Paper 3 - Causal Roadmap

Lorenzo Fabbri[1,2,*] Lea Maitre[1,2] Martine Vrijheid[1,2]

[1] Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
[2] Pompeu Fabra University, Barcelona, Spain

[*] Correspondence: Lorenzo Fabbri <lorenzo.fabbri@isglobal.org>

Table of contents

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

1 Formulate the research question(s)

The **aim** of the present study is to research the short-term effects of postnatal exposures 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) (Mark J. 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.

1.1 Descriptive analyses

The following was adapted from the R Workflow post of Frank Harrell.

Initial Workflow

The major steps we foreseen are:

- Improving variable names.
- Annotating and/or recoding variables with labels and units using the `upData` function from `Hmisc`.
- Viewing data dictionaries using the `contents` function from `Hmisc`.

Missing Data

It is important to understand the extent and patterns of missing data, starting with charting the marginal fraction of observations with `NA`s for each variable. We will achieve this by using the `naclus`, `naplot`, and `combplotp` functions from `Hmisc`. The `missChk` function from `qreport` uses these functions and others to produce a fairly comprehensive *missingness* report.

Data Overview

We will use the `dataOverview` function from `qreport` to produce a brief overview of the dataset(s):

- Number of distinct values.
- Number of `NA`s.
- An information measure that for numeric variables compares the information in the variable to that in a completely continuous variable with no ties.
- Degree of asymmetry of the variables' distribution.

- Modal variable value (most frequent value), and its own frequency.
- Minimum frequency value, and its own frequency.

Descriptive Statistics

We will use the `describe` function from `Hmisc` to get initial descriptive statistics and for quality controlling the data in a univariate fashion. We will further use the `summaryM` function to obtain a tabular summary of continuous and categorical variables.

Describing Variable Interrelationships

The most basic way to examine interrelationships among variables is to graphically depict a correlation matrix.

2 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.

3 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 shown in Figure 1.

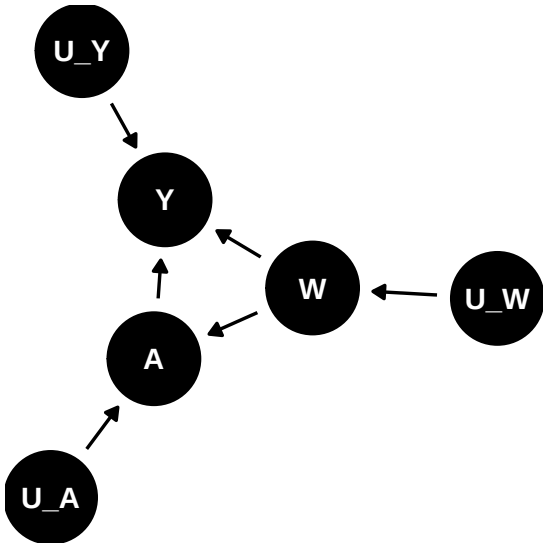

Figure 1: Causal graph corresponding to the structural equations.

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.

3.1 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. We will further stratify the analyses by sex, to account for potential sex-specific effects of the chemicals.

3.2 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 XWAS 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. We will further stratify the analyses by sex, to account for potential sex-specific effects of the chemicals.

3.3 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 XWAS 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 biomarker considered.

4 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

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

5 Estimation from data and statistical inference

To estimate the expected outcome under the observed exposure $\psi_0(A) = \mathbb{E}(Y)$, we will use the empirical mean outcome. For estimating the shift parameter $\psi_0(A^d)$, we will use TMLE, which typically uses the factorization of the observed data distribution into an outcome regression $\bar{Q}(A, W) = \mathbb{E}(Y|A, W)$ and an intervention mechanism $g(A|W) = P(A|W)$. Then, an initial estimate of the outcome regression is updated by a fluctuation that is a function of the intervention mechanism. In this study, we will use the implementation of (Díaz et al. 2021). Here, we summarize the algorithm for the shift parameter $\psi_0(A^d)$. We first define the density ratio

$$r(a, w) = \frac{g^d(a|w)}{g(a|w)}.$$

Estimation of the conditional density g is done by recasting it as a classification problem. It can be shown that

$$r(a, w) = \frac{P(\Lambda = 1|a, w)}{1 - P(\Lambda = 1|a, w)},$$

with Λ indicating whether each subject received the natural or shifted value of treatment. We refer to $P(\Lambda = 1|a, w)$ as $\bar{\lambda}$. To implement estimation of $\bar{\lambda}$ and $\bar{Q}$ flexibly, we will use the ensemble machine learning algorithm Super Learner (SL) (Mark J. van der Laan, Polley, and Hubbard 2007; Naimi and Balzer 2018). In the following, we present the steps necessary to estimate $\psi_0(A^d)$ and the associational parameter ψ_0^Δ. We note that our bounded continuous outcome will be **scaled** to $[0, 1]$, and that to handle missing covariates $W_{i,j}$, we will redefine the data to include δ_{ij}, where $\delta_{ij} = 1$ if the covariate was observed and 0 otherwise.

1. Define and calculate the density ratio under the specified shift $\hat{r}(a_i, w_i)$ for each subject $i = 1, \ldots, n$ from $\bar{\lambda}$ estimated via SL.
2. Generate initial conditional expectations for each subject, denoted $\bar{Q}(A, W)$.
3. Fit the following logistic regression on the observed (scaled) Y values, using the logit of the initial estimates as an offset
$$\text{logit}(\bar{Q}^*(A, W)) = \text{logit}(\bar{Q}(A, W)) + \epsilon$$
with weights $\hat{r}(a_i, w_i)$, calculating the estimated intercept $\hat{\epsilon}$, where $\bar{Q}^*(A, W)$ is now a targeted estimate of the conditional mean outcome.
4. Use the resulting $\hat{\epsilon}$ value to generate targeted predictions under a shift, $\bar{Q}^*(A^d, W)$:
$$\bar{Q}^*(A^d, W) = \text{logit}^{-1}\left[\text{logit}(\bar{Q}(A^d, W)) + \hat{\epsilon}\right].$$
5. With the updated estimates $\bar{Q}^*(A^d, W)$ in hand (after unscaling), define:
$$\hat{\psi}_{\text{tmle}} = \frac{1}{n}\sum_{i=1}^{n} \bar{Q}^*(A^d, W).$$

Using the empirical mean $\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, to estimate $\psi_0(A) = \mathbb{E}(Y)$, we obtain an estimate of the statistical association parameter $\psi_0^\Delta = \psi_0(A^d) - \psi_0(A)$ with

$$\hat{\psi}^\Delta = \hat{\psi}_{\text{tmle}} - \bar{Y}.$$

The 95% Wald-style confidence interval $\hat{\psi}^\Delta \pm 1.96\sqrt{Var(\hat{\psi}^\Delta)}$ is obtained from the influence curve.

Using 10-fold cross-validation (CV), SL builds the best weighted combination of predictions from the algorithms listed in @ref(tab:lmtp-specs). In case of repeated measures of the exposures, an `ID` variable will be used to make sure that the 10-fold CV splits will keep observations from the same individuals in the same split.

6 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.
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.
	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.

Effect	Approach	Description	Key sources of bias
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**. We will use TMLE in combination with SL, and a library of estimators including both simple parametric models (i.e., `glm`), and data-adaptive semi-parametric models. Specifically, we will make use of the `lmtp` R package, in combination with TMLE and SL. We will also compare the results obtained with TMLE with those obtained with a sequentially doubly robust estimator (SDRE). Table 3 provides detailed information about the specifications used for `lmtp`. TODO: specify the treatment and reference policies. The models will be adjusted for relevant confounders, as identified with `dagitty`. We hypothesize the presence of residual confounding, especially due to genetic and parental socio-economic (SE) factors, which would result in **exaggeration** of any true causal effect.
  - **Outcome negative control study**. We will try to identify outcomes in the same cohorts that would be affected by confounders that are relevant in the association of EDCs exposure to the clinical outcome of interest, but for which a biological/causal effect of the exposures is unlikely. We postulate that SE factors represent the main confounders for this association. We identified outcomes that are associated to SE factors, but that are not causally associated to the chemicals: having a car, and whether the child has their own room. A possible issue is represented by the fact that these

outcomes are measured on different scales. We hypothesized that a higher SE position is positively associated to, e.g., having a car, and negatively associated to the chemicals' concentrations. This would result in a (confounded) **inverse** relationship. We will thus test the assumptions regarding the relation between the SE factors with each outcome, and check whether these associations are in the anticipated direction.

  - **Literature search**. For each `outcome ~ chemical` association, we will perform a literature search to compare the obtained results. More precisely, we will adopt the following strategies:
    * PubChem (Kim et al. 2023). We will search each chemical, and report the results from the sections *Associated Disorders and Diseases* and *Chemical-Disease Co-Occurrences in Literature*. We will report the total number of hits for each section, the number of hits relevant for our outcome(s), and the number of articles for each hit.
    * The Comparative Toxicogenomics Database (CTD) (Davis et al. 2023). We will search each chemical, and report the results from the section *Diseases*. We will specifically look for *curated* associations. We will report the total number of hits, the number of hits relevant for our outcome(s), and the detailed information for each of these.

- `omic ~ chemical`
  - **Regression analysis**. As above.
  - **Cross-cohort comparison**. We will compare the regression models' results between subjects of different ethnic origins. We postulate that SE factors represent the main confounders for this association. Based on the associations between SE factors (e.g., maternal education) and the omic markers, we will compare the results of `omic ~ chemical` with the expected modified associations (e.g., weaker or stronger).
  - **Literature search**. As above.

- `outcome ~ omic`
  - **Regression analysis**. As above.
  - **Cross-cohort comparison**. We postulate that SE factors represent the main confounders for this association. We will thus test the association between SE factors and the omics, separately for the two sub-populations (e.g., white British and Pakistani in BiB (Wright et al. 2013)). If this association is smaller among Pakistani, if the association among white British were due to residual SE factor confounding, we would expect a **weaker** association among Pakistani.
  - **Mendelian Randomization (MR)**. A simplified directed acyclic graph (DAG) is shown in Figure 2. We will make use, ideally, of a weighted allele score of genetic variants known to be robustly associated with the omic markers as an instrumental variable (IV). We will use methods, including sensitivity analysis, to explore the possibility of bias due to: (i) weak instruments, and (ii) violation of the exclusion restriction criteria. Based on these, we will postulate whether the obtained results might be biased towards the **null** or not. By using MR, we assume that this approach tests *exposure* to omic markers across the whole of childhood. Specifically, we will perform the following analyses:
    * We will use the *standard* inverse-variance weighted (IVW) (Burgess, Butterworth, and Thompson 2013; Bowden et al. 2019) method for primary MR analyses.
    * We will use MR-Egger (Bowden, Davey Smith, and Burgess 2015), weighted median (Bowden et al. 2016), and MR-PRESSO (Verbanck et al. 2018) for sensitivity analyses. For *secondary* sensitivity analyses, we will use: Steiger filtering and/or bi-directional MR, Cochran's Q-statistics (to check for heterogeneity), and the Egger intercept (to check for horizontal pleiotropy [1]) (citations needed).

[1] In order to address the issue of horizontal pleiotropy, and to study the joint effects of multiple traits, another possibility is represented by the use of univariable MR for screening purposes, followed by multivariable (with clumping for linkage disequilibrium (LD)) MR.

We will make use of two-sample Mendelian Randomization (2SMR) (Davey Smith and Ebrahim 2003; Pierce and Burgess 2013; Davey Smith and Hemani 2014). We will thus employ results from two different types of genome-wide association studies (GWASs): a exposure genome-wide association study (GWAS), using The Metabolomics GWAS Server (Suhre et al. 2011; Shin et al. 2014), and a outcome GWAS, using the OpenGWAS project (Hemani et al. 2018; Elsworth et al. 2020). The effects of the selected genetic variants will then be harmonized. We will fit simple regression models to examine whether the selected instrumental variables (IVs) predict the relevant nodes of the DAG (i.e., exposures and outcomes), using the F-statistics. Furthermore, we will fit simple regression models of the *form* `confounder ~ SNP`, adjusted for the *top* genetic principal components (PCs), to test whether the single-nucleotide polymorphisms (SNPs) are separately associated with the confounders. As done in previous works, we will standardized all the considered outcomes. The results (both of the main and the sensitivity analyses) will be presented by means of forest plots, for the *significant* associations.

In the case of one-sample Mendelian Randomization (1SMR), we will make use of the following criteria to select the SNPs for both the exposures and the outcomes:

* We will favour *cis*-variants.
* We will *a priori* define a p-value threshold.
* We will perform LD clumping of the genetic variants.
* We will *a priori* define a minor allele frequency (MAF) threshold. Similarly, the models will be adjusted for the *top* genetic PCs.

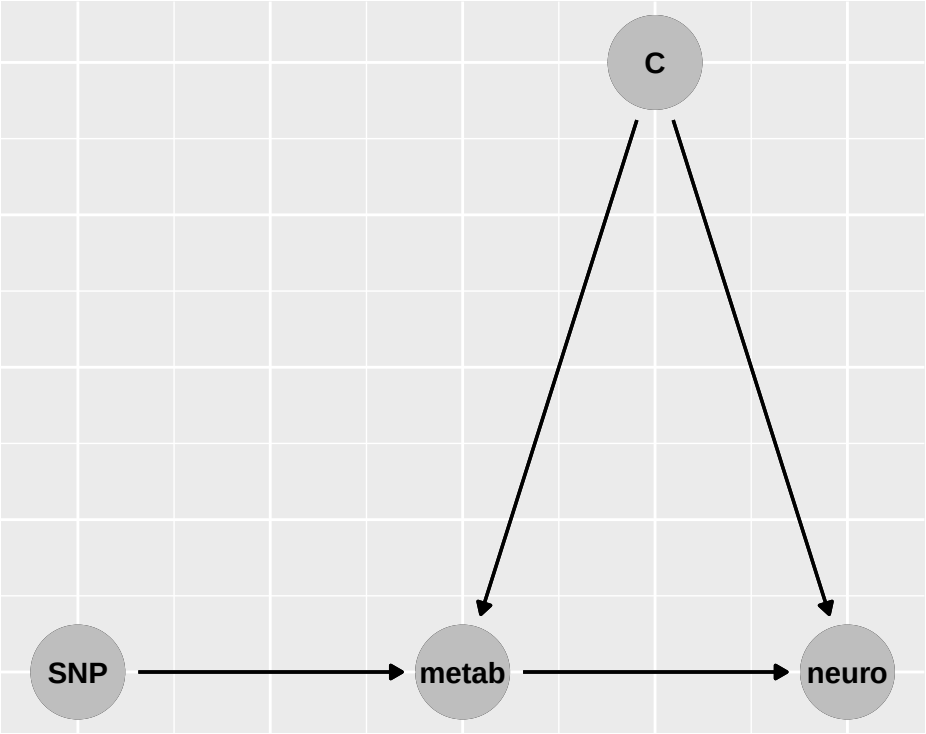

Figure 2: Simplified DAG for MR analyses of `outcome ~ omic` associations.

6.1 Mendelian Randomization: Assumptions

MR requires three IV assumptions (taken from Wikipedia):

- The genetic variant(s) used as instruments for the exposure is associated with the exposure (*relevance* assumption).
- There are no confounders of the association between the genetic variant(s) and the outcome (*independence* or *exchangeability* assumption).
- There is no pathway between the genetic variant(s) and the outcome, other than through the exposure (*exclusion restriction* or *no horizontal pleiotropy* assumption).

7 Interpretation and sensitivity analyses to inform a substantive conclusion

7.1 Reliable Causal Inference

The major aim of this PhD is to provide accurate and reliable causal effect estimates. We will thus follow the principles of **triangulation**. (Lawlor, Tilling, and Davey Smith 2016) provides a **minimum set of criteria** for use in triangulation:

- Two or more approaches with different and orthogonal sources of potential bias.
- These approaches all address the same causal question.
- The key sources of bias are explicitly acknowledged.
- The expected direction of the key sources of bias are made explicit.

Some **options** available for triangulation are (adapted from Lawlor, Tilling, and Davey Smith 2016):

- Exposure negative control study.
- Outcome negative control study.
- Multivariable analysis in observational data.
- Genetic instrumental variables (i.e., MR).
- Literature search.
- Cross-cohort comparison with different confounding structures.
- Method-based triangulation:
  - Comparison of *classical* (e.g., outcome regression, (A)IPW) and doubly-robust (e.g., TMLE) methods.
  - Comparison of frequentist and Bayesian methods. In this case, we will also compare the Bayesian posterior distribution of the parameter(s) of interest, with the distribution of the same parameter(s) obtained as a result of the uncertainty in the choice of the sample (i.e., with Bootstrapping).

While triangulation here is intended as *qualitative* triangulation, we will strive to make the obtained results as comparable as possible (e.g., by using the same units of change).

Nevertheless, causal inference is based on **untestable assumptions**. We will thus try to provide evidence of the *correctness* of these assumptions based on the data available:

- **Positivity assumption** (especially for continuous exposures):
  - We will define comparable, meaningful, and positivity-friendly policies (in the case of MTPs).
  - We will analyse the summary of the estimated propensity scores to check (near) violation of the positivity assumption.
- Sensitivity analysis to measure robustness to **unmeasured confounders**:
  - Broadly, an open question is whether the estimated treatment effect is biased due to confounding by unmeasured covariates. We will 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 the study. This

exercise illustrates how the effect estimates, and confidence interval bounds, change depending on the magnitude and direction of the hypothesized gap.

- **Propensity score**-based methods:
  - We will use balance plots of the scores to check the robustness of the outcome regression.
  - We will use balance plots of the scores for the entire dataset, for the re-weighted dataset, and separately by stratum (i.e., deciles of the propensity scores).

7.2 Reliable Statistical Inference

Considering the **high-dimensionality** of the data (wide datasets with more variables than observations), we foresee the following common problems:

- Irrelevance of significance testing.
- Unreliability of multiple comparison.
- Low power and a high false negative rate.
- Potential that the biomarkers under study are correlated with each other.
- Potential that small changes in data can lead to changes in the *winners*. We will use **resampling techniques** (e.g., bootstrapping) and obtain confidence intervals for variable importance (i.e., ranking and selection), to quantify the difficulty of variable selection. The confidence limits for the rank of each markers capture the stability of the rank itself.

A general problem in the field, is the unjustified reliance on linear models (e.g., multivariable linear regression). Some of the strategies we will adopt to relax the linearity assumptions, are:

- Regression splines.
- SL (Mark J. van der Laan, Polley, and Hubbard 2007; Naimi and Balzer 2018) and discrete SL (to deal with the potential of over-fitting) with a large enough library of learners, including BART (J. L. Hill 2011; J. Hill, Linero, and Murray 2020; Dorie et al. 2022), Bayesian Gaussian Processes.

A potential problem is the use of *transformed* variables:

- We will not categorize any of the used variables, in order to avoid loss of information.
- We will not rely on z-scores (e.g., zBMI) as outcomes or predictors.
- When using metabolomics data, we plan to use MFC during the discovery phase, and either concentrations or creatinine-adjusted values for validation.

We will perform extensive and complementary **sensitivity analyses**:

- Use of different adjustment sets provided by the `daggity` R package.
- In case of stratified analyses, we will check whether the stratification variable is, e.g., a collider.
- Complete-case sensitivity analyses.
- In the likely case that genetics might play a confounding role in the associations of interest, we will check whether the genotypes of the parents predispose the fetus to the outcome under investigation.
- In case of MR analyses, we will make use of MR-Egger and a weighted median method to explore and account for the impact of horizontal pleiotropy.

7.3 Replicability

In order to improve the **replicability** of our research, defined as the ability to obtain consistent results using the same data and code as the original study, we will use the following strategies and computational tools:

- `here`, to enable easy file referencing in project-oriented workflows.
- `tidylog`, to provide feedback about `dplyr` and `tidyr` operations.
- `labelled`, to set variables' labels.
- `dlookr`, to diagnose, explore, and transform data.
- `Hmisc`, for data checking and overview.
- `cli`, for errors, warnings, and messages.
- `acronymsdown`, to automatically handle acronyms inside RMarkdown documents.

We will also strive to be as critical as possible with respect to the expected **limitations** of our findings. Some common limitations include:

- Possibility of residual confounding when analysing observational data.
- Possibility of reverse causation due to the cross-sectional study design, especially for the associations between urine metabolites and the chemicals, and for the possibility that the outcome was measured before the exposures.
- If any of the covariates, exposures, or outcomes were self-reported, there is the possibility of under-reporting (e.g., of substance use).
- In case of long cohort studies, there is the possibility of selection bias since socio-economic and individual characteristics may affect the initial and continued participation in the study.
- Possibility that the analysis of individual exposures does not reflect their effect if considered as mixture.
- Possibility of measurement error of non-persistent chemicals (i.e., those characterized by episodic nature of exposure and short biological half-lives).
- Possibility of measurement error of the outcome. If the subjects under investigation are *too young*, there is the possibility of misclassification of the outcome (i.e., the subjects are too young to have developed the outcome of interest).
- Common methods bias.

8 Appendix

8.1 Data Checking

Before analyzing any dataset, we will perform the following *sanity checks*:

- Multi-omics data:
  - Check for (residual) **technical variability** (e.g., batch effects).
  - Explore eventual (residual) technical variability with **PCA** before and after pre-processing.
- Chemicals:
  - Check whether certain chemicals were **discretized** (e.g., DMDTP and DEDTP are binary variable in HELIX).
  - Check whether it makes sense to use *summary variables* (e.g., DEHP=MEHP+MEHHP+MEOHP+MECPP+M
    DiNP=MiNP+MHiNP).
- Covariates:

  - Check whether variables need to be transformed to **factors**, and if so, whether this transformation converts missing values to integers.

- All data:
  - Check whether different values where used for LOD, LOQ, and not detected.
  - Check whether any **transformation** was applied.
  - Check whether any **imputation** was performed.

8.2 Checklist A: Replicability

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
...	...	...

8.3 Checklist B1: `lmtp` package specifications

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

Is it possible to include models from the 'brms' R package?

Table 3: Specifications for the `lmtp` R package.

Argument	Setting	Default (Y/N)	Comment
`learners_outcome`	`SL.mean`	N	
`learners_outcome`	`SL.glm`	Y	
`learners_outcome`	`SL.gam`	N	
`learners_outcome`	`SL.hal9001`	N	
`learners_outcome`	`SL.bart`	N	
`learners_trt`	`SL.mean`	N	
`learners_trt`	`SL.glm`	Y	
`learners_trt`	`SL.gam`	N	
`learners_trt`	`SL.hal9001`	N	
`learners_trt`	`SL.bart`	N	
`folds`	10	Y	The number of folds to be used for cross-fitting
`.learners_outcome_folds`	10	Y	The number of CV folds for `learners_outcome`
`.learners_trt_folds`	10	Y	The number of CV folds for `learners_trt`

8.4 DAGs

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

There are a total of 19 minimal sufficient adjustment sets for estimating the direct effect. The first one is:

```
 [1] "age_child"                          "airPollution_child"
 [3] "airPollution_preg"                  "breastfeeding"
 [5] "bw"                                 "child_alcohol"
 [7] "child_diet"                         "child_smoking"
 [9] "edu_child"                          "familySEP"
[11] "gestational_age"                    "intelligence_SNPs"
[13] "maternalAlcohol_preg"               "maternalDiet_preg"
[15] "maternalIodine_preg"                "maternalIron_preg"
[17] "maternalSEP_preg"                   "maternalSmoking_preg"
[19] "maternal_folicAcid_preg"            "neuropsychologicalDiagnosis_child"
[21] "otherChemicals_child"               "otherChemicals_preg"
[23] "paternalSEP_preg"                   "paternalSmoking_preg"
[25] "qualityTesting_child"               "water_child"
[27] "water_preg"
```

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

There are a total of 19 minimal sufficient adjustment sets for estimating the direct effect. The first one is:

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

There are a total of 19 minimal sufficient adjustment sets for estimating the direct effect. The first one is:

References

Bowden, Jack, George Davey Smith, and Stephen Burgess. 2015. "Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection Through Egger Regression." *International Journal of Epidemiology* 44 (2): 512–25. https://doi.org/10.1093/ije/dyv080.

Bowden, Jack, Fabiola Del Greco M, Cosetta Minelli, Qingyuan Zhao, Debbie A Lawlor, Nuala A Sheehan, John Thompson, and George Davey Smith. 2019. "Improving the Accuracy of Two-Sample Summary-Data Mendelian Randomization: Moving Beyond the NOME Assumption." *International Journal of Epidemiology* 48 (3): 728–42. https://doi.org/10.1093/ije/dyy258.

Bowden, Jack, George Davey Smith, Philip C. Haycock, and Stephen Burgess. 2016. "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator." *Genetic Epidemiology* 40 (4): 304. https://doi.org/10.1002/gepi.21965.

Burgess, Stephen, Adam Butterworth, and Simon G. Thompson. 2013. "Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data." *Genetic Epidemiology* 37 (7): 658–65. https://doi.org/10.1002/gepi.21758.

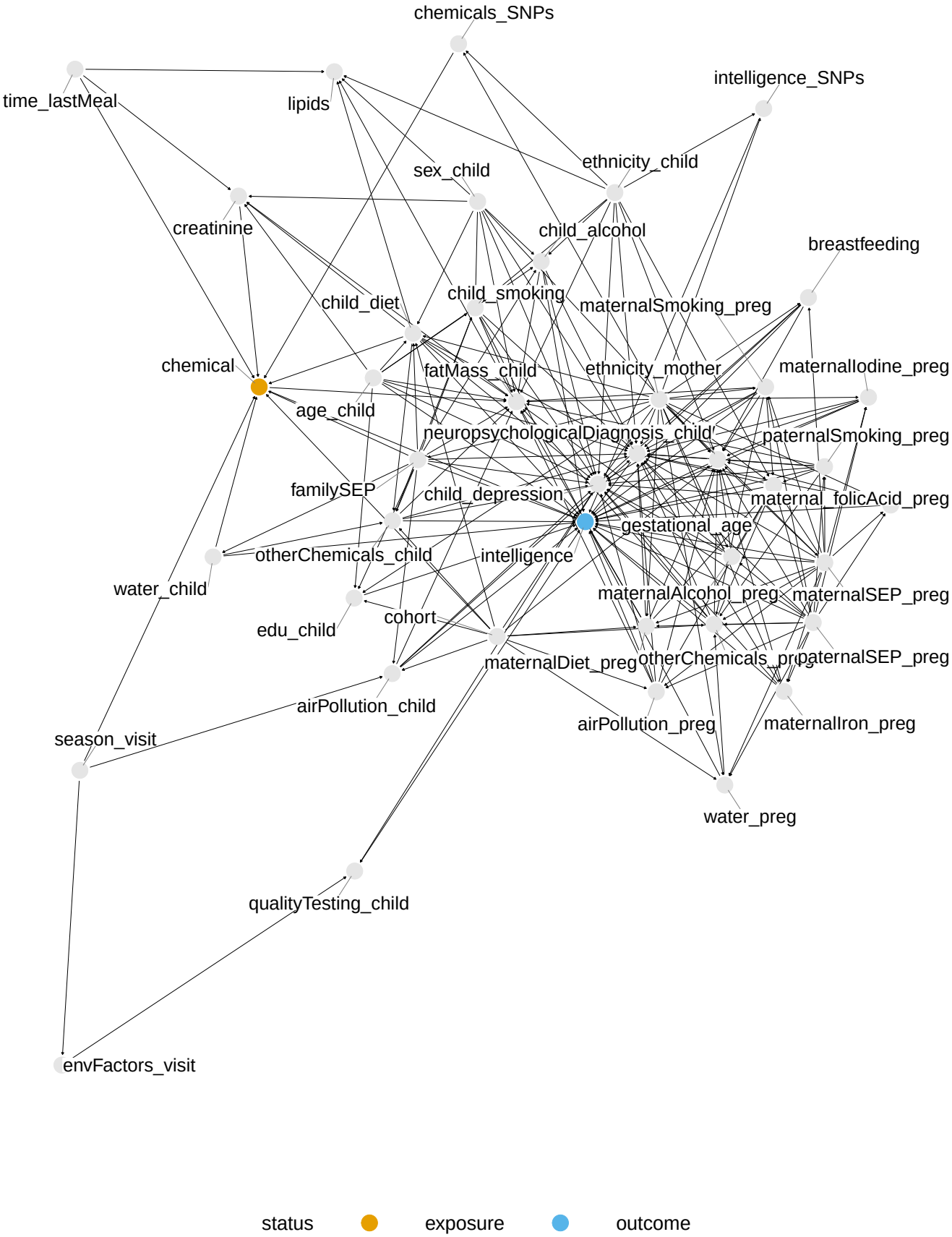

Figure 3: DAG for research question 1.

Davey Smith, George, and Shah Ebrahim. 2003. "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?*." *International Journal of Epidemiology* 32 (1): 1–22. https://doi.org/10.1093/ije/dyg070.

Davey Smith, George, and Gibran Hemani. 2014. "Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies." *Human Molecular Genetics* 23 (R1): R89–98. https://doi.org/10.1093/hmg/ddu328.

Davis, Allan Peter, Thomas C Wiegers, Robin J Johnson, Daniela Sciaky, Jolene Wiegers, and Carolyn J Mattingly. 2023. "Comparative Toxicogenomics Database (CTD): Update 2023." *Nucleic Acids Research* 51 (D1): D1257–62. https://doi.org/10.1093/nar/gkac833.

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.

Dorie, Vincent, George Perrett, Jennifer L. Hill, and Benjamin Goodrich. 2022. "Stan and BART for Causal Inference: Estimating Heterogeneous Treatment Effects Using the Power of Stan and the Flexibility of Machine Learning." *Entropy* 24 (12, 12): 1782. https://doi.org/10.3390/e24121782.

Elsworth, Ben, Matthew Lyon, Tessa Alexander, Yi Liu, Peter Matthews, Jon Hallett, Phil Bates, et al. 2020. "The MRC IEU OpenGWAS Data Infrastructure." August 10, 2020. https://doi.org/10.1101/2020.08.10.244293.

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.

Hemani, Gibran, Jie Zheng, Benjamin Elsworth, Kaitlin H Wade, Valeriia Haberland, Denis Baird, Charles Laurin, et al. 2018. "The MR-Base Platform Supports Systematic Causal Inference Across the Human Phenome." Edited by Ruth Loos. *eLife* 7 (May): e34408. https://doi.org/10.7554/eLife.34408.

Hill, Jennifer L. 2011. "Bayesian Nonparametric Modeling for Causal Inference." *Journal of Computational and Graphical Statistics* 20 (1): 217–40. https://doi.org/10.1198/jcgs.2010.08162.

Hill, Jennifer, Antonio Linero, and Jared Murray. 2020. "Bayesian Additive Regression Trees: A Review and Look Forward." *Annual Review of Statistics and Its Application* 7 (1): 251–78. https://doi.org/10.1146/annurev-statistics-031219-041110.

Kim, Sunghwan, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, et al. 2023. "PubChem 2023 Update." *Nucleic Acids Research* 51 (D1): D1373–80. https://doi.org/10.1093/nar/gkac956.

Laan, Mark J van der, Eric C Polley, and Alan E Hubbard. 2007. "Super Learner." *Stat. Appl. Genet. Mol. Biol.* 6 (September): Article25. https://doi.org/10.2202/1544-6115.1309.

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.

Naimi, Ashley I., and Laura B. Balzer. 2018. "Stacked Generalization: An Introduction to Super Learning." *European Journal of Epidemiology* 33 (5): 459–64. https://doi.org/10.1007/s10654-018-0390-z.

Pierce, Brandon L., and Stephen Burgess. 2013. "Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators." *American Journal of Epidemiology* 178 (7): 1177–84. https://doi.org/10.1093/aje/kwt084.

Shin, So-Youn, Eric B. Fauman, Ann-Kristin Petersen, Jan Krumsiek, Rita Santos, Jie Huang, Matthias Arnold, et al. 2014. "An Atlas of Genetic Influences on Human Blood Metabolites." *Nature Genetics* 46 (6, 6): 543–50. https://doi.org/10.1038/ng.2982.

Suhre, Karsten, So-Youn Shin, Ann-Kristin Petersen, Robert P. Mohney, David Meredith, Brigitte Wägele, Elisabeth Altmaier, et al. 2011. "Human Metabolic Individuality in Biomedical and Pharmaceutical Research." *Nature* 477 (7362, 7362): 54–60. https://doi.org/10.1038/nature10354.

Verbanck, Marie, Chia-Yen Chen, Benjamin Neale, and Ron Do. 2018. "Detection of Widespread Horizontal

Pleiotropy in Causal Relationships Inferred from Mendelian Randomization Between Complex Traits and Diseases." *Nature Genetics* 50 (5, 5): 693–98. https://doi.org/10.1038/s41588-018-0099-7.

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.

Wright, John, Neil Small, Pauline Raynor, Derek Tuffnell, Raj Bhopal, Noel Cameron, Lesley Fairley, et al. 2013. "Cohort Profile: The Born in Bradford Multi-Ethnic Family Cohort Study." *International Journal of Epidemiology* 42 (4): 978–91. https://doi.org/10.1093/ije/dys112.