



Grant Agreement 101155977

Milestone 4
Detailed protocol for the exposome analysis

Version 2.3 | October 2026

Milestone Title	Detailed protocol for the exposome analysis
Brief Description	Milestone 4 outlines the analysis protocol for T1.3, concerning the data-driven exposome analysis to identify socioenvironmental determinants of health behaviours and health of children and adolescents using existing data from four European studies (ABCD, BiB, BREATHE, WALNUTS)
Work Package No	1
Milestone No	4
Author(s)	V1.0: Paulina Jedynak, Xavier Basagaña, Payam Dadvand, Andrea Dietrich, Fabienne Fox, Darren Healy, Pieter Hoekstra, Jordi Julvez, Nerea Moreno, Carmen Peuters, Nikos Ntoumanis, Marlies Visser, Tanja Vrijkotte V2.3: Nerea Moreno, Darren Healy, Mai Nguyen, Nicolas Ayala, Huiyun Ding, María Arteaga, Augusto Anguita, Carmen Peuters, Payam Dadvand, Boris Cheval, Jordi Julvez
Milestone Due Date	June 30, 2025
Actual Delivery Date	V1.0: June 30, 2025 V2.3: October 1, 2026
Nature of the Milestone	Protocol
Dissemination Level	Openly accessible

Document History

Date	Version	Contributors
30/06/2025	1.0	Paulina Jedynak, Xavier Basagaña, Payam Dadvand, Andrea Dietrich, Fabienne Fox, Darren Healy, Pieter Hoekstra, Jordi Julvez, Nerea Moreno, Nikos Ntoumanis, Carmen Peuters, Marlies Visser, Tanja Vrijkotte
06/03/2026	2.0	Content update: Nerea Moreno, Darren Healy, Mai Nguyen, Nicolas Ayala, Huiyun Ding, María Arteaga, Augusto Anguita
06/08/2026	2.1	Review and revision: Mai Nguyen, Carmen Peuters, Jordi Julvez
18/09/2026	2.2	Revisions based on review: Nerea Moreno, Darren Healy, Mai Nguyen, Nicolas Ayala, Huiyun Ding
30/09/2026	2.2	Review: Boris Cheval
01/10/2026	2.3	Review and revision: Mai Nguyen

Table of Contents

1.	Introduction	1
2.	Aim of this milestone	1
3.	Study design and data sources	2
3.1.	Amsterdam Born Children and their Development (ABCD) Study	2
3.2.	Born in Bradford (BiB) Study	3
3.3.	BRain dEvelopment and Air polluTion ultrafine particles in sCHool childrEn (BREATHE) Study	3
3.4.	Walnuts Smart Snack (WALNUTs) Study	3
4.	Conceptual framework	3
5.	Descriptive analyses	6
5.1.	Summary statistics	6
5.2.	Data visualization	7
6.	Handling of the missing data	7
7.	Dimensionality reduction	7
8.	Analysis models	8
8.1.	Multi-exposure models	8
8.2.	Outcome-wide models	9
8.3.	Confounding factors	10
8.4.	Age categorization	10
9.	Secondary analyses	10
10.	Sensitivity analyses	11

11.	Reproducibility and reporting	11
11.1.	Software	11
11.2.	Reporting	11
12.	Planned analyses	11
12.1.	Proposed research project 1: Socioenvironmental determinants of physical activity and sedentary behavior in children and adolescents (leading beneficiary: ENS Rennes)	12
	Background	12
	Methodology	13
	References	14
12.2.	Proposed research project 2: Socioenvironmental determinants of sleep in children and adolescents (leading beneficiary: IISPV)	15
	Background	15
	Methodology	16
	References	18
12.3.	Proposed research project 3: Socioenvironmental determinants of dietary behavior in children and adolescents (leading beneficiary: IISPV)	19
	Background	19
	Methodology	19
	References	20
12.4.	Proposed research project 4: Socioenvironmental determinants of weight status in children and adolescents (leading beneficiary: IISPV)	20
	Background	20

Methodology	21
References	21
12.5. Proposed research project 5: Socioenvironmental determinants of subjective well-being in children and adolescents (leading beneficiary: ISGLOBAL)	22
Background	22
Methodology	23
References	23
12.6. Proposed research project 6: Socioenvironmental determinants of social well-being in adolescents (leading beneficiary: IISPV)	24
Background	24
Methodology	25
References	26

1. Introduction

The aim of Task 1.3 (T1.3) of Work Package 1 (WP1) of the YEAH! project is to perform data-driven exposome analyses to identify socioenvironmental factors associated with health behaviours and health of children and adolescents, using already existing study data. The analyses will be based on harmonised cohort data of four European studies: Amsterdam Born Children and their Development (ABCD, the Netherlands) (<https://doi.org/10.1093/ije/dyq128>), Born in Bradford (BiB, UK) (<https://doi.org/10.1093/ije/dyae037>), BRain dEvelopment and Air polluTion ultrafine particles in sChool childrEn (BREATHE, Spain) (<https://doi.org/10.1371/journal.pmed.1001792>) and Walnuts, Long-Chain Polyunsaturated Fatty Acids, and Adolescent Brain Development: Protocol for the Walnuts Smart Snack Dietary Intervention Trial (WALNUTS, Spain) (<https://doi.org/10.1016/j.eclinm.2023.101954>). More details on the harmonisation approach can be found in the [Milestone 2: Detailed protocol for data harmonisation](#).

Together with the findings in the T1.1 and T1.2, the results of the data-driven exposome analysis (T1.3) will lay the foundation for the evidence-based psychosocial intervention to promote health behaviours and health outcomes in children and adolescents.

The exposome, described as “the totality of human environmental exposures from conception onwards” (<https://doi.org/10.1158/1055-9965.EPI-05-0456>, <https://doi.org/10.1093/toxsci/kft251>), recognizes that individuals are exposed simultaneously to a multitude of different interrelated socioenvironmental factors and takes a holistic approach to the discovery of etiological factors for health-related outcomes. It provides a conceptual framework for the holistic study of multiple environmental determinants (socioenvironmental determinants in our case) and their combined effects. Thanks to the unique multi-cohort data coming from four established European cohorts with partially harmonized study designs, the proposed exposome analysis will bring new evidence on what are the most important socioenvironmental factors associated with child and adolescent health behaviours and health outcomes and guide the development of the YEAH! intervention.

2. Aim of this milestone

The aim of this document is to describe the data pre-processing, description, visualisation, and analysis steps to examine the relation between several socioenvironmental factors (exposome) simultaneously included in the models and child and adolescents' health behaviours and health outcomes. It will describe a set of potential statistical analysis strategies and tools for the data driven cross-sectional exposome analyses, carried separately for childhood and adolescence.

3. Study design and data sources

We will use available data from four cohorts (ABCD, BiB, BREATHE, WALNUTS) to assess potential socioenvironmental determinants of health behaviours and health outcomes at two key developmental stages: late childhood and mid-adolescence. The data available for at least two studies per age wave will be harmonized and pooled prior to future analyses (see Milestone 2: Detailed protocol for data harmonisation for more details). During the data pooling and harmonization across cohorts, we aimed to match the studied age ranges with the ones outlined in the YEAH project proposal (which is 8–10 years of age for childhood and 14–16 years for adolescence). Data availability resulted in a range of 7–12 (ABCD: 10–12y, BiB: 7–11y, BREATHE: 7–11y) and 12–16 years (ABCD: 15–16y, BiB: 12–15y, WALNUTS: 12–16y), respectively.

Together, the four studies provide rich, multi-level data – spanning from environmental exposures to neurodevelopment and socio-behavioural measures – providing a robust foundation for the analysis of potential health determinants and their interplays across childhood and adolescence. The four studies are briefly described below and additional details on participants' socio-demographic characteristics can be found in the publications cited below.

3.1. Amsterdam Born Children and their Development (ABCD) Study

The ABCD Study is an ongoing longitudinal multi-ethnic population-based birth cohort study started in pregnancy (<https://doi.org/10.1093/ije/dyq128>). The study investigates how lifestyle, medical, and environmental factors during pregnancy are associated with child health and development. A total of 12,373 pregnant women living in Amsterdam were approached between January 2003 and March 2004 during their first prenatal visit. Of these, 8,266 completed a pregnancy questionnaire covering sociodemographic characteristics, lifestyle, psychosocial stress, and obstetric history (phase 1). Participation rates were between 42 and 64% among foreign-born women and 77% among Dutch-born women. A total of 7,050 women gave consent for follow-up. Three months after birth, 5,131 mothers of liveborn singletons completed a second questionnaire. In comparison to the initial sample, this group tended to be older, higher educated, with fewer children, and were more frequently of Dutch origin. Follow-up thereafter focused exclusively on singleton children and included measurements such as growth data, school health checks, blood samples, medical records, and questionnaires concerning the health, behaviour, and development of both the child and the caregiver(s) at various time points. For the exposome analyses, we will use data from Phase 4 of the study, when data were collected when the children were 11–12 years old, and Phase 5 of the study, which took place when the children were 15–16 years old. These phases included questionnaires completed by children, mothers, fathers, and teachers.

3.2. Born in Bradford (BiB) Study

The BiB Study is based on a large, multi-ethnic birth cohort based at Bradford Royal Infirmary, UK. ~12,500 pregnant women (~13,500 children) were recruited between 2007 and 2011 (<https://doi.org/10.1093/ije/dyae037>). Follow-up measures included questionnaires, clinical and school-based assessments, and linkage to routine health and education records (<https://doi.org/10.12688/wellcomeopenres.20785.1>). The two age waves that are of interest for the YEAH! project are “Growing up” (7–11 years of age) and the subsequent “Age of Wonder” (12–15 years of age). The “Growing up” cohort includes ~9000 children and ~8400 mothers (<https://doi.org/10.1186/s12889-019-7222-2>); the ongoing “Age of Wonder” study aiming to track up to 30,000 adolescents (including those who participated in the original BiB study, and their peers) throughout secondary school into young adulthood with annual surveys and health measures. So far, the data include ~9000 adolescents (<https://doi.org/10.12688/wellcomeopenres.20785.1>).

3.3. BRain dEvelopment and Air polluTion ultrafine particles in scHool childrEn (BREATHE) Study

The BREATHE Study was conducted in Barcelona, Spain between January 2012 and February 2013 with the aim to evaluate the associations between traffic-related air pollutants (TRAPs) and noise pollution on mental health and neurodevelopmental outcomes in 2,745 children aged 7–11 years (<https://doi.org/10.1371/journal.pmed.1001792>). Outcomes include standardized neuropsychological assessments (e.g., attention, impulsivity, emotional regulation), cognitive tests, and behavioural questionnaires along with the omic markers of gene-environment interactions and neuroimaging data.

3.4. Walnuts Smart Snack (WALNUTs) Study

The WALNUTs study is a population-based, multi-school randomized controlled trial study conducted in Barcelona, Spain (<https://doi.org/10.1016/j.eclinm.2023.101954>). 771 healthy adolescents (aged 11–16 years) were assigned to either consume 30 g/day of walnuts or continue their usual diet over six months. Outcomes include standardized neuropsychological assessments (e.g., attention, impulsivity, emotional regulation), behavioural questionnaires, and biological markers (e.g., omega-3 PUFA status) measured at baseline and 6-month post-intervention (<https://doi.org/10.3389/fped.2021.593847>). For our analyses, we will use the baseline data only.

4. Conceptual framework

YEAH! project members involved in T1.1, T1.2, and T1.3 have developed a list of health (behaviour) outcomes and their potential determinants (see Milestone 2: Detailed protocol

for data harmonisation) relevant to the project objectives. In T1.3, we aim to investigate the importance of the socioenvironmental factors in relation to each of the outcomes of interest. The potential outcomes of interest for Task 1.3 are presented in Table 1. The socioenvironmental factors under study in T1.3, such as sociodemographics, living environment, and “urban exposome”, are presented in Table 2. Only variables that are available in at least **two cohorts** per age wave will be considered for the exposome analysis.

Table 1: A tentative list of health outcomes for the data-driven analyses.

Outcome family	Individual outcome
Health behaviours*	Physical activity
	Active commuting
	Sedentary behaviour
	Diet/ eating
	Sleep
Mental health	Anxiety symptoms
	Depressive symptoms
	Emotional problems
	Prosocial behavior
	Internalizing symptomatology
	Conduct problems
	Externalizing symptomatology
	Hyperactivity and Attentional problems
	Relationships with peers
	Total strengths & difficulties
	Loneliness and social isolation
	Coping and Optimism
	Perceived general physical health and General well-being
	Self-esteem
	Self-image and body image
	Disordered eating
	Somatic symptoms

Social Health	Perceived social support
Physical Health	Weight status: BMI/ obesity

*Variables included in this subdomain may potentially be included as also predictors of the mental health and weight status outcomes, given that health behaviours are thought to serve as a key mechanistic pathway linking exposure to health outcomes.

Table 2: A tentative list of socioenvironmental factors that are potentially important for child or adolescent health behaviours and health.

Determinant family	Individual determinant
Confounding factors	Cohort
	Age
	Sex
Sociodemographics	Puberty
	Ethnicity
	Immigration status
	Parental education
	Parental employment status
"Urban exposome"	Neighbourhood socioeconomic position
	Air pollution
	Noise pollution
	Light at night
	Temperature (heat)
	Green spaces
	Blue spaces
	Built environment
	Opportunities for physical activity (access to facilities like playgrounds, parks)
	Neighbourhood safety
	Neighbourhood walkability

	Urbanicity (building density, population density)
Inter-individual determinants	Leisure activities
	Social media use
	Social support
	Parent-for-child health behaviors - physical activity (PA), sedentary behavior (SB), and active commuting (AC)
	Parent-for-child health behaviors - risky behaviours
	Parent-for-child health behaviors - diet
	Parenting styles
Intra-individual determinants	Being bullied
	Emotion dysregulation
	Prosocial behavior
	Stress resilience / Control constructs
	Self-regulatory skills
	Other intra-individual determinants (popularity, perfectionism...)

5. Descriptive analyses

5.1. Summary statistics

The first step preceding data analysis will consist of calculation of the sample statistics by age group, cohort, and sex. For continuous variables, the following statistics will be calculated: mean, sd, median, min, max, skewness, kurtosis and interquartile ranges (IQRs). For categorical variables these will be frequency counts, percentages, and ranges (for ordered variables). And, missing rates for all variables.

Next, the distributions of determinants and outcomes will be compared across age groups, cohorts, and sex. For continuous determinants or outcomes, we will use independent samples t-test for normally distributed variables and Mann-Whitney U tests for non-normally distributed variables. For comparisons between categorical variables across groups, the chi-square test of independence will be used to assess whether distributions differ significantly across categories.

5.2. Data visualization

Data will be visualized with the most appropriate method (e.g., histograms, density plots, or boxplots/violin plots to represent the distributions of the variables; scatterplots to visualize relationships between pairs of variables; line graphs for time trends; or heatmaps for correlations). This will inform decisions on outliers, variable transformations, choice of confounders, dealing with correlated variables, and sensitivity analyses.

6. Handling of the missing data

Missingness in the outcome or determinant variables will be handled during data pre-processing preceding the analysis stage. If after the data cleaning stage (after e.g., drop-outs and other implausible values have been removed) a variable contains >25% missing values (a threshold previously used in e.g., the [LifeCycle project](#)), it could be decided to discard this variable from further analyses. Further decisions on missing values handling, also considering covariates, will be made after the inspection of the percentage and pattern of missing data.

7. Dimensionality reduction

Exposome studies are characterised by high dimensionality of exposure data and might require specific methods to reduce it while minimizing the loss of useful information regarding health associations and combined effects. The exposome dataset in the YEAH! project will consist of a substantial number of socioenvironmental determinants of health behaviors and health, with potentially high inter-variable correlations (e.g., between noise and air pollution exposures). One of the strategies to achieve dimensionality reduction is to apply either feature selection or feature extraction techniques. These precede the statistical methods that will subsequently be used to investigate the associations between a reduced set of components and the health outcome (e.g., by using the selected variables as the most relevant predictors of health outcomes in e.g., regression or classification models).

- *Feature selection methods* may include selecting based on the 1) correlation with the outcome of interest; 2) regularisation to shrink less relevant features' coefficients (e.g., LASSO, Elastic-Net regression); 3) methods based on feature importance which reflects the impact of a given feature on the model predictions through permutation, random forest, or regression coefficient.
- *Feature extraction methods* may include feature projection techniques (e.g., principal component analysis, independent component analysis, principal component pursuit, factor analysis, confirmatory factor analysis, etc.) to represent the data in a low dimensional space.

The choice of the most appropriate dimensionality reduction method will be based on the research question and/or specific variables under study.

8. Analysis models

8.1. Multi-exposure models

To investigate the relationship between multiple exposures and health outcomes, regression-based statistical methods that have been shown to perform well include the deletion/substitution/addition (DSA) algorithm, Graphical Unit Evolutionary Stochastic Search (GUESS), elastic net (ENET) and sparse partial least-squares (sPLS) regression (<http://dx.doi.org/10.1289/EHP172>). Other multi-index approaches that could be used are Bayesian Multiple Index Model (BMIM, <https://arxiv.org/pdf/2101.05352>) or Bayesian Group Index Regression (BGIR, <https://doi.org/10.3390/ijerph18073486>). We briefly outline the different methods below:

- Deletion-Substitution-Addition (DSA) is a stepwise model selection algorithm for linear regression that iteratively searches for the best model by deleting, substituting, or adding covariates. The model space is constrained by user-defined limits on interaction terms, polynomial degrees, and model size. The process starts from a simple intercept-only model and selects the final model by minimizing the root mean squared error (RMSE) using 5-fold cross-validation.
- Graphical Unit Evolutionary Stochastic Search (GUESS) is a Bayesian variable selection method that explores combinations of covariates to identify models that best predict the outcome. It uses an evolutionary Monte Carlo algorithm combining multiple chains and genetic operations to efficiently search through a vast model space. Variables are selected based on their marginal posterior probability of inclusion (MPPI), with a conservative threshold applied to control false positives. Coefficients can be estimated using ridge regression on the selected variables, as direct Bayesian coefficient estimation could be computationally intensive.
- Elastic Net (ENET) is a regression method combining two penalty techniques. It uses (1) a weighted mixture of the least absolute shrinkage and selection operator (LASSO) for variable selection, hence promoting sparsity; and (2) ridge penalties for reducing sensitivity to multicollinearity (accommodating for correlated variables), hence promoting consistent model estimates.
- Sparse Partial Least Squares (sPLS) is a supervised dimension reduction method that builds new summary variables (components) as linear combinations of the original predictors, chosen to explain as much as possible of the remaining covariance between the predictors and the outcome. It promotes sparsity by applying a penalty that sets small coefficients to zero (soft thresholding), allowing for variable selection while retaining predictive performance.
- Bayesian Multiple Index Model (BMIM) combines the dimension reduction and interpretability of linear index models (such as weighted quantile sum regression, WQSR) and flexible exposure-response modelling of response surface methods (such as Bayesian kernel machine regression). The approach simultaneously

estimates the index component weights (with variable selection) and a potentially complex, high-dimensional exposure–response relationship between the indices and health outcome. Thus, it allows non-linear effects of the indices and interactions between them.

- Bayesian Group Index Regression (BGIR) is a Bayesian equivalent to WQSR that does not use directionality constraints and allows multiple indices (based on groups of exposures) in the same model. With both BMIM and BGIR, the original exposures are reduced to a set of indices reflecting the importance of the association between each exposure and health outcome.

When deciding which model to use, we will consider model interpretability (simple, sparse models might be preferred), predictive performance and robustness of variable selection (based on previous studies with a similar design and variables set), handling of multicollinearity (depending on our correlation results), as well as availability of an up-to-date software package.

To control for Type I error inflation due to multiple comparisons, the False Discovery Rate (FDR) correction will be applied. This approach controls the proportion of false positives across all tested associations between determinants and outcomes (Benjamini & Hochberg, 1995).

8.2. Outcome-wide models

For the interrelated outcome domains (such as neurodevelopmental scores derived from e.g., Strengths and Difficulties Questionnaire, SDQ) and to examine the effects of multiple exposures across multiple outcomes simultaneously (i.e., many to many associations), we will run adjusted outcome-wide analyses. One possible method we would use is mixed-response sparse reduced-rank regression, which is a multivariate-output approach that first applies canonical correlation analysis to find canonical variables before applying multivariate regression (Anguita-Ruiz et al., 2024).

A few conditions to take into account:

- We should restrict to the variables present for all cohorts when possible, better than imputing a whole cohort.
 - In any case, non-parametric imputation methodologies should be employed (e.g., missforest)
- The RRR (reduced rank regression) methodology can mitigate outcome-specific missingness. <30% missing rates can be employed in the outcome variables.
- mRRR (mixed-outcome RRR) focuses on capturing shared structure, by using all available and different types of intercorrelated outcomes.

- Then, sRRR (sparse RRR) adds variable selection. It uses sparsity to select a subset of relevant exposures and keeps only exposures that contribute meaningfully to the shared patterns.

8.3. Confounding factors

Each analysis will be adjusted for relevant confounders selected “*a priori*” based on literature and expert knowledge. Among the confounders there will be: cohort, sex, and age, as well as other variables relevant to the particular outcome studied.

8.4. Age categorization

To ensure consistent and non-overlapping age classification across cohorts, the analytical sample was restricted to children aged 7–11 years and adolescents aged 13–16 years. 12-year-olds were excluded because this age was classified differently across cohorts: as childhood in ABCD but as adolescence in BiB and WALNUTS. Excluding 12-year-olds therefore ensured that participants of the same chronological age were not assigned to different developmental stages based solely on cohort-specific classification, thereby improving the comparability and consistency of the age-group analyses.

9. Secondary analyses

A set of secondary analyses may be run to complement the main analyses. This may help to draw conclusions, and pinpoint potential sources of variance. Potential secondary analyses are:

- An exposome-wide association study (ExWAS) to investigate bivariate associations for comparing our results with previous single-determinant studies. ExWAS fits one linear or logistic regression model per each outcome-determinant pair, adjusting for a relevant set of confounders. To control for multiple comparisons, we will apply an appropriate correction to the p-value threshold,
- A cohort-stratified analysis to check robustness of results across datasets and to identify potential differences between cohorts. We might also evaluate the between-cohort heterogeneity using Cochran's Q test and the I^2 statistic (<https://doi.org/10.1002/sim.1186.12111919>),
- An exploration of the sex-specific effects by e.g., adding an interaction term between each determinant and child's sex or by performing analyses stratified for sex.

10. Sensitivity analyses

To test the robustness of our results, the following sensitivity analyses may be performed, if relevant:

- Complete case analyses to compare the results with those based on the imputed covariates,
- Comparison of results for different imputation strategies (unless multiple imputation is applied),
- Additional adjustment for factors not included in the main analyses but being potentially relevant covariates.

11. Reproducibility and reporting

11.1. Software

We will create R scripts integrated in a version-controlled platform (e.g., Github).

11.2. Reporting

STROBE (<https://www.strobe-statement.org>) and STROBE-RECORD (<https://www.record-statement.org>) guidelines for observational studies will be followed for concise and transparent reporting.

12. Planned analyses

The Task 1.3 working institutions (IISPV, ISGlobal, and ENS Rennes) have the following **research projects** planned, under the scope of exposome analyses:

1. Socioenvironmental determinants of physical activity and sedentary behavior in children and adolescents (leading beneficiary: ENS Rennes)
2. Socioenvironmental determinants of sleep in children and adolescents (leading beneficiary: IISPV)
3. Socioenvironmental determinants of dietary behavior in children and adolescents (leading beneficiary: IISPV)
4. Socioenvironmental determinants of weight status in children and adolescents (leading beneficiary: IISPV)
5. Socioenvironmental determinants of subjective well-being in children and adolescents (leading beneficiary: ISGLOBAL)
6. Socioenvironmental determinants of social well-being in adolescents (leading beneficiary: IISPV)

12.1. Proposed research project 1: Socioenvironmental determinants of physical activity and sedentary behavior in children and adolescents (leading beneficiary: ENS Rennes)

Background

Insufficient physical activity (PA) in childhood and adolescence represents a major public health concern. Regular PA contributes to the prevention of cardiometabolic risk and supports physical and mental health during youth (Biddle et al., 2019). However, approximately 80% of school-aged children worldwide fail to meet the recommended 60 minutes of moderate-to-vigorous physical activity per day, with particularly low levels observed among girls (Aubert et al., 2021; Guthold et al., 2020). Given that PA behaviors established during youth tend to track into adulthood (Telama, 2009), early-life determinants of PA are important targets for prevention strategies.

Emerging evidence suggests that socioenvironmental contexts shape opportunities, motivations, and constraints for engaging in PA. Factors such as access to green and blue spaces, neighborhood walkability, air pollution, traffic density, safety, urban design, and socioeconomic conditions may jointly influence children's movement behaviors (Nieuwenhuijsen, 2016; Sallis et al., 2016). Yet, most studies have examined these exposures in isolation, with less attention to their co-occurrence and interdependence within the broader environmental context.

Within the exposome framework, this project aims to characterize associations between a broad range of social and physical environmental characteristics and movement behaviors across childhood and adolescence. The exposome concept emphasizes the comprehensive characterization of environmental exposures across the life course and provides a framework for considering multiple exposure domains simultaneously (Wild, 2005). By adopting a holistic, data-driven approach, this study examines multiple correlated environmental, neighborhood, and individual-level characteristics simultaneously to identify socioenvironmental features associated with sedentary behavior in children and physical activity in adolescents (Gorman et al., 2021).

This research will address key gaps by:

- i) examining multiple correlated social and physical environmental characteristics within a common exposome framework;
- ii) identifying environmental, neighborhood, and individual-level characteristics associated with daily television viewing in children and weekly physical activity frequency in adolescents using a two-stage variable-selection and regression approach; and
- iii) exploring heterogeneity in the selected characteristics and their associations across sex and cohort strata.

Methodology

Study Population

Data will be drawn from harmonized European cohorts participating in the YEAH! project, including ABCD, BREATHE, BiB, and WALNUTS. The cohorts represent diverse European environmental and sociodemographic contexts and include participants assessed during childhood and adolescence.

The analytical populations will depend on the availability and harmonization of the movement behavior outcomes. The childhood analysis of daily television viewing includes participants from the ABCD, BiB, and BREATHE cohorts. The adolescent analysis of weekly physical activity frequency is restricted to the ABCD and BiB cohorts because of limited usable observations for the harmonized physical activity outcome in WALNUTS.

Outcome Assessment : Physical Activity

The analyses focus on two movement behavior outcomes that were available for multicohort analysis after harmonization.

For children, the outcome is daily television viewing duration, measured as the average number of minutes spent watching television per day. Television viewing was used as an indicator of sedentary behavior and was analyzed across the ABCD, BiB, and BREATHE cohorts.

For adolescents, the outcome is weekly frequency of physical activity, measured as the number of days per week participants engaged in physical activity. This analysis was conducted using participants from the ABCD and BiB cohorts.

The childhood and adolescent analyses therefore examine different but behaviorally related dimensions of movement. Sedentary behavior and physical activity are distinct behaviors and should not be interpreted as direct opposites, although greater time spent in sedentary activities may reduce the time available for physical activity. Accordingly, comparisons between the childhood and adolescent analyses are interpreted cautiously because they involve different developmental stages, movement behavior outcomes, cohort compositions, and analytical samples.

Exposure Assessment: Social and Physical Exposome

The exposome analysis includes harmonized environmental, neighborhood-level, and individual-level characteristics representing multiple dimensions of the social and physical environment.

Statistical Analysis

To address the high dimensionality and correlation structure of the exposome data, a two-stage analytical framework combining Elastic Net (ENET) regularization with post-selection multivariable regression will be used.

Stage 1: Elastic Net Variable Selection

ENET regularized regression will be used to select candidate environmental, neighborhood-level, and individual-level characteristics associated with each movement behavior outcome. ENET is particularly suitable for exposome data because it can accommodate a relatively large number of correlated predictors while reducing overfitting through penalization (Zou & Hastie, 2005).

Variable selection will be performed separately for the pooled population and within sex- and cohort-stratified analyses. Sex, age, and cohort will be specified as predefined unpenalized covariates where applicable and will therefore be included in the models independently of ENET variable selection.

Stage 2: Post-Selection Multivariable Regression

Variables retained by ENET will subsequently be entered into multivariable linear regression models to estimate the direction and magnitude of their associations with the movement behavior outcomes while accounting for the other selected characteristics. False Discovery Rate (FDR) correction will be applied to account for multiple comparisons (Benjamini & Hochberg, 1995).

The pooled analyses will be complemented by sex-stratified and cohort-specific analyses to examine whether the characteristics selected by ENET and their associations with movement behaviors vary across population groups and study settings. Differences observed across strata will be interpreted as patterns of heterogeneity rather than evidence of sex-specific or cohort-specific effects because formal interaction tests will not be performed.

Sensitivity Analysis

The robustness of the regression findings will be evaluated by comparing conventional linear regression estimates with robust regression models. This comparison will assess the potential influence of outlying or influential observations on the estimated associations.

References

Aubert S et al. (2021). Global prevalence of physical activity for children and adolescents; inconsistencies, research gaps, and recommendations: A narrative review. *Int J Behav Nutr Phys Act*.

Benjamini Y and Hochberg Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol.

Biddle SJH et al. (2019). Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. Psychol Sport Exerc.

Gorman S et al. (2021). Exposomes and metabolic health through a physical activity lens: A narrative review. J Endocrinol.

Guthold R et al. (2020). Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1.6 million participants. Lancet Child Adolesc Health.

Nieuwenhuijsen MJ. (2016). Urban and transport planning, environmental exposures and health—New concepts, methods and tools to improve health in cities. Environ Health.

Sallis JF et al. (2016). Physical activity in relation to urban environments in 14 cities worldwide: A cross-sectional study. Lancet.

Telama R. (2009). Tracking of physical activity from childhood to adulthood: A review. Obes Facts.

Wild CP. (2005). Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev.

Zou H and Hastie T. (2005). Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol.

12.2. Proposed research project 2: Socioenvironmental determinants of sleep in children and adolescents (leading beneficiary: IISPV)

Background

Healthy sleep is essential for optimal development during childhood and adolescence. Adequate sleep supports emotional regulation, behavior, cognitive functioning, and learning, while insufficient or poor-quality sleep has been associated with a wide range of adverse outcomes including anxiety, depression, behavioral problems, and impaired academic performance (Félix & Candeias, 2025; Gregory & Sadeh, 2012; Dewald et al., 2010).

www.yeah-project.eu Yeah Project

15



Emerging evidence suggests that sleep health is shaped not only by individual behaviors but also by broader social and environmental contexts. Factors such as neighborhood disadvantage, environmental noise, air pollution, housing conditions, and family socioeconomic status may influence children's sleep through multiple biological, psychological, and behavioral pathways (Johnson et al., 2018; Billings et al., 2019). For example, stressful or disadvantaged neighborhood environments have been associated with shorter sleep duration and increased sleep disturbances among youth (Singh & Kenney, 2013; Bagley et al., 2018). At the household level, family functioning, parental stress, and socioeconomic conditions may also affect bedtime routines, sleep environments, and children's emotional security, all of which contribute to sleep outcomes (El-Sheikh & Kelly, 2017; Bagley et al., 2015).

Sleep has increasingly been recognized as a key social determinant of health, reflecting broader structural inequalities and environmental conditions. Disparities in sleep health may therefore reflect unequal exposure to adverse social and physical environments (Hale et al., 2020). Addressing these disparities requires a better understanding of how multiple socioenvironmental factors jointly influence sleep patterns in children and adolescents. However, most previous studies have examined environmental exposures individually, without considering the complex interplay between social and physical environmental factors. This approach may overlook the combined or interactive effects of exposures that commonly co-occur in real-world environments.

Within the exposome framework, this project aims to investigate the combined and independent effects of the social and physical exposome on sleep outcomes among children and adolescents. By adopting a comprehensive, data-driven approach, this study seeks to identify environmental exposure patterns and contextual profiles associated with favorable or unfavorable sleep health.

This research will address key gaps by:

- i) jointly examining multiple social and physical environmental determinants of sleep;
- ii) identifying key environmental domains associated with sleep duration; and
- iii) exploring potential social inequalities in sleep outcomes among children and adolescents.

Methodology

Study Population

Data will be drawn from harmonized European child cohorts, including ABCD, BREATHE, BiB, and WALNUTS. These cohorts include children from diverse European settings and encompass both cross-sectional and longitudinal data collected across childhood and adolescence. Participants aged approximately 7–16 years will be included in the analyses.

www.yeah-project.eu Yeah Project

16

Outcome Assessment: Sleep duration

Sleep outcomes will be derived from harmonized questionnaire measures and, where available, objective assessments. The primary outcome is sleep duration.

Exposure Assessment: Social and Physical Exposome

The determinants include harmonized environmental, neighborhood-level, and individual-level characteristics representing multiple dimensions of the social and physical environment. For example:

Physical environment

- Air pollution exposure
- Environmental noise
- Green and blue space availability
- Urban density and land use characteristics

Social environment

- Family socioeconomic status
- Household conditions
- Psychosocial stressors and family context

Statistical Analysis

To address the high dimensionality and potential multicollinearity among socioenvironmental exposures, a multi-step statistical strategy will be applied.

First, **Elastic Net** will be used for variable selection. This approach combines the strengths of LASSO and ridge regression and is well suited for identifying relevant predictors in datasets with correlated determinants.

Then, **post-selection effects estimation** will be conducted using regression models to estimate the association between selected determinants and sleep duration. Depending on the distribution of the outcome and presence of influential observations, **robust regression models** will be implemented to improve the stability of estimates and reduce sensitivity to outliers.

All models will be adjusted for key confounders such as cohort, age, sex, and other relevant individual characteristics.

To account for heterogeneity across cohorts, cohort-stratification analysis will be performed. Stratifications by sex will also be carried out to account for sex differences.

References

Bagley, E. J., Fuller-Rowell, T. E., Saini, E. K., Philbrook, L. E., & El-Sheikh, M. (2018). Neighborhood economic deprivation and social fragmentation: Associations with children's sleep. *Behavioral Sleep Medicine*, 16(6), 542–552.

Bagley, E. J., Kelly, R. J., Buckhalt, J. A., & El-Sheikh, M. (2015). What keeps low-SES children from sleeping well: The role of presleep worries and sleep environment. *Sleep Medicine*, 16(4), 496–502.

Billings, M. E., Hale, L., & Johnson, D. A. (2019). Physical and social environment relationship with sleep health and disorders. *Chest*, 156(2), 427–438.

Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Medicine Reviews*, 14(3), 179–189.

El-Sheikh, M., & Kelly, R. J. (2017). Family functioning and children's sleep. *Child Development Perspectives*, 11(4), 264–269.

Félix, A., & Candeias, A. (2025). Sleep as a developmental process: A systematic review of cognitive, emotional, and behavioral outcomes in children aged 6–12 years. *Clocks & Sleep*, 7(4), 66.

Gregory, A. M., & Sadeh, A. (2012). Sleep, emotional and behavioral difficulties in children and adolescents. *Sleep Medicine Reviews*, 16(2), 129–136.

Hale, L., Troxel, W., & Buysse, D. J. (2020). Sleep health: An opportunity for public health to address health equity. *Annual Review of Public Health*, 41, 81–99.

Johnson, D. A., Billings, M. E., & Hale, L. (2018). Environmental determinants of insufficient sleep and sleep disorders: Implications for population health. *Current Epidemiology Reports*, 5(2), 61–69.

Singh, G. K., & Kenney, M. K. (2013). Rising prevalence and neighborhood, social, and behavioral determinants of sleep problems in US children and adolescents, 2003–2012. *Sleep Disorders*, 2013, Article 394320, 1–15.

12.3. Proposed research project 3: Socioenvironmental determinants of dietary behavior in children and adolescents (leading beneficiary: IISPV)

Background

Diet in childhood and adolescence is a cornerstone of healthy development and a major public health priority across Europe. Suboptimal dietary habits during these formative years contribute to nutrient deficiencies, excess weight gain, and the early emergence of cardiometabolic risk factors, as well as impacts on cognitive and emotional development (Parajuli and Prangthip, 2025). Meanwhile, adherence to a healthy dietary pattern, such as the Mediterranean diet, has been linked with lowered cardiometabolic risk (Dubey et al., 2025). Unfortunately, many European adolescents continue to consume poor-quality diets; in 20 of the 31 countries that had data on fresh fruit consumption, less than half of the children consumed fruit every day.³ This is even worse for vegetable intake, with less than 1/3 of children consuming vegetables on a daily basis (WHO European Childhood Obesity Surveillance Initiative, 2025).

Most research has examined individual social or environmental determinants of diet in isolation. In reality, children and adolescents are simultaneously exposed to a complex web of socio-environmental factors—including neighbourhood characteristics, food environments, social norms, and economic conditions—that interact to shape dietary intake (World Health Organization, 2024; Sarwar et al., 2024; Mesch et al., 2025).

Within the exposome framework, this project will use a data-driven approach to characterize the combined and independent effects of diverse socio-environmental determinants on dietary quality and food intake among children and adolescents. By jointly modeling multiple social, and environmental exposures in an exposure-wide approach, the project will better reflect real-world determinants of diet and identify sensitive developmental periods during which specific exposures exert greater influence.

Methodology

The study will use harmonised cohort data from BiB and ABCD (childhood) and ABCD and WALNUTs (adolescence) and advanced, exposome-oriented statistical methods (Multi-task Elastic Net) to generate robust evidence on the contribution of socioenvironmental determinants of diet in young populations. This outcome-wide exposure-wide analysis will be carried out using multi-task elastic net regression. Models will be adjusted for key confounders (age, sex, and cohort), while sensitivity analyses will be used to assess the robustness of the results. Stratified analyses will also be carried out to assess sex- and cohort-specific findings.

References

Dubey, V.P. et al. (2025) 'Association between adherence to the Mediterranean diet and metabolic syndrome in children and adolescents: a systematic review and meta-analysis', *BMJ Nutrition, Prevention & Health*, 8.

Mesch, A. et al. (2025) 'The food environment through the eyes of adolescents: exploring socioecological factors and opportunities for healthy and sustainable food choices', *BMC Public Health*, 25, 2177.

Parajuli, J. and Prangthip, P. (2025) 'Adolescent nutrition and health: a critical period for nutritional intervention to prevent long term health consequences', *Current Nutrition Reports*, 14, 116.

Sarwar, S. et al. (2024) 'Associations between adolescent food environments and dietary intake', *Journal of Nutrition Education and Behavior*, 56, pp. 869–880.

World Health Organization (2024) 'The inequality epidemic: low-income teens face higher risks of obesity, inactivity and poor diet'. Available at: <https://www.who.int/europe/news/item/23-05-2024-the-inequality-epidemic--low-income-teens-face-higher-risks-of-obesity--inactivity-and-poor-diet> (Accessed: February 2026).

World Health Organization (2025) WHO European Childhood Obesity Surveillance Initiative (COSI): report on the sixth round of data collection, 2022–2024. Available at: <https://www.who.int/europe/publications/i/item/WHO-EURO-2025-11788-51560-78769> (Accessed: February 2026).

12.4. Proposed research project 4: Socioenvironmental determinants of weight status in children and adolescents (leading beneficiary: IISPV)

Background

One-quarter of children in Europe are living with overweight or obesity (World Health Organization, 2025), with the issue recognised as a significant public health challenge partly due to its impact on cardiometabolic health later in life (Drozd et al., 2021; Valerio et al., 2024). Living with overweight and obesity has also been linked to mental health difficulties (Ortiz-Marrón et al., 2026). Adiposity is not driven by individual behaviours alone; instead, it reflects a complex interplay of socioenvironmental exposures, behaviours, and host biology (Jeannot et al., 2015; Vandevijvere et al., 2023; Kleiser et al., 2009). Socioenvironmental determinants – including limited access to nutritious food, obesogenic built environments,

reduced opportunities for physical activity, and socioeconomic disadvantage – are major contributors to childhood obesity risk (Vandevijvere et al., 2023).

Within the exposome framework, this project will apply a data-driven, exposure-wide approach (preferably Elastic Net) to characterise the combined and independent effects of multiple socio-environmental exposures on BMI in childhood and adolescence. By jointly modelling different social and environmental factors, the project will better reflect real-world determinants of adiposity and identify sensitive developmental periods.

Methodology

The study will use harmonised cohort data from ABCD, BiB and BREATHE (childhood) and ABCD, BiB, and WALNUTs (adolescence) and advanced, exposome-oriented statistical methods to generate robust evidence on the contribution of socioenvironmental determinants of BMI in young populations. This exposure-wide analysis will be carried out using elastic net regression. Models will be adjusted for key confounders (age, sex, and cohort), while sensitivity analyses will be used to assess the robustness of the results. Stratified analyses will also be carried out to assess sex- and cohort-specific findings.

References

Drozd, D. et al. (2021) 'Obesity and cardiometabolic risk factors: from childhood to adulthood', *Nutrients*, 13.

Jeannot, E., Mahler, P., Elia, N., Cerruti, B. and Chastonay, P. (2015) 'Sociodemographic and economic determinants of overweight and obesity for public-school children in Geneva State, Switzerland: a cross-sectional study', *International Journal of Preventive Medicine*, 6, 39.

Kleiser, C., Schaffrath Rosario, A., Mensink, G.B., Prinz-Langenohl, R. and Kurth, B.-M. (2009) 'Potential determinants of obesity among children and adolescents in Germany: results from the cross-sectional KiGGS study', *BMC Public Health*, 9, 46.

Ortiz-Marrón, H. et al. (2026) 'Association between obesity and mental health problems among Spanish children aged 9 and 12 years: the ELOIN study', *BMC Public Health*, 26, 857.

Valerio, G. et al. (2024) 'Cardiometabolic risk in children and adolescents with obesity: a position paper of the Italian Society for Pediatric Endocrinology and Diabetology', *Italian Journal of Pediatrics*, 50, 205.

Vandevijvere, S. et al. (2023) 'Upstream determinants of overweight and obesity in Europe', *Current Obesity Reports*, 12, pp. 417–428.

World Health Organization (2025) WHO European Childhood Obesity Surveillance Initiative (COSI): report on the sixth round of data collection, 2022–2024. Available at: <https://www.who.int/europe/publications/i/item/WHO-EURO-2025-11788-51560-78769> (Accessed: February 2026).

12.5. Proposed research project 5: Socioenvironmental determinants of subjective well-being in children and adolescents (leading beneficiary: ISGLOBAL)

Background

Mental health problems that emerge during childhood or adolescence place a significant burden on public health, resulting in long-term consequences for educational attainment, social functioning and adult health (Schlack et al., 2021; Wang et al., 2025). There is increasing evidence that socioenvironmental exposures, including air pollution, noise, light at night, temperature, green and blue spaces, and neighborhood characteristics, play an important role in shaping neurobehavioural and emotional development during sensitive periods of development (Polemiti et al., 2024; Dadvand et al., 2015; Domínguez et al., 2024; Zubizarreta-Arruti et al., 2025). However, most existing studies have examined these exposures in isolation, despite their co-occurrence and complex interrelationships.

Within the exposome framework, this project aims to characterize the combined and independent effects of multiple socioenvironmental exposures on child and adolescent mental health, as assessed using Strengths and Difficulties Questionnaire (SDQ)–derived outcomes. The SDQ provides a well-validated, multidimensional measure of emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior, making it particularly suitable for investigating heterogeneous mental health profiles across developmental stages.

This research will address key gaps by: i) jointly modeling multiple environmental and social exposures to better reflect real-world exposure patterns; ii) identifying key socioenvironmental determinants in two developmental periods relevant to mental health outcomes; and iii) assessing stratified associations between environmental exposures and emotional and behavioral difficulties.

The study will use harmonised cohort data and advanced, exposome-oriented statistical methods to generate robust evidence on the contribution of socioenvironmental determinants to mental health inequalities in young populations.

Methodology

This project will adopt an exposome-based analytical approach:

Multivariate statistical methods will be employed to address the high dimensionality and correlation structure of exposures and outcomes. In particular, multivariate sparse reduced-rank regression (msRRR) will be applied to simultaneously model multiple exposures and multiple SDQ domains, enabling identification of key intercorrelated exposures associated with distinct emotional and behavioral difficulties.

The models will be adjusted for relevant confounders, and sensitivity analyses will then be used to assess the robustness of the results.

References

Dadvand, P. et al. (2015) 'Green spaces and cognitive development in primary schoolchildren', *Proceedings of the National Academy of Sciences of the United States of America*, 112(26), pp. 7937–7942. Available at: <https://doi.org/10.1073/pnas.1503402112>.

Domínguez, A. et al. (2024) 'Childhood exposure to outdoor air pollution in different microenvironments and cognitive and fine motor function in children from six European cohorts', *Environmental Research*, 247, 118174. Available at: <https://doi.org/10.1016/j.envres.2024.118174>.

Polemiti, E. et al. (2024) 'How does the macroenvironment influence brain and behaviour—a review of current status and future perspectives', *Molecular Psychiatry*, 29(10), pp. 3268–3286. Available at: <https://doi.org/10.1038/s41380-024-02557-x>.

Schlack, R. et al. (2021) 'The effects of mental health problems in childhood and adolescence in young adults: results of the KiGGS cohort', *Journal of Health Monitoring*, 6(4), pp. 3–19. Available at: <https://doi.org/10.25646/8863>.

Wang, Z. et al. (2025) 'Global, regional, and national burden of mental disorders among adolescents and young adults, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021', *Translational Psychiatry*, 15, 397. Available at: <https://doi.org/10.1038/s41398-025-03623-w>.

Zubizarreta-Arruti, U. et al. (2025) 'Associations between air pollution and surrounding greenness with internalizing and externalizing behaviors among schoolchildren', *Child and Adolescent Mental Health*, 30, pp. 149–158. Available at: <https://doi.org/10.1111/camh.12772>.

12.6. Proposed research project 6: Socioenvironmental determinants of social well-being in adolescents (leading beneficiary: IISPV)

Background

Social support is a fundamental determinant of health and well-being across the life course. During childhood and adolescence, supportive relationships with family members and friends play a crucial role in shaping emotional development, psychological resilience, and social functioning. High levels of perceived social support have been associated with better psychological well-being and lower levels of depression among children and adolescents (Chu et al., 2010; Rueger et al., 2016).

Beyond individual and interpersonal characteristics, broader socioenvironmental contexts may shape opportunities for developing supportive social relationships. Neighborhood characteristics and socioeconomic conditions can influence adolescents' social experiences and access to supportive networks within their communities. For example, neighborhood features such as residential greenness has been linked to adolescents' perceived social support, suggesting that local environments can play a role in shaping the availability and perception of social support (Piera Pi-Sunyer et al., 2025). Neighborhood and school socioeconomic disadvantage have been linked to adolescents' depression, with perceived social support moderating these associations. Specifically, higher levels of perceived social support are associated with lower depressive symptoms across levels of neighborhood disadvantage, although the protective effect of social support appears stronger among adolescents living in less disadvantaged neighborhoods (Huang et al., 2020).

Social ties and supportive relationships are embedded within broader social structures, where socioeconomic position and social stratification shape opportunities for forming and maintaining networks of support (Kawachi & Berkman, 2001). Family socioeconomic conditions may influence adolescents' social relationships and wellbeing through mechanisms such as access to resources and parenting stress (Wang & McLeroy, 2023). Economic hardship can also contribute to increased family conflict and strain within households, potentially affecting the quality and stability of supportive relationships during childhood and adolescence.

Building on this perspective, understanding adolescents' social support requires considering the complex interplay of multiple socioenvironmental influences that operate simultaneously across different domains of daily life. Children and adolescents are exposed to a wide range of social, physical, and contextual conditions – from family environments and peer contexts to neighborhood and broader socioeconomic settings – that may jointly shape opportunities for developing and maintaining supportive relationships. Capturing

these interconnected influences requires an integrative framework capable of examining how diverse environmental exposures combine to influence perceived support.

Within the exposome framework, this project aims to investigate how multiple social and physical environmental determinants jointly shape perceived social support among adolescents. By adopting a comprehensive, data-driven approach, the study seeks to identify socioenvironmental contexts associated with higher or lower levels of perceived support across family and peers.

This research will address key gaps by:

- i) jointly examining multiple socioenvironmental determinants of social support;
- ii) identifying environmental domains associated with perceived social support levels; and
- iii) exploring potential social inequalities in perceived support among adolescents.

Methodology

Study Population

Data will be drawn from harmonized European cohorts, including ABCD, BiB, and WALNUTS. These cohorts include adolescents from diverse European urban and regional contexts and contain both cross-sectional and longitudinal data collected across developmental stages. Adolescents of around 13–16 years old will be included in the analyses.

Outcome Assessment: Social Support

Perceived social support will be assessed using the **Multidimensional Scale of Perceived Social Support (MSPSS)** questionnaire. The primary outcome will be the **total MSPSS score**, representing overall perceived social support. Secondary analyses may examine domain-specific support, including family support and peer support.

Exposure Assessment – Social and Physical Exposome

The determinants include harmonized environmental, neighborhood-level, and individual-level characteristics representing multiple dimensions of the social and physical environment. For example:

Physical environment

- Green and blue space availability
- Sport facilities availability
- Air pollution exposure

- Environmental noise
- Urban density and land use characteristics

Social environment

- Family socioeconomic conditions
- Household characteristics
- Psychosocial stressors

These environmental domains may influence opportunities for social interaction, community engagement, and the development of supportive relationships.

Statistical Analysis

To address the high dimensionality and potential correlations among socioenvironmental exposures, a structured analytical approach will be applied.

First, **Elastic Net** will be used for variable selection. This approach allows the identification of key predictors while accounting for correlations between determinants.

Then, **post-selection effects estimation** will be conducted using regression models to estimate the association between selected determinants and perceived social support outcomes. Depending on the distribution of the MSPSS score and the presence of influential observations, **robust regression models** will be applied to improve the stability of estimates and reduce sensitivity to outliers.

All models will be adjusted for relevant individual-level confounders such as cohort, age, sex, and other key demographic characteristics.

To account for heterogeneity across cohorts, cohort-stratification analysis will be performed. Since there are both biological and social differences between males and females, stratifications by sex will also be carried out.

References

Chu, P. S., Saucier, D. A., & Hafner, E. (2010). Meta-analysis of the relationships between social support and well-being in children and adolescents. *Journal of Social and Clinical Psychology*, 29(6), 624–645.

- Huang, Y., Edwards, J., & Laurel-Wilson, M. (2020). The shadow of context: Neighborhood and school socioeconomic disadvantage, perceived social integration, and the mental and behavioral health of adolescents. *Health & Place*, 66, 102425.
- Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban Health*, 78(3), 458–467.
- Piera Pi-Sunyer, B., Bignardi, G., García-Baquero, G., Dadvand, P., Vrijheid, M., Guxens, M., Esplugues, A., Delgado-Saborit, J. M., Ibarluzea, J., Blakemore, S.-J., & Subiza-Pérez, M. (2025). Investigating the relationship between neighbourhood characteristics, perceived social support and psychological wellbeing in Spanish adolescents. *Scientific Reports*, 15, Article 38845.
- Rueger, S. Y., Malecki, C. K., Pyun, Y., Ayccock, C., & Coyle, S. (2016). A meta-analytic review of the association between perceived social support and depression in childhood and adolescence. *Psychological Bulletin*, 142(10), 1017–1067.
- Wang, Y. C., & McLeroy, A. M. (2023). Poverty, parenting stress, and adolescent mental health: The protective role of school connectedness reexamined. *Children and Youth Services Review*, 153, 107127.