Paper 3 - Causal Roadmap

Lorenzo Fabbri

Table of contents

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

Step 0: Formulate the research question(s)

The **aim** of the present study was to research the...

The study **population** was based on..., consisting of N =... Inclusion criteria were:...

Measured **exposures** consisted of...

The **primary outcome** was... **Secondary outcomes** included...

Descriptive analyses

Step 1: 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 assumed 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.

Step 2: Specify a causal model and causal quantity of interest

We specified the following nonparametric 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 $(f_{W_t}, f_{A_t}, f_{Y_t})$ were the nonparametric structural equations, and $(U_{W_t}, U_{A_t}, U_{Y_t})$ were 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.
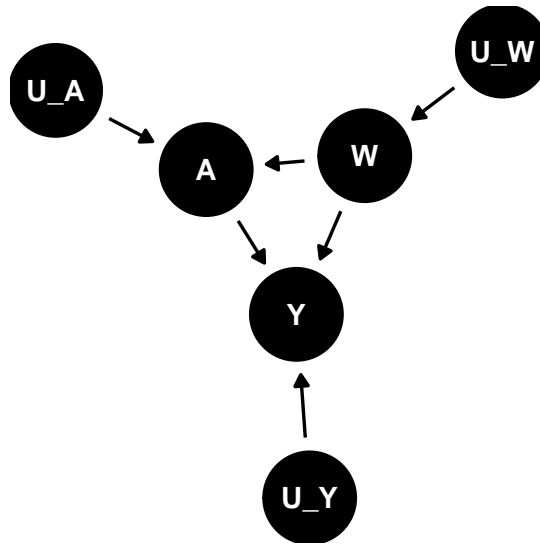

Figure 1: Causal graph corresponding to structural equations.

For each time index, we specified our causal parameter ϕ as...

As **primary analysis**, we estimated....

In the following sections, we present the simplified DAG specific to each research question.
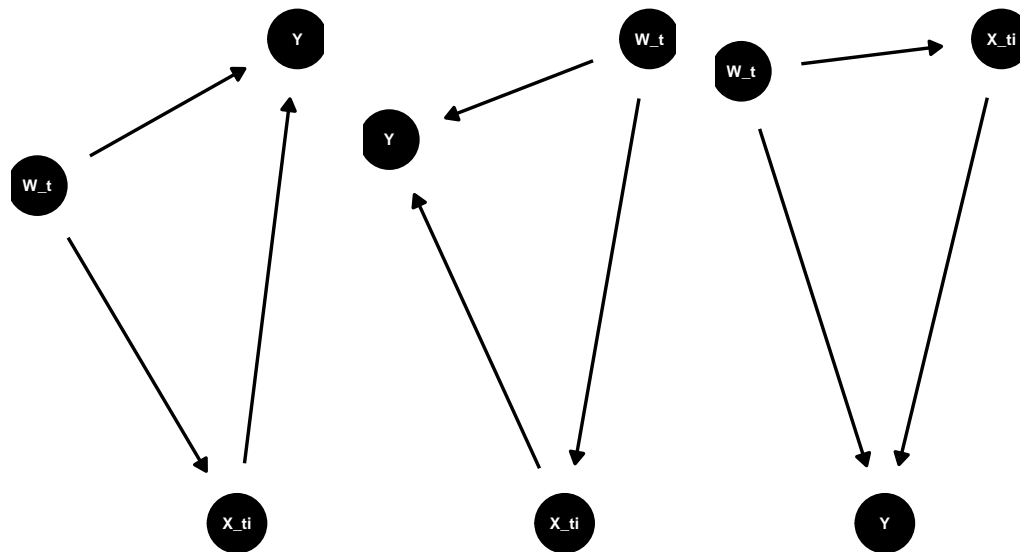
Research question 1:...

Research question 2:...

Research question 3:...

RQ1   RQ2   RQ3

Step 3: Identification and the statistical estimand

For the causal parameter ϕ, 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. In our context, this assumption would be violated if, for example, an unmeasured variable influenced both the exposure levels and the outcome. We cannot guarantee that this assumption holds, and therefore limit our interpretations to statistical associations rather than causal effects.
- **Positivity**:....
- **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 $Y(a) = Y$. 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..., is given by

$$...,$$

Therefore, our statistical estimand of interest was

$$....$$

Step 4: Estimation from data and statistical inference

Step 5: 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,
```
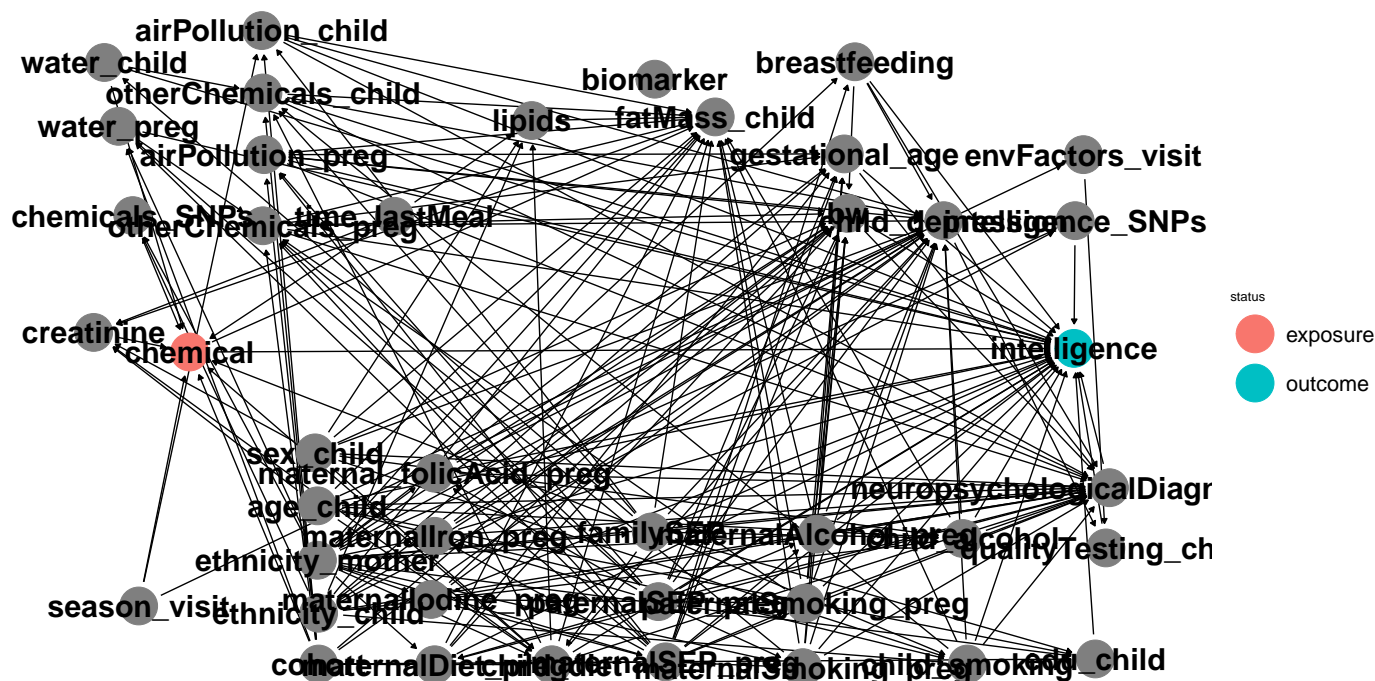
Figure 2: DAG for Research question 1.

  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