Environment International 173 (2023) 107864

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Full length article

Modelling socioeconomic position as a driver of the exposome in the first 18 months of life of the NINFEA birth cohort children

Chiara Moccia[a,*], Costanza Pizzi[a], Giovenale Moirano[a], Maja Popovic[a], Daniela Zugna[a], Antonio d'Errico[a], Elena Isaevska[a], Serena Fossati[b], Mark J. Nieuwenhuijsen[b], Piero Fariselli[c], Tiziana Sanavia[c], Lorenzo Richiardi[a], Milena Maule[a]

[a] *Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO Piemonte, Turin, Italy*
[b] *ISGlobal (Barcelona Institute for Global Health), Barcelona, Spain*
[c] *Department of Medical Sciences, University of Turin, Turin, Italy*

ARTICLE INFO

Handling Editor: Shoji Nakayama

Keywords:
Exposome
Socioeconomic position
Life course epidemiology
Health inequalities
Environmental epidemiology

ABSTRACT

Background: The exposome drivers are less studied than its consequences but may be crucial in identifying population subgroups with unfavourable exposures.
Objectives: We used three approaches to study the socioeconomic position (SEP) as a driver of the early-life exposome in Turin children of the NINFEA cohort (Italy).
Methods: Forty-two environmental exposures, collected at 18 months of age (N = 1989), were classified in 5 groups (lifestyle, diet, meteoclimatic, traffic-related, built environment).
We performed cluster analysis to identify subjects sharing similar exposures, and intra-exposome-group Principal Component Analysis (PCA) to reduce the dimensionality. SEP at childbirth was measured through the Equivalised Household Income Indicator.
SEP-exposome association was evaluated using: 1) an Exposome Wide Association Study (ExWAS), a one-exposure (SEP) one-outcome (exposome) approach; 2) multinomial regression of cluster membership on SEP; 3) regressions of each intra-exposome-group PC on SEP.
Results: In the ExWAS, medium/low SEP children were more exposed to greenness, pet ownership, passive smoking, TV screen and sugar; less exposed to NO_2, NO_X, PM_{25abs}, humidity, built environment, traffic load, unhealthy food facilities, fruit, vegetables, eggs, grain products, and childcare than high SEP children.
Medium/low SEP children were more likely to belong to a cluster with poor diet, less air pollution, and to live in the suburbs than high SEP children.
Medium/low SEP children were more exposed to lifestyle PC1 (unhealthy lifestyle) and diet PC2 (unhealthy diet), and less exposed to PC1s of the built environment (urbanization factors), diet (mixed diet), and traffic (air pollution) than high SEP children.
Conclusions: The three approaches provided consistent and complementary results, suggesting that children with lower SEP are less exposed to urbanization factors and more exposed to unhealthy lifestyles and diet. The simplest method, the ExWAS, conveys most of the information and is more replicable in other populations. Clustering and PCA may facilitate results interpretation and communication.

1. Introduction

Every individual experiences a multitude of exposures from different sources, beginning before birth and accumulating throughout life. The human exposome provides a representation of this longstanding exposure history (CDC - Exposome and Exposomics - NIOSH). The exposome concept considers three broad intertwined and overlaying domains: 1- the general external domain, including factors such as urban–rural environment, climate factors and societal factors, 2- the specific external domain, including diet and lifestyles assessed at the individual level, and 3- the internal domain, that encompasses metabolic processes, stress responses, and ageing (Wild, 2012; Wild, 2005). Several exposome-

* Corresponding author. Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin, Via Santena 7, 10126 Turin, Italy.
E-mail address: chiara.moccia@unito.it (C. Moccia).

https://doi.org/10.1016/j.envint.2023.107864
Received 23 June 2022; Received in revised form 2 March 2023; Accepted 2 March 2023
Available online 7 March 2023

related research initiatives aim to combine all environmental hazards in order to study how they relate with adverse outcomes (Siroux et al., 2016). Pregnancy and birth cohorts, in particular, offer a unique opportunity to perform exposome studies in highly susceptible periods of life, such as the early infancy.

In this context, understanding how an unequal distribution of environmental exposures creates specific population subgroups that share similar exposome patterns is crucial. Different factors can have a role in creating population subgroups with unfavourable exposures. Better insight into such factors, hereafter referred to as "drivers of the exposome", could suggest preventive strategies and decrease disease burden. The concept of environmental inequality (Brulle and Pellow, 2006) puts forward the socioeconomic position (SEP), educational level and area of residence as exposome drivers that may create differential exposure to environmental risks in the population (Ganzleben and Kazmierczak, 2020), although the hypothesis that more disadvantaged are more exposed to environmental hazards does not always hold (Vrijheid et al., 2012). Patterns of environmental inequalities are likely to be uneven in different communities and geographical areas. In North America, low-socioeconomic communities seem to experience higher air pollutants concentration, while this association is different elsewhere. In Europe (Hajat et al., 2015), it has been found that polluted central areas are mostly inhabited by wealthy communities in some cities and by disadvantaged people in others (Unequal exposure and unequal impacts — European Environment Agency). The Human Early Life Exposome (HELIX) (Vrijheid et al., 2014) project observed that the social determinants of the urban exposome of European pregnant women differed in different cities (Robinson et al., 2018). Air pollutants levels were higher among pregnant women of low SEP in Bradford, Nancy, and Valencia, whereas the opposite was observed in Oslo, Poitiers, and Sabadell. Similarly, high levels of chemical contaminants have been reported both in low or high socioeconomic groups (Montazeri et al., 2019).

Another study on the Spanish population-based birth cohort INMA found generally weak and inconsistent associations between social and educational classes and environmental pollutants measured in maternal serum during pregnancy and in cord blood. Moreover, social class explained less than 5% of the variability in pollutants concentrations (Vrijheid et al., 2012). A positive association between belonging to a higher level of socio-economic position and consumption of fruits, vegetables, white meat, fish and eggs was found in children participants to the Portuguese National Food, Nutrition and Physical Activity Survey (Correia et al., 2020).

In the general exposome research, studies may include all the three exposome domains together (Haddad et al., 2019). Because each specific group of components carries a biological fingerprint in relation to a health outcome, the integration of all the exposome domains may provide a broader perspective and could help identifying specific groups of components that may be under-studied (Haddad et al., 2019).

In the context of the study of exposome, interest is more focused on the association with modifiable factors, with the goal of identifying components on which to intervene to reduce exposome disparities. The socio-exposome (Senier et al., 2017) concept has been proposed to serve as a bridge between eco-social theories (Krieger, 2001) and the concept of the exposome (Wild, 2012) and, within this framework, different subgroups of the exposome have been analysed, such as air pollutants, environmental chemicals or urban environment (Vrijheid et al., 2012; Robinson et al., 2018; Montazeri et al., 2019). Recently, Sum and colleagues (Sum et al., 2022) implemented an exposome approach to investigate the association between 8 SEP indicators and 134 exogenous and endogenous exposures measured during pregnancy in a setting with no clear SEP geographical segregation. In addition to specific associations (such as the relevant role of paternal variables), their work highlights how SEP-exposome relationships are complex, non-linear and context-specific, and how they may be modified by ethnicity and nativity (Sum et al., 2022).

The joint analysis of components from different exposome domains at the same time provides a broad perspective and involves many technical challenges (Santos et al., 2020). These include the heterogeneity of a large number of variables, with some exposome subgroups being more homogeneous and correlated (e.g. air pollutants) and others including dissimilar, less correlated, variables (e.g. lifestyles), and the need to integrate different data sources, implying different measurement scales and precision (Santos et al., 2020). To address the high dimensionality and to characterise exposome patterns, data-driven dimensionality reduction techniques (such as Principal Component Analysis (PCA) and factor analysis) and methods for grouping of observation (like cluster analysis) have been proposed (Santos et al., 2020; Stafoggia et al., 2017). When studied in relation to the drivers of the exposome, the patterns derived from the aforementioned techniques can improve both data synthesis and interpretation, two limits of traditional Exposome Wide Association Study (ExWAS), where a separate regression model of each exposome variable on the driver is fitted. Although previous SEP-exposome papers utilized some of the approaches, most of them focused on PCA, leaving out other possible techniques (Robinson et al., 2018; Montazeri et al., 2019). Furthermore, the use of a global PCA when the set of variables is high-dimensional and heterogeneous may not be advantageous for the interpretation of results. This study aims to fill the gap in the SEP-exposome literature by applying and comparing three different and complementary approaches to analyse the relationship between SEP at birth, measured through a standardised indicator of material resources, as a potential driver of the exposome. It focuses on the characterization of the exposome, with particular attention to the general and specific external exposome domains, in 18-month-old children of the NINFEA birth cohort, living in the city of Turin, Italy.

2. Methods

2.1. Study population

The NINFEA study (https://www.progettoninfea.it) is an Italian web-based multi-purpose mother–child cohort aimed at exploring the relationships between exposures starting in early-life and long-term health outcomes (Richiardi et al., 2007). Approximately 7500 pregnant women participated in the study between 2005 and 2016. Mothers were asked to complete the baseline questionnaire during pregnancy and invited to respond to other seven follow-up questionnaires, at 6 and 18 months after delivery, and when their children turn 4, 7, 10, 13, and 16 years of age. The NINFEA study was approved by Ethical Committee of the San Giovanni Battista and CTO/CRF/Maria Adelaide Hospital of Turin, and all the participating mothers gave informed consent at enrolment and at each follow-up. Details on the cohort have been published before (Firestone et al., 2015; Blumenberg et al., 2018).

The current study population included 2289 children living in the area of Turin, whose mothers completed the NINFEA 18-month questionnaire. Turin is a metropolitan city in the Piedmont region, in north-western Italy. The city's population is approximately 900,000 residents, with the most densely populated area in the city centre (10617 pop/km^2), partly surrounded by hills (1853 pop/km^2) and partly surrounded by suburbs with a relatively high level of industrialization (3667pop/km^2). Supp. Fig. 1 shows the distribution of the degree of urbanisation (Degree of urbanisation classification) of the territory of Turin municipality, with 90% of the surface of the area classified as an urban centre.

Out of 2289 eligible children, 1989 children with complete information on the driver (SEP) and on variables potentially influencing participation in the NINFEA study (namely maternal age, maternal parity and maternal country of birth, as explained below) were included in the analyses.

2.2. SEP assessment

SEP at childbirth was measured through the EHII (Equivalised Household Income Indicator), a standardised cross-cohort comparable indicator of material resources. The EHII measures the total disposable monthly household income, standardised for the household size and composition (Pizzi et al., 2020). To overcome the difficulties of measuring disposable income through a questionnaire, the EHII was developed within the LifeCycle project (Jaddoe et al., 2020), using external data from the 2011 "pan-European Union Statistics on Income and Living Conditions" (EUSILC) survey, and individual data from the cohorts (parental age, cohabitation status, educational level, country of birth, occupation, house size and type, and family size). For this study, the main analyses are based on the tertiles of the distribution of the equivalised total disposable household income from the 2011 Italian EUSILC population used to derive the EHII (i.e. all households with at least one child (16-years old or younger) and his/her mother, and excluding households with 8 or more members or with very atypical structure, such as two or more family units living together). Participants with an EHII in the first two tertiles were classified as having a medium/low SEP, while those with an EHII in the upper tertile as having a high SEP (Supp. Fig. 2). Medium and low SEP children were grouped together because of the low proportion of children with a low SEP in the cohort (3.2%) (Table 2). We performed additional analyses using deciles of the internal EHII distribution.

2.3. Children exposome assessment

Early life exposome variables were selected a priori and classified in five subgroups according to the exposome domains of Wild's classification (Wild, 2012). Thus, the 18-month exposome considered here consists of lifestyle and diet components for the specific external domain, and meteoclimatic, traffic-related, and built environment components for the general external domain.

The lifestyle group includes 4 variables: exposure to passive smoking, pet ownership, TV screen time (1h30min per day or more), and childcare attendance. All lifestyle variables were binary, and were obtained from questionnaires. If a variable was measured at two points in time in the first 18 months of age (the first 6 months and from 6 to 18 months of age), we combined the two measures and defined the variable as ever/never in the first 18 months.

The diet group includes 13 variables derived from questionnaires. Two are dichotomous: ever breastfed in the first 18 months of age (assessed at 6 and 18 months), and introduction of solid food at 4 months of age or later (assessed at 6 months). The other 11 variables were assessed at 18 months of age only, and included the consumption of vegetables, fruit, dairy products, fish, meat, eggs, grain products, potatoes, sweet beverages and sugar products, measured in servings per day.

The last three exposome groups, describing various aspects of the urban environment, are based on the geo-coded residential address at the age of 18 months, and were generated within the Lifecycle project (de Castro Pascual et al., 2021). The meteoclimatic group includes 4 continuous variables: humidity percentage, temperature, UV irradiance DNA damaging dose, and land surface temperature. The traffic-related group includes 9 continuous variables: NO_2, NO_X, PM_{10}, $PM_{2.5}$, PM_{coarse}, $PM_{absorbance}$, night noise level (obtained using European road traffic noise maps), inverse distance to the nearest road, and traffic load of major roads. The built environment group includes 12 continuous variables: building density, facility richness (the number of different facility types - like restaurants, shops, medical centres, schools, libraries -, divided by the maximum potential number of facility types), connectivity density (number of intersections that are not dead-ends), Shannon index (a quantitative measure of land-use diversity), walkability, blue spaces, green spaces, population density, number of facilities related to unhealthy food, normalised difference vegetation index (NDVI, a measure of greenness), length of public bus lines, and number of bus stops.

Urban time-varying factors, such as exposure to air pollutants and meteorological parameters, were available as daily mean and were averaged to obtain a unique value for different time intervals, such as pregnancy, 1st and 2nd year of life (de Castro Pascual et al., 2021). The value for the second year of life was used as a proxy for the exposure at 18 months of age.

2.4. Statistical analysis

Continuous variable transformations (squared root or logarithmic) were applied to approach normality (details are shown in Table 1). If normality was not achieved, variables were dichotomised using the median of their distribution as the cut-off. Continuous exposome variables were first mean centred and scaled by standard deviation (SD).

All the analyses were adjusted for the following covariates, potentially influencing participation in the NINFEA study (Pizzi et al., 2012): maternal age at delivery (continuous, years), maternal parity at the delivery of the index child (0, 1, 2 +) and maternal nativity (born in Italy vs. born outside Italy).

A graphical representation of the overall analysis plan, described in detail in the following sections, is shown in Fig. 1.

2.4.1. Exposome characterization

2.4.1.1. Correlation matrix of variables. We estimated the pairwise correlations of the entire exposome dataset, showing the results with a heat map. Pearson, polychoric or polyserial correlations were calculated between all variable pairs using the *hetcor* function in the polycor R package (Fox, 2021).

2.4.1.2. Cluster analysis. We performed a cluster analysis to group subjects sharing a similar exposome pattern. Spectral clustering (Von Luxburg, 2007), which is a graph-based algorithm that can detect clusters of arbitrary shapes, such as non-convex clusters, was applied using the *mspec* function in the SpectralClMixed R package that accounts for mixed data (continuous and categorical) (Mbuga and Tortora, 2021). We imposed a priori a minimum number of three clusters, to avoid reducing the complexity and variety of the exposome into a dichotomous variable. The within-cluster sum of squares and the between-divided by the total within- clusters sum of squares were then used as internal validation indexes to estimate the optimal number of clusters. All subjects with missing values in at least one exposome variable were excluded.

2.4.1.3. Intra-group principal component analysis. We performed a Principal Component Analysis (PCA) to derive a summary indicator describing each exposome group (lifestyle, diet, meteoclimatic factors, traffic-related characteristics and built environment). We used the *prcomp* function in the stats R package for the PCA in the meteoclimatic group to handle continuous variables and the *FAMD* function in the Factoshiny R package to perform a Factor Analysis in families with mixed data (Pagès, 2004). We excluded observations with missing values within each group before performing the PCA, leading to a different number of subjects in each group. We retained the first principal components (PCs) that explained at least 20% of the variance for each exposome group. To illustrate how the original variables contribute to different PCs, we estimated the correlation of each exposome group with its retained PCs in heat maps, using the *hetcor* function in the polycor R package (Fox, 2021).

2.4.2. Approaches to analyse the association between SEP and the exposome

Three analytical approaches were used to evaluate the association

Table 1
Descriptive table of the variables included in the analyses that were chosen to represent the children exposome in the NINFEA population under study (N = 1989).

N = 1989	Exposome family	(N) and total % of missing in the exposome family	Applied transformation	Exposome variable	Description	N or mean[b]	% or SD[b]	% of Missing	Assessment method	Variable type	Time frame
1	Lifestyle	(26) 1.3%	–	pets	Furry pet ownership	664	33.4%	1.3%	questionnaire	binary	6–18 months
2			–	smoke	Passive smoking	306	15.4%	2.8%	questionnaire	binary	6–18 months
3			–	childcare	Child attending a day care centre within the first 18 months of life	950	47.8%	7.9%	questionnaire	binary	6–18 months
4			–	tv	Child's tv watching duration > 1h30min	298	15.0%	7.8%	questionnaire	binary	18 months
5	Diet	(13) 0.7%	–	breastfeeding	Child ever breastfed	1811	91.1%	1.3%	questionnaire	binary	6–18 months
6			sqrt	vegetables	Vegetables without potatoes[a]	1.12	0.33	7.8%	questionnaire	continuous	18 months
7			sqrt	fruit	Fruit[a]	1.14	0.38	7.1%	questionnaire	continuous	18 months
8			sqrt	dairy	Milk and milk products[a]	1.44	0.33	6.8%	questionnaire	continuous	18 months
9			sqrt	fish	Fish and fish products[a]	0.51	0.19	7.1%	questionnaire	continuous	18 months
10			sqrt	meat	Meat and meat products[a]	0.91	0.26	7.8%	questionnaire	continuous	18 months
11			sqrt	eggs	Egg and egg products[a]	0.39	0.14	8.0%	questionnaire	continuous	18 months
12			sqrt	grain	Grains and grain products[a]	1.87	0.35	7.3%	questionnaire	continuous	18 months
13			sqrt	pulses	Legumes, nuts and their products[a]	0.53	0.24	8.6%	questionnaire	continuous	18 months
14			sqrt	potatoes	Potatoes[a]	0.56	0.25	7.4%	questionnaire	continuous	18 months
15			dic	sweet beverages	Sugar-sweetened beverages[a]	1145	57.6%	7.4%	questionnaire	binary	18 months
16			dic	sugar products	Sugar, sugar products, chocolate products and confectionery[a]	873	43.9%	8.0%	questionnaire	binary	18 months
17			dic	solid food	Solid food introduction ≥ 4 months of age	1673	84.1%	8.1%	questionnaire	binary	6 months
18	Meteoclimatic	(163) 8.2%	–	humidity	% Humidity	69.46	2.12	8.1%	Local meteorologic stations	continuous	2nd year of life
19			–	temperature	Annual average of daily mean temperature, °C	13.0	0.49	8.1%	Local meteorologic stations	continuous	2nd year of life
20			–	UV	Annual average of daily mean UV irradiance DNA damaging dose, kJ/m^2	1.04	0.05	8.1%	Satellite based observation	continuous	2nd year of life
21			–	surface temperature	Annual average of daily mean land surface temperature, °C	22.20	1.69	8.4%	Satellite based observations	continuous	2nd year of life
22	Traffic	(160) 8.0%	–	NO_2	NO_2, $\mu g/m^3$	47.87	9.87	8.1%	Escape LUR model[c]	continuous	2nd year of life
23			log	NO_x	NO_X, $\mu g/m^3$	4.42	0.24	8.1%	Escape LUR model[c]	continuous	2nd year of life
24			–	PM_{10}	PM_{10}, $\mu g/m^3$	40.17	6.55	8.1%	Escape LUR model[c]	continuous	2nd year of life
25			–	PM_{25}	$PM_{2.5}$, $\mu g/m^3$	23.91	2.95	8.1%	Escape LUR model[c]	continuous	2nd year of life
26			–	PM_{coarse}	PM coarse, $\mu g/m^3$	14.76	3.20	8.1%	Escape LUR model[c]	continuous	2nd year of life

(*continued on next page*)

Table 1 (*continued*)

N = 1989	Exposome family	(N) and total % of missing in the exposome family	Applied transformation	Exposome variable	Description	N or mean[b]	% or SD[b]	% of Missing	Assessment method	Variable type	Time frame
27			–	$PM_{25absorbance}$	PM 25 absorbance, μg/m^3	2.81	0.53	8.1%	Escape LUR model[c]	continuous	2nd year of life
28			–	night noise	Night noise level, dB(A)	55.34	5.39	8.1%	EC Directive 2002/49/ EC Noise Maps[d]	continuous	2nd year of life
29			log	distance road	Inverse distance to the nearest road (m^{-1})	−2.42	0.89	8.1%	Open StreetMaps[e]	continuous	2nd year of life
30			dic	traffic major load	Traffic load of major roads within a buffer of 100 m, 1000 veh/d m	917	46.1%	8.1%	Local traffic monitroing maps	binary	2nd year of life
31	Built environment	(160) 8.0%	–	buldings	Building density (area of building cover (km^2) / area of buffer (km^2) within buffer of 300 m ± SD	370681.9	109772.3	8.0%	European Settlement Map 2017[f]	continuous	2nd year of life
32			sqrt	facilities	Number of different facility types divided by the maximum potential number of facility types within a 300 m buffer	0.35	0.11	8.0%	NAVTEQ 2012[g]	continuous	2nd year of life
33			–	connectivity	Connectivity density (number of intersections/ km^2) within a buffer of 300 m ± SD	176.52	68.21	8.0%	NAVTEQ 2012[g]	continuous	2nd year of life
34			–	shannon	Land use SEI within 300 m buffer ± SD	0.44	0.11	8.0%	Urban Atlas[h]	continuous	2nd year of life
35			–	walkability	Walkability within 300 m buffer ± SD	0.38	0.05	8.0%	Urban Atlas/ NAVTEQ 2012[g] / Global Human Settlement Map[i]	continuous	2nd year of life
36			–	blue spaces	Is there blue space > 5,000 m^2 within 300 m buffer?	226	11.4%	8.0%	Urban Atlas[h]	continuous	2nd year of life
37			–	green spaces	Is there green space > 5,000 m^2 within 300 m buffer?	1446	72.7%	8.0%	UrbanAtlas[h]	continuous	2nd year of life
38			dic	population	Population density, Inhabitants/km^2 ± SD	1237	62.2%	8.3%	Global Human Settlement Map[i]	binary	2nd year of life
39			dic	unhealthy food facilities	number of facilities related to unhealthy food divided by the area of the 300 m buffer	1109	55.8%	8.0%	NAVTEQ 2012[g]	binary	2nd year of life
40			log	NDVI	NDVI values within a buffer of 300 m ± SD	−1.47	0.30	8.0 %	Satellite based observations	continuous	2nd year of life
41			sqrt	bus lines	Length of public bus lines within 300 m buffer ± SD	77.35	18.21	8.7%	Google Transit Feeds[l]	continuous	2nd year of life
42			sqrt	bus stops	Bus stops within 300 m buffer ± SD	5.50	1.37	8.6%	Google Transit Feeds[l]	continuous	2nd year of life

Abbreviations: PM, particulate matter; NO, nitrogen oxides; UV, ultraviolet; NDVI, normalized difference vegetation index.

[a] Servings-per-day.

[b] Statistics are evaluated after applied transformations, i.e. squared root or log-transformation, to approach normality. If normality was not achieved, variables were dichotomized using the median of their distribution as the cut-off. The 4th column in Table 1 reports the transformation chosen for each variable.

[c] Escape LUR model: http://www.escapeproject.eu/manuals/ESCAPE_Exposure-manualv9.pdf.

[d] EC Directive 2002/49/EC Noise Maps: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:189:0012:0025:EN:PDF.

[e] Open StreetMaps: https://www.openstreetmap.org/.

[f] European Settlement Map 2017 https://land.copernicus.eu/pan-european/GHSL/european-settlement-map.

[l] Google Transit Feeds https://developers.google.com/transit/gtfs.

[h] Urban Atlas https://land.copernicus.eu/local/urban-atlas.

[g] NAVTEQ 2012: https://www.gbcnet.net/archive/index.php/t-52135.html.

[i] Global Human Settlement Map https://ghsl.jrc.ec.europa.eu/.

Table 2
Descriptive table of the driver (SEP) and the main adjustment covariates used in the analyses for the NINFEA population under study (N = 1989).

N = 1989	Name	Description	N (%) or Mean (SD)	Median (IQR)
Exposure	EHII	Indicator of log-equivalised total disposable monthly household income	7.45 (0.23)	7.49 (0.30)
	SEP			
		low	64 (3.2%)	
		medium	508 (25.5%)	
		high	1417 (71.2%)	
Adjustment covariates	maternal age at delivery	Mean maternal age at childbirth, years	33.5 (4.1)	
	maternal parity at the delivery of the index child	Maternal parity for the index child		
		0	1521 (76.5%)	
		1	390 (19.6%)	
		2+	78 (3.9%)	
	maternal nativity	Mother's born outside Italy	103 (5.2%)	

between SEP and the exposome, in which SEP was always used as the independent variable, and the exposome as the outcome. The three approaches differed on how the exposome variables were treated: (1) separate regression models for each single exposome variable; (2) multinomial regression model of the exposome cluster membership; and (3) linear regression models of selected intra-exposome-group PCs. For all analyses we considered the conventional p-value cut-off of 0.05, and for the first approach we included the adjusted p-value after the correction for multiple testing using Benjamini and Hochberg false discovery rate (FDR).

(1) Exposome Wide Association Study (ExWAS)

With a "one exposure – one outcome" approach, we fitted a separate regression model of each exposome variable, the outcome, on the SEP binary indicator (medium/low vs. high), the independent variable, adjusting for confounders. Typically, the exposome-wide association studies (ExWAS (Juarez et al., 2014)) aim at evaluating the association between the exposome and a health outcome by considering, one at a time, the relationship between an exposure variable and the outcome. In the ExWAS approach proposed here, the association goes from the SEP (the driver) to the exposome variables (the outcome). Linear regressions were fitted for continuous outcomes, and logistic regressions for dichotomous outcomes. P-values adjusted for multiple comparisons were calculated using the Benjamini and Hochberg false discovery rate (FDR). We drew two distinct volcano plots for continuous and dichotomous variables to visualise the associations.

As a sensitivity analysis, the ExWAS was also carried out using the SEP as a continuous exposure, based on the deciles of the internal NINFEA EHII distribution.

(2) Association between SEP and cluster membership

Cluster analysis groups subjects that share similar exposome patterns. Thus, cluster membership was used as a dependent variable in a multinomial logistic regression model to explore the association with medium/low SEP group, with the high SEP group as the reference. We used the *multinom* function of the nnet R package (Venables and Ripley, 2002).

(3) Association between SEP and intra-group PCs

To study the association between SEP and the exposome groups (lifestyle, diet, meteoclimatic factors, traffic-related characteristics and built environment), each PC was regressed on SEP in separate models.

3. Results

A total of 1989 children with 42 early life exposome variables at 18 months of age were included in this study (Table 1 and Table 2). The proportion of children with medium/low SEP in the study population was 29% (Supp. Fig. 2), mostly consisting of participants with medium SEP. Table 2 shows that only 3.2% of participant had a SEP in the lowest tertile of the Italian equivalised total household disposable income distribution. Most of the mothers were nulliparous (77%), and only 5% of mothers were born outside Italy (Table 2). The distribution of the exposome variables is reported in Table 1.

3.1. Exposome characterization

3.1.1. Correlation matrix

Fig. 2 shows the pairwise correlation between the 42 exposome variables. Overall, the correlation coefficients were stronger within than between exposome groups, particularly within the traffic-related variables. A positive correlation pattern was present in the food consumption group, even though none of the correlation coefficients exceeds 0.35. Air pollutants exhibited positive correlation coefficients with urbanization factors (0.02 to 0.36), particularly evident for NO_2 (0.34 for buildings, 0.36 for facilities and 0.28 for connectivity) and $PM_{25absorbance}$ (0.27 for buildings, 0.30 for facilities, 0.22 for connectivity), and negative correlations with blue (−0.58 to −0.09) and green spaces (−0.23 to 0.05), Shannon index (−0.39 to −0.04), and NDVI (−0.34 to −0.10).

Air pollutants correlated positively with DNA-damage UV (0.16 to 0.28) and negatively with both humidity (−0.10 to −0.34) and temperature (−0.18 to −0.24). We observed a small positive correlation between passive smoking and pet ownership (0.26) and an inverse correlation between childcare attendance and TV screen exposure in the lifestyle group (−0.21).

3.1.2. Cluster analysis

The cluster analysis was based on the complete case approach including 1440 children (72.4% of the study population). The optimal sum of squares suggested two as the optimal number of clusters (Supp. Fig. 3), less than the minimum number of three clusters that we had decided a priori to avoid reducing the exposome to a binary variable. As the next best choice was three, three clusters were used in the analysis.

Fig. 3A and B describe the three different groups of subjects according to their exposome patterns obtained from the cluster analysis. The plot in Fig. 3A compares the mean level in each cluster for continuous variables, the normalised mean of the whole study population being 0. The plot in Fig. 3B shows the prevalence of binary variables in each cluster after subtracting the prevalence in the study population.

Cluster 2 and, even more so, cluster 1 are characterised by high urbanization levels (building density, facilities, connectivity), below average Shannon and NDVI indexes, above average traffic from major roads, bus lines and bus stops, population density and unhealthy food facilities, suggesting that, given the characteristics of the city of Turin, subjects in cluster 1 and 2 likely live in central, well served areas of the city. Air pollution levels were however different: higher than average for cluster 2, suggesting a residence close to the city's main arteries, and lower than average for cluster 1, suggesting a residence in a green,

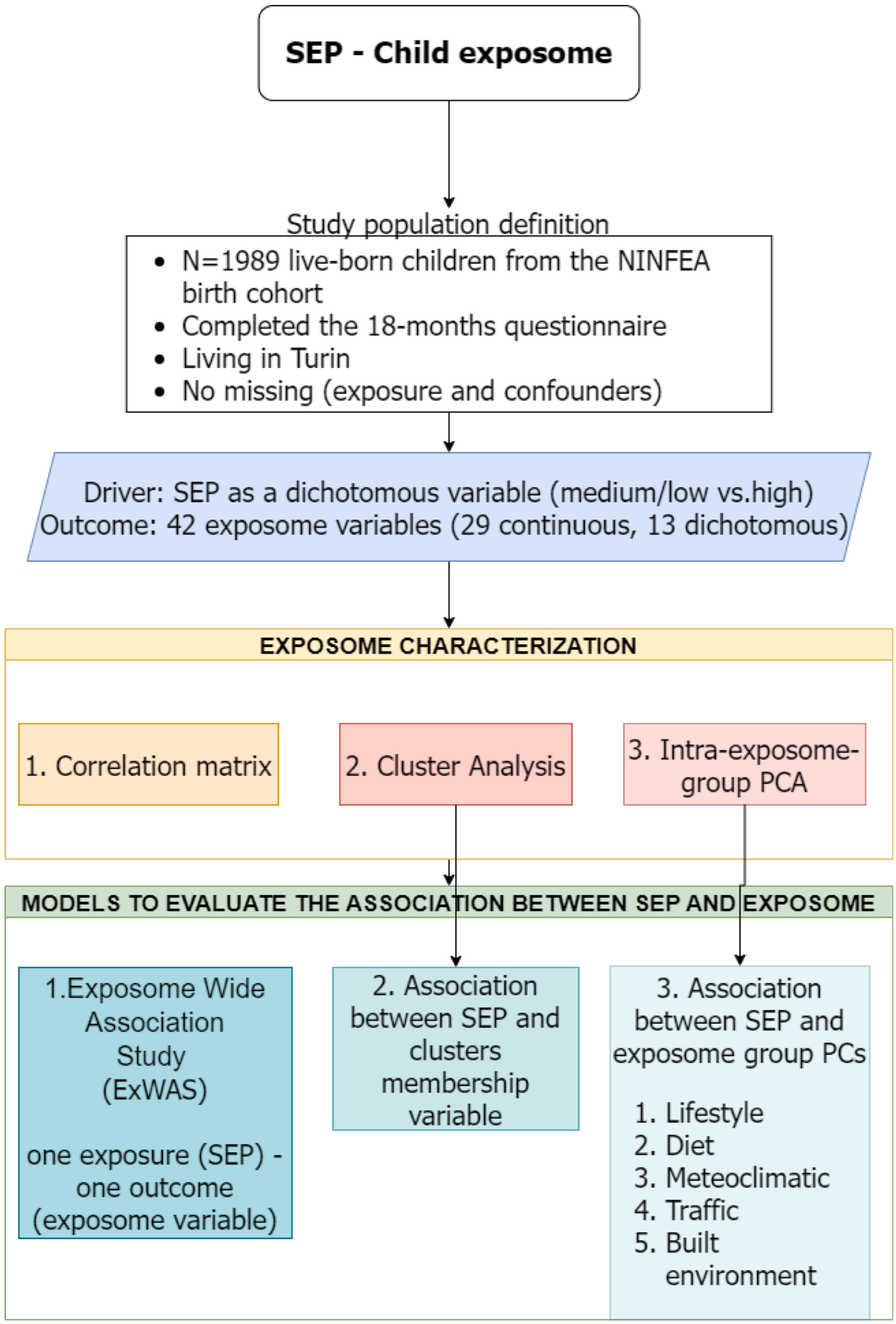

Fig. 1. Workflow of the overall analysis plan.

central and service-rich area. Cluster 3 identifies children with lower than average levels of air pollutants and urbanisation factors (buildings, facilities, connectivity). The urbanization pattern of cluster 3, rich in green spaces, less polluted, less urbanised and less populated, with less traffic and unhealthy food facilities, suggests a more suburban residential area. This interpretation is supported by the geographical distribution of children's residential addresses in the three clusters (Fig. 4).

The differences in diet are smaller than those for the variables representing urbanisation and pollution. Cluster 1 appears to capture the healthiest diet, with above average fruit, vegetables and below average dairy products, fish, meat, grain products, potatoes, sugar products and sweet beverages. Cluster 2 and 3 are characterised by lower than average consumption of vegetables, fruit, pulses, grain products and potatoes, higher than average consumption of dairy products, fish, meat, eggs and higher consumption of sweet beverages but lower consumption of sugar products.

Regarding lifestyle factors, individuals in cluster 3 were more frequently pet owners, more exposed to TV screens, less likely to be exposed to passive smoking and less likely to have attended childcare before the age of 18 months. Children in cluster 1 were more likely to

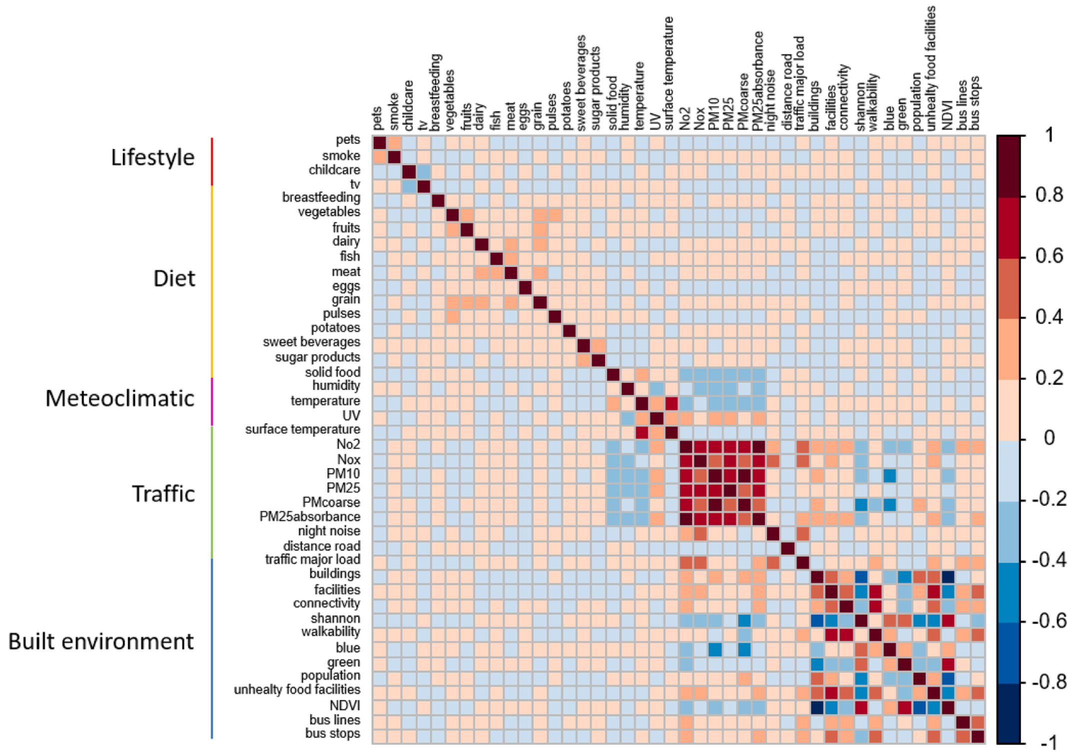

Fig. 2. Heat map showing positive (red) and negative (blue) pairwise correlations (Pearson's, polychoric or polyserial correlation, depending on the type of the variables) between the variables chosen to represent the children exposome. See Table 1 for short names of the exposures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

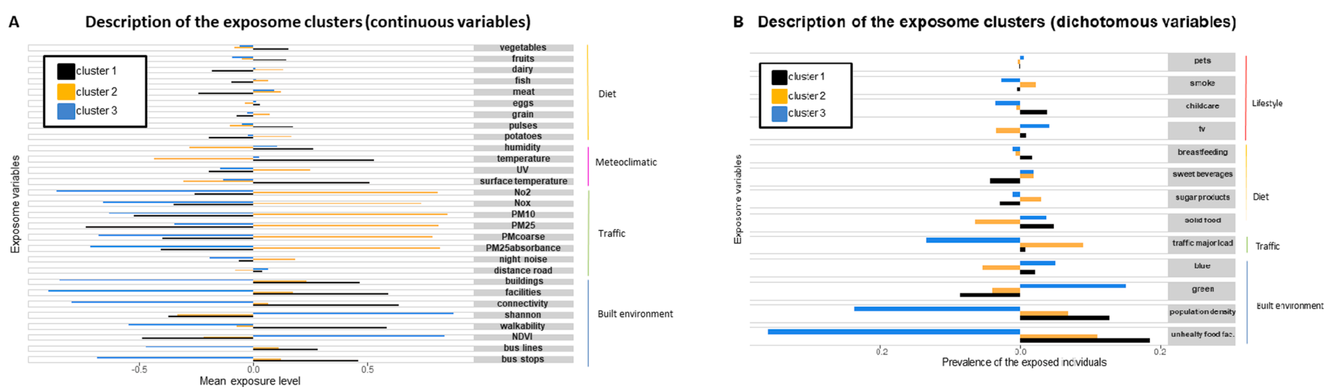

Fig. 3. Description of the three exposome clusters (N = 1440). Clusters are identified by different colors (black, orange, blue). In A, the length of the bar represents the mean variable level within the cluster, the overall mean being 0. In B, the length of the bar represents the prevalence of exposed individuals for each exposome variable within the cluster, after subtracting the prevalence of exposed individuals in the whole study population. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

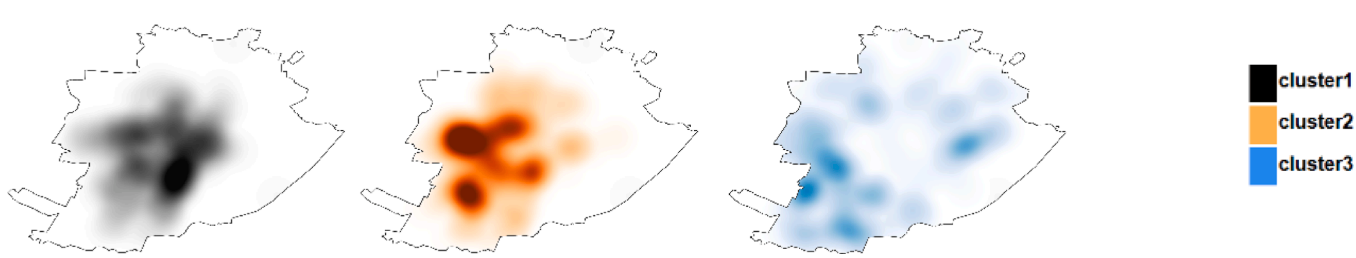

Fig. 4. Map showing the kernel density of geocoded addresses of residence of NINFEA children in the Turin city area, divided in the three clusters (black = cluster 1, orange = cluster 2, blue = cluster 3) obtained from the cluster analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

have attended childcare and were more exposed to TV screens, while children in cluster 2 watched less TV and were more exposed to passive smoking.

3.1.3. Intra-group PCA

To explain at least 20% of intra-group variance, we selected the first two PCs for the diet group and only the first PC for the other exposome groups (Table 4). Fig. 5 and Supp. Table 2 show that the first PC in the lifestyle group is highly positively correlated with pet ownership (0.61), exposure to passive smoking (0.64), and TV screen exposure time (0.63), while it shows an inverse correlation with childcare attendance (−0.61), capturing mostly unhealthy lifestyles. The first PC in the diet group is positively correlated with many food items, including healthy (vegetables (0.61) and fruits (0.50)) and unhealthy ones (meat (0.45), sweet beverages (0.20)), while the second PC seems to capture mostly unhealthier dietary habits, with positive correlation with meat (0.53) and inverse correlation with breastfeeding (−0.47), vegetables (−0.43) and pulses (−0.44). The first PC in the meteoclimatic group is highly and inversely correlated with temperature-related variables (surface temperature −0.81, temperature −0.87). The first PC in the traffic group shows a high positive correlation with particulate matters (from 0.81 to 0.86), while the first PC of the built environment group seems to represent highly urbanised areas, showing a high correlation with facility richness (0.81), building density (0.80), connectivity index (0.63) and unhealthy food facilities (0.78) and an inverse correlation with Shannon index (−0.70), green spaces (−0.63) and NDVI (−0.78).

3.2. Association between SEP and exposome

(1) Exposome Wide Association Study

The ExWAS analysis results are shown in Fig. 6A and B, and in Supp. Table 1. Fig. 6A shows the standard deviation (SD) difference in the exposome levels for continuous variables in individuals with a medium/low SEP, compared with those with a high SEP. Odds ratios (OR, medium/low SEP vs. high SEP) are shown for dichotomous variables (Fig. 6B).

With respect to those with high SEP, children with medium/low SEP were exposed to higher levels of Shannon index (β 0.14, CI 0.03;0.24) and NDVI (β 0.12, CI 0.01;0.23), and lower levels of fruit (β −0.19, CI −0.30; −0.09), vegetables (β −0.14, CI −0.24; −0.03), eggs (β −0.17, CI −0.27; −0.06) and grain products consumption (β −0.12, CI −0.23; −0.02), lower levels of NO_2 (β −0.24, CI −0.35; −0.14), NO_x (β −0.13, CI −0.24; −0.03), PM_{25abs} (β −0.12, CI −0.22; −0.01), humidity (β −0.11, CI −0.21; −0.00), and built environment (connectivity index (β –0.22, CI −0.33; −0.12), facilities richness (β −0.20, CI −0.31; −0.10), walkability (β −0.20, CI −0.30; −0.09), building density (β −0.14, CI −0.25; −0.04)).

Fig. 6B shows that children with medium/low SEP were more frequently pet owners (OR 1.57, CI 1.27;1.94), were more exposed to passive smoking (OR 1.62, CI 1.23;2.11), TV screen (OR 1.76, CI 1.34;2.32) and sugar products (OR 1.32, 1.06;1.65) than children with high SEP. Conversely, medium/low SEP was associated with lower exposure to unhealthy food facilities (OR 0.69, CI 0.56;0.86) and traffic load of major roads (OR 0.68, CI 0.55;0.85), and to a lower probability of attending childcare in the first 18 months of age (OR 0.46, CI 0.37;0.57) than children with high SEP.

Table 3
Multinomial logistic regression of cluster membership on SEP. Estimates are adjusted for maternal nativity (born in Italy vs. born outside Italy), maternal age at delivery (years) and maternal parity at the delivery of the index child (0, 1, 2 +).

Adjusted for maternal nativity, age, parity (n = 1440)	**Cluster 2** vs. Cluster 1 OR (95% CI)	**Cluster 3** vs. Cluster 1 OR (95% CI)
Medium/low SEP vs high SEP	1.28 (0.94;1.73)	1.67 (1.22;2.31)

Table 4
Association between SEP and the intra-exposome-group Principal Components. The first column reports the number of subjects in each group after excluding observations with missing values; the second column the intra-group variance explained by each component. The third column shows the estimated coefficients of the linear regression of each PC on SEP (medium/low vs. high) and their 95% confidence intervals. Estimates are adjusted for maternal nativity (born in Italy vs. born outside Italy), maternal age at delivery (years) and maternal parity at the delivery of the index child (0, 1, 2+).

Exposome family	N (% of total population)	% of variance explained by the PC	Adjusted Estimates Coefficients β (95%CI)
PC1 Lifestyle	1816 (91.3)	30.97	0.52 (0.40;0.63)
PC1 Diet	1652 (83.1)	16.67	−0.19 (-0.36;-0.02)
PC2 Diet		10.68	0.16 (0.03;0.30)
PC1 Meteoclimatic	1821 (91.6)	47.93	0.01 (-0.14;0.16)
PC1 Traffic	1825 (91.8)	51.06	−0.29 (-0.52;0.07)
PC1 Built environment	1805 (90.7)	36.15	−0.43 (-0.65;-0.21)

In the sensitivity analysis in which SEP was modelled as a continuous variable (Supp. Fig. 4, Supp. Table 3), findings were fully consistent with the main analyses (with SEP modelled as a dichotomous variable) in terms of the direction of the associations of SEP with all the exposome variables. As expected, effect sizes were greater and the precision of effect estimates was lower when SEP was treated as a dichotomous than a continuous variable. In terms of p-values, the number of outcomes for which nominal and FDR adjusted p-values were below 0.05 was larger when SEP was treated as a dichotomous than as a continuous variable.

(2) Association between SEP and cluster membership

Table 3 shows the results of the association study between SEP and cluster membership. Children with medium/low SEP have higher odds of belonging to cluster 2 (poor diet, highest levels of air pollution, and living in central areas) (OR 1.28; 95%CI 0.94–1.73) and cluster 3 (poor diet, lowest levels of pollution, living in sub-urban greener areas) (OR 1.67; 95%CI 1.22–2.31) compared to children with high SEP. The sensitivity analysis treating SEP as a continuous variable in deciles (Supp. Table 4) showed consistent results.

(3) Association between SEP and intra-group PCs

Table 4 shows the associations of SEP with the six intra-group PCs. Compared to those with high SEP, children with medium/low SEP were more likely to be exposed to higher level of the lifestyle PC1 (“unhealthy” lifestyle) (β = 0.52; 95%CI 0.40–0.63) and diet PC2 (“unhealthy” diet) (β = 0.16; 95%CI 0.03–0.30), and less likely to be exposed to the built environment PC1 (urbanization factors) (β = −0.43; 95%CI −0.65; −0.21), diet PC1 (mixed diet) (β = −0.19; 95%CI −0.36; −0.02), and traffic PC1 (air pollution) (β = −0.29; 95%CI −0.52; 0.07). Consistent results were obtained when SEP was modelled as a continuous variable in deciles (Supp. Table 5).

4. Discussion

In this study, we proposed three different methods to investigate the role of a distal driver on the early-life exposome (18 months of age), within the Turin participants in the Italian NINFEA birth cohort. We explored 42 different environmental and behavioral components of the general and specific external exposome domains, grouped as lifestyle, diet, meteoclimatic factors, traffic-related characteristics, and built

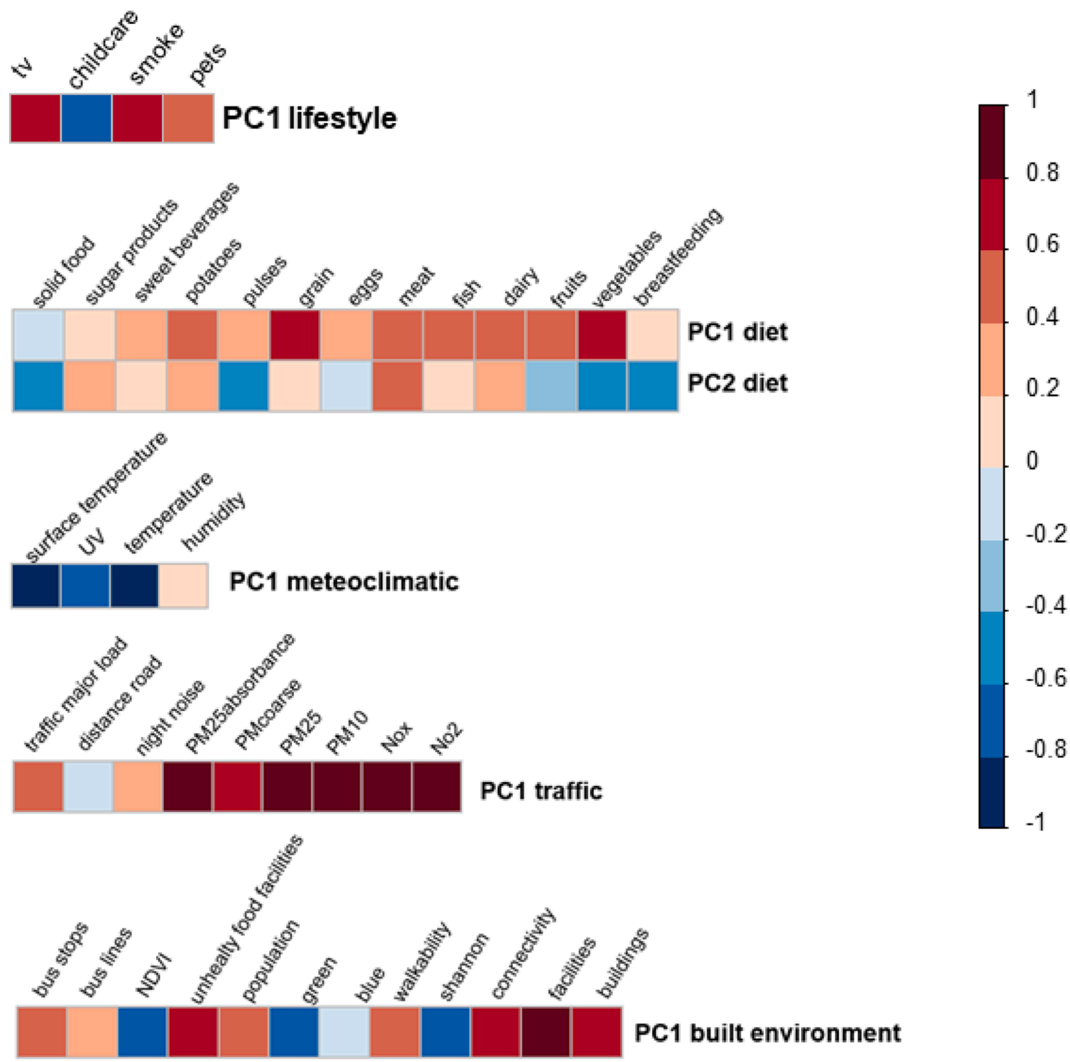

Fig. 5. Heat maps illustrating positive (red) or negative (blue) correlations (Pearson's or polyserial correlations, depending on the type of the variables) between each exposome family and its retained PCs (explaining at least 20% of the intra-group variance). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

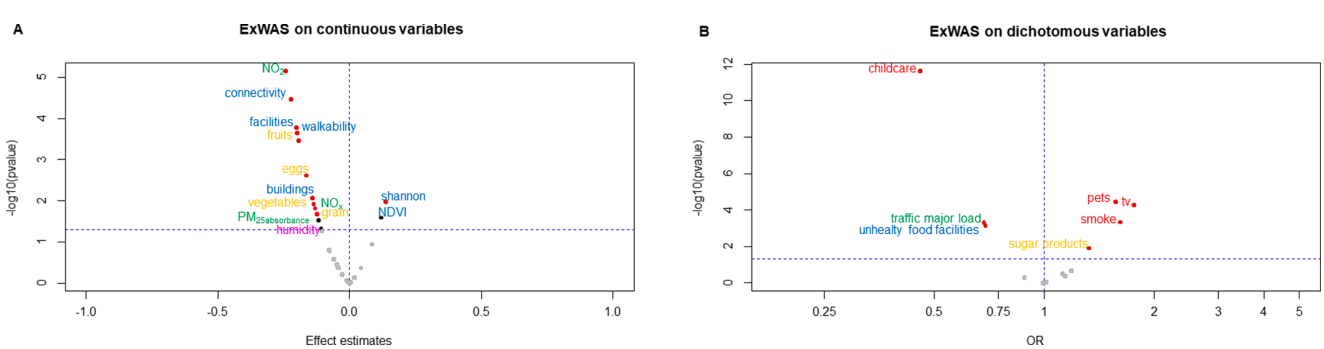

Fig. 6. Volcano plots of the Exposome Wide Association Study (ExWAS) analysis between SEP and early life exposome (N = 1989). Y-axis shows the strength of evidence against the null hypothesis ($-\log_{10}$ p-value) and x-axis shows the effect size (medium–low SEP children vs. high SEP children). Exposome variables with FDR adjusted p-values below 0.05 are highlighted in red. Exposome variables with nominal p-values below 0.05 but FDR adjusted p-values above 0.05 are highlighted in black. The horizontal dotted blue line represents the nominal p-value threshold of 0.05. Estimates are adjusted for maternal nativity, parity and age at delivery. In A (continuous variables), the effect size is expressed as the mean difference measured in standard deviations (SD). In B (dichotomous variables), the effect size is expressed as the OR on a logarithmic scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

environment.

We chose SEP as the distal driver of the exposome because of its potential in explaining health inequalities of social origin. SEP is a composite construct that encompasses both resource-based measures, like income and wealth, and prestige-based measures, like education and social status (Krieger, 2022). It has been shown that different SEP measures have different relationships with the exposome (Sum et al., 2022). Family income and wealth are difficult to measure through questionnaires. We took advantage of a standardised indicator of material resources, the EHII, currently used in several birth cohorts throughout Europe, which overcomes the limitations of other commonly used individual SEP measures, like maternal education (Khalatbari-Soltani et al., 2022).

We applied three different methods to assess the SEP-exposome association: (1) one exposure – one outcome approach (ExWAS), (2) cluster analysis, to create homogeneous groups of subjects with regard to the exposome patterns, and (3) PCA as a dimensionality reduction technique on a priori defined exposome subgroups.

The ExWAS approach showed that, in the NINFEA Turin cohort, being born in a family with a medium/low SEP increases the probability of living in greener areas with a higher diversity index, with less air pollution, fewer buildings, facilities, connectivity and walkability with respect to children born in a high-SEP family. On the other hand, children with medium/low SEP are more likely exposed to unhealthy lifestyles (passive smoking, TV screen, less childcare attendance) and unhealthy diet (higher sugar product consumption, less vegetables and fruit). These results on lifestyles and diet are mostly consistent with the cluster analysis results (with the exception of passive smoking). Children with medium/low SEP have higher odds of belonging to cluster 3 (overall characterised by poor diet, the lowest levels of air pollution, and living in sub-urban greener areas), with respect to cluster 1 (low levels of pollution, better diet and living in central areas). Consistently, with the intra-group PCA, we found that medium/low SEP children tended to have higher levels of unhealthy lifestyles and diet, and lower levels of urbanization and traffic-related factors than high SEP children.

These results are plausible when considering the population distribution in the city of Turin, where wealthier people tend to live in central areas and more deprived subjects tend to live in the suburbs. The map of the kernel density estimation of geocoded residential addresses of children in the three clusters shows that clusters 1 and 2 are located in the central areas of the city. Knowledge of the spatial and social structure of Turin indicates that children in cluster 1 live in the richer central areas surrounded by parks and blue areas (not visible in the map), while children in cluster 2 live in the central areas close to major traffic arteries (not visible in the map) and rich in services. Urbanization factors in cluster 2 may have both a detrimental effect on children's development, due to air pollution, noise, and reduced green spaces and a beneficial effect, due to the presence of various facilities and good services, public transport facilities and connectivity. The simultaneous action of factors with opposite effects (harmful and beneficial) on health is a common but problematic situation when studying the human exposome. This may lead to the paradoxical effect of observing implausible associations between adverse exposures and improved health outcomes (or vice versa), which should be considered when interpreting results for specific outcomes. In our attempt to characterize the children exposome, we realised that while several individual behavioral disease risk factors are likely to cluster together and to be negatively associated with socioeconomic position, the environmental exposures related to urbanization, air pollution and other residence-related factors may be very specific to the social and contextual factors. Our findings might support the enhancement of urban planning to reduce environmental risks and promote supportive strategies to decrease behavioral risk factors, tailored to specific characteristics of population subgroups.

In this article, we have selected three approaches to summarise the relation between SEP and the complex exposome structure that are relatively easy to implement and offer interpretable results. These are not exhaustive of all possibilities but represent conceptually different options that achieve different goals. We have not proposed a head-to-head comparison between these approaches but rather showed their specificities in grasping different aspects, and we think that they may reach their maximum potential through their integration, always bearing in mind that the choice about the most suitable method eventually depends on the specific research question. In our setting, the three methods found consistent results despite their analytical differences.

The ExWAS approach, although based on the one exposure – one outcome paradigm, has substantial advantages, thanks to its outcome-wide structure. In the traditional ExWAS approach, independent regression models are fitted on potentially correlated exposures, making it difficult to know if the estimated effect is due to a specific exposure or to another correlated exposure ignored in the analysis. This has repercussions on the high false discovery proportion (Agier et al., 2016). However, in the ExWAS approach proposed here, since we are dealing with a distal driver and a set of potentially highly correlated outcomes (the exposome variables), we enjoy the advantage that, in principle, we can control for confounding between the single exposure (driver) and all the outcomes (exposome) simultaneously. This is possible if the set of confounders includes all variables temporally preceding the exposure that could influence the exposure (VanderWeele, 2017). This overcomes the problem of the correlation structure between the different components of the exposome.

Although transparent, replicable and useful when we want to speculate on the association between SEP and a specific exposome variable, this approach has its limitations when trying to capture the salient features of the exposome and somehow summarise its complex structure.

To this aim, techniques to reduce the dimensionality of the exposome can be more helpful.

We applied a dimensionality reduction method (PCA) and a method for clustering observations. They provide different solutions to the same problem: investigating the variability in the dataset in order to find a way to explain it through fewer components. The intra-exposome-group PCA aims to compress the exposome variables into fewer components on the basis of their correlation, capturing as much as possible variability within their specific groups. Thus, the evaluation of the association between the SEP and exposome is conceptually similar to the ExWAS but using fewer variables (one or two for each exposome group) instead of all the exposome variables. It allows us to speculate on the role that the SEP has had in driving that exposome-group pattern (e.g., dietary) in our population. Applying the PCA within each exposome group allowed us to improve the interpretability of the resulting patterns, still retaining information on which group of the original variables is driving the associations. This approach, however, did not account for the between-group variability and only partially overcame the problem of the small proportion of variance explained by the first components (especially in not strongly correlated subgroups, like diet).

Clustering aims to separate the subjects into mutually exclusive groups based on individual differences in the exposome variables. The association with SEP provides insights into the role that SEP has had in driving the grouping of children and it is useful to identify population subgroups on whom to plan an intervention. However, it does not tell us anything about which variables characterizing the clusters are mostly affected by the driver. Thus, investigating the relationship between a distal driver and the exposome considered as a multi-outcome context gives us the opportunity to investigate two different problems: if the SEP is a determinant of the exposome subgroup patterns in a population and if there are identifiable population subgroups with a higher risk of adverse exposures.

4.1. Strengths and limitations

The wide range and diverse nature of the general and specific external exposome components gave us the opportunity to apply different strategies to analyse the exposome as a multi-outcome context with a holistic approach. The internal exposome domain was not considered in our analysis, but the proposed approaches could be adopted in a study that also includes this aspect. Moreover, we are aware that our selection is not exhaustive and that some components of the external exposome have not been included in this analysis. Our objective however is substantially methodological and can thus be seen as an exploration of methods useful in the study of distal drivers of the multi-dimensional early-life exposome. Other works, including all three exposome domains and more comprehensive in the exogenous and endogenous exposures considered are better suited to unravel the specific challenges of SEP-exposome associations in birth cohort studies (Sum et al., 2022; Robinson et al., 2015; Jiang et al., 2018).

The findings presented here are exploratory in nature and we suggest caution in a causal interpretation of the role of SEP as a distal driver of the environmental exposome. Although missing data proportions in our analyses were less than 10% and hence their impact in terms of selection bias is likely to be minimal, the presence of unmeasured confounding cannot be ruled out. Moreover, it would be difficult to exclude violations of the consistency assumption, which entails that the exposure (here, SEP) is defined with enough specificity that different variants of it do not have different effects on the outcome (here, the exposome) (Rehkopf et al., 2016).

In addition, SEP can be and has been (Wild, 2012) considered as part of the exposome, since a household's material and intellectual resources can justifiably be included among the exposures capable of influencing future health status. However, in this study, we attempted to investigate if familial SEP at birth may play a role in shaping the general and specific external exposome of children at 18 months of age.

The population included in this study may not be representative of the general Turin population because of under-recruitment of low-SEP women. This may be an issue from a descriptive point of view but in this study we still had enough variability to characterise the exposome patterns of participating children and to study the association between SEP and the early childhood exposome (Pizzi et al., 2012). Finally, another limiting point is that, although several of the considered environmental exposome components may vary in time, we ignored their longitudinal dimension and only included single point measures or averages, thus hiding the possible presence of peaks and limiting the variance.

5. Conclusion

This study shows that, in the city of Turin, SEP is associated with many early-life exposome characteristics. In particular, people with lower SEP are less exposed to air pollution and more exposed to unhealthy lifestyles and diet than people with higher SEP. Specificities, potential, and pros and cons of three different analysis strategies were examined. The ExWAS approach conveys most of the information and is more replicable in other populations, although the cluster analysis and the intra-group PCA may facilitate the interpretation and communication of the results. In the setting analysed here, the results obtained with the three approaches were fairly consistent. They provided complementary information for the assessment of the socioeconomic determinants of the exposome, an under-explored topic that could lead to the identification of population subgroups with higher risk of adverse exposures.

CRediT authorship contribution statement

Chiara Moccia: Investigation, Formal analysis, Methodology, Visualization, Writing – original draft. **Costanza Pizzi:** Investigation, Data curation, Writing – review & editing. **Giovenale Moirano:** Data curation, Visualization, Investigation, Writing – review & editing. **Maja Popovic:** Data curation, Supervision, Writing – review & editing. **Daniela Zugna:** Writing – review & editing. **Antonio d'Errico:** Writing – review & editing. **Elena Isaevska:** Data curation, Writing – review & editing. **Serena Fossati:** Data curation, Writing – review & editing. **Mark J. Nieuwenhuijsen:** Data curation, Writing – review & editing. **Piero Fariselli:** Methodology, Writing – review & editing. **Tiziana Sanavia:** Methodology, Writing – review & editing. **Lorenzo Richiardi:** Conceptualization, Funding acquisition, Project administration, Methodology, Supervision, Writing – review & editing. **Milena Maule:** Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors are grateful to all the participants of the NINFEA cohort. We are grateful to Xavier Basagaña for insightful comments.

Funding

The NINFEA study was partially funded by the Compagnia San Paolo Foundation. This research was partially funded by the Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca – MIUR) under the programme "Dipartimenti di Eccellenza 2018–2022", by the European Union's Horizon2020 research and innovation programme ATHLETE, grant agreement number 874583. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2023.107864.

References

Agier, L. et al., 2016. A systematic comparison of linear regression–based statistical methods to assess exposome-health associations. Environ Health Perspect. 124(12): 1848-1856. doi: 10.1289/EHP172. Epub 2016 May 24. PMID: 27219331; PMCID: PMC5132632.

Blumenberg, C., Zugna, D., Popovic, M., Pizzi, C., Barros, A.J.D., Richiardi, L., 2018. Questionnaire breakoff and item nonresponse in web-based questionnaires: multilevel analysis of person-level and item design factors in a birth cohort. J. Med. Internet Res. 20 (12). doi: 10.2196/11046.

Brulle, R.J., Pellow, D.N., 2006. Human health and environmental Inequalities. Annu. Rev. Public Heal. 27, 103–127. https://doi.org/10.1146/annurev.publhealth.27.021405.102124.

CDC - Exposome and Exposomics - NIOSH Workplace Safety and Health Topic. Available from: <https://www.cdc.gov/niosh/topics/exposome/default.html> (accessed Apr. 12, 2022).

Correia, D., Severo, M., Lopes, C., 2020. The role of socio-economic factors in food consumption of Portuguese children and adolescents: results from the National Food, Nutrition and Physical Activity Survey 2015–2016. Br. J. Nutr. 124 (6), 591–601. https://doi.org/10.1017/S0007114520001373.

Rehkopf, D.H., Maria Glymour, M., Osypuk, T.L. The consistency assumption for causal inference in Social Epidemiology: when a rose is not a rose. doi: 10.1007/s40471-016-0069-5.

de Castro Pascual, M., Fossati, S., Nieuwenhuijsen, M., Vrijheid, M. Protocol for integrated urban environment stressorsgeneration in LifeCycle (WP3 – Task 3.3).

Degree of urbanisation classification - 2011 revision - Statistics Explained. Available from: <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Degree_of_urbanisation_classification_-_2011_revision> (accessed Nov. 02, 2022).

Firestone, R., et al., 2015. Internet-based birth-cohort studies: is this the future for epidemiology? JMIR Res. Protoc. 4 (2), e71.

Fox, J., 2021. polycor: Polychoric and Polyserial Correlations,. [Online]. Available from: <https://cran.r-project.org/package=polycor>.

Ganzleben, C., Kazmierczak, A., 2020. Leaving no one behind - understanding environmental inequality in Europe. Environ. Heal. A Glob. Access Sci. Source 19 (1), 1–7. https://doi.org/10.1186/S12940-020-00600-2/PEER-REVIEW.

Glossary A glossary for social epidemiology N Krieger. Accessed: Feb. 10, 2022. [Online]. Available from: <www.jech.com>.

Haddad, N., Andrianou, X.D., Makris, K.C. A scoping review on the characteristics of human exposome studies. doi: 10.1007/s40726-019-00130-7.

Hajat, A., Hsia, C., O'Neill, M.S., 2015. Socioeconomic disparities and air pollution exposure: a global review. Curr. Environ. Heal. Rep. 2 (4), 440–450. https://doi.org/10.1007/S40572-015-0069-5/TABLES/3.

Jaddoe, V.W.V., et al., 2020. The LifeCycle Project-EU Child Cohort Network: a federated analysis infrastructure and harmonized data of more than 250,000 children and parents. Eur. J. Epidemiol. 35 (7), 709–724. https://doi.org/10.1007/S10654-020-00662-Z/TABLES/3.

Jiang, C., et al., 2018. Dynamic human environmental exposome revealed by longitudinal personal monitoring. Cell 175 (1), 277–291.e31. https://doi.org/10.1016/J.CELL.2018.08.060.

Juarez, P.D., et al., 2014. The public health exposome: a population-based, exposure science approach to health disparities research. Int. J. Environ. Res. Public Health 11 (12), 12866–12895. https://doi.org/10.3390/IJERPH111212866.

Khalatbari-Soltani, S., Maccora, J., Blyth, F.M., Joannè, C., Kelly-Irving, M. Measuring education in the context of health inequalities. doi: 10.1093/ije/dyac058.

Krieger, N., 2001. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int. J. Epidemiol. 30 (4), 668–677. https://doi.org/10.1093/IJE/30.4.668.

Montazeri, P., et al., 2019. Socioeconomic position and exposure to multiple environmental chemical contaminants in six European mother-child cohorts. Int. J. Hyg. Environ. Health 222 (5), 864. https://doi.org/10.1016/J.IJHEH.2019.04.002.

Pagès, J., 2004. Volumen 27 N o 1. Págs. 1 a 26. Junio.

Pizzi, C., et al., 2012. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. J. Epidemiol. Community Health 66 (11), 976–981. https://doi.org/10.1136/jech-2011-200065.

Pizzi, C. et al., 2020. Measuring child socio-economic position in birth cohort research: the development of a novel standardized household income indicator. Int. J. Environ. Res. Public Health 17 (5). doi: 10.3390/ijerph17051700.

Richiardi, L., Baussano, I., Vizzini, L., Douwes, J., Pearce, N., Merletti, F., 2007. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. Eur. J. Epidemiol. 22 (12), 831–837. https://doi.org/10.1007/s10654-007-9194-2.

Robinson, O., et al., 2015. The pregnancy exposome: multiple environmental exposures in the INMA-Sabadell birth cohort. Environ. Sci. Technol. 49 (17), 10632–10641. https://doi.org/10.1021/ACS.EST.5B01782/SUPPL_FILE/ES5B01782_SI_002.XLSX.

Robinson, O. et al., 2018. The urban exposome during pregnancy and its socioeconomic determinants. Environ. Health Perspect. 126 (7). doi: 10.1289/EHP2862.

Santos, S., et al., 2020. Applying the exposome concept in birth cohort research: a review of statistical approaches. Eur. J. Epidemiol. 35 (3), 193–204. https://doi.org/10.1007/S10654-020-00625-4/FIGURES/1.

Mbuga, F., Tortora, C., 2021. Spectral Clustering of Mixed-Type Data. doi: 10.3390/stats5010001.

Senier, L., Brown, P., Shostak, S., Hanna, B. The socio-exposome: advancing exposure science and environmental justice in a post-genomic era. doi: 10.1080/23251042.2016.1220848.

Siroux, V., Agier, L., Slama, R. The exposome concept: a challenge and a potential driver for environmental health research. doi: 10.1183/16000617.0034-2016.

Stafoggia, M., Breitner, S., Hampel, R., Basagaña, X., 2017. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the state of the science. Curr. Environ. Heal. Reports 4 (4), 481–490. https://doi.org/10.1007/s40572-017-0162-z.

Sum, K.K., et al., 2022. The socioeconomic landscape of the exposome during pregnancy. Environ. Int. 163, 107205 https://doi.org/10.1016/J.ENVINT.2022.107205.

Unequal exposure and unequal impacts — European Environment Agency. Available from: <https://www.eea.europa.eu/publications/unequal-exposure-and-unequal-impacts> (accessed Jan. 18, 2022).

VanderWeele, T.J., 2017. Outcome-wide epidemiology. Epidemiology 28 (3), 399–402. https://doi.org/10.1097/EDE.0000000000000641.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth. Springer, New York.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat Comput 17, 395–416. https://doi.org/10.1007/s11222-007-9033-z.

Vrijheid, M., et al., 2012. Socioeconomic status and exposure to multiple environmental pollutants during pregnancy: evidence for environmental inequity? J. Epidemiol. Community Heal. 66 (2), 106–113. https://doi.org/10.1136/JECH.2010.117408.

Vrijheid, M., et al., 2014. The human early-life exposome (HELIX): project rationale and design. Environ. Health Perspect. 122 (6), 535. https://doi.org/10.1289/EHP.1307204.

Wild, C.P., 2005. Complementing the genome with an 'exposome': the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomarkers Prev. 14 (8), 1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456.

Wild, C.P., 2012. The exposome: from concept to utility. Int. J. Epidemiol. 41 (1), 24–32. https://doi.org/10.1093/IJE/DYR236.