ATHLETE
The Exposome from Evidence to Translation

# Power the use of innovative methods for predicting health and disease in exposome studies

**Alicia Guillien[1], PhD and Aline Davias[1]**
[1]INSERM 1209, Equipe d'épidémiologie environnementale – Grenoble

Slides préparées par Charline Warembourg[2], PhD
[2]INSERM 1085, Institut de Recherche en Santé, Environnement, Travail - Rennes

ISGlobal
Barcelona Institute for Global Health

**April 9th, 2024**

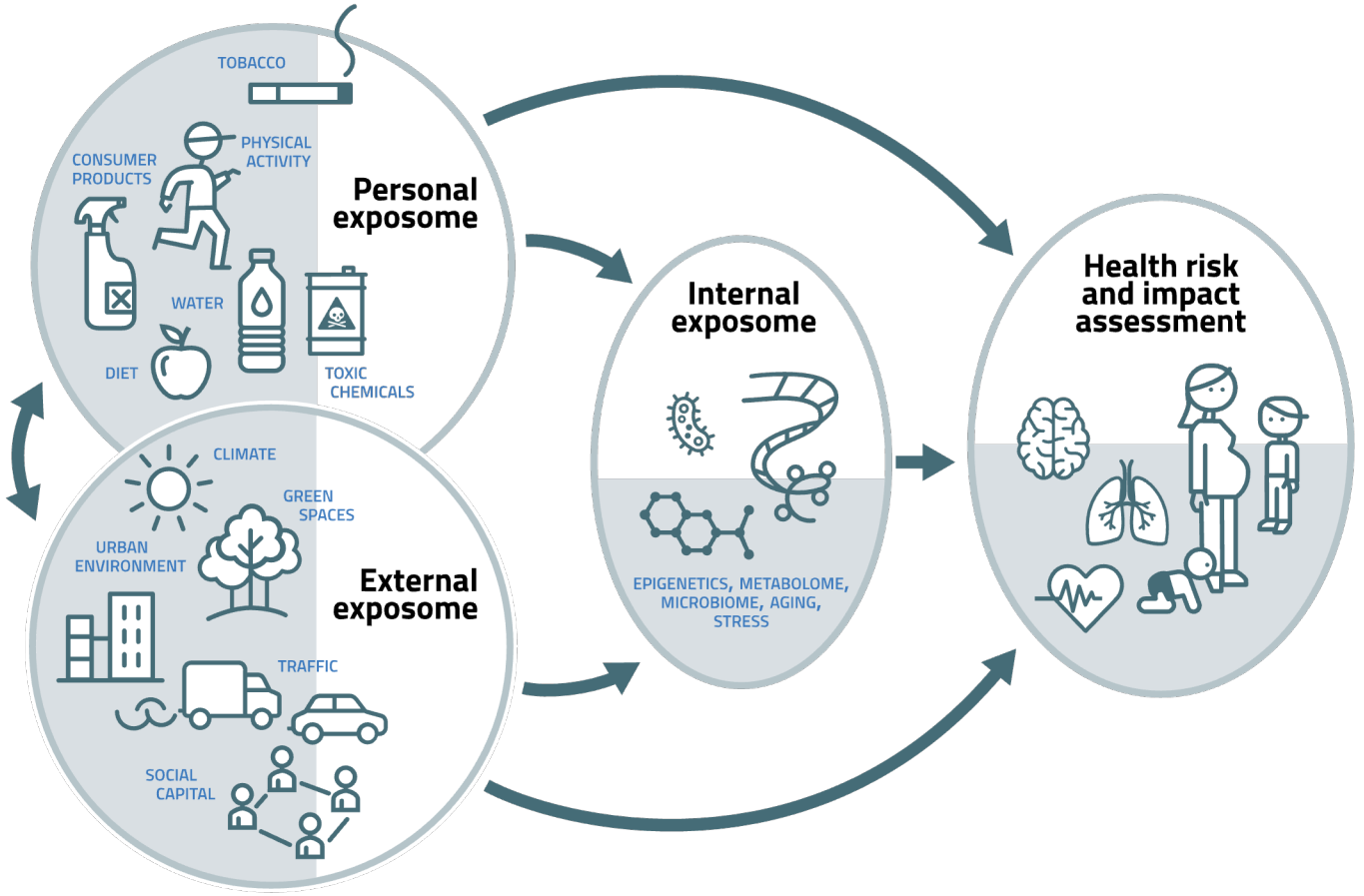
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874583

# L’Exposome

# Le concept de l'exposome

*The totality of environmental exposures (meaning all non-genetic factors) that a person experiences, from conception onwards.*

*Chris Wild, 2005.*



# Approche classique en épidémiologie environnementale: association mono-exposition

- Sélection des résultats (positifs) → Biais de publication
- Pas de correction sur les tests multiples (1 article par exposition ou par famille d'expositions)
- Pas de prise en compte des co-expositions
- Manque de consideration des "effets cocktail"

## Approche Exposome

Vision plus holistique des effets des expositions environnementales sur la santé humaine

# Pourquoi s’intéresser aux expositions dès la grossesse ?

- Période de vulnérabilité importante
- De nombreuses maladies chroniques ont des origines fœtales/développementales (concept DOHaD)
- Période clé en matière de prévention en Santé Publique

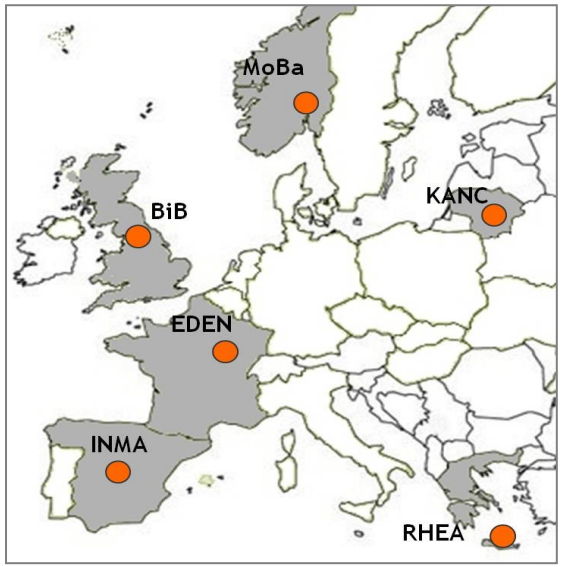
irset
Institut de recherche en santé, environnement et travail

# Le projet Helix

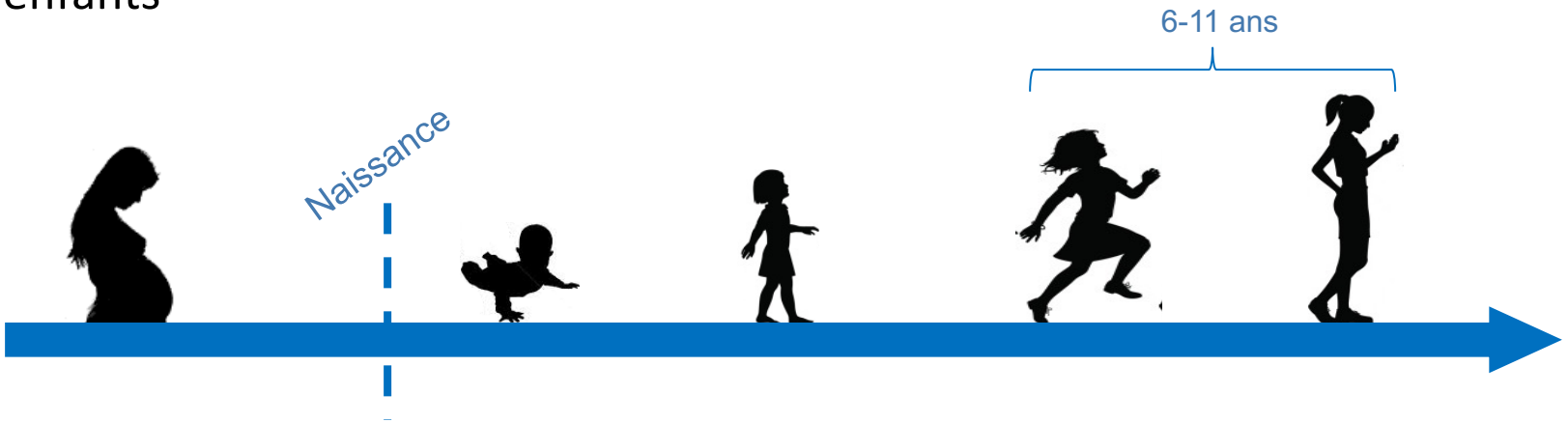
# Le projet HELIX, *Human Early-life Exposome*

- Un des premiers projets Exposome (2013-2017; IP : M. Vrijheid, ISGlobal)

- Six cohortes mères-enfants
  - Cohorte entière : données de santé préexistantes, **n=32000** paires mères-enfants
  - Sous-cohorte : suivi standardisé des enfants à l'âge de 6-11 ans (examen clinique, prélèvements biologiques, …), **n=1301** paires mères-enfants




*Vrijheid et al, EHP 2014*



# Composantes de l'exposome étudiées

## Exposome individuel

- **Habitudes de vie** – alimentation, activité physique, sommeil, tabac passif…
- **Facteurs socio-éco –** Composition du foyer, score de ressources...
- **Air intérieur-** PMs, NO2, benzène, toluène
- **Sous-produits de chloration de l'eau**

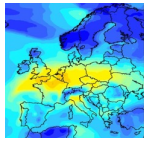
**54 variables d'exposition**
**- 16 expositions prénatales**
**- 38 expositions postnatales**

- Questionnaires

- Capteurs personnels

- Modélisation (questionnaire x SIG)

X



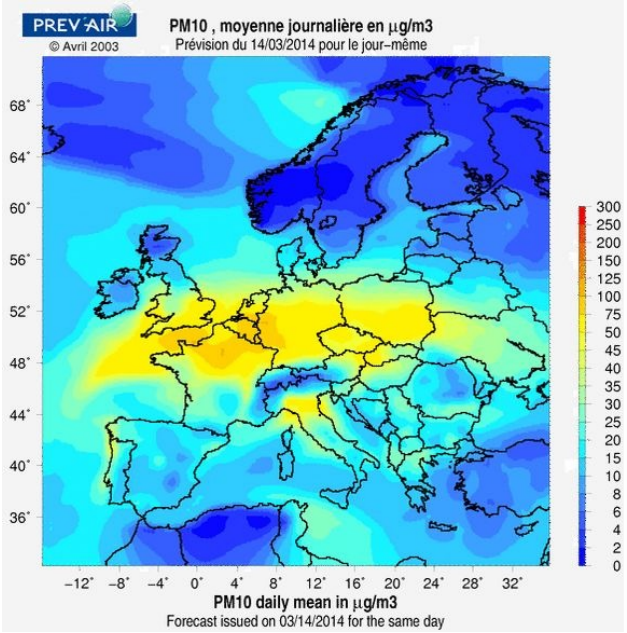
# Composantes de l'exposome étudiées

## Exposome externe (urbain)

- **Pollution de l'air -** $NO_2$, $PM_{10}$, $PM_{2.5}$, absorbance
- **Environnement bâti -** Densité de population, densité du bâti, marchabilité, connections, …
- **Espaces naturels -** Espaces verts, Espaces bleus, NDVI
- **Bruit -** 24h (Lden), Nuit (Ln)
- **Trafic routier -** Proximité, densité du trafic,…
- **Météorologie** - Température, humidité, pression, UV
- **Indice de déprivation sociale**

**69 variables d'exposition**
**- 23 expositions prénatales**
**- 46 expositions postnatales**

- Système d'information géographique (SIG)
- Géolocalisation des domiciles des participants pendant la grossesse et l'enfance et des écoles des enfants
- Croisement avec cartes existantes



# Composantes de l’exposome étudiées

## Exposome interne

- **Polluants organiques persistants** – PCBs, OCs, PBDEs, PFASs
- **Pesticides organophosphorés**
- **Phtalates**
- **Phénols –** Bisphenol A, Parabens, Triclosan, Benzophenone-3
- **Métaux** - Mercure, Plomb, Cadmium
- **Fumée du tabac**

**99 variables d’exposition**
**- 49 expositions prénatales**
**- 50 expositions postnatales**

Biomarqueurs

# Composantes de l'exposome étudiées

## Exposome individuel

- **Habitudes de vie** – alimentation, activité physique, sommeil, tabac passif…
- **Facteurs socio-éco –** Composition du foyer, score de ressources...
- **Air intérieur-** PMs, NO2, benzène, toluène
- **Sous-produits de chloration de l'eau**

**54 variables d'exposition**

## Exposome externe (urbain)

- **Pollution de l'air -** $NO_2$, $PM_{10}$, $PM_{2.5}$, absorbance
- **Environnement bâti -** Densité de population, densité du bâti, marchabilité, connections, …
- **Espaces naturels -** Espaces verts, Espaces bleus, NDVI
- **Bruit -** 24h (Lden), Nuit (Ln)
- **Trafic routier -** Proximité, densité du trafic,…
- **Météorologie** - Température, humidité, pression, UV
- **Indice de déprivation sociale**

**69 variables d'exposition**

## Exposome interne

- **Polluants organiques persistants** – PCBs, OCs, PBDEs, PFASs
- **Pesticides organophosphorés**
- **Phtalates**
- **Phénols –** Bisphenol A, Parabens, Triclosan, Benzophenone-3
- **Métaux** - Mercure, Plomb, Cadmium
- **Fumée du tabac**

**99 variables d'exposition**

**= Plus de 200 facteurs mesurés pendant la grossesse et l'enfance**

# Paramètres de santé d'intérêt

**Issues de grossesse** : Poids de naissance, durée de gestation

**Anthropométrie**: Poids, taille, IMC, circonférence abdominale…

**Pression artérielle** : Systolique et diastolique

**Neurodéveloppement**: Cognition, langage, comportement

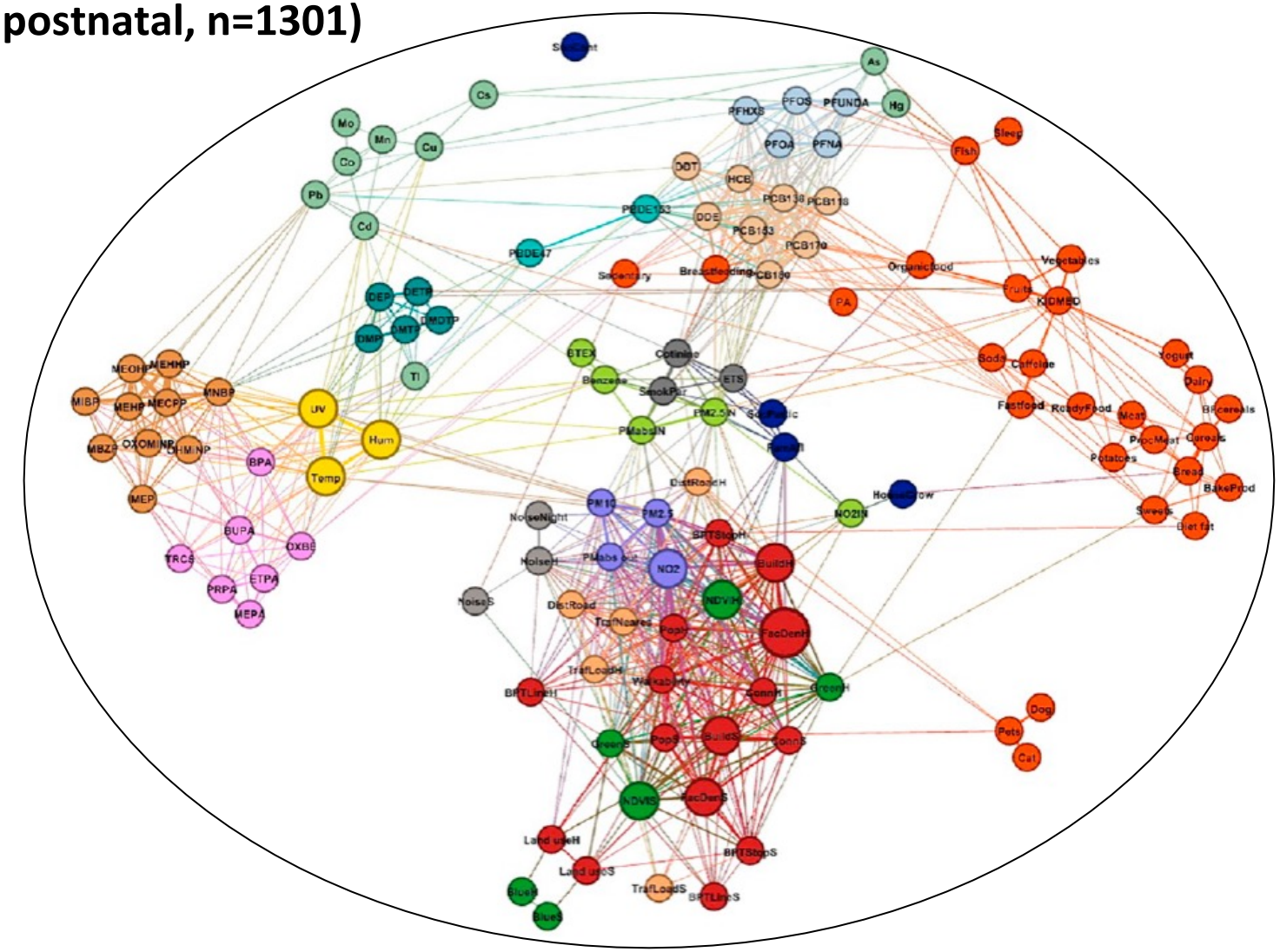
**Santé respiratoire** : Asthme, sifflement, fonction pulmonaire

**A l'âge de 6-11 ans**

**Examen et questionnaires standardisés**

# Structure de la corrélation de l'exposome

**(Exposome postnatal, n=1301)**



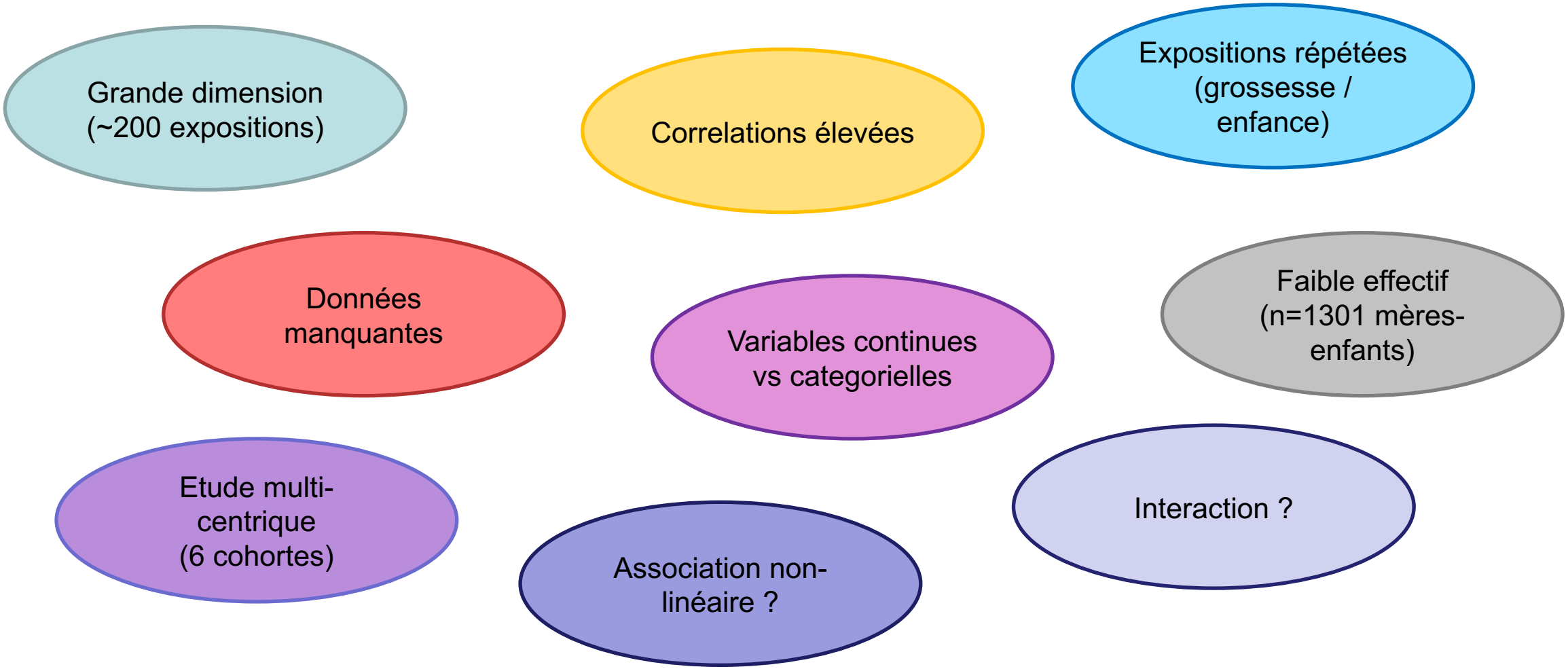

| |
|---|
| Atmospheric pollutants |
| Surrounding natural spaces |
| Meteorological |
| Built Environment |
| Traffic |
| Road traffic noise |
| OCs |
| PBDEs |
| PFASs |
| Metals |
| Phthalates |
| Phenols |
| OP Pesticides |
| Tobacco Smoke |
| Indoor air pollution |
| Lifestyle |
| Socio-economic capital |

*Tamayo-Uria et al. 2018*

Corrélations intra-famille élevées / Corrélations inter-famille faibles

# Etudier les associations entre exposome et santé :

## Un réel défi statistique



# Etudier les liens entre exposome et santé:

## Différentes approches pour différents objectifs

**Mono-exposition**

**Multi-exposition**

### Exposome-wide association study (ExWAS)

Pour estimer les associations exposition-par-exposition

Ex : Modèles de Régression avec correction pour les tests multiples

### Sélection de variables

Pour identifier les expositions qui prédisent au mieux le paramètre de santé d'intérêt et obtenir des estimations ajustées sur les co-expositions

Ex : Lasso, Elastic Net, …

### Clustering

Pour identifier des groupes de sujets qui partagent des profils d'expositions similaires

Ex : ACP, classification hierarchique

# Exposome et santé de l'enfant

Exposome

Exposome urbain

Exposome chimique

*Agier, Int J Epidemiol 2020* — *Nieuwenhuijsen, EHP 2019* → Poids de naissance

*Julvez, Environ Pollut 2021* — *Maitre, Environ Int 2021*

*Jedynak, Sci Total Environ 2021* — *Binter, Environ Int 2021* → Neurodéveloppement

*Warembourg, JACC 2019* — *Warembourg, Environ Int 2020*

*De Prado-Bert, Environ Res 2022* → Pression artérielle

*Agier, Lancet Plan Health 2019* → Fonction pulmonaire

*Granum, Environ Int 2020* → Allergie

*Vrijheid, EHP 2020* → Indice de masse corporelle

**Marqueurs OMIC**

**(multi-omic, telomere length, Epigenetic aging)**

*Cadiou, Env Int 2020* — *de Prado-Bert, Environ Int 2021*

*Clemente, EHP 2019*

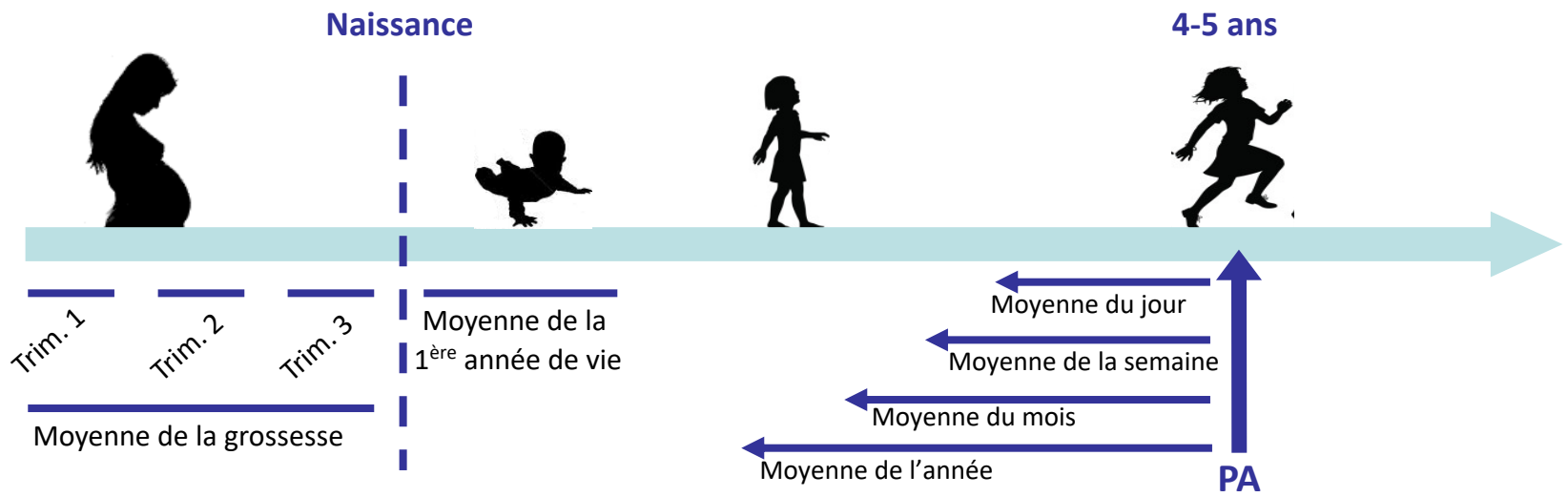
irset
Institut de recherche en santé, environnement et travail

Liste non exhaustive

# Exposome et pression artérielle de l’enfant

# Exposome urbain et pression artérielle

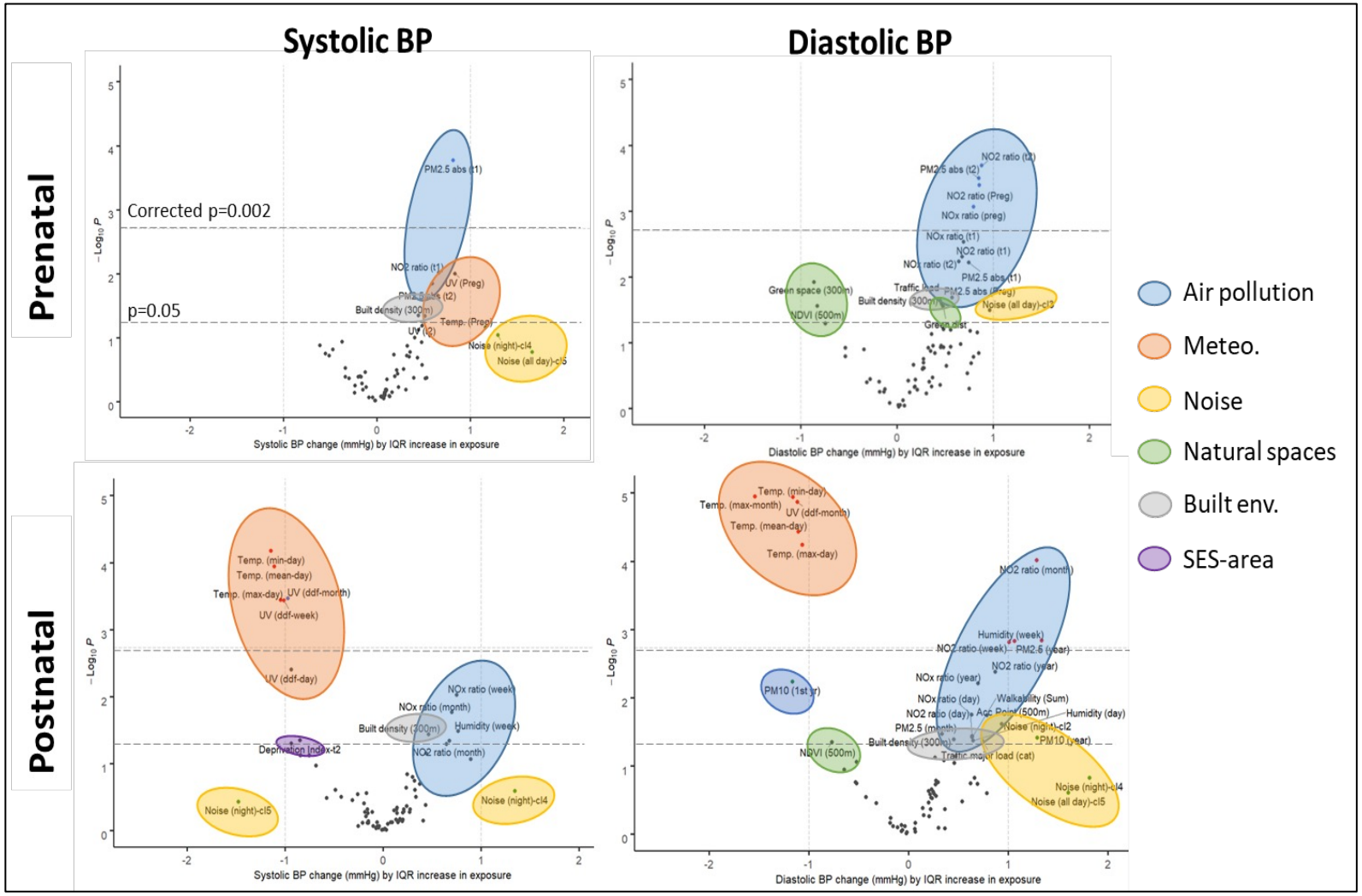
- Cohorte entière HELIX
  - N=4,700
  - Pression artérielle à 4-5 ans
  - Restreint à l'exposome urbain




*Warembourg, Environ Int 2020*

# Exposome urbain et pression artérielle à 4-5 ans (n=4,700)

**Exposome-wide association study (ExWAS) = Mono-exposition**




*Warembourg, Environ Int 2020*

All analyses are adjusted for cohort, maternal age, maternal education, maternal BMI, parity, parental country of birth, child age, sex and height

# Exposome urbain et pression artérielle à 4-5 ans (n=4700)

## Sélection de variables (DSA) = Multi-expositions

↗ Pollution de l'air, température (pendant grossesse), Bruit, Densité du bâti

↘ Température (la veille), Connectivité

↘↗ Niveau socio-économique local

All analyses are adjusted for cohort, maternal age, maternal education, maternal BMI, parity, parental country of birth, child age, sex and height

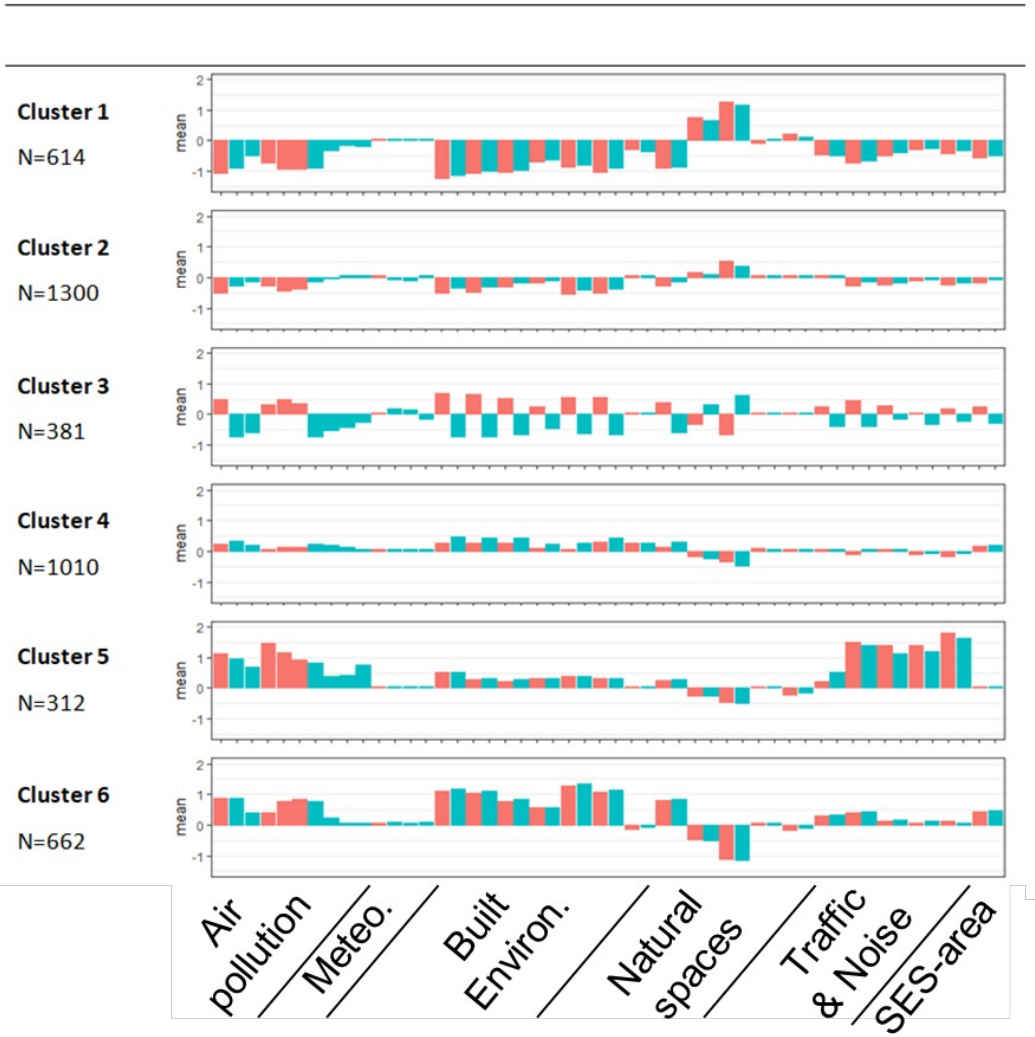
| | IQR or n (%) | % of DSA selection | Beta [95%CI] * |
|---|---|---|---|
| **SBP** | | | |
| Temperature (Day before BP) | 11.6ºC | 98 | -1.1 [-1.7; -0.6] |
| SES-area (at 4-5 years old) | | 78 | |
| 1st tertile (less deprived) | 1366 (32) | | Ref. |
| 2nd tertile | 1559 (36) | | -0.9 [-1.8; -0.0] |
| 3rd tertile (most deprived) | 1354 (32) | | 0.2 [-0.9; 1.2] |
| 24-h noise (Preg.) | | 22 | |
| <55 dB | 1418 (33) | | Ref. |
| 55-60 dB | 996 (23) | | 0.2 [-0.7; 1.0] |
| 60-65 dB | 1200 (28) | | 0.4 [-0.6; 1.3] |
| 65-70 dB | 494 (12) | | 0.1 [-1.2; 1.3] |
| >70 dB | 171 (4) | | 1.4 [-1.1; 3.9] |
| Built density (300m, at 4-5 years old) | 154998 $m^2$ built/$km^2$ | 22 | 0.8 [0.2; 1.4] |
| Temperature (Preg.) | 7.8 ºC | 20 | 1.2 [-0.0; 2.5] |
| Connectivity (300m, at 4-5 years old) | 128 intersections/$km^2$ | 16 | -0.5 [-1.1; 0.1] |
| **DBP** | | | |
| Temperature (Day before BP) | 11.6ºC | 100 | -1.1 [-1.6; -0.6] |
| $NO_2$ (Preg.) | 9.1 µg/$m^3$ | 82 | 0.7 [0.3; 1.2] |
| $PM_{2.5}$ (Year before BP) | 3.8 µg/$m^3$ | 64 | 0.9 [0.1; 1.8] |
| 24-h noise (Preg.) | | 24 | |
| <55 dB | 1418 (33) | | Ref. |
| 55-60 dB | 996 (23) | | 0.4 [-0.4; 1.3] |
| 60-65 dB | 1200 (28) | | 0.8 [-0.0; 1.9] |
| 65-70 dB | 494 (12) | | 0.5 [-0.6; 1.7] |
| >70 dB | 171 (4) | | 0.2 [-2.1; 2.5] |
| Temperature (Preg.) | 7.8ºC | 12 | 0.6 [-0.6; 1.7] |

*Warembourg, Environ Int 2020*

irset
Institut de recherche en santé, environnement et travail

# Exposome urbain et pression artérielle à 4-5 ans (n=4700)

## Clustering

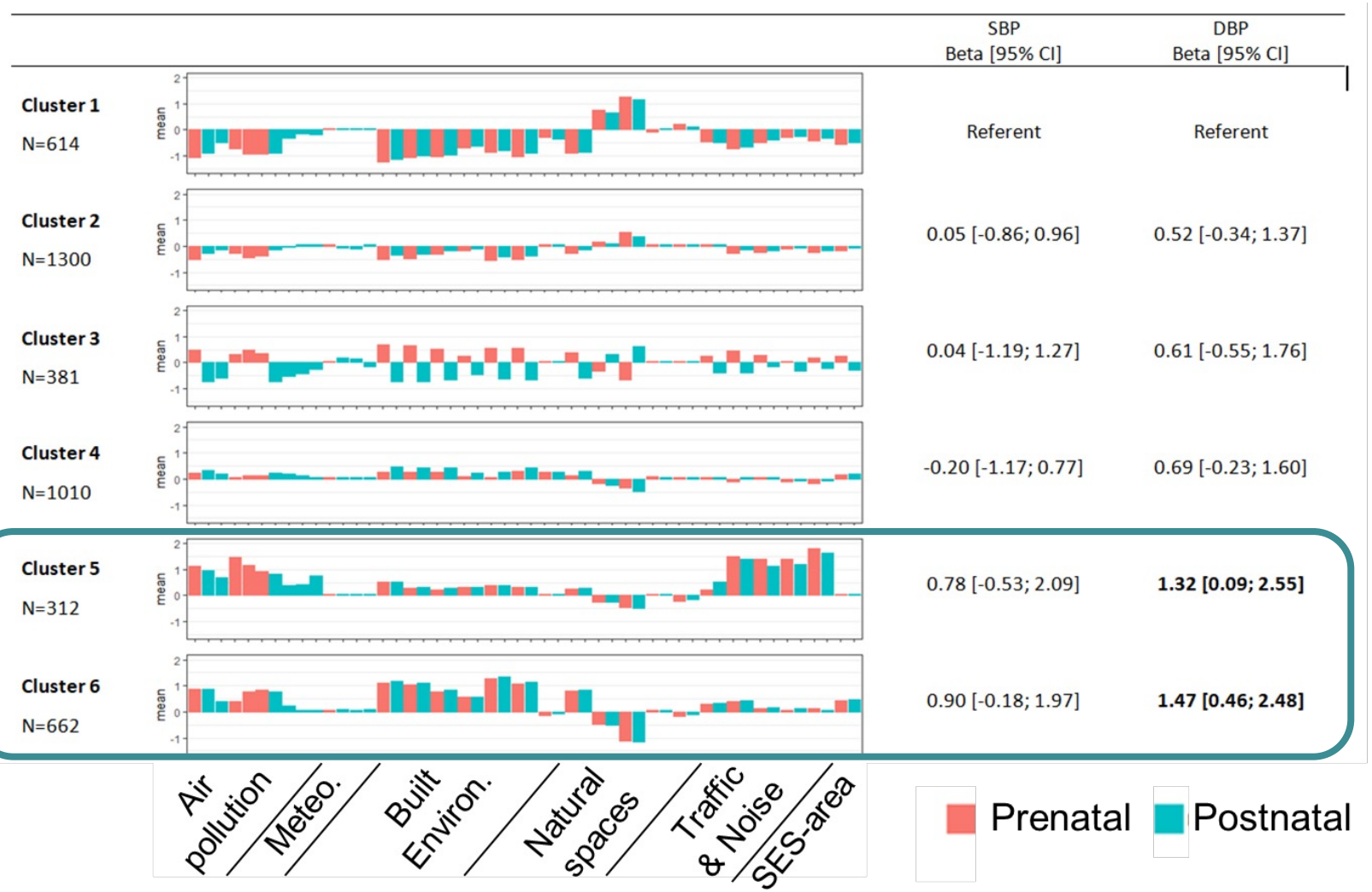



All analyses are adjusted for cohort, maternal age, maternal education, maternal BMI, parity, parental country of birth, child age, sex and height

*Warembourg, Environ Int 2020*

# Exposome urbain et pression artérielle à 4-5 ans (n=4700)

## Clustering




~25% de la population d'étude

All analyses are adjusted for cohort, maternal age, maternal education, maternal BMI, parity, parental country of birth, child age, sex and height

*Warembourg, Environ Int 2020*

# The Exposome data challenge

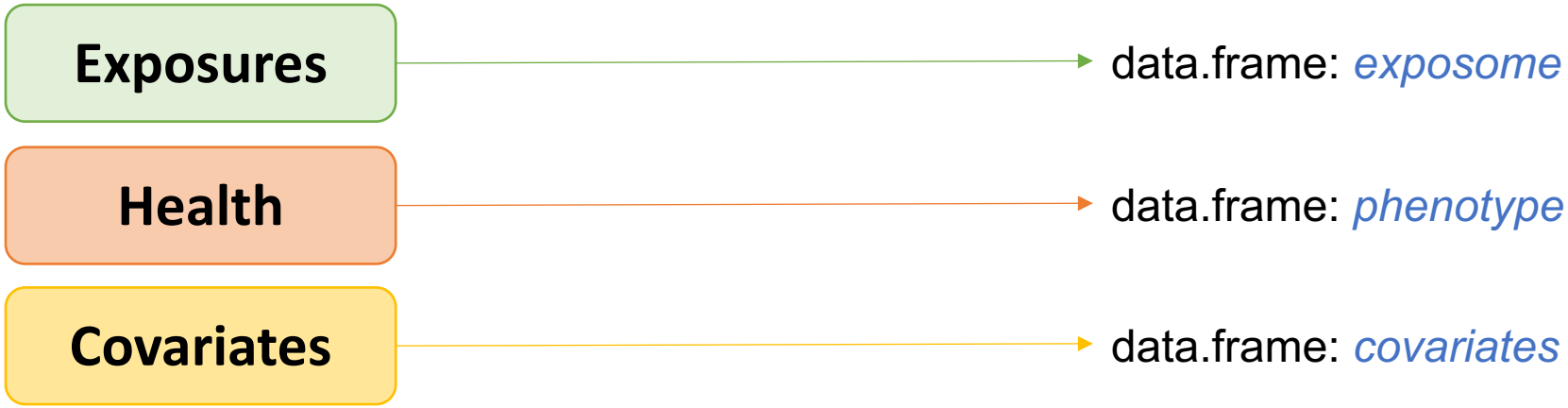
# The Exposome data challenge

- Event created in the framework of the ATHLETE project: https://athleteproject.eu/
- Organized by ISGlobal, Barcelona
- Simulated data (based on the HELIX project) publicly available to challenge researchers on statistical tools to study exposome-health associations

ATHLETE
The Exposome from Evidence to Translation

# Organization of the datasets

**Data available here:** *https://github.com/guilliea/DataChallenge_Helix*




Simulated data

Variable transformations (e.g., log2) → See codebook

Imputation of missing data

# Exposures

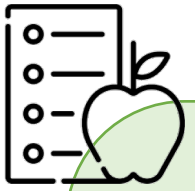
>200 environmental factors assessed during **pregnancy** and **childhood**

## Outdoor exposures
**(Geographic Information System)**

Air pollution*
Noise†
Built environment†
Natural spaces†
Traffic
Meteorology*
Water DBP
Indoor air

## Chemicals
**(blood or urine biomarkers)**

Organochlorines
PBDE
PFAS
Metals
Phthalates
Phenols
Organophosphate pesticides



## Lifestyles
**(questionnaires)**

Smoking

Diet

Physical activity

Social and economic capital

Sleep

* Postnatal exposures available within different time window

† Postnatal exposures available at different location: home and school

# Health outcomes

6 health outcomes

At birth or at the time of the children follow-up (6-11yo)

**Continuous variables**

Birth weight

Body mass index at 6-11yo

**Categorical variables**

Asthma at 6-11yo (binary)

Body mass index at 6-11yo (4 categories)

**Count variables**

Intelligence quotient at 6-11yo
Total correct answers (RAVEN test)

Neuro behavior at 6-11yo
Internalizing and externalizing problems (CBCL scale)

# Covariates, potential confounders

Maternal and child data

**Maternal data**

Cohort of inclusion

Age

Education

Pre-pregnancy body mass index

Weight gain during pregnancy

Parity

**Child data at birth**

Sex

Gestational age

Year of birth

Native origin

**Child data at 6-11yo**

Age

Weight

Height

# Data summary


Exposures ++

Health

Covariates

Exposures ++

Health

Covariates

Omics ++++

# Codebook

https://github.com/guilliea/DataChallenge_Helix/blob/main/data/codebook.xlsx

**Exposures** **Health** **Covariates**

| | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | variable_name | domain | family | subfamily | period | location | period_postnatal | description | var_type | transformation | labels |
| 2 | h_abs_ratio_preg_Log | Outdoor exposures | Air Pollution | PMAbsorb | Pregnancy | Home | NA | abs value (extrapolated back in time u | numeric | Natural Logarithm | PMabs |
| 3 | h_no2_ratio_preg_Log | Outdoor exposures | Air Pollution | NO2 | Pregnancy | Home | NA | no2 value (extrapolated back in time u | numeric | Natural Logarithm | NO2 |
| 4 | h_pm10_ratio_preg_No | Outdoor exposures | Air Pollution | PM10 | Pregnancy | Home | NA | pm10 value (extrapolated back in time | numeric | None | PM10 |
| 5 | h_pm25_ratio_preg_No | Outdoor exposures | Air Pollution | PM2.5 | Pregnancy | Home | NA | pm25 value (extrapolated back in time | numeric | None | PM2.5 |
| 6 | hs_no2_dy_hs_h_Log | Outdoor exposures | Air Pollution | NO2 | Postnatal | Home | Day before examir | no2 value (extrapolated back in time u | numeric | Natural Logarithm | NO2(day) |
| 7 | hs_no2_wk_hs_h_Log | Outdoor exposures | Air Pollution | NO2 | Postnatal | Home | Week before exan | no2 value (extrapolated back in time u | numeric | Natural Logarithm | NO2(week) |
| 8 | hs_no2_yr_hs_h_Log | Outdoor exposures | Air Pollution | NO2 | Postnatal | Home | Year before exami | no2 value (extrapolated back in time u | numeric | Natural Logarithm | NO2(year) |
| 9 | hs_pm10_dy_hs_h_Non | Outdoor exposures | Air Pollution | PM10 | Postnatal | Home | Day before examir | pm10 value (extrapolated back in time | numeric | None | PM10(day) |
| 10 | hs_pm10_wk_hs_h_Non | Outdoor exposures | Air Pollution | PM10 | Postnatal | Home | Week before exan | pm10 value (extrapolated back in time | numeric | None | PM10(week) |
| 11 | hs_pm10_yr_hs_h_Non | Outdoor exposures | Air Pollution | PM10 | Postnatal | Home | Year before exami | pm10 value (extrapolated back in time | numeric | None | PM10(year) |
| 12 | hs_pm25_dy_hs_h_Non | Outdoor exposures | Air Pollution | PM2.5 | Postnatal | Home | Day before examir | pm25 value (extrapolated back in time | numeric | None | PM2.5(day) |
| 13 | hs_pm25_wk_hs_h_Non | Outdoor exposures | Air Pollution | PM2.5 | Postnatal | Home | Week before exan | pm25 value (extrapolated back in time | numeric | None | PM2.5(week) |
| 14 | hs_pm25_yr_hs_h_Non | Outdoor exposures | Air Pollution | PM2.5 | Postnatal | Home | Year before exami | pm25 value (extrapolated back in time | numeric | None | PM2.5(year) |
| 15 | hs_pm25abs_dy_hs_h_L | Outdoor exposures | Air Pollution | PMAbsorb | Postnatal | Home | Day before examir | pm25 absorbance value (extrapolated | numeric | Natural Logarithm | PMabs(day) |
| 16 | hs_pm25abs_wk_hs_h_ | Outdoor exposures | Air Pollution | PMAbsorb | Postnatal | Home | Week before exan | pm25 absorbance value (extrapolated | numeric | Natural Logarithm | PMabs(week) |
| 17 | hs_pm25abs_yr_hs_h_L | Outdoor exposures | Air Pollution | PMAbsorb | Postnatal | Home | Year before exami | pm25 absorbance value (extrapolated | numeric | Natural Logarithm | PMabs(year) |
| 18 | h_accesslines300_preg_ | Outdoor exposures | Built environr | Access | Pregnancy | Home | NA | Meters of public transport mode lines | numeric | Dichotomous | Access_ lines |
| 19 | h_accesspoints300_preg | Outdoor exposures | Built environr | Access | Pregnancy | Home | NA | Number of bus public transport mode | numeric | Natural Logarithm | Access_stops |
| 20 | h_builtdens300_preg_Sc | Outdoor exposures | Built environr | Building d | Pregnancy | Home | NA | Building density (m2 built/km2) within | numeric | Square root | Building |
| 21 | h_connind300_preg_Sqr | Outdoor exposures | Built environr | Connectiv | Pregnancy | Home | NA | Connectivity density (number of inters | numeric | Square root | Connectivity |
| 22 | h_fdensity300_preg_Log | Outdoor exposures | Built environr | Facility | Pregnancy | Home | NA | Number of facilities present divided by | numeric | Natural Logarithm | Facility_dens |

# Your objective

# Your objective

**To apply statistical method(s) to study the associations between exposure to a wide range of environmental factors and child health, considering**:

- The dimension of the dataset
- The high correlations between exposures
- The study center effect
- Potential confounding factors
- And any other particularities of the data (categorical variables, causal relation between exposures,  non-linear associations, repeated data …)

# Ressources

## HELIX project

Data inventory: https://www.projecthelix.eu/index.php/es/data-inventory

Tamayo-Uria I, et al. The early-life exposome: Description and patterns in six European countries. Environ Int. 2019.

Maitre L, et al. Human Early Life Exposome (**HELIX**) study: a European population-based exposome cohort. BMJ Open. 2018.

Vrijheid M, et al. The human early-life exposome (**HELIX**): **project** rationale and design. Environ Health Perspect. 2014

## Exposome data challenge

Dataset: https://github.com/guilliea/DataChallenge_Helix/tree/main

Data description: https://docs.google.com/document/d/1ul3v-sIniLuTjFB1F1CrFQIX8mrEXVnvSzOF7BCOnpQ/edit

## Contact

alicia.guillien@univ-grenoble-alpes.fr; aline.davias@univ-grenoble-alpes.fr