HEALS

Health and Environment-wide Associations based on Large population Surveys

FP7-ENV-2013- 603946

http://www.heals-eu.eu/

D11.2 Report on the application of the exposure modelling framework to population studies covered in Stream 5

WP11: Integration of time- and spatially resolved data
Data and model synthesis

Version number 1.0

Lead beneficiary: 4 - USTUTT

Date: 17 May 2017

Nature: R - Report

Dissemination level: PU - Public

Document Information

Grant Agreement Number	ENV-603946	**Acronym**	HEALS
Full title	Health and Environment-wide Associations based on Large population Surveys		
Project URL	http://www.heals-eu.eu/		
EU Project Officer	Tuomo Karjalainen,- Tuomo.KARJALAINEN@ec.europa.eu		

Deliverable	**Number**	11.2	**Title**	Report on the application of the exposure modelling framework to population studies covered in Stream 5
Work Package	**Number**	11	**Title**	Integration of time- and spatially resolved data: Data and model synthesis

Delivery date	**Contractual**	M42	**Actual**	M44
Status	Draft ☒		Final ☐	
Nature	Demonstrator ☐ Report ☒ Prototype ☐ Other ☐			
Dissemination level	Confidential ☐ Public ☒			

Author (Partners)	Christian Schieberle (USTUTT), Naixin Li (USTUTT), Rainer Friedrich (USTUTT)			
Responsible Author	Christian Schieberle		**Email**	cs@ier.uni-stuttgart.de
	Partner	USTUTT	**Phone**	+49 711 685 87847

Table of Contents

List of Figures

List of Abbreviations

CI	Confidence Interval
DEHP	diethylhexyl phthalate
DIDP	diisodecyl phthalate
DINP	diisononyl phthalate
EDMS	Environmental Data Management System
EFSA	European Food Safety Authority
EIONET	European Environment Information and Observation Network
EMEP	European Monitoring and Evaluation Programme
EMF	Electro-magnetic fields
ITR	Italian Twin Registry
LCET	Life-course Exposure Trajectory
MTUS	Multi-national Time-Use Study
SILC	Statistics on Income and Living Conditions

1 Scope of this study

The aim of this report is to present the current progress of applying the probabilistic exposure modelling framework developed in Stream 3 on the population studies covered in Stream 5. As a showcase scenario to be presented in this document, we selected a subsample of individuals covered by the Italian Twin Registry (ITR).

The ITR carries out epidemiological studies on health related topics and has enrolled about 26,000 twins since 2000. Due to the regulation implemented for ITR, the data collected remained anonymous within the exposure study applied and the address of the individuals were not given. Instead, information about age, gender, education level, geographic information of the hometown and traffic intensity are available. We selected a subset of 556 individuals to be analyzed in this study. The individuals cover an age range from 19 to 73 years, about six in ten are female. Education level ranges from below high school level over high school level to individuals holding university or college degrees.

Based on the data given on individual level, we estimated the past life course of each of the sample individuals separately. We applied the probabilistic modelling framework developed in Stream 3. The model estimates the past life course of individuals and links it with time-activity patterns and an activity-exposure matrix to establish a life exposure trajectory model (see HEALS Deliverable 9.2B and HEALS Deliverable 11.1 for detailed descriptions of the approaches).

We estimated the past exposure of each individual to 12 stressors across several exposure routes. Specifically, we assessed: inhalation of nitrogen dioxide, fine particles and ozone; exposure to UV radiation and electromagnetic fields; exposure to mould; exposure to selected metals via ingestion, namely arsenic, chromium and lead; exposure to phthalates (DEHP, DIDP and DINP) via inhalation, mouthing and ingestion. Analyzing the resulting individual exposure estimates of the 556 Italian individuals, we identified substantial differences in exposure for some of the stressors. We grouped the individuals based on their socio-economic background and found that some of the differences can be linked to circumstances that are characteristic of certain vulnerable groups.

2 Methodology: Life Course Exposure Trajectories (LCET)

In the scope of HEALS we aim to estimate exposure retrospectively and prospectively as well for individuals as for vulnerable groups. To achieve this, we developed a modelling framework which assesses the determinants of past and future life trajectories. The model was defined in a probabilistic manner to allow for incorporation of uncertainties during all stages of the assessment. The overall approach and a prototype implementation of the life course trajectory estimation was presented in detail in HEALS Deliverable 9.2B and is summarized in section 2.1.

Information collected at individual or group level is used to reconstruct past and estimate future life trajectories. We developed this approach as part of a 3-level methodology in which

- Level 1: Estimate life trajectories of main activities (years in a lifetime)
- Level 2: Estimate time-activity within the year (minutes in a day)
- Level 3: Estimate exposure at location/activity level

For a more detailed description the reader is referred to the aforementioned deliverables D9.2B and D11.1.

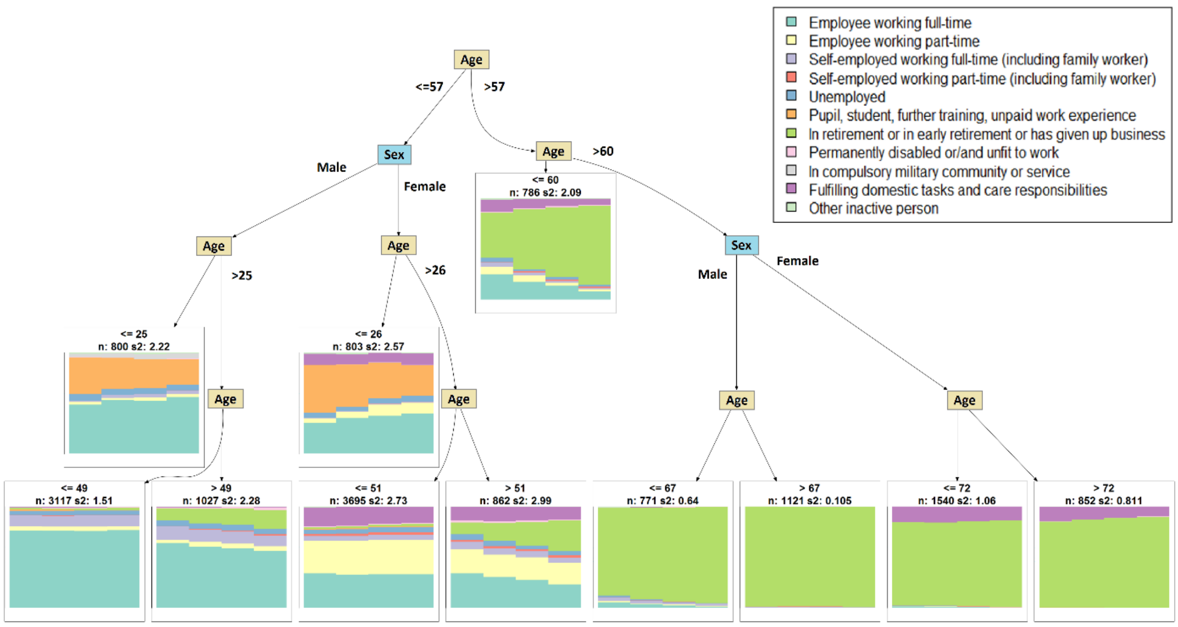

Figure 1: Example of a simplified regression tree of life trajectories determined using Optimal Matching. The data used is based on longitudinal data from 4-year surveys in Austria. The self-assessed economic situation (see leg-end for the eleven potential values) was regressed against two covariates only, namely age and sex. The covariates explain 37% of the variability.

2.1 Probabilistic model of life course trajectories

In social sciences, sequence analysis is a key approach to study life trajectories (Studer & Ritschard (2016), Widmer & Ritschard (2009)). It allows to identify trajectory patterns that account for all states of interest and, more importantly, transitions between states. For the first time we used this analysis approach in the field of exposure science to develop a probabilistic life exposure trajectory model that links life trajectories with time-activity patterns and an activity-exposure matrix to establish a life exposure trajectory model.

Note, that this is only the first level of the aforementioned 3-level methodology. In summary, the basic algorithm of the simulation process works as follows: Start at the current stage in life of the individual. For simplicity we focus here on age and sex only, where sex remains constant (but is an important determinant of the life-course) and age will increase in 1-year steps. In each step, the individual 'enters' a regression tree of life stages and a subsequent (or preceding) stage is assessed. This corresponds to one of the leaf nodes of the tree.

An example of a simplified regression tree is shown in Figure 1. The current situation determines the next (prospective) or the past years (retrospective). These are selected from the sequences that are grouped within each of the leaf nodes of the regression tree. Therefore, the individual's life course is probabilistically estimated using a Markov Chain that is constructed from the inner leaf-node states. In a last step the age of the individual is adjusted (i.e. increased for prospective estimates and decreased for retrospective analysis). Once an age below 16 is reached, a separate model is used for the assessment which we will not further describe in this document.

Figure 2: Distribution plot of retrospective and prospective life trajectories of 1000 simulations each. For an individual that is (a) 34-yr old woman, full-time employee, (b) 34-yr old woman, full-time self-employed, (c) 34-yr old man, full-time employee, (d) 34-yr old man, full-time self-employed. The dashed line separates the retrospective from the prospective trajectories.

An exemplary result is shown in Figure 2. The plots show the distribution of retrospective and prospective life trajectories (ranging from age 16 to 81) of a number of simulations (n=1000 each) for 4 different 'stereotype' individual. In detail, it shows (a) a 34-years old woman, full-time employed, (b) a 34-years old woman, full-time self-employed, (c) a 34-years old man, full-time employed, and (d) a 34-years old man, full-time self-employed. The dashed line separates the retrospective from the prospective trajectories.

2.2 Estimates of day-to-day exposure

When developing the framework specific emphasis was put on properly taking account of temporal and spatial variation as well as substance group-specifics and subgroup-specific characteristics within the population. These characteristics include socioeconomic variables, time-activity pattern and relevant microenvironments which can be covered by the datasets from EU-SILC and MTUS (see D11.1 for details).

EU-SILC provides annual pan-European detailed microdata on income, poverty, social exclusion and living conditions since 2004 (EUROSTAT, 2016). MTUS records the time-activity pattern of the diarists as well as socioeconomic variables and microenvironments for 10 European countries since 1985 (Fisher and Gershuny, 2015). A methodology was developed to fuse the two datasets based on common existing variables (e.g. age, income, education level) to generate a comprehensive dataset that inherits the information from both.

The fused data were combined with exposure data including stressor levels and model parameters based on population subgroup-specific characteristics (region, degree of urbanization and socio-economic data). Due to limitations in determining exact locations of individuals a distribution of stressor concentration and model parameters was determined which best describes the individual exposure and its uncertainty.

A detailed description of the estimation of day-to-day exposure is given in HEALS Deliverable D11.1.

2.3 Retrospective estimation of external exposure

Finally, we linked the approaches described in sections 2.1 and 2.2 to determine individual external exposure estimates over the life course. In this study, we conducted retrospective estimates only. For each of the 556 individuals, we determined numerous potential life course trajectories (i.e. transitions between 'main activities') and linked these with estimated time-activity patterns and a related activity-exposure matrix, eventually determining daily exposure patterns. Obviously, all of these estimates are uncertain in nature but can be constrained when knowing specifics about the individual. In this study, we constrained based on information given about an individual's age, sex, location and level of education during the ITR study period.

The results for a random individual are given in Figure 3. The female individual was 73 years old when being visited in the year 2009 in the context of the ITR study. At that time she reached a formal level of education below high school level.

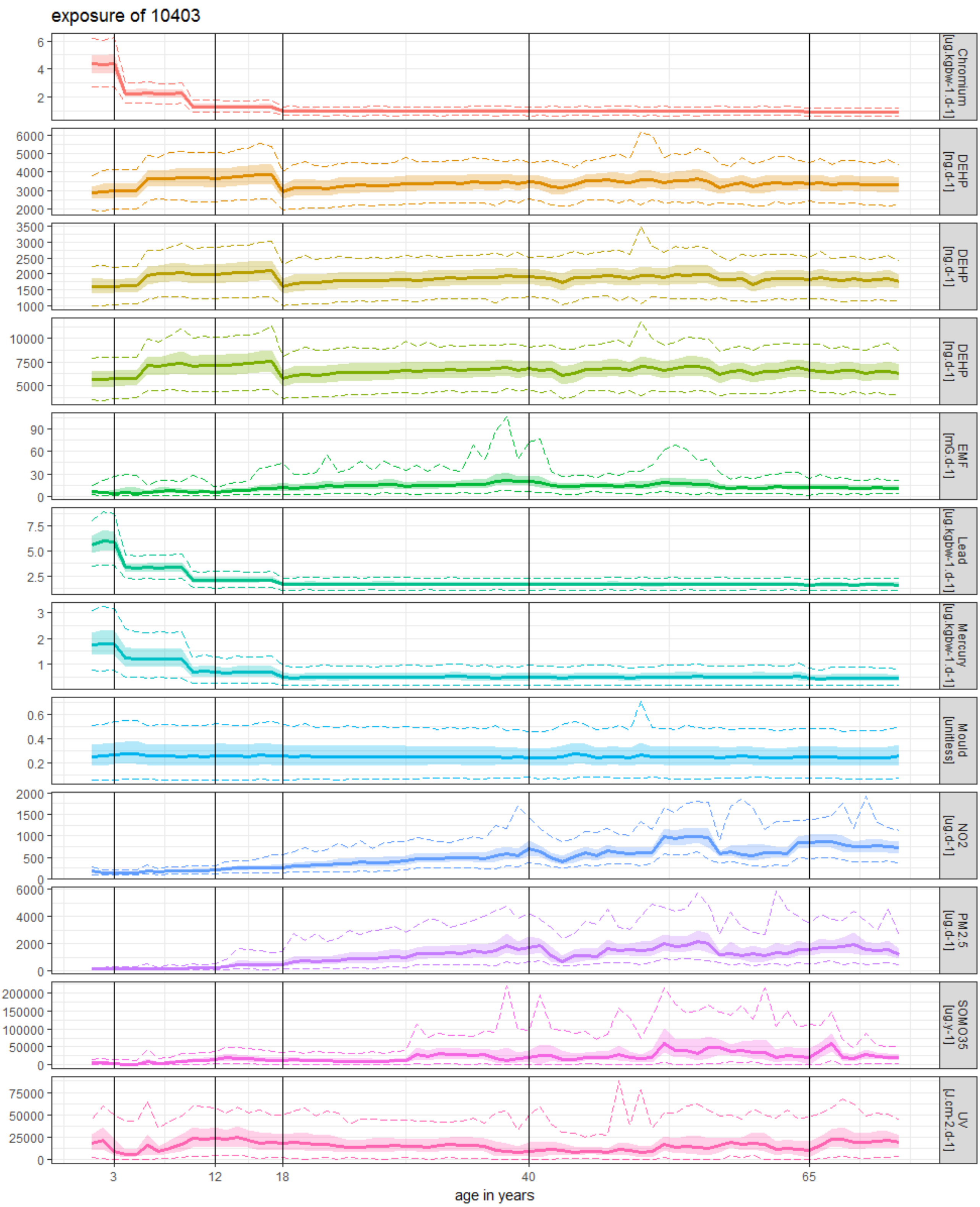

Figure 3: Exemplary plot of the retrospective external exposure estimation of a random individual. The woman was 73 years old at begin of the study. The plot shows past exposure estimates for 12 stressors (solid line = median; shaded area = 25% to 75% quantile; dashed lines = 95% CI). The solid vertical lines represent transition between pre-defined stages that we later aggregated to exposure windows.

3 Population and environmental data used

3.1 Data from population study: ITR subset

The population data applied in this report is an anonymized subset of the Italian Twins Registry (ITR). The ITR carries out epidemiological studies on health related topics and has enrolled about 26,000 twins since 2000. In this deliverable, we selected 556 individuals from ITR to apply the exposure modeling methodology developed within Stream 3. Due to the regulation implemented for ITR, the data collected should stay anonymous and the address of the individuals are not given. Instead, information about age, gender, education level, geographic information of their hometown and traffic flow are available. Figure 4 shows the distribution of sex, age, level of education, and traffic intensity across the sample data. Age ranges from 19 to 73 years, with people in their 20s being over-represented. About 60% of the individuals are female. Besides, more than half of the samples (56%) are living by streets with continuously or frequently vehicles passing by. A predominant amount of the individuals have reached a formal level of high school education or lower. This may very likely be an artifact caused by the skewed age distribution.

3.2 Environmental data

Environmental data was taken from the Environmental Data Management System (EDMS) developed in HEALS to integrate data on emissions of stressors, concentrations of toxic substances in environmental media (outdoor and indoor air, soil, water), in food and in drinking water and external exposures to environmental hazards.

3.2.1 Air pollution

Annual background concentration maps are important input for modelling exposure to air pollution. In this deliverable, we aim to estimate retrospectively the life-long exposure of people to multiple air pollutants. Thus it is necessary to simulate the temporal development of pollutant concentration. The simulation is traced back to the 1930s since the oldest individual in the sample is 73 years old when being visited in the late 2000s. Depending on data availability we utilized interpolated air quality maps from EIONET (Horálek et al. 2007) or modelled data, especially for periods before 1980.

The exposure to air pollution occurring indoors plays an important role since people spend a large proportion of time indoors. For modelling the PM2.5 levels within indoor environments, we assume the concentration indoors is under steady state and apply the mass balance model from Dockery and Spengler (1981). Detailed information with respect to data that have been used for modelling can be found in HEALS deliverable D11.1.

3.2.2 Food consumption patterns and food contamination

Dietary exposure to chemicals can be assessed at individual level by multiplying the average daily consumption for each food item with the corresponding average chemical content and summing up the respective intakes throughout the diet. The food consumption is gathered from the EFSA Comprehensive European Food

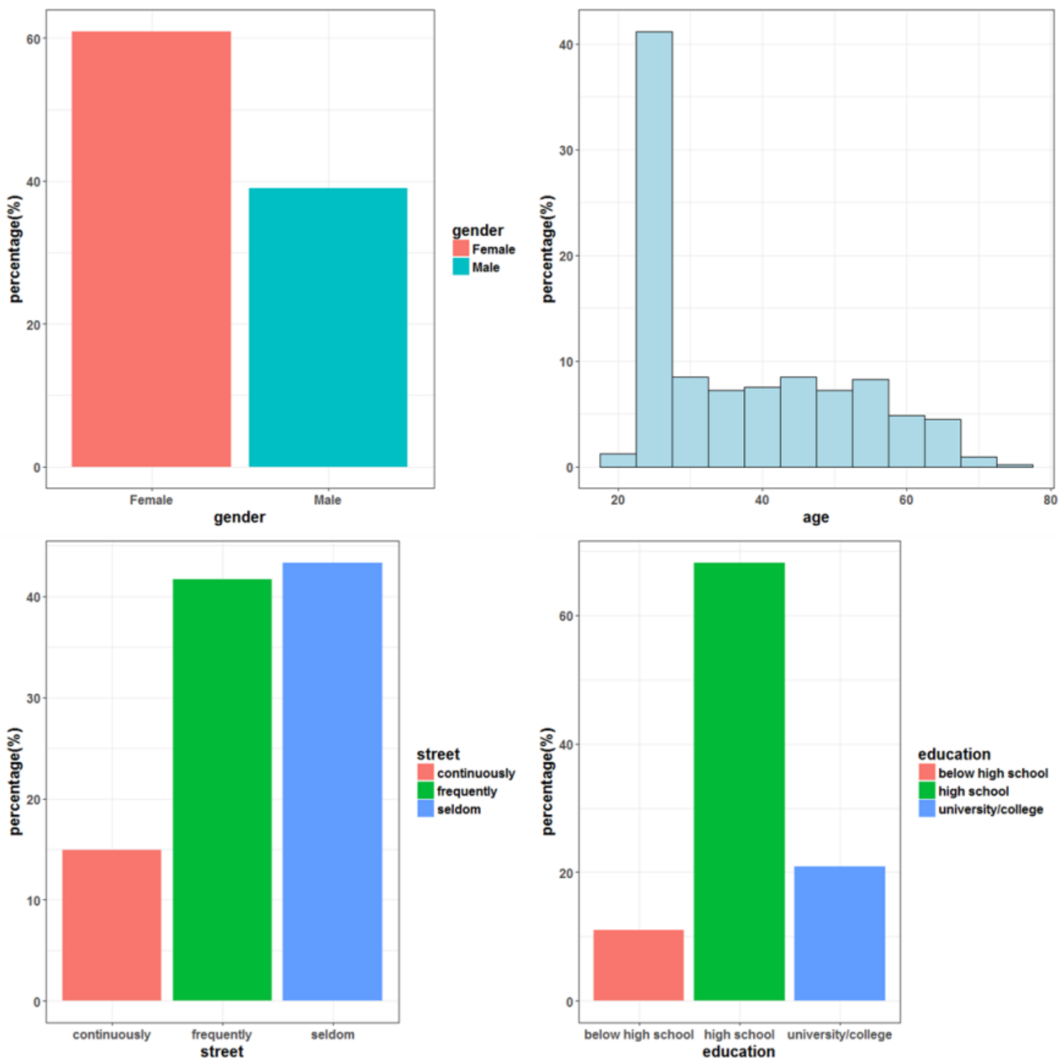

Figure 4: Distribution of gender, age, traffic flow and education within the subset sample of the Italian Twin Registry (ITR) population data. The sample comprises about 60% of the samples are female (upper left), age ranges from 19 to 73 years with a strong peak of people in their early 20s (upper right), more than half of the samples (56%) are living near streets with frequent vehicle pass-by (lower left), and – due to age peak – a majority has attended high school and one fifth attended university or college.

Consumption Database, which is a source of information on food consumption across Europe. Detailed statistical data is given for each food category (FoodEx1) and age group for 19 European countries. The food contamination data are collected from EFSA literatures and data is given by food category system that is exactly the same, or highly similar to FoodEx1. According to the suggestion of the EFSA Guidance on Standard Sample Description for Food and Feed (EFSA 2010), the lower bound (LB) and upper bound (UB) approach should be used for chemicals likely to be present in the food (EFSA 2014), which represents the minimum possible and maximal possible value, respectively. For this report, we focused on chromium, lead and mercury.

3.2.3 UV

UV bandwidth was defined according to WMO standard (WMO 2008), with UVA ranging from 315 nm to 400 nm; and UVB from 280 nm to 315 nm. Small amounts of UV are beneficial for people and essential in producing vitamin D (WHO 2017). But when people are exposed to excessive UV radiation, this may lead to malignant melanomas, other skin cancers and cataracts (WHO 2006). The UV data is generated from European Centre for Medium-Range Weather Forecasts (ECMWF), ERA-20CM dataset. The monthly mean downward UV radiation at the surface is available at a 0.125 by 0.125 grid from 1900 to 2010. Similar to air pollutant concentration data, the UV data are aggregated based on NUTS2 region and degree of urbanization.

3.2.4 Mould

Mould is rather difficult to quantify and only limited Europe-wide data can be accessed. For this report we generated the likelihood of mould occurrence in individuals' households by interpreting the EU-SILC variable "leaking roof, damp walls/floors/foundation, or rot in window frames or floor" and linking it to socio-economics. The data are aggregated by NUTS2 regions and degree of urbanization and a probabilistic model was built to simulate the possibility that certain individuals are exposed to mould to a certain extent in specific micro-environments.

3.2.5 EMF

The general public is exposed to electromagnetic fields (EMF) generated by an increasing variety of electrical and electronic devices and installations. The rapid increase in mobile telecommunications and the growing range of personal, domestic, commercial and medical equipment have considerably increased the number of sources of EMF exposure and are significantly changing the level, type and pattern of everyday exposure of the public. In order to properly assess the health risk associated to radiation exposure during different stages of human life in HEALS, a comprehensive estimation of all radiation (ionizing and non-ionizing) exposure events from conception to death is needed. Most people are exposed to magnetic fields that average less than 2 milligauss (mG), although individual exposures vary considerably. A detailed list and description of studies used within HEALS are given in deliverable D11.1.

3.2.6 Phthalates

Phthalates are a family of chemicals that are mainly used as plasticizers and widely found in cosmetics, personal care products, pharmaceuticals, medical devices and paints. People are exposed to phthalates through inhalation, oral uptake (food intake, mouthing) and dermal uptake. Evidence has found that phthalates are related to various health effects such as increased risk of allergic disease and asthma.

Within the scope of HEALS we conducted a literature review on studies that investigated phthalates exposure (see HEALS Deliverable D11.1). These studies cover most of the European countries and samples were taken from various vulnerable groups, including infants, children, adolescents, mothers (pregnant women), young men and senior citizens. Samples taken from urine, serum, dust, gas were analyzed and the minimum, maximum, mean and median value can be available for each vulnerable group.

4 Results

We analysed two groups of people that belong to two 'generations'. The first group (Y=young) is aged below 40 at the time of the ITR study. It consists of 342 individuals (61.5%). The second group (E=elderly) is aged 40 and older at the time of the study (214 individuals, 38.5%). In particular, we investigate the differences in exposure during pre-defined age ranges between the two generations. The four exposure windows investigated are: Infancy up to the age of 3, childhood from age 4 to 11, adolescence from age 12 to 17, and adulthood below the age of 40. The generations were selected to identify both differences in human behaviour and differences in stressor levels between two groups of people that spent specific life stages during different periods in time.

We present correlograms showing the matrix of pairwise Pearson correlation during each exposure window for both generations separately (see Figure 5 and Figure 6). We intend to identify how, and to what extent, the individuals' age, sex and education level correlate with the estimated external exposure to the stressors addressed in this study. Also, we use the correlogram to identify hints on co-exposures and their strength during the respective exposure window. The main findings are as follows.

For both generations exposure to fine particles (PM2.5) during adulthood before 40 is correlated with being a woman (r_Y = -0.80 and r_E = -0.90; sex is coded: 1=female, 2=male; see also bottom row of Figure 6). There is also positive correlation between PM2.5 and nitrogen dioxides (NO_2) which both to a large extent are emitted during combustion processes ($r_Y = r_E$ = 0.38). However, we do not find a comparably high correlation between sex and NO_2 during adulthood in neither of the two generations. Figure 7 shows the distribution of individual median exposure to PM2.5. Individuals were combined into age groups with respect to their age at the beginning of the ITR study (coloured). The grid shows female (F) and male (M) individuals over and estimates for specific exposure windows from infancy to adulthood. We can observe the same trend described above with higher, in fact increasing, relative exposure of women during adulthood. However, we observe that the levels across the age groups are consistently lower for recent generations. For men, peak exposure levels were experienced during their adulthood before 40 (18-39 years) by those aged between 36 and 65 at study start and during their later adulthood (40-64 years) by those aged between 56 to 75 years. This appears to be related to generally higher outdoor levels of PM2.5 during that period. Women seem to experience 2 to 4 times higher levels of PM2.5 exposure than men and those being older do so to an even higher extent (see Figure 7). This effect can be linked to specific activity patterns more prevalent in women of a certain age which expose them to indoor source of fine particles (e.g. cooking). These results should be considered preliminary and need careful interpretation of the underlying modelling assumptions.

For ozone (SOMO35) levels, we observe the opposite in both generations: There appears to be a negative correlation between aggregate exposure to fine particles and to ozone (e.g. during adulthood: r_Y = -0.46 and r_E = -0.62). Also, there is indication of correlation between exposure to SOMO35 and being a man (e.g. during adulthood: r_Y = 0.75 and r_E = 0.80). It is shown in Figure 9 that there is a generally higher level of exposure for men (M) as opposed to women (W). The exposure to ozone is mainly due to conducting outdoor activities, thus we cannot observe any similar indoor activity related pattern specific to exposure windows or age groups as we can for women with respect to exposure to fine particles.

With respect to metal exposure through ingestion, we were only able to utilize data stratified by age, and neither by sex nor level of education. Thus, obviously, we do not find any correlations between sex and levels

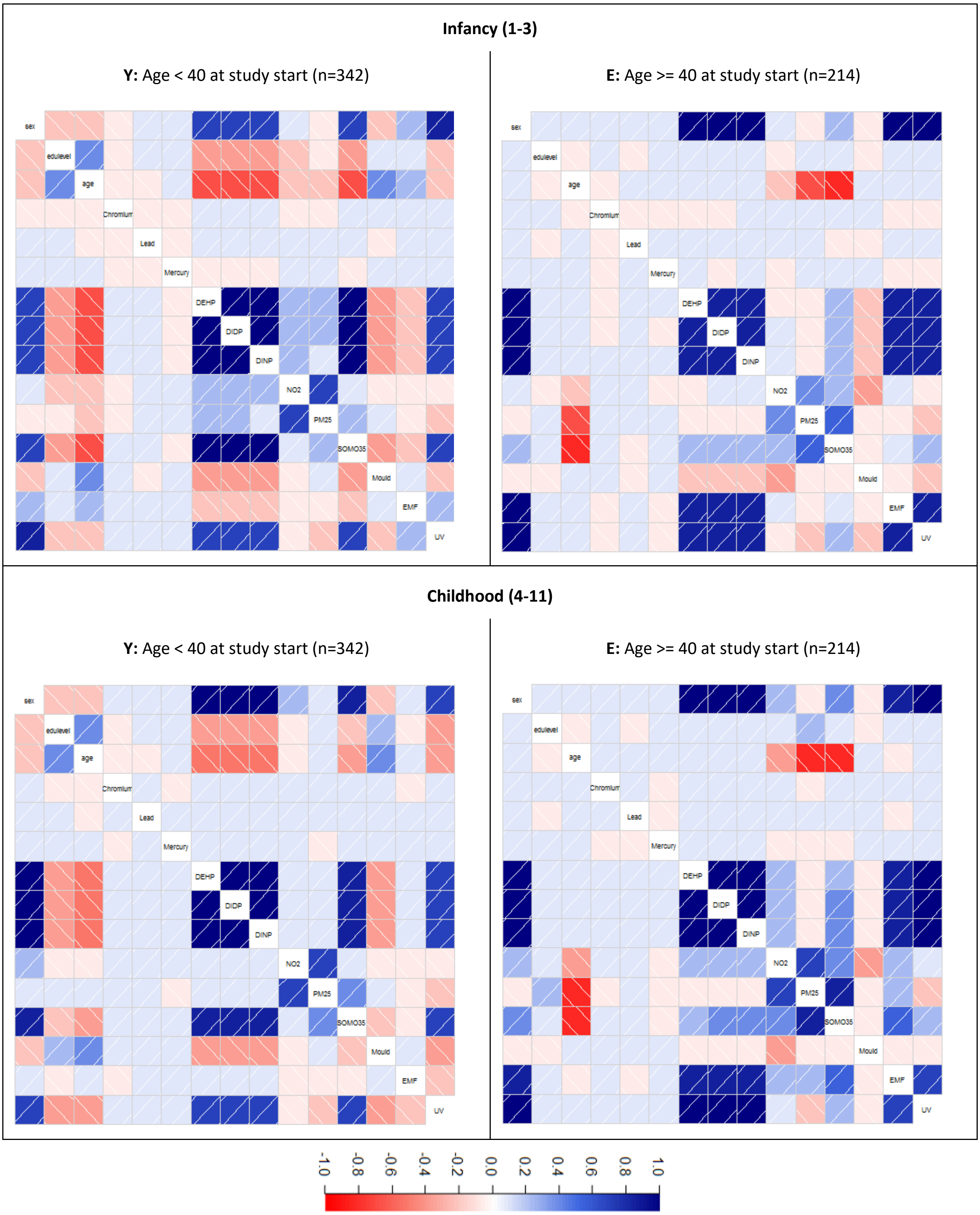

Figure 5: Correlograms of two generations (Y=young, E=elderly), each during infancy and childhood. The correlation matrix gives pairwise Pearson correlations (Sex: 1=female, 2=male; Education: 0=below high school, 1=high school, 2=university/college).

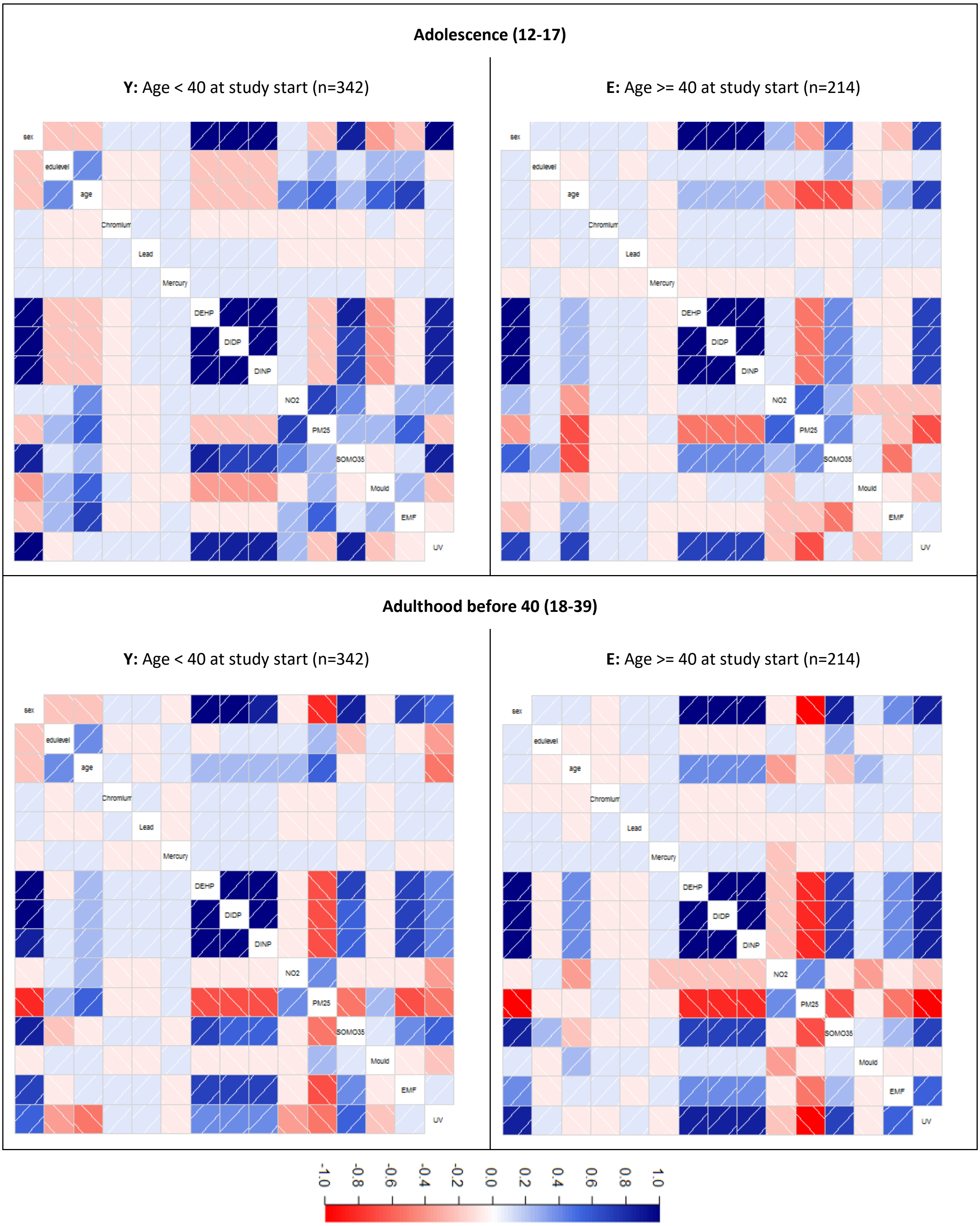

Figure 6: Correlograms of two generations (Y=young, E=elderly), each during adolescence and adulthood below 40. The correlation matrix gives pairwise Pearson correlations (Sex: 1=female, 2=male; Education: 0=below high school, 1=high school, 2=university/college).

of chromium, lead or mercury via ingestion when looking at metal exposure levels given per-kilogram body-weight. We do find as well that there is no direct indication of general co-exposure between the aforementioned metals.

For all generations and across all exposure windows we can observe strong co-correlations across phthalates considered in this study (see DEHP, DIDP and DINP in in the correlation matrices shown in Figure 5 and Figure 6). We took account of the following routes of phthalate exposure: mouthing, inhalation and food consumption. Phthalate exposure appears to be higher for men across all exposure windows (e.g. for DEHP during adulthood: $r_Y = 0.88$ and $r_E = 0.93$). This may seem surprising but can be linked to many factors, including different food consumption patterns, different inhalation rates, and higher average body weight. However, a more subtle analysis of the modelling results is needed to identify the explanatory factors linked to route-specific levels of exposure.

For the younger generation (age < 40 at study start), there is some indication that during the pre-adulthood exposure windows (infancy, childhood and adolescence, see Figure 5 and Figure 6), there is a small tendency of general negative correlation between being exposed to phthalates and the level of education at study start (e.g. for DEHP: $r_Y = -0.18$, whereas $r_E = 0.04$). In this group there is also a negative correlation with the age of the individual at study start (e.g. for DEHP: $r_Y = -0.22$). We do not see these correlation patterns in the older generation (e.g. for DEHP: $r_E = 0.05$). The effects needs closer investigation as they cannot be explained by the gender imbalance which is in fact representative for the sample: Within group Y, 61.4% of the individuals are women (210 of 342), whereas the older generation consists of 60.3% women (129 of 214). The original sample consists of 61% women (339 of 556).

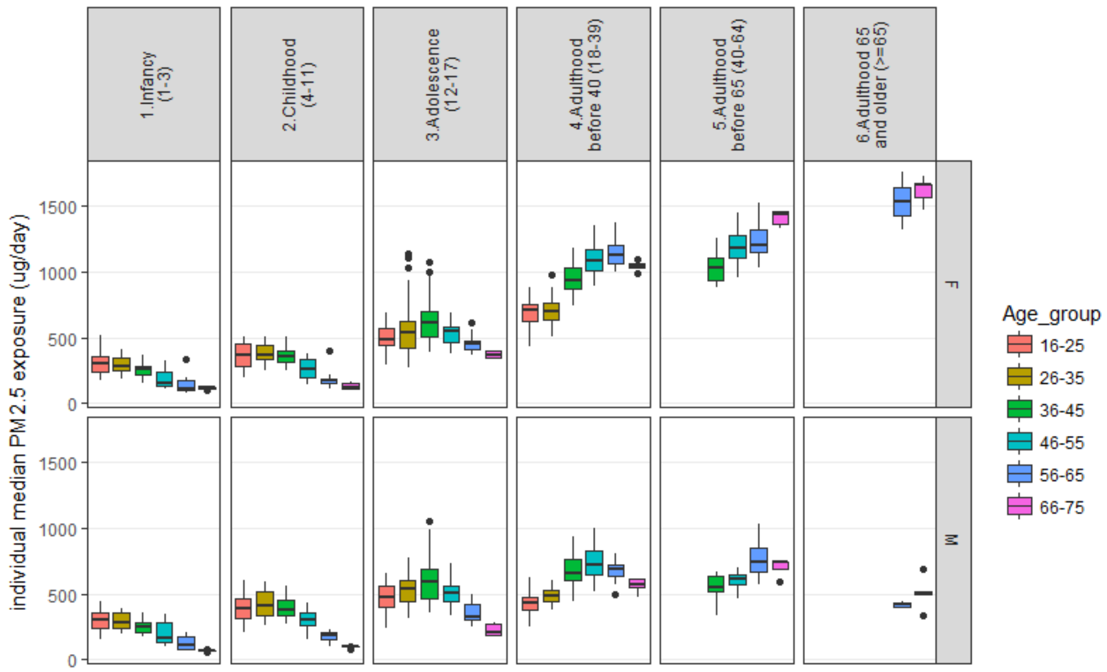

Figure 7: Individual median exposure to fine particles (PM2.5) in average micrograms per day. Individuals are grouped into age groups at the start of the study (see colors in legend). The grid shows female (F) and male (M) individuals over and estimates for specific exposure windows from infancy to adulthood.

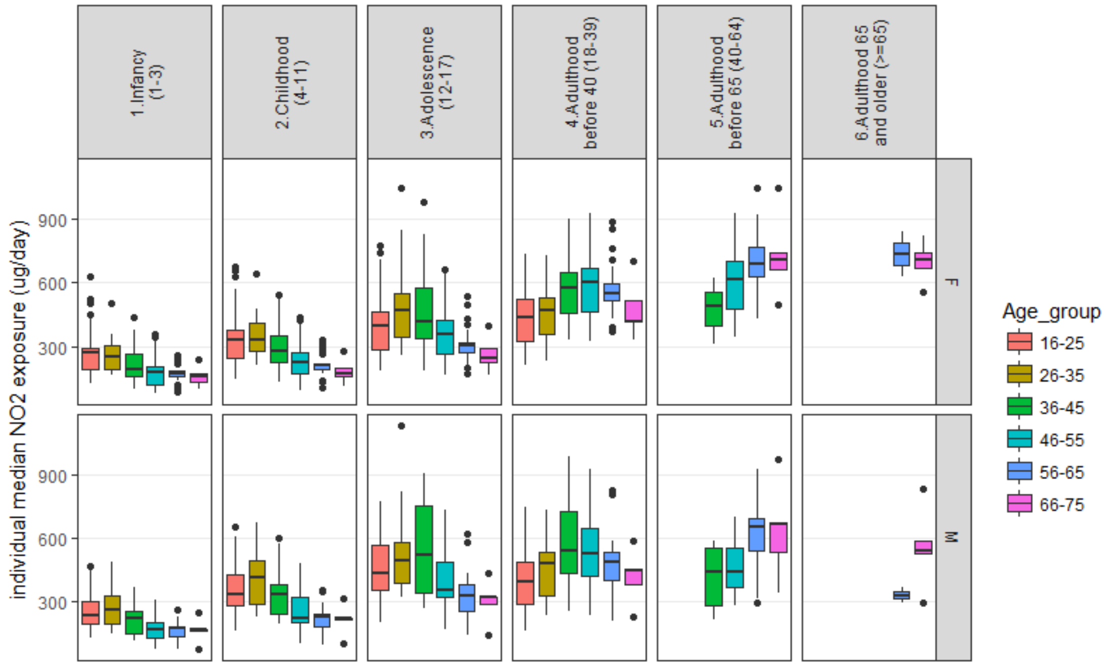

Figure 8: Individual median exposure to nitrogen dioxide (NO_2) in average micrograms per day. Individuals are grouped into age groups at the start of the study (see colors in legend). The grid shows female (F) and male (M) individuals over and estimates for specific exposure windows from infancy to adulthood.

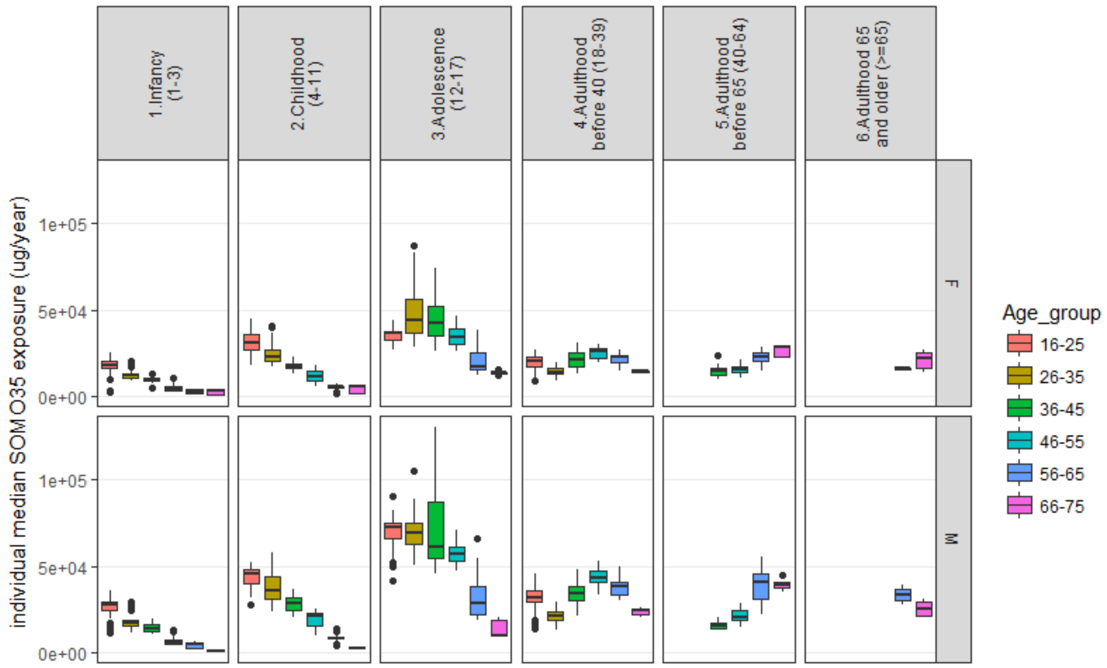

Figure 9: Individual median exposure to ozone (SOMO35) in average micrograms per year. Individuals are grouped into age groups at the start of the study (see colors in legend). The grid shows female (F) and male (M) individuals over and estimates for specific exposure windows from infancy to adulthood.

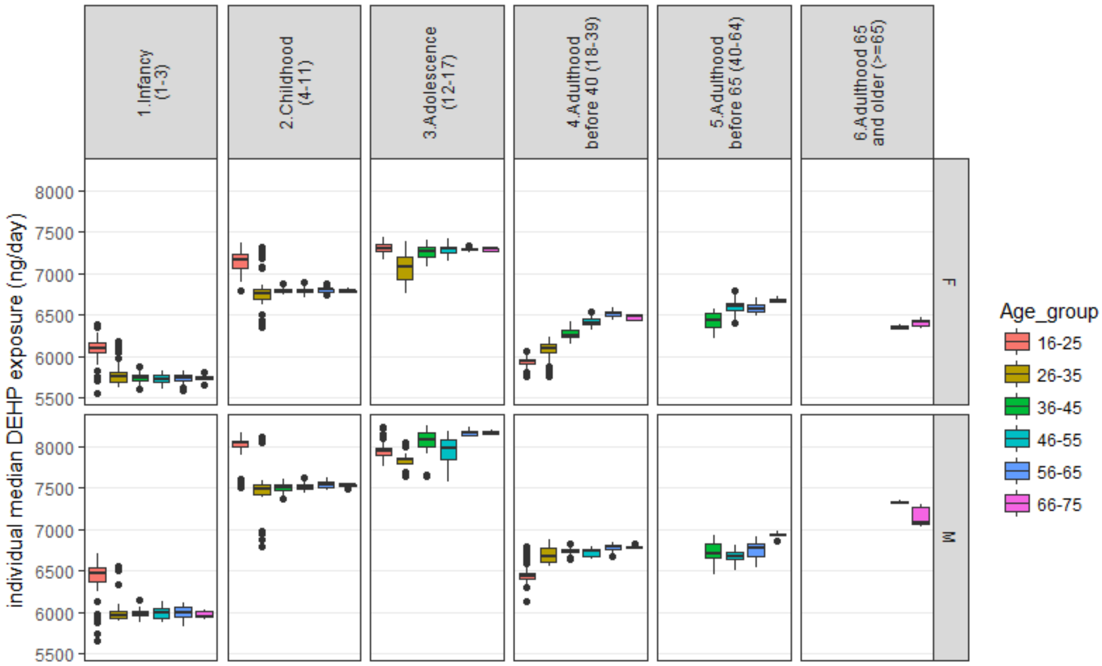

Figure 10: Individual median exposure to diethylhexyl phthalate (DEHP) in average nanograms per day. Individuals are grouped into age groups at the start of the study (see colors in legend). The grid shows female (F) and male (M) individuals over and estimates for specific exposure windows from infancy to adulthood.

5 Conclusion

In this report we showed the application of the probabilistic external exposure assessment framework developed in Stream 3 to a subsample of the Italian Twin Registry (ITR). As a result we were able to estimate retrospective exposure to 12 stressors via several exposure routes on an individual level.

We analysed the linkage between the socio-economic and exposure characteristics of an individual and of certain groups to present a showcase of the modelling framework. The approach will be applied to other studies identified in Stream 5.

Literature

Dockery, DW & Spengler, JD 1981, Indoor-outdoor relationships of respirable sulfates and particles. *Atmospheric Environment*, vol. 15, no.3, pp. 335-343.

EUROSTAT n.d., *Territorial typologies*. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Territorial_typologies . [20 March 2016].

EUROSTAT (2016): Glossary: EU statistics on income and living conditions (EU-SILC). Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_statistics_on_income_and_living_conditions_(EU-SILC). [15 March 2016].

Fisher, K. & Gershuny, J. (2015): Multinational Time Use Study, User`s Guide and Documentation, Version 7-updated. Available from: http://www.timeuse.org/sites/ctur/files/9727/mtus-user-guide-r7-february-2015.pdf . [20 March 2016].

Horálek, J, Denby, B, de Smet, P, de Leeuw, F, Kurfürst, P, Swart, R & van Noije, T 2007, *Spatial mapping of air quality for European scale assessment, ETC/ACC Technical Paper 2006/6.* Available from: http://acm.eionet.europa.eu/docs/ETCACC_TechnPaper_2006_6_Spat_AQ.pdf. [20 March 2016].

Studer, M. and Ritschard, G. (2016): What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. In: Journal of the Royal Statistical Society, Series A, 179, Part 2, pp. 481–511.

Widmer, E.D. & Ritschard, G. (2009): The de-standardization of the life course: Are men and women equal? Advances in Life Course Research, 14 (1–2), pp. 28–39.