounding structures.	Confounders are the same in the populations being compared: for observed confounders, differences between the two populations should be established. There are different sources of bias (over and above different confounding structures), for example differential misclassification of exposure or outcome. Measurement of the exposure and outcome, and the quality of these measurements, should be the same or very similar in the populations being compared.

Effect	Approach	Description	Key sources of bias
	Literature search	Aims to validate the obtained results with data from the published scientific literature.	Positive-results bias, a type of publication bias. Use of different nomenclatures for, e.g., metabolites.
	Qualitative polygenic risk score	TODO	TODO
Omic → outcome	Multivariate analysis in observational data	Application of multivariable regression to observational data.	Residual confounding. Reverse causality. Misclassification of exposure is related to the outcome, or vice versa.
	Cross-cohort comparison	Compares results between two or more populations in different contexts that result in different confounding structures.	Confounders are the same in the populations being compared: for observed confounders, differences between the two populations should be established. There are different sources of bias (over and above different confounding structures), for example differential misclassification of exposure or outcome. Measurement of the exposure and outcome, and the quality of these measurements, should be the same or very similar in the populations being compared.
	Mendelian Randomization (MR)	Instrumental variable (IV) is one or more genetic variant(s) that have been shown to robustly relate to the exposure.	Violation of the exclusion restriction criteria, as a result of horizontal pleiotropy, is likely to be the main source of bias: using multiple genetic IVs that likely have different (unrelated) paths to the exposure, and employing recently developed sensitivity analyses to these, can test and control (to some extent) for this violation. Population stratification produces confounding: this may be avoided by using ethnically homogeneous populations, and/or controlling for principal components that reflect different population subgroups. With increasing availability of results from large-scale genome-wide association studies and application of two-sample MR to these, weak instrument bias is less likely, and when it occurs would bias towards the null.

For each research question, and each approach, we will perform the following analyses:

- `outcome ~ chemical`
  - Regression analysis.
  - Outcome negative control study.
  - Literature search.
  - Qualitative polygenic risk score (PRS).
- `omic ~ chemical`
  - Regression analysis.
  - Cross-cohort comparison.
  - Literature search.
  - Qualitative PRS.
- `outcome ~ omic`
  - Regression analysis.
  - Cross-cohort comparison.
  - Mendelian Randomization (MR).

MR: Assumptions

Interpretation and sensitivity analyses to inform a substantive conclusion

Highlight possible **limitations**:

- When considering only one exposure, we do not consider the effect of the others on that exposure and the outcome;
- When we consider only one outcome, we do not consider the effect of the other outcomes on that one;
- Although the interpretation of ψ^{causal} is clear, how closely ψ^{stat} matches it merits discussion. One considers the plausibility of each of the identifying assumptions in turn.

We performed **sensitivity analysis** to measure robustness to unmeasured confounders, based on the following considerations:

- Broadly, an open question is whether the estimated treatment effect is biased due to confounding by unmeasured covariates. We can examine how the substantive conclusion would be impacted under a range of presumed causal bias, $\delta = \psi^{\text{causal}} - \psi^{\text{stat}}$. That is, how *strong* would a particular confounder (or group of confounders) have to be in order to change the conclusions of this study? In a worst case scenario, how vulnerable are the study's results to many or all unobserved confounders acting **together**, possibly non-linearly? The exercise illustrates how the effect estimates, and confidence interval bounds change, depending on the magnitude and direction of the hypothesized gap;
- We implemented the following methods to conduct sensitivity analysis:
  - Selection bias and stratification...
  - Placebo treatment (replacement of the actual treatment with a random variable);
  - We performed leave-1-out analyses to assess the influence of single subjects on the effect of interest.

We took into consideration the following points to *validate* our results:

- We evaluated whether the obtained results are of any relevance from a public health point of view.

Appendix

Checklist A: Reproducibility

Table with random seed, names, description, and version numbers of all software packages.

Name	**Version**	**Description**
Random seed	NA	Will be set to X
`R`	4.1.2	Statistical programming environment
...	...	...

Checklist B1: `xxx` package specifications

Table providing values for all non-data arguments, and brief rationale when departing from the default specification.

Argument	**Setting**	**Default** (Y/N)	**Comment**
...	...	...	...

DAGs

Research question 1:...

The minimal sufficient adjustment sets for estimating the direct effect of X on Y are:

```
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_diet, child_smoking, edu_child, familySEP,
  gestational_age, intelligence_SNPs, maternalAlcohol_preg,
  maternalDiet_preg, maternalIodine_preg, maternalIron_preg,
  maternalSEP_preg, maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  qualityTesting_child, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_diet, child_smoking, edu_child, ethnicity_child,
  ethnicity_mother, familySEP, gestational_age, maternalAlcohol_preg,
  maternalDiet_preg, maternalIodine_preg, maternalIron_preg,
  maternalSEP_preg, maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  qualityTesting_child, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, envFactors_visit, familySEP, gestational_age,
  intelligence_SNPs, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg,
```

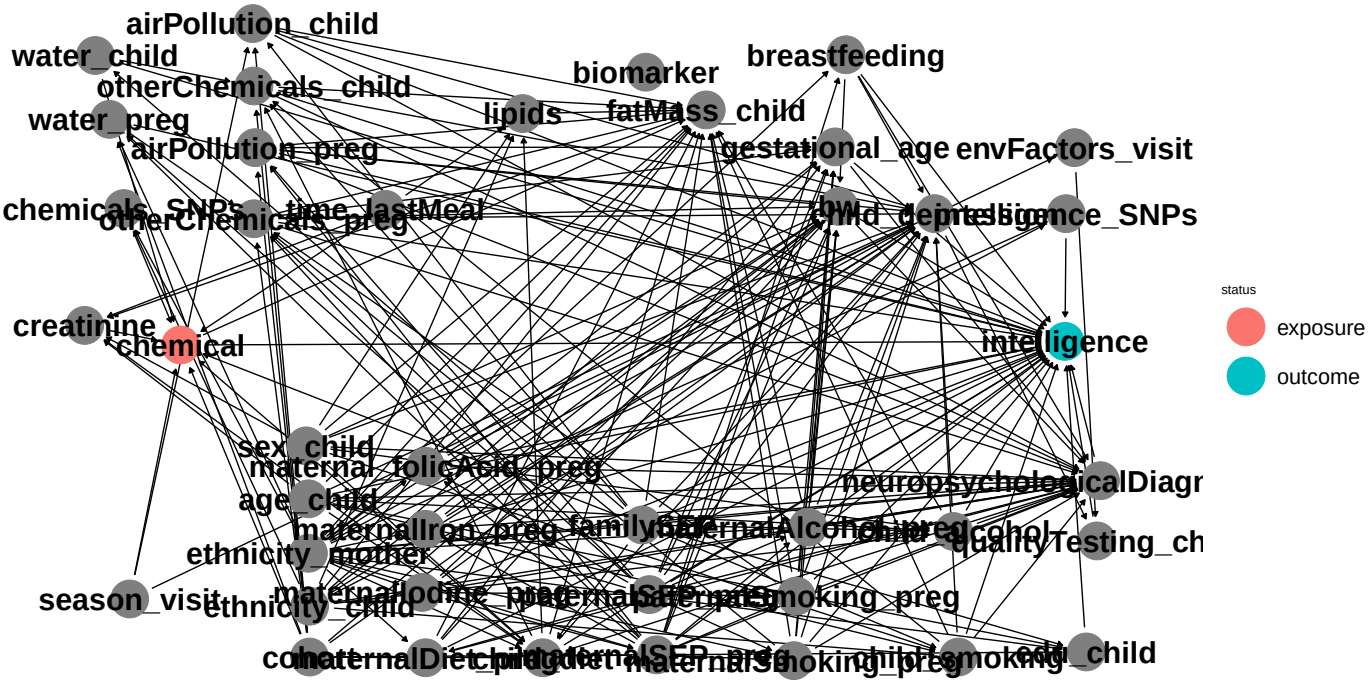

Figure 2: DAG for Research question 1.

```
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, envFactors_visit, ethnicity_child, ethnicity_mother,
  familySEP, gestational_age, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_diet, child_smoking, edu_child, ethnicity_child,
  ethnicity_mother, familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  qualityTesting_child, sex_child, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, envFactors_visit, ethnicity_child, ethnicity_mother,
  familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  sex_child, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, familySEP, gestational_age, intelligence_SNPs,
  maternalAlcohol_preg, maternalDiet_preg, maternalIodine_preg,
```

```
  maternalIron_preg, maternalSEP_preg, maternalSmoking_preg,
  maternal_folicAcid_preg, neuropsychologicalDiagnosis_child,
  otherChemicals_child, otherChemicals_preg, paternalSEP_preg,
  paternalSmoking_preg, season_visit, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, ethnicity_child, ethnicity_mother, familySEP,
  gestational_age, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  season_visit, water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking,
  edu_child, ethnicity_child, ethnicity_mother, familySEP,
  maternalAlcohol_preg, maternalDiet_preg, maternalIodine_preg,
  maternalIron_preg, maternalSEP_preg, maternalSmoking_preg,
  maternal_folicAcid_preg, otherChemicals_child, otherChemicals_preg,
  paternalSEP_preg, paternalSmoking_preg, season_visit, sex_child,
  water_child, water_preg }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_diet, child_smoking, cohort, ethnicity_mother,
  familySEP, gestational_age, intelligence_SNPs, maternalAlcohol_preg,
  maternalDiet_preg, maternalIodine_preg, maternalIron_preg,
  maternalSEP_preg, maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  qualityTesting_child, water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_diet, child_smoking, cohort, ethnicity_child,
  ethnicity_mother, familySEP, gestational_age, maternalAlcohol_preg,
  maternalDiet_preg, maternalIodine_preg, maternalIron_preg,
  maternalSEP_preg, maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  qualityTesting_child, water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking, cohort,
  envFactors_visit, ethnicity_mother, familySEP, gestational_age,
  intelligence_SNPs, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking, cohort,
  envFactors_visit, ethnicity_child, ethnicity_mother, familySEP,
  gestational_age, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSEP_preg,
  maternalSmoking_preg, maternal_folicAcid_preg,
  neuropsychologicalDiagnosis_child, otherChemicals_child,
  otherChemicals_preg, paternalSEP_preg, paternalSmoking_preg,
  water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding,
  child_alcohol, child_diet, child_smoking, cohort, ethnicity_child,
  ethnicity_mother, familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalIodine_preg, maternalIron_preg, maternalSmoking_preg,
```

```
  maternal_folicAcid_preg, otherChemicals_child, otherChemicals_preg,
  paternalSEP_preg, paternalSmoking_preg, qualityTesting_child,
  sex_child, water_child }
{ age_child, airPollution_child, airPollution_preg, child_alcohol,
  child_depression, child_diet, child_smoking, cohort, ethnicity_child,
  ethnicity_mother, familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalSmoking_preg, otherChemicals_child, otherChemicals_preg,
  paternalSmoking_preg, qualityTesting_child, sex_child, water_child }
{ age_child, airPollution_child, airPollution_preg, child_alcohol,
  child_diet, child_smoking, cohort, envFactors_visit, ethnicity_child,
  ethnicity_mother, familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalSmoking_preg, otherChemicals_child, otherChemicals_preg,
  paternalSmoking_preg, sex_child, water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking, cohort,
  ethnicity_mother, familySEP, gestational_age, intelligence_SNPs,
  maternalAlcohol_preg, maternalDiet_preg, maternalIodine_preg,
  maternalIron_preg, maternalSEP_preg, maternalSmoking_preg,
  maternal_folicAcid_preg, neuropsychologicalDiagnosis_child,
  otherChemicals_child, otherChemicals_preg, paternalSEP_preg,
  paternalSmoking_preg, season_visit, water_child }
{ age_child, airPollution_child, airPollution_preg, breastfeeding, bw,
  child_alcohol, child_depression, child_diet, child_smoking, cohort,
  ethnicity_child, ethnicity_mother, familySEP, gestational_age,
  maternalAlcohol_preg, maternalDiet_preg, maternalIodine_preg,
  maternalIron_preg, maternalSEP_preg, maternalSmoking_preg,
  maternal_folicAcid_preg, neuropsychologicalDiagnosis_child,
  otherChemicals_child, otherChemicals_preg, paternalSEP_preg,
  paternalSmoking_preg, season_visit, water_child }
{ age_child, airPollution_child, airPollution_preg, child_alcohol,
  child_diet, child_smoking, cohort, ethnicity_child, ethnicity_mother,
  familySEP, maternalAlcohol_preg, maternalDiet_preg,
  maternalSmoking_preg, otherChemicals_child, otherChemicals_preg,
  paternalSmoking_preg, season_visit, sex_child, water_child }
```

Bibliography

Díaz, Iván, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. 2021. "Nonparametric Causal Effects Based on Longitudinal Modified Treatment Policies." *Journal of the American Statistical Association* 0 (0): 1–16. https://doi.org/10.1080/01621459.2021.1955691.

Haneuse, S., and A. Rotnitzky. 2013. "Estimation of the Effect of Interventions That Modify the Received Treatment." *Statistics in Medicine* 32 (30): 5260–77. https://doi.org/10.1002/sim.5907.

Laan, Mark J. van der, David Benkeser, and Oleg Sofrygin. 2018. "Targeted Minimum Loss-Based Estimation." In *Wiley StatsRef: Statistics Reference Online*, 1–8. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat07908.

Lawlor, Debbie A, Kate Tilling, and George Davey Smith. 2016. "Triangulation in Aetiological Epidemiology." *International Journal of Epidemiology* 45 (6): 1866–86. https://doi.org/10.1093/ije/dyw314.

Muñoz, Iván Díaz, and Mark van der Laan. 2012. "Population Intervention Causal Effects Based on Stochastic Interventions." *Biometrics* 68 (2): 541–49. https://doi.org/10.1111/j.1541-0420.2011.01685.x.

Vrijheid, Martine, Rémy Slama, Oliver Robinson, Leda Chatzi, Muireann Coen, Peter van den Hazel, Cathrine Thomsen, et al. 2014. "The Human Early-Life Exposome (HELIX): Project Rationale and Design." *Environmental Health Perspectives* 122 (6): 535–44. https://doi.org/10.1289/ehp.1307204.