HEALS

Health and Environment-wide Associations based on Large population Surveys

FP7-ENV-2013- 603946

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

D16.2: REPORT ON EXPOSOME RESULTS AND OF THE ENVIRONMENT-WIDE APPROACH REGARDING ASSESSMENT OF THE ENVIRONMENTAL DETERMINANTS OVERWEIGHT, OBESITY AND DIABETES

Version 1 (30/09/2018)

Lead beneficiary: SU

Date: 23/11/2018

Nature: Report

Dissemination level: 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**	16.2	**Title**	REPORT ON EXPOSOME RESULTS AND OF THE ENVIRONMENT-WIDE APPROACH REGARDING ASSESSMENT OF THE ENVIRONMENTAL DETERMINANTS OVERWEIGHT, OBESITY AND DIABETES
Work Package	**Number**	16	**Title**	Obesity and childhood diabetes - link with endocrine disruptors

Delivery date	**Contractual**	M40	**Actual**	23/11/2018
Status	Draft ☐		Final ☒	
Nature	Demonstrator ☐ Report ☒ Prototype ☐ Other ☐			
Dissemination level	Confidential ☐ Public ☒			

Author (Partners)	Sanyal[1], Isabella Annesi-Maesano[1]			
Responsible Author	Isabella Annesi-Maesano		**Email**	isabella.annesi-maesano@inserm.fr
	Partner	SU	**Phone**	+33 6 08 61 51 32

Document History

Name	Date	Version	Desciption
SU	23/11/2018	1	Submitted

1 Introduction

1.1 Aim and Scope of the study

The Environmental-wide association studies (EWAS) to analyse the exposomic effect of environmental stressors on overweight, obesity and diabetes is encompassed within the WP16. The main aim of this WP is to provide relevant data to unravel the relationship between body burden from endocrine disrupting chemicals (EDCs) and the onset of overweight, obesity and diabetes. The emergence of various cardiovascular diseases, future risks of obesity and cancer is often associated with a high body mass index (BMI) in childhood (Baker *et al*., 2007; Han *et al*., 2010). Studies like Holtcamp (2012) and La Merrill and Birnbaum (2011) have focused on the possibility of certain xenobiotic chemical exposure to be obesogenic, which may eventually lead to obesity and type-2 diabetes. Huang *et al*. (2007) suggests that, the effect of such obesogenic exposure is highest during pregnancy. Besides disrupting several hormonal pathways, prenatal exposure to EDCs can alter the epigenetic control of gene expression, which consequently affects the developmental programming causing obesity or other diseases. Epidemiological studies on the effect of EDCs in case of prenatal exposure has mainly focused on gestational tobacco exposure (Thayer *et al.*, 2012) and dichloro-diphenyldichloroethylene (DDE) (La Merrill and Birnbaum, 2011; Delvaux *et al*., 2014; Lee *et al*., 2014; Valvi *et al.*, 2014) as the most consistent EDCs. Taking into account a review of the past literatures, this study aims at effectively evaluating the association between the biomarker concentration of EDCs and the metabolic diseases occurring in children.

Additionally, the direct effect of environmental exposure leading to obesity and diabetes is also a pivotal aspect of our study. Several literature have investigated on the probable causal linkage between environmental chemicals like arsenic, phthalates, persistent organic pollutants (POPs), polychlorinated biphenyls (PCBs); and type 2 diabetes (for example, Patel *et al.*, 2010; Chen *et al.*, 2010; Lee et al., 2010; Rylander et al., 2005). With the increase in prevalence of obesity, there has been a rising interest on evaluating the potential risk of environmental factors related to obesity (Thayer *et al.*, 2012). Although improper diet and lack of physical activities are the major reasons for obesity, metabolic processes can also get affected by

exposure to certain environmental stressors leading to weight gain. The EWAS study allows for a systematic evaluation of several environmental factors that can be prominently related to diabetes and obesity.

Population effects from exposures to endocrine disruptors on metabolism is studied using different cohorts from which data of interest are available. Mother-child cohorts where both exposure (fetal and/or neonatal) and weight outcomes have been measured was considered, in order to assess whether prenatal exposure to endocrine disrupting compounds in food plays a role in the development of childhood obesity and related disorders later in life, as well as in the occurrence of gestational diabetes. These data are derived from birth cohorts that include mother-child pairs in several European countries. Additionally, general population studies along with the birth-cohorts are considered for estimating the overall exposomic effect of environmental stressors on obesity and diabetes. These cohorts include individuals of different age groups. The subsequent analyses, based on the accrued data, focused on a multidisciplinary approach that combined epidemiology, neonatology, endocrinology, toxicology, analytical chemistry and risk assessment, which are integrated as the HEALS approach.

This WP aims at validating and helping to refine the development of the HEALS methodological framework and platform. Data allowing the exposome assessment, namely: biological samples (e.g. umbilical cord blood, urine) useful for -omics analyses; information about lifestyle and health from questionnaires and follow-up studies, individual and environmental geocoded data for the assessment and management of temporally-spatially resolved data and models, are assembled together for this study.

1.2 Objectives

The general objectives of this study are:

• Provide relevant data/samples for testing the hypothesis of exposure-effect relationships of endocrine disruptors to metabolic disorders like overweight, obesity and diabetes.

• Quantify the link between exposures to endocrine disruptors and obesity and diabetes.

The specific objectives of this study are:

• Identify major gaps in knowledge in the relationships between endocrine disruptors and overweight, obesity and diabetes by analysing existing data available in HEALS that focuses on internal and external exposures and their effects on metabolic disorders such as overweight, obesity and diabetes;

• Assess internal exposure, investigate hypotheses on mechanisms of action using the Stream 2 methodological framework on –omics and epigenetics and assess biomarkers of exposure, effect and susceptibility through available data;

• Assess external exposure in link with Stream 3 and its associations with overweight, obesity and diabetes after taking into account potential confounders and interactions (lifestyle, physical activity, social class, smoking, air pollution) through the analysis of data from existing cohorts and studies;

• Assess the spatio-temporal relationships between internal and external environmental exposures and health data through the analysis of data from existing cohorts and studies geocoded when possible;

• Apply the HEALS methodology to validate a model strategy for the establishment of causal links between exposures and overweight, obesity and diabetes to be applied to available data (general population, singletons, twins' studies).

2 Data

2.1 Extraction of data from pre-existing cohorts:

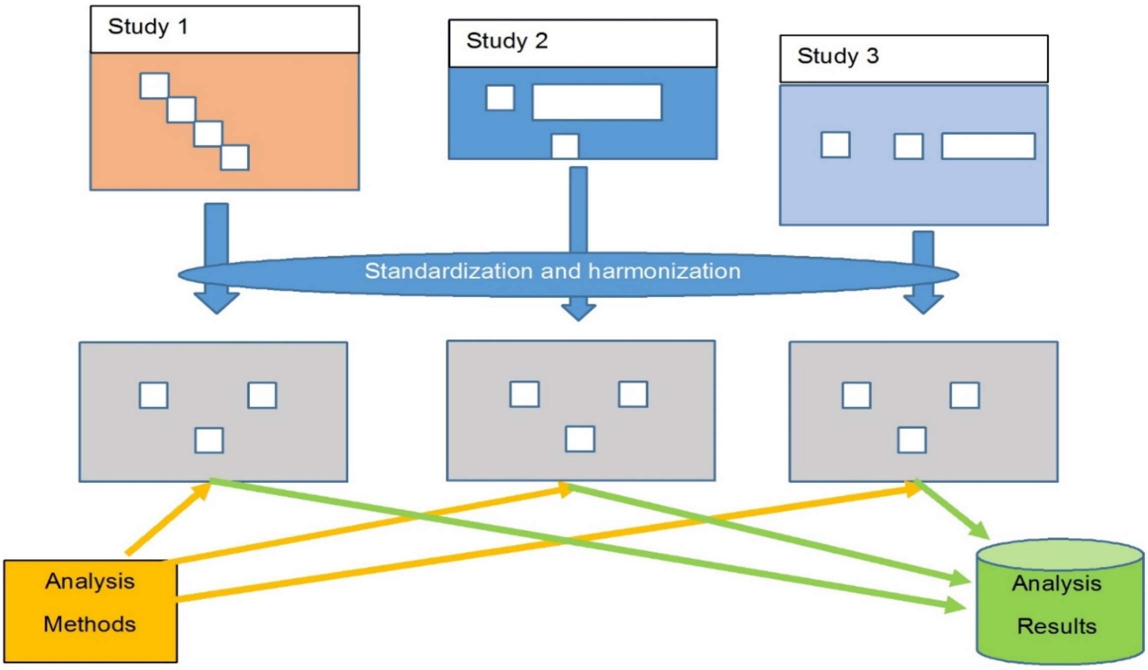
As mentioned, data for this study is assimilated from different European studies that include birth cohorts as well as general population studies.

2.1.1 Standardization of data:

In order to harmonize the data sets from different cohort studies, a standardized format of the variables were created in an excel file. The variables included in this harmonization file pertained to WP 14, 15 and 16, with several categories of variables in separate worksheets of the same file. The individual worksheets include 'Identification and individual characteristics', 'Health data by questionnaire and medications used', 'Factors and exposures', 'Exposure assessments',

‘Neurodevelopment and neuropsychological outcome’, ‘Additional Data’ and ‘Comments regarding the preparation of the data’. For each of the variables, a description, a common variable name, the format of the variables (character/numeric categorical/numeric binary) with the maximum number of digits or letters; along with its domain, core and role are included in the file. This file was sent to the different partners who are responsible for the individual cohort studies, such that they can retrieve the information on the required variables from their studies using the standardized format.

Figure 1: Standardization and harmonization of data



2.1.2 Contacting the partners:

The partners were contacted via email, with the request for providing their data in the specific standardized format. In addition to the data harmonization file, a ‘Data Request and Transfer Agreement Form’ was also prepared and attached to the email. This agreement form was duly signed by the work-package leaders of WP 14-16 and each party in possession of the data. For the purpose of data protection and confidentiality, the individual partners were requested to send their data sets encrypted by email to the members responsible for the mentioned work-packages of

the HEALS project. The partners were first contacted in December 2015. However, several follow-up emails were sent to the partners regarding the transfer of data.

2.1.3 Singleton cohort studies from which data was requested:

Several pre-existing European cohorts were considered for the analysis. While data has been obtained from most of the cohorts, some are still at the preparatory stage and would be considered for other analyses within the project. The following table indicates the singleton cohorts considered and the data obtained so far.

Table 1: Pre-existing Singleton cohorts

Cohorts considered	Data obtained
REPRO-PL	√
PI1	√
PI2	√
IMCA	√
SEASD	√
Indoor School CCM	√
PEGGS	
RESPIRA	
ARPA	
PHIME-CRO	√
PHIME-SI	√
HBM-SI	
DEMOCOPHES SI	
Generation XXI (G21)	√
EPITeen	√
Sinphonie	
EDEN	√
FERMA	
6 Cities Study	
CHISQ200	√
MAAS	√
MoBa	

2.1.4 A brief description of the individual cohort studies

While many of the considered cohorts are birth cohorts, with information on mother and child; some studies are based on general population samples. Below are the descriptions of some of the important cohorts that are considered for our analysis.

2.1.4.1 Birth Cohorts:

EDEN:

Country: France

General Overview: The EDEN study is the first French cohort study, which is based on the pre and early postnatal determinants of child health and development. This study involves a number of INSERM groups that specializes in epidemiology and obstetricians, and is in collaboration with paediatricians from the Nancy and Poitiers Universities. INSERM is the ‘promoter’ of the EDEN study. It has received approval from the ethnic committee (CCPPRB) of Kremlin Bicêtre on December 12th, 2002.

Study Design: EDEN is a longitudinal study that follows children as early as the second trimester of pregnancy until 5 years of age. Recently most of the families agreed to a longer follow-up. Data are also collected from mother and father, with clinical examinations, and there is a large collection of stored blood samples.

The study is carried out in two centres, the maternity hospital in Nancy and Poitiers. The study was proposed to all women presenting at the prenatal clinic before the 24th week of amenorrhea (WA).

Enrolment in the study started in Poitiers in February 2003 and in Nancy in September 2003, and lasted until January 2006. Two thousand and two women were enrolled.

The post-partum study included Interviewer- and self–administered questionnaires to the mother; clinical examination of the mother; clinical examination of the father with biology and self-administered questionnaire; specific clinical examination of the new born.

Subsequently, the follow-up of the child was carried out at the first-year of the child, age one, age two, age three, age four and age five.

Aim: The general purpose of the study is to analyse the developmental plasticity of the foetal and early life period when exposed to environmental toxics. It further intends to provide insight into the potential mechanisms and establish the contribution of these phenomena to later health and development of the child in comparison with other environmental factors operating during childhood.

In regard to the child’s health and development, specific impetus is provided on the biological and clinical aspects, that involves the growth and development of the adipose tissue, the masculine genital tract, teeth, as well as immune, respiratory, cardiovascular and metabolic functions behavioural, cognitive and psychomotor development.

Generation XXI:

Country: Portugal

General Overview: Generation XXI is the first Portuguese birth cohort and intends to characterize prenatal and postnatal development, identifying its determinants in order to understand the state of health in childhood and later in adolescence and adult age. Thus, Generation XXI aims at attaining novel and useful knowledge for understanding the Portuguese reality.

Study Design: The recruitment of 8647 newborns occurred between April 2005 and August 2006 at the five public maternity units in the metropolitan area of Porto. Interviewers collected data on socio-demographics, clinical and reproductive history, prenatal care and about lifestyle and behaviours before and during pregnancy. Both parents’ weight and height were measured and detailed anthropometrics were performed on the new-borns. The medical records were assessed and for a subgroup of participants’ blood samples from the mother, father and the umbilical cord were collected. This project aims to follow these children up to adulthood.

At 6, 15 and 24 months, sub samples of babies were examined and at 4 years of age, the whole cohort was invited to participate. This evaluation occurred between April 2009 and August 2011. During this period 7300 children were evaluated.

Between April 2012 and March 2014, all participants were once again invited to participate at 7 years of age, which led to the evaluation of 6900 children.

At this moment, a new wave of evaluation is ongoing at a time when the children have already attained 10 years of age.

Aim: Project Generation XXI is the first Portuguese cohort study that aims at identifying pregnancy characteristics and the early stages of childhood that relate to the development and health at subsequent stages of life. Thus, the main objective of this study is the acquisition of new knowledge in order to scientifically explain the current health challenges of children in Portugal.

This study intends to better understand the multiple aspect of growth and child development; and the early challenges in the health of children. This would enable faster intervention and better prognosis of various health disorders (like eye and hearing problems, or oral health).

REPRO_PL:

Country: Poland

General Overview: The Polish Mother and Child Cohort Study (REPRO_PL) is a multi-centre prospective cohort study conducted in different regions of Poland. The study is mainly designed to identify the exposures that may influence pregnancy outcome and children's health. The verification of such exposures is done by biomarker measurements and notification of any changes in exposure levels.

Study Design: The final cohort comprise of 1800 mother-child pairs to be recruited within 4-year period (2007–2011). The recruitment and all scheduled visits are conducted in maternity units or clinics in the districts participating in the study.

The children are followed-up until the age of 2 years. The incidence of upper and lower respiratory tract infections, middle ear diseases, and the symptoms of allergy among the children are identified. BSID-III is used for the assessment of neurodevelopment in children.

Aim: The aim of the study is to evaluate the impact of exposure to different environmental factors on pregnancy outcome and children's health. Specific research hypotheses refer to the role of heavy metals, exposure to phthalates, polycyclic aromatic hydrocarbons (PAHs) and environmental tobacco smoke (ETS) in the aetiology of intrauterine growth retardation (IUGR), preterm delivery (PD) and

the risk of respiratory diseases, allergy and poor mental and physical development. This study also intends to explain the role of oxidative stress and nutritional status of the pregnant women. The impact of occupational exposures and stressful situations on pregnant mothers will also be evaluated from the questionnaire data.

The results of the study would eventually determine the levels of child prenatal and postnatal exposure in several areas of Poland, in order to study the impact on pregnant mothers and subsequently the health and neurodevelopment of the child.

MAAS:

Country: United Kingdom

General Overview: The Manchester Asthma and Allergy Study (MAAS) is an unselected, population-based prospective cohort study that investigates the relationship between genetic predisposition and indoor environment in the development of atopy and asthma. It is set in the urban and rural catchment areas of Wythenshawe and Stepping Hill Hospitals.

Study Design: The study follows a cohort of 1640 children. A trained nurse and a clinician screened all mothers and their partners at the antenatal visits. Along with the completion of the questionnaire (a validated ISAAC questionnaire for symptoms of asthma, eczema, hay fever and food allergies), consent was taken from the couples for undergoing skin prick testing (SPT) to 4 most common inhalant allergens (house dust mite [HDM], cat, dog and grasses). The duration of the recruitment was between October 1995 and July 1997. Altogether 5890 mothers attended 654 antenatal clinics. Of these 1306 were not interested in the study and 1058 failed to meet the entry criteria (e.g. more than 16 weeks pregnant, approximately 18 years of age, etc.). Therefore, a total of 3526 mothers and 2034 fathers were skin tested. The children had their lung function tested along with the skin testing for allergies. In addition, most of the children provided blood for further allergy testing and for genetic studies.

Aim: This birth cohort study is designed to comprehend the development of asthma and other atopic disorders in a cohort of children. Since the complex etiologies of asthma or allergies involve both genetic and environmental conditions, the study was initiated from the prenatal stage (8-10 weeks of pregnancy), and subsequently the

children were assessed around their 1^{st}, 3^{rd}, 5^{th}, 8^{th} and 11^{th} birthdays. Further follow-up assessments are underway.

PHIME:

Country: 22 countries (35) within the EU's Sixth Framework Programme for Research and Technological Development. For the HEALS project, Croatia and Slovenia are considered.

General Overview: PHIME (Public health impact of long-term, low-level mixed element exposure in susceptible population strata) was an Integrated Project within EU´s Sixth Framework Programme for Research & Technological Development, during the period 2006-2011. The background of the study was a renewed interest in toxic metals and the toxic effects in susceptible individuals of Europe. Such exposures may have a role in the etiology of common clinical diseases, as well as sub-clinical effects. Further, novel methods (analytical, effect markers, molecular biology and epidemiology) had been developed, which facilitated better assessment of health risks.

Within the framework of the PHIME cohort, several studies have mainly dealt with the impact of environmental stressors on pregnant mother and child, as they constitute a potential risk group.

Study Design: To undertake an international comparison of metal exposure in children and women, blood samples from 1,363 children from six European (Croatia, Czech Republic, Poland, Slovakia, Slovenia, and Sweden) and three non-European countries (China, Ecuador, and Morocco) were collected. Subsequently, the level of mercury in the children's blood was analysed. Additionally, the blood samples of 480 women across Europe were also taken into account to identify traces of mercury, platinum, palladium and rhodium elements. For the HEALS project, PHIME data for Croatia and Slovenia has been considered.

Aim: PHIME focused on producing scientific data, through close cooperation between research groups with knowledge, methods and study groups that complement each other. A closely cooperating working party was formed, which planned and executed collaborative studies and discussed findings, resulting in more than 180 articles in peer-reviewed (mostly international) journals.

A wealth of important, novel information has been produced from these studies. It has been found that the toxic effect of methyl mercury on the central nervous system of foetuses and the myocardium of adults is markedly modified by nutrition. Arsenic and manganese, ingested mainly through drinking water and food, affect development and health of foetuses, infants and children. Lead exposure is toxic to children´s central nervous system at very low exposures.

Gene-environment interaction is important in metal toxicity. Thus, the metabolism (toxicokinetics) of mercury, arsenic, and lead is modified by genetics, as is toxicodynamics of arsenic, lead, cadmium and manganese. This should be considered in risk assessment, as the risk may vary between individuals, and between populations with different gene frequencies. Taken together, the scientific results contributed by PHIME have brought new, important insights to the health impact from toxic metals and risk assessment on children.

MoBa:

Country: Norway

General Overview: The Norwegian Mother and Child Cohort Study (MoBa) is a unique study of Norway, which delineates the causes of diseases among the mothers and children. It intends to look into the effect of toxins, infections, dietary factors, work stress or other factors (exposures) on the health conditions during pregnancy and early childhood, as well as on the child’s later health.

Study Design: MoBa began to recruit pregnant women in 1999. Fathers were also invited. In 2008, the goal of more than 100 000 pregnancies was reached. Biological material and questionnaire data have been collected since the 17th week of pregnancy. There was an eventual follow-up of the mothers and their children over time.

Aim: The MoBa research is specific and is based on defined problems that take into account the prevalence and diversity of causal factors and the incidence of diseases. On adding together the prevalence of many serious conditions (like cancer, stillbirth, severe congenital malformations, diabetes, arthritis, severe psychiatric and neurological diseases), it is observed that many people are affected even at a young

age. The study aims at investigating the causal chain of diseases and the perception of health as a concept with many dimensions.

2.1.4.2 General population samples:

EPITeen:

Country: Portugal

General Overview: The EPITeen project (Epidemiological Health Investigation of Teenagers in Porto) was initiated in 2003. The participants for this study were adolescents who were born in 1990, and attending public and private schools in Porto in the 2003-2004 academic year.

Study Design: The study includes 2943 participants who were evaluated at 13, 17 and 21 years of age. These participants will be evaluated throughout their life span. An evaluation for the 4^{th} time in underway. The evaluation process consist of collection of relevant information through questionnaires or objective measurements like weight, height, respiratory function, blood pressure, blood analyses, bone mineral density etc.

Aim: The aim of this study is to understand the habits and behaviours of adolescence that is reflected in adult health conditions. The information, thus obtained, would eventually lead to health promotion planning by meeting several scientific goals. Data accrued from the EPITeen project, till date, has already been used to answer several scientific research questions through the publication of papers in peer-reviewed journals.

CHIS2000:

Country: Spain (Catalonia)

General Overview: Catalonia is a Mediterranean region of 32,000 km^2 in southern Europe, with a total population of 6.5 million inhabitants as of 2002. Its major economic activities include agriculture and industry. The cohort is based on a public health survey (n = 8,400) that was conducted by the Government of Catalonia in 2002, including a health exam and blood tests (n = 2,100).

Study Design: The survey provided a valuable and representative sample of the general population in terms of age (18-74 years), sex and socio-demographic conditions. Information on body mass index, diabetes was obtained from the health exam. Information on the demographic variables, such as age, sex, place of birth, educational level and parity in women was obtained from face-to-face interviews that were conducted between October 2001 and April 2002. Social class was estimated through the household occupational status based on the Spanish Occupational Classification (Grupo de trabajo SEE-SEMFC, 2000). Further details are available in Juncà et al. (2003) and Porta et al. (2010).

Aim: The study aims to understand the health conditions and the socio-economic impact on the well- being of the Catalonians.

CCM, IMCA, PISA1, PISA2, SEASD:

Country: Italy

General Overview: The Pulmonary Environmental Epidemiology-EPAP Unit of the Institute of Clinical Physiology, CNR, Pisa (head dr. Viegi), previously performed a large longitudinal study on Italian general population sample living in the urban and sub-urban area of Pisa (Central Italy) with the objective of studying the natural history of respiratory diseases (i.e. Chronic Obstructive Pulmonary Disease - COPD, asthma and allergic rhinitis) and its risk factors.

Study Design: In 1985-88, a randomized, stratified family cluster sample (N=3855; age 5-90 year) was selected to participate in the first cross-sectional study (PISA 1) based on an interviewer-administered questionnaire on respiratory symptoms/diseases and risk factors.

In 1991-93, the same sample, along with the new family-members, (N=2841; age 8-97 year) were invited to participate in the second cross-sectional survey (PISA 2), based on the same questionnaire as well as lung function tests, bronchial responsiveness challenge to methacholine, allergological evaluation through skin prick tests, total IgE and mutagenetic determinations (sister chromatid exchanges, micronuclei), hemoglobin and DNA adducts to benzo(a)pyrene).

In 2009-2011, the same sample, along with the new family-members, (N=1620; age 18-103 year) were invited to participate in the third cross-sectional survey performed a European project (IMCA2- Indicators for Monitoring COPD and Asthma in the EU). The subjects filled in a questionnaire on respiratory symptoms/diseases and risk factors and a sub-sample performed lung function test, blood sample analysis, blood pressure, pulse-oximetry, weight and height measurements.

In 1997-1998, non-smoking women performing PISA 2 study were invited to participate in the cross-sectional SEASD study ("Epidemiological Study on Environment and Health in Women from four Italian Areas") aimed at investigating the differences in demographic, environmental and medical characteristics between women exposed and unexposed to environmental tobacco smoke (ETS). The subjects (N=639; age 13-99 year) filled in a questionnaire on respiratory symptoms/diseases and risk factors, passive smoke exposure and preventive lifestyle behaviours, performed blood sample, urine and saliva analyses and blood pressure measurements.

The CCM Indoor-School project ("Exposure to indoor pollutants: guidelines for the assessment of risk factors in school environment and definition of measurements for school-children and adolescents respiratory health protection") involved school-children of 7 Italian Regions (Friuli Venezia Giulia, Lombardia, Toscana, Lazio, Puglia, Sicilia, Sardegna) and 8 cities (Pisa, Milano, Sondrio, Udine, Bari, Cagliari, Palermo, Roma) in the period 2010-2014.

Aim: The aim of the project was to assess the relationship between children respiratory health and indoor air quality (IAQ) in the school environment. The parents filled in a questionnaire on the children respiratory symptoms/diseases and risk factors exposure (N=2573; age 5-15 year); a subsample of the children performed skin prick test and spirometry. Environmental measurements ($PM_{2.5}$, NO_2, VOC) were performed inside and outside the classrooms.

The cities that were involved in the studies for the considered cohorts are situated at different countries of Europe. The map below illustrates the main cities considered for the cohort studies in their respective geographic locations.

Figure 2: Main cities considered in the different studies

2.1.5 Data Cleaning

The preparation of the final harmonized data by amalgamating the datasets obtained from the pre-existing studies involved thorough data cleaning processes. This involved the identification of inaccurate, incomplete, irrelevant or duplicate data, and their subsequent corrections or removal. For this the members responsible for different cohorts were contacted for plausible explanation on certain ambiguities in the data sets, and they were eventually modified and cleaned. Although this process involved a significant amount of time, several inconsistencies in the data sets were detected and were removed. This led to an improvement of the data quality to a great extent. The final data set consisted of 28608 individuals.

2.2 Environmental Exposures

2.2.1 Environmental exposure data from the pre-existing cohorts

The acquired data from the pre-existing cohort studies have provided information on environmental exposures at the individual level. The EDEN study of France includes data for pollutants like benzene, carbon-monoxide (CO), volatile organic compounds (VOC), nitrogen-dioxide (NO_2), sulphur-dioxide (SO_2) and particulate matter (PM_{10}), for the time period of 2003-2006. In case of the Italian data, CCM study provides information on outdoor and indoor $PM_{2.5}$, PM_{10} and VOCs (in both home and classroom) for 2010-2014. While information on SO_2 is available for PISA2 cohort study, no information on pollutants is available for IMCA, PISA1 and SEASD. Data on PM_{10} for 2007 is also available for the Polish study REPRO-PL.

Regarding other environmental stressors, information on toxic metals and chemicals like polychlorinated biphenyls (PCBs) and hexachlorobenzene (HCB), presence of heavy metal (like Hg, Pb, Cd, As etc.) in the plasma or cord blood of mother, phthalates and other EDCs in urine samples were obtained from different cohort studies. Information on maternal smoking habits was present in EPITeen, Generation 21, REPRO-PL, CCM, PHIME and EDEN. In addition, smoking habits of father as well as passive tobacco exposure is also available in most of the cohorts that are used for our analysis. Although smoking habits have not been included in our analysis, the level of cotinine in plasma, saliva and urine are obtained from some of the studies and considered as an EDC.

2.2.2 Environmental Data Management system (EDMS)

In addition to environmental data obtained from the pre-existing cohorts, the EDMS database platform of HEALS also stores exposure data as collected in Task 8.1. This database, which comprises of both HEALS data and other associated existing databases are retrieved through suitable query script. The standard database package of the EDMS has been implemented to perform interoperability in data storage with HEALS geodatabase as developed in WP12.

The EDMS data that could be accessed presently is at the aggregate level for different cities, which coincide with the cities included in the health study; spanning

across a period of 1976-2012. But the time span varies across different cities, owing to the different availability of data. The exposure data from the EDMS include information on particulate matter (PM_{10}, $PM_{2.5}$), carbon monoxide, ozone, nitrogen-dioxide, toxic metals and several other priority pollutants are included from the EDMS data base.

Table 2 represents the cities considered from the EDMS database with the trend in the annual means of major air pollutants from 2006-2011. The reason for considering the specific time frame and the major five pollutants for all the 14 European cities is the maximum number of observations that could be accrued to represent the trend on the pollutants. Moreover, the latest years depict the recent trend in air pollution in these cities of Europe.

As can be seen, the overall trend of sulphur-dioxide (SO_2) has increased by more than 5% over the 6 year period. An increase in SO_2 level is seen for the Greek cities of Athens and Thessaloniki, as well as in Palermo, Porto, Lodz and London. A significantly high trend in SO_2 rise has been observed in London, which results in an overall increasing trend in pollutants in this city. Industrial activities like wood pulp conversion into paper, coal and oil combustion and incineration of refuse are some of the prominent factors that influences the level of SO_2 in the atmosphere. The hygroscopic nature of this gas causes it to react with atmospheric humidity that produces sulphurous and sulphuric acid that constitutes in the acid rain. Consequently, this damages soil, water pipes, flora and fauna alike. Moreover, this pollutant is highly dangerous for human health and can cause several respiratory and cardiovascular disorders. Therefore, industrial emissions should be appropriately monitored in order to recede the level of this harmful stressor in the environment.

An increase in the level of Ozone (O_3) is observed for Athens, Thessaloniki, Zagreb and Manchester, while a rise in PM_{10} is observed for Athens and Lodz. A rise in CO and NO_2 is witnessed in case of Porto, while pollution in Regensburg depicts a notable rise in CO. Although all these atmospheric pollutants have significant health impacts leading to morbidity and mortality, its genetic and epigenetic effect on offsprings due to gestational exposure is yet to be discerned.

Table 2: Trend of major pollutants over the period 2006-2011

Region	CO (mg/m3)	NO2 (µg/m3)	O3 (µg/m3)	PM10 (µg/m3)	SO2 (µg/m3)	**Overall**
Athens	→	→	↗	↗	↗	↗
Thessaloniki	↘	→	↗	→	↗	↗
Roma	→	→	→	→	→	→
Palermo	↘	→	→	→	↗	→
Pisa	→	→	→	→	→	→
Paris	↘	→	→	→	→	→
Regensburg	↗	→	→	→	→	→
Porto	↗	↗	→	→	↗	↗
Tarragona	→	→	→	↘	→	→
Zagreb	→	→	↗	↘	↘	→
Ljubljana	↘	→	→	→	→	→
Lodz	→	→	→	↗	↗	↗
London	→	→	→	→	↗	↗
Manchester	→	↘	↗	↘	→	→
Overall	→	→	→	→	↗	→

→ : Not more than 5% change over the 6 year period

↘ : More than 5% decrease over the 6 year period

↗ : More than 5% increase over the 6 year period

For this purpose, other stressors like tobacco exposures and presence of harmful chemicals and metals in the maternal serum or urine has been collected in several of the cohorts that we have considered for our study.

2.2.3 Linking health data to environmental data

Since the information on environmental stressors obtained from the cohort studies are at an individual level, we have used this data for our analysis. The missing data for the pollutants are imputed by the multiple imputation methodology.

However, for our present study, we have not used the EDMS data. As already discussed, the EDMS data is at an aggregated level and the desired pollutants and duration is not harmonious to our health data. It was initially postulated that

geographic coordinates would be used for each observation, in order to link the EDMS data to the harmonized health dataset. By this method, all data would be geo-referenced based on the same geographic projection system. However, since the health data from many of the cohort studies does not provide the geographic coordinates of the individuals for privacy reasons, the two datasets could be linked only at an aggregate level.

Additionally, the information on cities are not provided by some of the cohort members. Therefore, the information on cities for those studies are retrieved from the websites of the cohort surveys undertaken. Moreover, the pollution data from the EDMS is represented as annual averages. Since the aggregation of exposure data as a yearly average at the city level is a prominent approximation, we have omitted the EDMS data for our present analysis. Nevertheless, the analysis will be furthered using the EDMS data at a later stage. We aim at exploring the EDMS data in greater details, such that it could be utilized for a sensitivity analysis of our results at hand.

2.2.3.1 Limitations with geo-referencing

Due to ethical reasons and confidentiality issues that arises to protect the identities of the individuals, many studies do not provide the geographic co-ordinates of the individuals from their respective studies. The access restriction to specific details of the research subjects varies for different cohorts. However all studies maintain specific procedure such that the exact address location of the individuals cannot be deciphered to protect their privacy.

The table below indicates whether information is available for specific geographic references of each of the cohort studies, which includes GPS coordinates, address, city names and postal code. As can be seen, the information on GPS longitudes and latitudes are not present for some of the cohorts. In addition, the table also provides the respective countries included in the studies and the corresponding study periods.

Table 3: Geographic and time references of individual cohorts under study

Cohort	Countries	Study Period	GPS coordinates	Address	City	Postal Code
REPRO-PL	Poland	2006-ongoing	×	×	√	×
PI1	Italy	1985-1988	√	×	√	√
PI2	Italy	1991-1993	√	√	√	√
IMCA	Italy	2009-2011	√	√	√	√
SEASD	Italy	1997-1998	√	×	√	√
Indoor School CCM	Italy	2010-2014	×	×	√	√
PHIME-CRO	Croatia	2006-2011	√	×	×	×
PHIME-SI	Slovenia	2006-2011	√	×	×	×
Generation XXI (G21)	Portugal	2005-2006, 2009-2012, 2012-2014	×	×	×	×
EPITeen	Portugal	2007-2013	×	×	×	×
EDEN	France	2003-2006	×	×	√	×
CHIS2000	Spain	2001-2002	×	×	√	√
MAAS	United Kingdom	1997-2003	×	×	×	×

Present: √

Not Present: ×

3 Imputation Methodology

Both health data and air pollution data in our study is fraught with the problem of missing values. This problem of missing data is commonly seen in case of clinical epidemiology. Since mishandling of missing data can lead to biased or incorrect statistical results, we have done a thorough examination on the structure of the data such that the most relevant missing value imputation method is used. The potential challenges posed by missing observations can be averted with the appropriate methodology and its implementation.

The missing data pattern is categorized into three types, which includes, missing completely at random (MCAR), missing not at random (MNAR) and missing at random (MAR) (Rubin, 1976). MCAR denotes the probability of an observation being missing is independent of the observed and unobserved values of the data. In mathematical notation, $Pr(V^{mis} \mid V^{uob}, V^{ob})=Pr(V^{mis})$; where V^{mis} is the missing observation and V^{uob} and V^{ob} are the unobserved and observed data respectively. However, in case of MAR, the missingness of a variable depends on the observed data and not on the unobserved data. So, in this case, $Pr(V^{mis} \mid V^{uob}, V^{ob})=Pr(V^{mis} \mid V^{ob})$. When the former two categories fail to hold, the missing data type is said to be MNAR. This implies that the missingness of the data is not at random and it actually depends on the unobserved data. That is, $Pr(V^{mis} \mid V^{uob}, V^{ob})=Pr(V^{mis} \mid V^{uob})$. Mostly, in case of epidemiology, the missingness is mainly of MAR type.

3.1 Different methodologies for minimizing missing data

Simple imputation techniques are more pertinent when the number of missing observations are not large. Simple imputation techniques include mean or median imputation, regression-based imputation, sensitivity analysis considering the worst or best value, or a more simplified complete case analysis. However a complete case analysis is not representative of the whole data and it leads to lack of precision of the data due to reduced statistical power and too large standard error. On the other hand, the other methods may lead to too small standard errors and consequently an overestimation of precision of results.

Nevertheless, the large number of missing values in our data set impels us to use a suitable multiple imputation (MI) technique. This imputation scheme is more appropriate for our analysis, as the data set has mostly MAR type of missing data, and each missing value is imputed by a set of probable values from the conditional distribution of the missing data, with the observed data considered as given. It is more appropriate as it allows asymptotically unbiased estimation, with the considerably weaker assumption of MAR. It mitigates the problem of extreme standard errors as obtained from traditional methods, with the aim of providing unbiased estimations from the available information. Although the multiple imputation technique is more viable when using the MAR assumption, this methodology is also

appropriate while handling both MCAR and MNAR. The statistical theory behind the multiple imputation technique is explained by Carpenter and Kenward (2013).

Likelihood based approach, probability weightings and full Bayesian models are alternative methods to multiple imputation technique. Since the maximum likelihood (ML) technique uses all observed data in order to impute the missing observations, variables which are associated with the missingness can be beyond the ones that are included in the analysis model. Additionally, more complex algorithms may be required by the missing predictor variables. However, in case of the inverse probability weighting (IPW), the complete cases are being weighted by the plausibility of it being observed. Since a subset of the available information is being used, this technique is considered to be less accurate than the ML methodology. On the other hand, the full Bayesian data estimates both missing data and analysis model simultaneously. In case of availability of prior evidences, this method is of advantage and it approximates multiple imputation method. But this method may have convergence issues and can lead to complexities in case of missing covariates.

Therefore, in consideration of all the imputation techniques, the MI can be considered as the most appropriate technique for epidemiological analysis. The main advantages of this method which provides a leverage over all other imputation techniques are, the allowance of the inclusion of auxiliary variables, imputation of missing values for both outcomes and the covariates, recognition of uncertainty associated with the missingness and the provision of a suitable framework for sensitivity analysis.

3.1.1 Multiple imputation methodology

The MI methodology that we have used for the imputation of our incomplete data set is a simulation based approach, whereby it replaces missing values with multiple sets of simulated outcomes such that it holds valid statistical inferences (Rubin, 1996).

Before performing the imputation technique, we reduce our main data set with 879 variables into a sub sample of the data having 223 variables, which are relevant to our line of study based on obesity and diabetes. However, owing to the large number of missing data, variables with more than 70% missing values were deleted. This

resulted in a reduction of the data, with 70 variables in the final study. The final data comprised of the environmental factors, demographic characteristics and the response variables.

Although the credibility of MI broadly depends on infinite number of imputations (N), this method is also statistically robust with finite N. For a robust statistical inference with 95% relative efficiency (RE) and 50% missing information, the total M is suggested to be 5 (Rubin, 1987; Van Buuren et al., 1999). However, the number of imputations not only depends on the missing values, but also on the type of model and data used. The evaluation of the sampling error from MI estimates may lead to the determination of the number of imputations required (White et al., 2011). For our analysis, we have used 10-15 imputations of the different variables, such that the sampling error is reduced.

4 Statistical analysis

The Environment Wide Association Study (EWAS) of the HEALS project is based on the exposomic approach at an individual level. The study aims at evaluating the effect of endogenous exposures (measured by-omics) and environmental exposures (measured using environmental monitoring systems and estimated models) on the health conditions of individuals.

Although the Genome-wide association studies (GWAS) investigates the genetic components and identifies numerous single nucleotide polymorphisms linked with diseases, the common human phenotypes remain unclear. Thus the EWAS approach takes into account the human exposure to environment using a high through-put and unbiased approach, similar to genetic effects in case of GWAS.

For the empirical analysis, a Survey-Weighted Logistic Regression (WLR) is used. Since the harmonized data set is obtained from several cohort studies, the data involves unequal weightings, clustering and stratification. Hence a standard logistic regression is not pertinent in this context. Since WLR takes into account a number of weights and design effects to fit correlated data, we have applied this methodology for our analysis. For further details on the models of weighted logistic regression, please refer to Wilson et al. (2015).

A Bonferroni correction was used for the analysis. Due to the risk of making type 1 errors for multiple statistical tests, the Bonferroni correction is used to adjust the probability values. Therefore, this method is used to counteract the problem of multiple comparisons.

The health outcomes considered for our study includes Diabetes (Type 2 diabetes mellitus) and obesity (based on bmi level). We have also investigated the indirect effect of prenatal exposure to environmental stressors. For this, we have considered a subsample of the data, where only the birth cohorts are taken into account. Furthermore, we have also looked into the effect of maternal exposure on gestational diabetes.

For the purpose of EWAS study, 62 unique environmental assays are considered. This environmental loci is measured across the cases that account for diabetes and obesity, as well as the control factors. The environmental factors are either continuous or categorical variables. They can be systematically categorized as air pollutants, polychlorinated biphenyls, brominated diphenyl ether, metals, phthalates, cotinine and allergens.

The model is adjusted for age, sex, work position and educational level. For our analysis, we consider the environmental variables that are present in the pre-existing cohort studies. The advantage of using this data is that, it is available at individual level. However, it is fraught with several missing values, which are imputed using the multiple imputation technique.

For the statistical analysis, STATA (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.) and R software (R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) has been used.

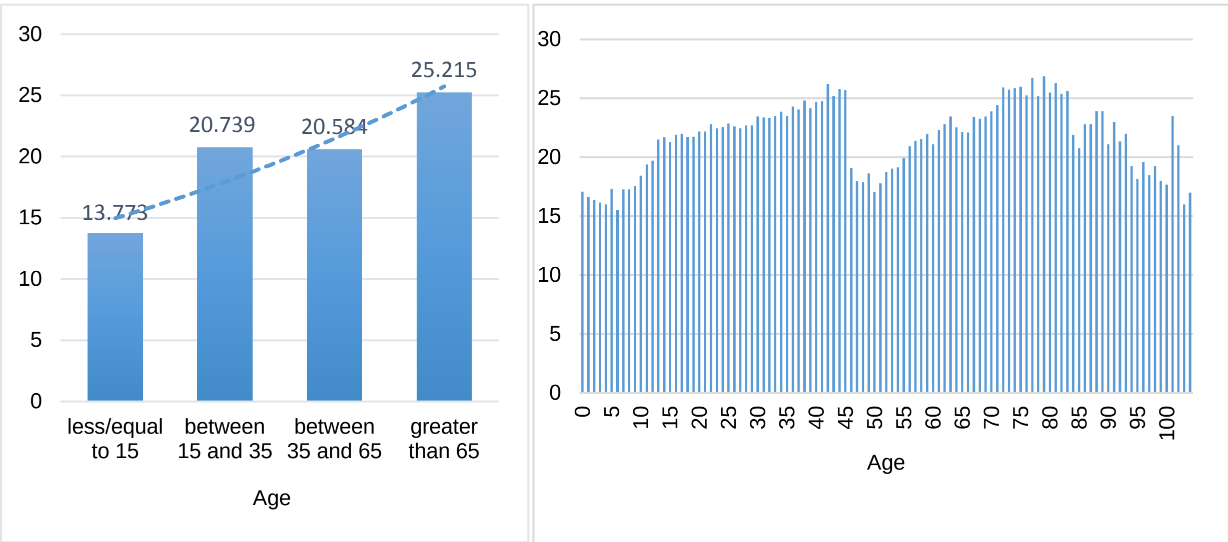
4.1 Descriptive Statistics

Our analysis mainly deals with the effect of 62 environmental stressors on type 2 diabetes and obesity. A glycated hemoglobin (A1C) test or a fasting blood sugar test were performed to detect the presence of Type 2 diabetes. An A1C greater than or equal to 6.5%, or a blood sugar level of 7 mmol/L or higher after an overnight fasting is used as the threshold level for detecting diabetes. The variable diabetes in

denoted in binary (0/1). On the other hand, bmi is measured as the ratio of weight (in kg) and height squared (in metres). Overweight is used as the predictor variables, which is categorized as 1 when bmi is greater than or equal to 25.0. Similarly, individuals are considered obese if their bmi is greater than or equal to 30.

From our data, it has been observed that 3.18% of individuals have diabetes. Concurrently, 6.71% of individuals were found to be obese (bmi>=30) and 17.37% of individuals being overweight (25<=bmi<30). However this percentages of population includes individuals of all ages and gender. Figure 3 demonstrate the average bmi across different ages and age categories. It is observed that the average bmi increases with age, with the highest average bmi accounted for age above 65 years. However the general European population in our study did not exhibit any sign of obesity.

Fig 3: BMI Average



In order to take into account observational variations that may occur between male and female, we also investigate on bmi levels based on gender categories along with age. It was seen that the average bmi does not vary prominently with gender classification. It is generally seen that women tend to have a higher percentage of body fat, while men are likely to have a greater proportion of muscle. Since BMI does

not distinguish fatty tissue from lean mass, we cannot infer on the gender specific body fat. Figure 5 represents the box plots for age and gender categories.

Fig 4: BMI based on Age and Gender categories

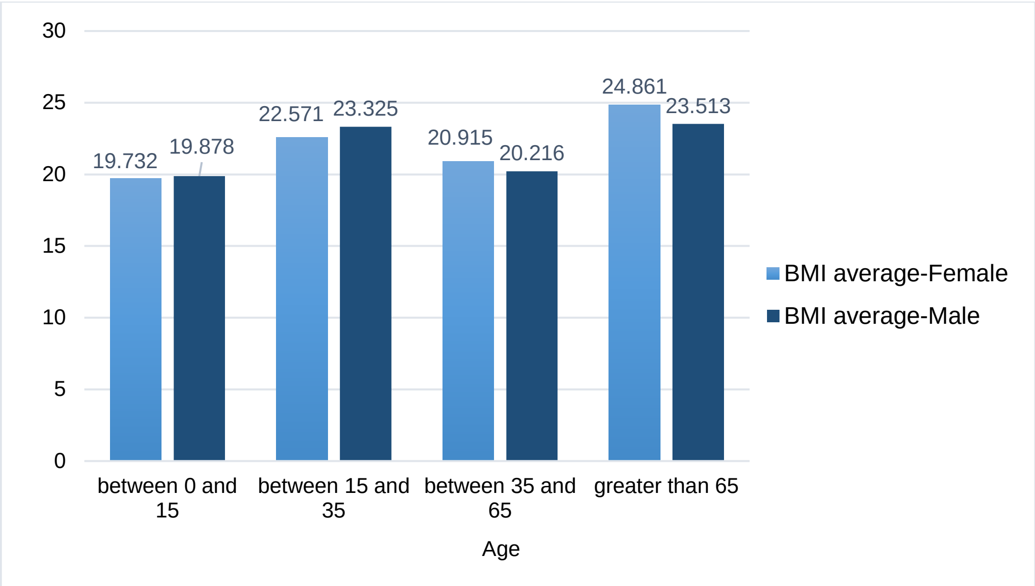
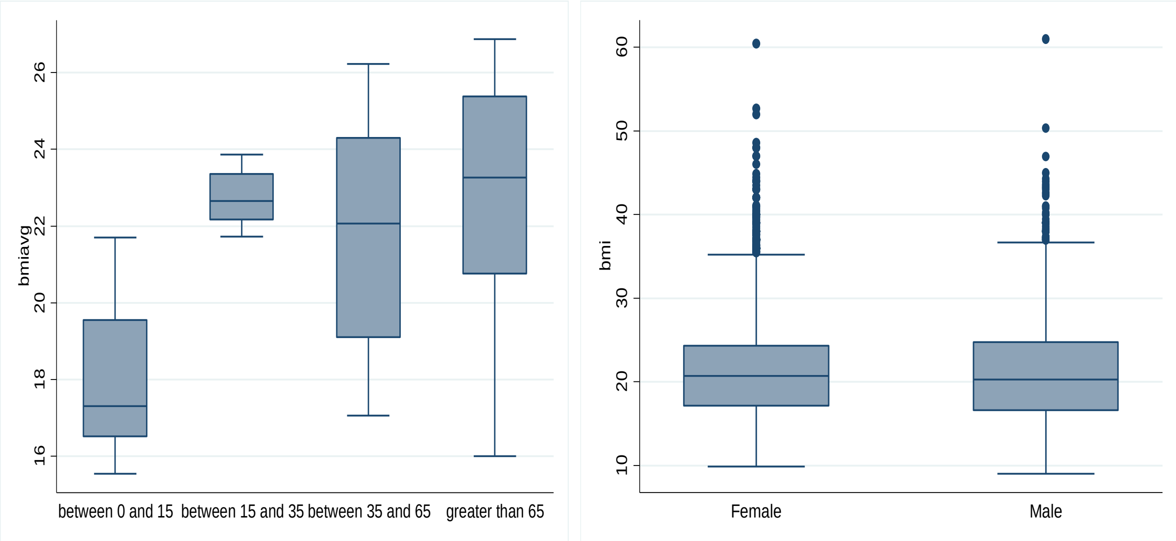


Fig 5 : Boxplot of BMI Average over age and gender categories



As can be seen in figure 6, the percentage of individuals with bmi>=25 increased with age, with 52.31% and 16.64% of individuals aged over 65 years are found to be overweight and obese respectively. At the same time, 11.16% of individuals aged over 65 years are found to suffer from type 2 diabetes mellitus (Figure 7). In general, type 2 diabetes is diagnosed more often among the elderly population, due to the loss of muscle mass and increase in body weight. Type 2 diabetes is generally associated with excess body weight, and it is usually diagnosed in individuals with high blood pressure or high level of cholesterol.

Fig 6: Percentage of overweight and obesity for each age category

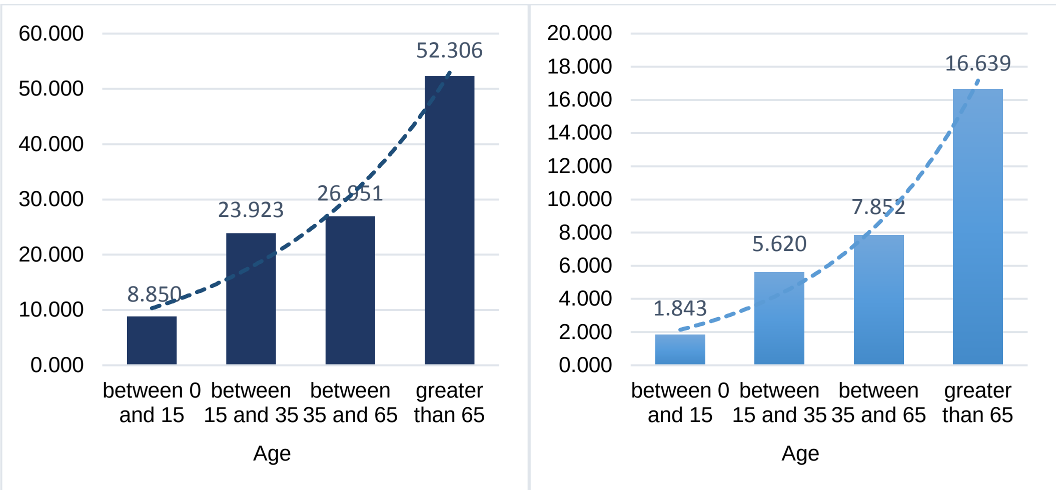
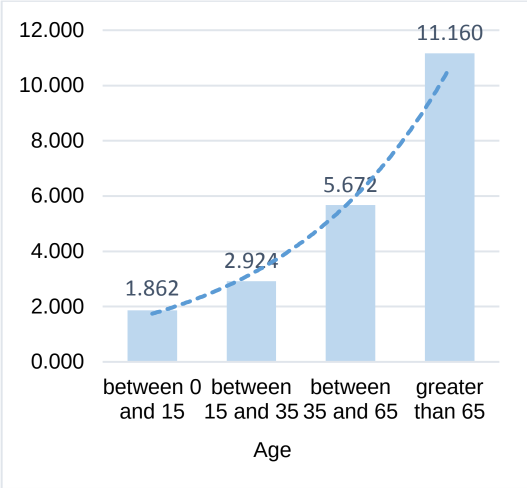


Fig 7: Percentage of individuals with diabetes for each age category



5 Empirical results

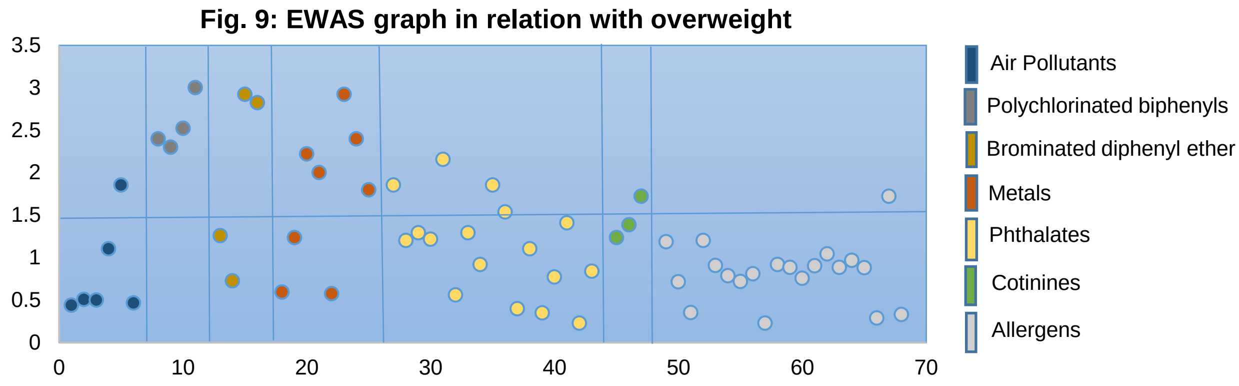
For the empirical analysis, we have estimated the effect of the array of air pollutants and environmental stressors on obesity and diabetes. Our analysis is based on an overall EWAS estimation to evaluate the effect of environmental pollutants on obesity and diabetes. In order to consider the indirect effect of environmental stressors on offspring due to the complex gene-environment interaction in maternal physiology, we have also analyzed a subsample of the data that is accrued from the birth-cohorts. An investigation on the effect of EDCs on gestational diabetes was also done.

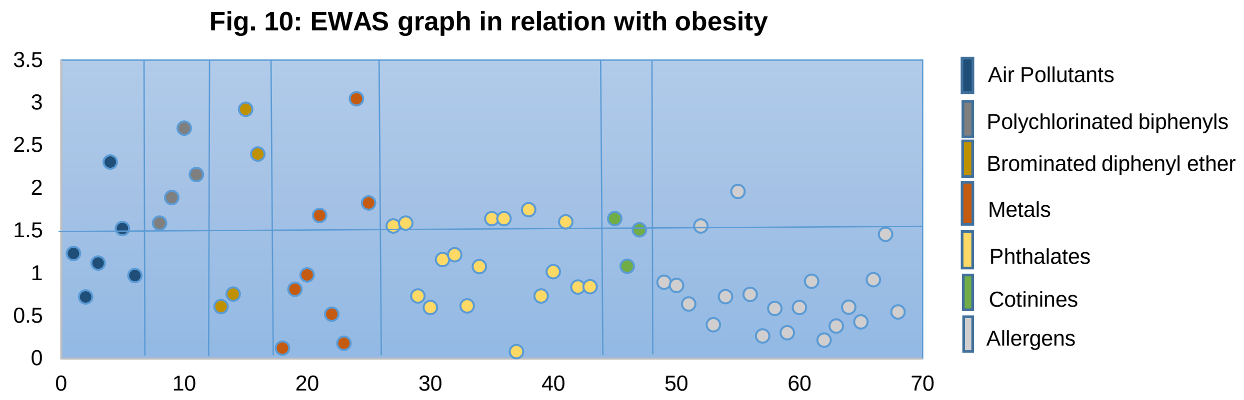
5.1 An overall analysis to investigate the direct effect of environmental exposures

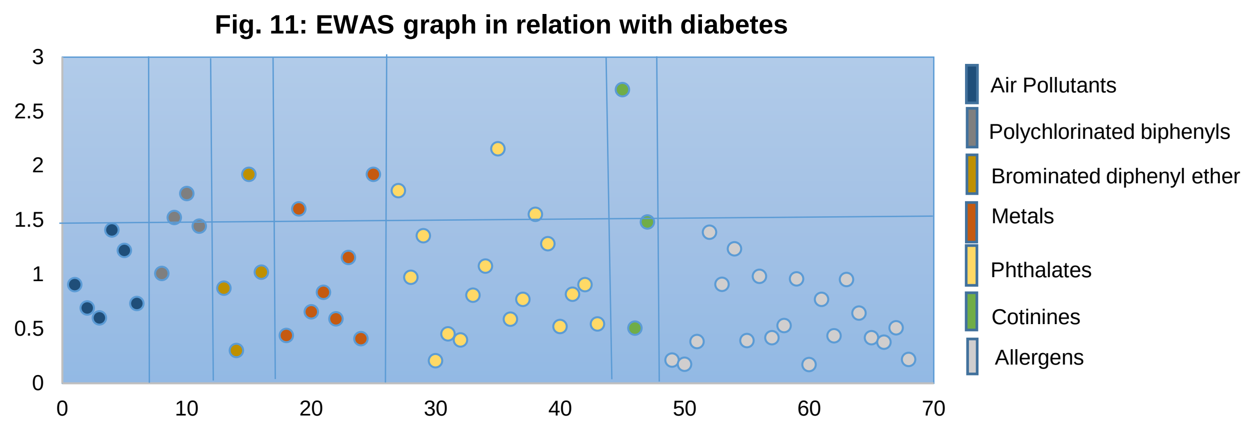
The overall analysis was based on the data from all cohorts, which includes both birth cohorts and general population studies. We have systematically assessed the impact of each of the environmental stressors on overweight, obesity and diabetes. Many of the pollutants were found to be significantly associated with excess weight gain or diabetes.

Manhattan plot inspired graphical representations (figures 9, 10 and 11) are used to illustrate the effect of the considered stressors on our outcome variables respectively. The y-axis of the graphs represent –log(p) values, with a cut-off level of

1.5. The stressors having a –log(p) value above 1.5 are considered to be significant, which are subsequently presented in table 4, 5 and 6. The gamut of environmental factors are classified using different colours for each of the environmental groups.







Although overweight, obesity and diabetes are essentially linked with biological and behavioral factors, our estimated results suggest that environmental factors play a preponderant role as well. Among the air-pollutants considered, it can be seen from Table 4, 5 and 6 that outdoor NO_2 and PM_{10} have a significant association with overweight and obesity. While outdoor NO2 was significantly associated with both overweight (adjusted OR is 1.109, p-value=0.014) and obesity (adjusted OR is 1.276, p-value=0.030), outdoor particulate matter was also found to strongly affect obesity (adjusted OR is 1.174, p-value=0.005). Likewise, outdoor exposure to PM_{10} was significantly related to type 2 diabetes (adjusted OR is 1.015, p-value=0.039). Epidemiological studies have confirmed that, persistent exposure to air-pollutants can accentuate metabolic disorders, which may lead to insulin resistance causing diabetes (Eze *et al.*, 2014; Eze *et al.*, 2015).

Exposure to polychlorinated biphenyls (PCBs) is found to have a very prominent and statistically significant association with overweight, obesity and diabetes. PCB118 has the strongest effect on overweight (OR=4.894, p-value=0.004) and obesity (OR=2.687, p-value=0.026), in comparison to the other significantly related PCBs. Estimated results also suggest diabetes to be significantly affected by PCB138 and PCB153 (OR=1.189, p-value=0.030; OR=1.194, p-value=0.018). This persistent organic pollutant is found extensively in the environment and can cause metabolic syndrome leading to obesity and diabetes (Thayer *et al.*, 2012).

However, Brominated Diphenyl Ethers (BDE153 and BDE209) were found to be slightly negatively related to overweight and obesity. However BDE153 was significantly unrelated with diabetes (OR=1.000, p-value=0.012). A study by Suvorov *et al.* (2009) suggest an increase in body weight of male adult rats by exposure to low doses of BDE47 at the perinatal stage, whereas a decrease in bodyweight was observed for mice offspring. However most studies have not reported any change in body weight due to BDE47 (Ta *et al.*, 2011).

Regarding toxic metals, it is a clinical consensus that toxic metals can essentially lead to metabolic abnormalities by causing oxidative stress or by substituting essential micronutrients and essential metals. In our analyses, it was observed that overweight was negatively associated with copper (OR=0.860, p-value=0.006), cadmium (OR=0.706, p-value=0.004) and lead (OR=0.979, p-value=0.016).

Similarly, cadmium (OR=0.868, p-value=0.001) and lead (OR=0.964, p-value=0.015) were negatively associated with obesity. However zinc had a positive associated with overweight and obesity. Subsequently, diabetes was found to have a significantly negative relation with manganese (OR=0.944, p-value=0.025), but a positive relation with lead (OR=1.052, p-value=0.012). This conforms to a study (Afridi *et al.*, 2008) that reported a higher concentration of lead in plasma for diabetic patients as compared to non-diabetic patients.

Odds ratio was found to be consistently but marginally high for overweight and obesity, in case of phthalate metabolites. Phthalates are essentially diesters of phthalic acid, which are widely used in personal care items and household products. Phthalates can enter the human system through air exposure, body contact or food and water intake. Subsequently certain biochemical reactions like antiandrogenic effects may occur, resulting in obesity. According to our analyses, Mono-Butyl Phthalate (MBP) OR=1.053, p-value=0.018) and Mono-2-Ethyl-5-Oxohexyl Phthalate (MEOHP) (OR=1.047, p-value=0.025) (had highest association with obesity as compared to other phthalates. In case of diabetes, the highest odds ratio for phthalates was observed for Mono-Ethyl Phthalate (MEP) (OR=1.386, p-value=0.017), which confirms to their relation with glucose metabolism and diabetes prevalence.

Table 4: Highly statistically significant environmental factors associated with overweight

Environmental Classification	Environmental Factor	Odds Ratio	P-value
Air Pollutants	Outdoor NO2	1.109	0.014
Polychlorinated Biphenyls	PCB118	4.894	0.004
	PCB138	2.003	0.005
	PCB153	2.093	0.003
	PCB180	1.915	0.001
Brominated Diphenyl Ethers	BDE153	0.987	0.001
	BDE209	0.997	0.002
Metals	Copper	0.860	0.006
	Zinc	1.066	0.010
	Selenium	1.001	0.001

	Cadmium	0.706	0.004
	Lead	0.979	0.016
Phthalates	Mono-Ethyl Phthalate (MEP)	1.018	0.014
	Mono-Benzyl Phthalate (MBzP)	1.019	0.007
	Monoisobutyl Phthalate (MiBP)	1.000	0.014
	Mono-n-butyl Phthalate (MnBP)	1.013	0.029
Cotinines	Cotinine in Saliva	0.891	0.019
Allergens	DPTERO pollens	1.004	0.019

Table 5: Highly statistically significant environmental factors associated with obesity

Environmental Classification	Environmental Factor	Odds Ratio	P-value
Air Pollutants	Outdoor NO2	1.276	0.030
	Outdoor PM10	1.174	0.005
Polychlorinated Biphenyls	PCB118	2.687	0.026
	PCB138	1.325	0.013
	PCB153	1.311	0.002
	PCB180	1.781	0.007
Brominated Diphenyl Ethers	BDE153	0.980	0.001
	BDE209	0.995	0.004
Metals	Zinc	1.098	0.021
	Cadmium	0.868	0.001
	Lead	0.964	0.015
Phthalates	Mono-Ethyl Phthalate (MEP)	1.007	0.028
	Mono-2-Ethylhexyl Phthalate (MEhP)	1.000	0.026
	Monoisobutyl Phthalate (MiBP)	1.006	0.023
	Mono-n-Butyl Phthalate (MnBP)	1.034	0.023

	Mono-Butyl Phthalate (MBP)	1.053	0.018
	Mono-2-Ethyl-5-Oxohexyl Phthalate (MEOHP)	1.047	0.025
Cotinines	Cotinine in Saliva	0.959	0.031
	Cotinine in Plasma	0.899	0.023
Allergens	Bla g 2	1.013	0.028
	Der f 1	1.001	0.011

Table 6: Highly statistically significant environmental factors associated with diabetes

Environmental Classification	Environmental Factor	Odds Ratio	P-value
Air Pollutants	Outdoor PM10	1.015	0.039
Polychlorinated Biphenyls	PCB138	1.189	0.030
	PCB153	1.194	0.018
Brominated Diphenyl Ethers	BDE153	1.000	0.012
Metals	Manganese	0.944	0.025
	Lead	1.052	0.012
Phthalates	Mono-Ethyl Phthalate (MEP)	1.386	0.017
	Monoisobutyl Phthalate (MiBP)	1.011	0.007
	Mono-Butyl Phthalate (MBP)	1.002	0.028
Cotinines	Cotinine in Plasma	1.064	0.002

Tobacco consumption, mostly by smoking, has been ascertained by the level of cotinine in plasma, saliva and urine. A negative association confirms that cotinine acts as a stimulant to metabolism and appetite suppressant, consequently leading to

lesser prevalence of obesity. However cotinine in plasma is found to be positively affecting diabetes.

No prominent association was found with allergens and obesity or diabetes. Among all the allergens considered, only Blag 2 was found to be slightly related to obesity (OR=1.013, p-value=0.028). Although studies suggest a strong correlation between obesity and allergic diseases (Gorgievska-Sukarovska *et al.*, 2008), it is yet to be deciphered whether allergy causing agents can potentially cause any biochemical and pathological effect that may eventually lead to increase in weight.

5.2 An analysis using the birth cohorts to investigate the effect of maternal exposure on birth-weight:

The indirect effect of maternal exposure to EDCs, affecting in utero development of offspring and other neonatal complications, is taken into account. A comprehensive weighted logistic regression suggested that both indoor and outdoor ambient air pollution may lead to the reduction in birth-weight. Prenatal exposure to toxic metals was found to have a prominent effect on newborns. The effect of selenium (OR=0.681, p-value=0.001) on low birth-weight was substantial. Lead was also found to be affecting the foetal growth and weight at birth. Trans-placental circulation of heavy metals can hinder fetal growth to some extent (Caserta *et al.*, 2013). Several phthalate metabolites were also found to be significantly related to birth-weight. Fetal metabolic dysfunction due to phthalate metabolites was evidenced by Ashley-Martin *et al.* (2014). However an in-depth analyses on the effect of different phthalate groups on neonatal growth is yet to be examined. Nevertheless, our study indicates an inverse association between maternal phthalate exposure and birth-weight, which is most prominent for Mono-3-Carboxypropyl phthalate (MCPP) (OR=0.843, p-value=0.031). Cotinine found in the plasma of pregnant women have a greater incidence of low birth-weight. Several studies have emphasized on the birth outcomes due to smoking habits of mother (Meghea *et al.*, 2012; Moga *et al.*, 2008). However we will focus on this issue for our concerned studies individually, as it is beyond the scope of this deliverable.

Table 4: Highly statistically significant environmental factors associated with overweight at birth

Environmental Classification	Environmental Factor	Odds Ratio	P-value

Air Pollutants	Indoor NO2	0.955	0.004
	Outdoor NO2	0.957	0.007
Metals	Zinc	1.000	0.010
	Selenium	0.681	0.001
Phthalates	Mono-2-Ethylhexyl-Phthalate (MEhP)	1.000	0.003
	Mono-n-Octyl Phthalate (MOP)	0.965	0.021
	Mono-Butyl Phthalate (MBP)	0.988	0.012
	Mono-3-Carboxypropyl Phthalate (MCPP)	0.993	0.023
Cotinine	Cotinine in Plasma	0.974	0.022

Table 4: Highly statistically significant environmental factors associated with obesity at birth

Environmental Classification	**Environmental Factor**	**Odds Ratio**	**P-value**
Air Pollutants	Outdoor PM10	0.931	0.005
	Outdoor NO2	0.950	0.022
Metals	Lead	0.975	0.013
Phthalates	Mono-2-Ethylhexyl-Phthalate (MEhP)	1.000	0.026
	Mono-n-Octyl Phthalate (MOP)	0.986	0.024
	Mono-Butyl Phthalate (MBP)	0.997	0.024
	Mono-3-Carboxypropyl Phthalate (MCPP)	0.843	0.031
Cotinine	Cotinine in Plasma	0.991	0.015

Although pre-natal exposure to organic pollutants like PCBs and BDEs were found to have an effect on birth-weight in a Swedish cohort study (Lignell *et al.*, 2013), no such evidence could be obtained in our analysis.

Our findings confirm to the fact that, the toxicological effects due to EDCs can play a vital role in the physiological alteration of fetal development. Inspite of the selective absorption by the placenta, certain environmental stressors were able to cross the placental barrier, thereby affecting the perinatal stage of offspring (Needham *et al.*, 2011).

5.3 An analysis using the birth cohorts to investigate the effect of maternal exposure on gestational diabetes

Studies have indicated a risk propensity to type 2 diabetes mellitus due to exposure to air pollutants. A meta-analysis conducted by Balti *et al.* (2014) found prominent effect of NO2 and PM2.5 on gestational occurrence of type-2 diabetes mellitus. Malmqvist *et al.* (2013) had ascertained that higher exposure to NO_x at the first trimester of pregnancy can lead to gestational diabetes mellitus. However studies like Hooven *et al.* (2009) indicates no association between air pollutants and gestational diabetes.

In our study, we found a prominent effect of indoor and outdoor NO_2 and indoor particulate matter (PM_{10} and $PM_{2.5}$) on gestational diabetes. The probable biological reason for this relation may be due to inflammation of visceral adipose tissue, increase in hemoglobin A1c (HbA1c) level and autoimmune tone alteration, that may increase resitance to insulin (Esposito *et al.*, 2016).

Table 4: Highly statistically significant environmental factors associated with gestational diabetes

Environmental Classification	Environmental Factor	Odds Ratio	P-value
Air Pollutants	Indoor PM10	1.084	0.015
	Indoor NO2	1.078	0.024
	Indoor PM2.5	1.060	0.003
	Outdoor NO2	1.081	0.018
Metals	Cadmium	2.383	0.013
Phthalates	Mono-Ethyl Phthalate (MEP)	1.106	0.007
	Mono-2-Ethyl-5-Hydroxyhexyl Phthalate (MEHHP)	1.102	0.010
	Mono-Benzyl Phthalate (MBzP)	1.106	0.004
	Mono-Butyl Phthalate (MBP)	1.097	0.026
	Mono-3-Carboxypropyl Phthalate (MCPP)	1.011	0.021
Cotinine	Cotinine in Plasma	1.011	0.028

Among heavy metals, cadmium was found to have a high impact on gestational diabetes (OR=2.383, p-value=0.013). According to a study by Peng *et al.* (2015), heavy metals, especially arsenic, impacts gestational diabetes. Additionally, two

subsequent studies from China (Xing *et al.*, 2018; Liu *et al.*, 2018) concluded a positive association between cadmium in plasma of pregnant women and gestational diabetes mellitus.

Several phthalate metabolites were also found to have significant impact on gestational diabetes. We discovered substantial associations for Mono-Benzyl Phthalate (MBzP) (OR=1.106, p-value=0.004) and Mono-Ethyl Phthalate (MEP) (OR=1.106, p-value=0.007). A study by James-Todd *et al.* (2016) suggested that a decrease in glucose tolerance is significantly affected by MEP in the second trimester of pregnancy, whereas Di-2-ethylhexyl ether was associated with decreased odds of impaired glucose tolerance.

Cotinine in plasma was also found to be associated with increased odds of gestational diabetes risk (OR=1.011, p-value=0.028).

6 Further work needed to be done

With the comprehensive results that we have obtained, we would further our analysis with the inclusion of additional data. Therefore, the data accumulation process is still in progress and is updated from time to time. The application of new data may provide more enriching results for our analyses.

Additionally, we may use other statistical techniques like the Cox Proportional Hazard model, using the time dependent and fixed covariates.

6.1 Data accumulation

6.1.1 EDMS Data

The Environmental Data Management System (EDMS) is a HEALS platform created to collect and link environmental databases, which can be used by other work packages. After the evaluation of the data quality and applicability, the data is formatted in compliance with the INSPIRE Directive.

We aim at performing additional analyses using the environmental data from the EDMS. However, the EDMS data, at hand, is not at the individual level, but has been approximated at city level, with annual averages of the pollutants. A comprehensive exposure dataset would be further used to perform sensitivity for our analysis.

Additionally the studies, which have information on the geographic coordinates of the individuals, can also be geo-referenced with the environmental data from EDMS. The EDMS data may be subject to the geographic projection system, by which each observation can be located for point and polygonal spatial information. We can then apply this dataset for our individual level study.

6.1.2 Data from the MoBa birth cohort study:

The collection of data from the Norwegian Mother and Child Cohort study (MoBa) has been completed. Data has been successfully harmonized and collected based on all the MoBa questionnaires (Week 15 of pregnancy-Questionnaire for mother, Week 15 of pregnancy-Questionnaire for father, Week 22 of pregnancy-Questionnaire about mother diet, Week 30 of pregnancy-Questionnaire about mental and physical health, Questionnaire at 6 months, Questionnaire at 18 months, Questionnaire at 36 months, Questionnaire at 5 years, Questionnaire at 7 years and Questionnaire at 8 years). This data will be incorporated with the main data set. Subsequently we will re-run the analyses.

However, it may be noted that the data collected from the MoBa study are questionnaire based data, and no clinical data is included in the study. However we would require clinical data for environmental stressors, like phthalates, pcb, bde, cotinine in blood, serum or urine for the exposomic study. Additionally, no data was collected for atmospheric pollution like PM_{10}, NO_2 and SO_2, for Norway. We would consult with the EDMS database to check for the availability of some exposure data from Norway.

6.2 Other statistical methodologies

The EWAS study that we have conducted so far is based on weighted logistic regression, using a bonferroni correction. The missing data was imputed using a multiple imputation technique.

However we aim at using other statistical models, like the Cox proportional hazard model on our data set to analyze survival-time outcomes on one or more predictors.

6.3 More detailed analysis

Although the EWAS study undertaken in this deliverable is quite comprehensive, given the data at hand and the problem of missing values. However certain other research avenues can be explored in greater details. For example, the analyses on gestational diabetes can be investigated for the different trimesters of pregnancy. Additionally, sex-specific differences can also be considered.

While considering the birth cohorts, we have not determined whether in utero and childhood exposure to EDCs can affect metabolic dysfunctions at a later stage of life. This can be further investigated as a continuation of our study.

Nevertheless, any detailed analysis will also require us to take into account the availability of data from the different cohorts.

7 Conclusion

The main purpose of this research analysis was to investigate the effect of EDCs on obesity and diabetes, using the pre-existing cohort studies from Europe. We have analyzed this issue by considering the direct effect of these pollutants on the total population derived from the different cohorts. Prenatal exposure is also taken into account, where the maternal exposure to environmental contaminants can affect the fetal development and the subsequent perinatal weight of the offsprings at birth. In addition, a supplementary analysis is also performed to investigate if gestational diabetes is affected by environmental exposures.

Although genetic susceptibility, sedentary and dietary habits, and certain demographic factors can largely affect metabolic functioning of individuals, environmental factors also play a complementary role in this respect. Our study, based on EWAS approach, confirms health related outcomes based on metabolic factors due to exposure to various environmental contaminants. Based on our analyses, it was confirmed that air borne pollutants, chemicals, organic compounds and cotinine levels can affect the metabolic system of human physiology.

Air pollutants like NO_2, PM_{10} and $PM_{2.5}$ are found to impact metabolism resulting in obesity or diabetes. Due to the emerging study in this area, it needs to be clinically confirmed whether these pollutants can affect the insulinotropic mechanism, beta cell functions as well as functioning of hormones like glucagen (Rajagopalan *et al.*, 2012). Phthalates, cotinine and certain heavy metals are also found to have a

significant impact on obesity and diabetes, both directly and fetal exposure. Gestational diabetes was also affected by some of these environmental stressors. Phthalates are ubiquitous chemicals that binds to human proliferator receptor alpha and gamma, that causes changes in hormones related to glucose metabolism and adipogenesis (Desvergne et al., 2009). On the other hand, direct ingestion or trans-placental flow of heavy metals can also prominently affect weight and insulin tolerance. Studies suggest that an increase in certain metals in human plasma can increase adiposity, whereas some metals may lead to opposite effects (Katzen-Luchten, 2007; Huang et al., 2007). Organic chemicals like PCBs and BDEs were seen to affect human metabolism directly. However, contrary to previous findings of Lignell et al. (2013), no significant effect was seen in the birth-weight due to maternal exposure to these pollutants.

Inspite of the detailed and systematic analyses performed, our study also has certain limitations. One of the major hindrance faced was the prominent number of missing values. Hence our analyses depends on imputed data to some extent. Further, other factors like the technical aspects of the exposure assessment mechanisms and robustness of the individual cohort data have not been taken into consideration.

Inspite of certain loopholes, the novelty of our study is the large data that is accrued from population across different regions in Europe. Although a growing body of research studies investigate the involvement of environmental stressors on obesity and diabetes, none has undertaken an analysis on such vast and diverse spectrum of data. Our findings are crucial and can have direct implications on human health, as well as maternal and child care.

References

Balti, E.V., Echouffo-Techeugui, J.B., Yako, Y.Y., Kengne, A.P. (2014) Air pollution and risk of type 2 diabetes mellitus: A systematic review and meta-analysis. *Diabetes Research and Clinical Practice* 106:161-172.

Ashley-Martin, J., Dodds, L., Arbuckle, T.E., Ettinger, A.S., Shapiro, G.D., Fisher, M. (2014) A birth cohort study to investigate the association between prenatal phthalate and bisphenol A exposures and fetal markers of metabolic dysfunction. *Environmental Research* 13:84.

Afridi, H.I., Kazi, T.G., Kazi, N., Jamali, M.K., Arain, M.B., Jalbani, N., Baig, J.A., Sarfraz, R.A. (2008) Evaluation of status of toxic metals in biological samples of diabetes mellitus patients. *Diabetes Research and Clinical Practice* 80(2):280-288.

Baker, J.L., Olsen, L.W., & Sorensen, T.I. (2007) Childhood body-mass index and the risk of coronary heart disease in adulthood. *New England Journal of Medicine* 357: 2329-2337.

Carpenter, J., & Kenward, M. (2013) Multiple Imputation and Its Applications. New York, NY: John Wiley & Sons.

Caserta, D., Graziano, A., Lo Monte, G., Bordi, G., Moscarini, M. (2013) Heavy metals and placental fetal-maternal barrier; A mini-review on the major concerns. *European Review for Medical and Pharmacological Sciences* 17:2198-2206.

Chen, Y., Ahsan, H., Slavkovich, V., Peltier, G.L., Gluskin, R.T., Parvez, F., et al. (2010). No association between arsenic exposure from drinking water and diabetes mellitus: a cross-sectional study in Bangladesh. *Environmental Health Perspective* 118:1299–1305.

Delvaux, I., Van Cauwenberghe, J., Den Hond, E., Schoeters, G., Govarts, E., Nelen, V., et al. (2014) Prenatal exposure to environmental contaminants and body composition at age 7-9 years. *Environmental Research* 132:24-32.

Desvergne, B., Feige, J.N., Casals-Casas, C. (2009) PPAR-mediated activity of phthalates: A link to the obesity epidemic? *Molecular and Cellular Endocrinology* 304:43-48.

Eze, I.C., Schaffner, E., Schikowski, T., Adam, M., Imboden, M., et al. (2014). Long-term air pollution exposure and diabetes in a population-based Swiss cohort. *Environment International* 70C:95-105.

Eze, I.C., Hemkens, L.G., Bucher, H.C., Hoffmann, B., Schindler, C., Kunzli, N., et al. (2015). Association between ambient air pollution and diabetes mellitus in Europe and North America: systematic review and metaanalysis*. Environmental Health Perspective* 123(5):381-389.

Esposito, K., Petrizzo, M., Maiorin, M.I., Bellastella, G., Giugliano, D. (2016) Particulate matter pollutants and risk of type 2 diabetes: A time for concern? *Endocrine* 51:32-37.

Gorgievska-Sukarovska, B., Lipozencic, J., Susac, A. (2008). Obesity and Allergic Diseases. *Acta Dermatovenerologica Croatica* 16(4):231-235.

Grupo de trabajo de la Sociedad Española de Epidemiologia y de la Sociedad Española de Medicina Familiar y Comunitaria (2000) Una propuesta de medida de la clase social. *Atención Primaria* 25:350-363.

Han, J.C., Lawlor, D.A., & Kimm S.Y.S. (2010) Childhood obesity. *Lancet* 375:1737-1748.

Hindorff, L.A., Senthupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. & Manalio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences U.S.A.* 106:9362-9367.

Holtcamp, W. (2012) Obesogens: an environmental link to obesity. *Environmental Health Perspectives* 120:a62-a68.

Huang, L., Yu, Y.Y., Kirschke, C.P., Gertz, E.R., Lloyd, K.K. (2007) Znt7 (Slc30a7)-deficient mice display reduced body zinc status and body mass accumulation. *The Journal of Biological Chemistry* 282:37053-37063.

Huang, R.C., Burke, V., Newnham, J.P., Stanley, F.J., Kendall, G.E., Landau, L.I., et al. (2007) Perinatal and childhood origins of cardiovascular disease. *International Journal of Obesity* 31:236-244.

James-Todd, T.M., Meeker, J.D., Huang, T., Hauser, R., Ferguson, K.K., RichEdwards, J.W. (2016) Pregnancy urinary phthalate metabolite concentrations and gestational diabetes risk factors. *Environment International* 96:118-126.

Juncà, S., Guillén, M., Aragay, J.M., Brugulat, P., Castell, C., et al. (2003) Methodological aspects in the evaluation of health and risk-reduction objectives of Health Plan for Catalonia for the year 2000. *Medicina Clinica (Barcelona)* 121:10-19.

Katzen-Luchenta, J. (2007) The declaration of nutrition, health, and intelligence for the child-to-be. *Nutrition and Health* 19: 85-102.

La Merrill, M., & Birnbaum, L.S. (2011) Childhood Obesity and Environmental Chemicals. *Mount Sinai Journal of Medicine* 78:22-48.

Lignell, S., Aune, M., Darnerud, P.A., Hanberg, A., Larsson, S.C., Glynn, A. (2013) Prenatal exposure to polychlorinated biphenyls (PCBs) and polybrominated diphenyl ethers (PBDEs) may influence birth weight among infants in a Swedish cohort with background exposure: a cross-sectional study. *Environmental Health* 12:44.

Liu, W., Zhang, B., Huang, Z., Pan, X., Chen, X., Hu, C., et al. (2018) Cadmium Body Burden and Gestational Diabetes Mellitus: A Prospective Study. *Environmental health Perspective* 126(2):027006.

Lee, D.H., Porta, M., Jacobs, D.R. Jr., & Vandenberg, L.N. (2014) Chlorinated persistent organic pollutants, obesity, and type 2 diabetes. *Endocrine Reviews* 35:557-601.

Lee, D.H., Steffes, M.W., Sjodin, A., Jones, R.S., Needham, L.L. & Jacobs, J.D.R. (2010) Low dose of some persistent organic pollutants predicts type 2 diabetes: a nested case–control study. *Environmental Health Perspective* 118:1235–1242.

Malmqvist, E., Jakobsson, K., Tinnerberg, H., Rignell-Hydbom, A., Rylander, L. (2013) Gestational diabetes and preeclampsia in association with air pollution at levels below current air quality guidelines. *Environmental Health Perspectives* 121:488-493.

Meghea, C.I., Rus, D., Rus, I.A., Summers Holtrop, J., Roman, L. (2012) Smoking during pregnancy and associated risk factors in a sample of Romanian women. *European Journal of Public Health* 22:229-233.

Moga, M., Preda, G.H. (2008) Smoking in pregnancy. *Journal of Environmental Protection and Ecology* 9:566:573.

Needham, L.L., Grandjean, P., Heinzow, B., Jorgensen, P.J., Nielsen, F., Patterson, D.G., Sjodin, A., Turner, W.E., Weihe, P. Partition of environmental chemicals between maternal and fetal blood and tissues. *Environmental Science and Technology* 45:1121-1126.

Patel, C.J., Bhattacharya, J. & Butte A.J. (2010) An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *Plos One* 5:e10746.

Peng, S., Liu, L., Zhang, X., Heinrich, J., Zhang, J., Schramm, K.W., Huang, Q., Tian, M., Egani, S.A., Shen, H. (2015) A nested case-control study indicating heavy metal residues in meconium associate with maternal gestational diabetes mellitus risk. *Environmental Health* 14:19.

Porta, M., Gasull, M., Puigdomènech, E., Gari, M., de Basea, M.B., et al. (2010) Distribution of blood concentrations of persistent organic pollutants in a representative sample of the population of Catalonia. *Environment International* 36:655-664.

Rajagopalan, S., Brook, R.D. (2012) Air Pollution and Type 2 Diabetes Mechanistic Insights. *Perspectives in Diabetes* 61:3037-3045.

Rubin, D.B. (1976) Inference and missing data. *Biometrika* 63:581-592.

Rubin, D.B. (1987) Multiple imputation for non-response in surveys. New York: Wiley.

Rubin, D.B. (1996) Multiple imputation after 18+ years. *Journal of the American Statistical Association* 91:473-489.

Rylander, L., Rignell-Hydbom, A. & Hagmar, L. (2005). A cross-sectional study of the association between persistent organochlorine pollutants and diabetes. *Environmental Health* 4:28.

Suvorov, A., Takser, L. (2010). Global gene expression analysis in the livers of rat offspring perinatally exposed to low doses of 2,2',4,4'-tetrabromodiphenyl ether. *Environmental Health Perspective* 118(1):97-102.

Ta, T.A., Koenig, C.M., Golub, M.S., Pessah, I.N., Qi, L., Aronov, P.A., Berman, R.F. (2011). Bioaccumulation and behavioral effects of 2,2',4,4'-tetrabromodiphenyl ether (BDE-47) in perinatally exposed mice. *Neurotoxicology and teratology* 33(3):393-404.

Thayer, K.A., Heindel, J.J., Bucher, J.R., & Gall, M.A. (2012) Role of environmental chemicals in diabetes and obesity: a National Toxicology Program workshop review. *Environmental Health Perspective* 120:779-789.

Van den Hooven, E.H., Jaddoe, V.W., de Kluizenaar, Y., Hofman, A., Mackenbach, J.P., Steegers, E.A., Miedema, H.M., Pierik, F.H. (2009) Residential traffic exposure and pregnancy-related outcomes: A prospective birth cohort study. *Environmental Health Perspective* 8:59.

Valvi, D., Mendez, M.A., Garcia-Esteban, R., Ballester, F., Ibarluzea, J., Goňi, F., et al. (2014) Prenatal exposure to persistent organic pollutants and rapid weight gain and overweight in infancy. *Obesity (Silver Spring)* 22:488-496.

Van Buuren, S., Boshuizen, H.C., & Knook, D.L. (1999) Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 18:681-694.

Wilson, J.R. & Lorenz, K.A. (2015) Weighted Logistic Regression Model. In: Modeling Binary Correlated Responses using SAS, SPSS and R. ICSA Book Series in Statistics. Vol 9. Springer, Cham.

White, I.R., Royston, P. & Wood, A.M. (2011) Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 30:377-399.

Xing, Y., Xia, W., Zhang, B., Zhou, A., Huang, Z., Zhang, H., et al. (2018) Relation between cadmium exposure and gestational diabetes mellitus. *Environment International* 113 :300-305.