This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement N°603946

HEALS

HEALS

Health and Environment-wide Associations based on Large population Surveys

FP7-ENV-2013- 603946

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

Deliverable 3.2 – Report on conceptual framework of HEALS

WP 3 Definition of methodological framework

Version number 2

Lead beneficiary: AUTH
Date: 30/09/2014
Nature: R
Dissemination level: PU

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**	3.2	**Title**	Report on conceptual framework of HEALS
Work Package	**Number**	3	**Title**	Definition of methodological framework

Delivery date	**Contractual**	M12	**Actual**	M22-15/07/2015
Status	Draft ☐		Final ☑	
Nature	Demonstrator ☐	Report ☑	Prototype ☐	Other ☐
Dissemination level	Confidential ☐	Public ☑		

Author (Partners)	Name (Organization) AUTH			
Responsible Author	Denis Sarigiannis		**Email**	d.a.sarigiannis@gmail.com
	Partner	**AUTH**	**Phone**	+30 2310 994562

Document History

Name	Date	Version	Desciption
AUTH	28/02/2015	1	Draft report
AUTH	30/04/2015	2	Final report

TABLE OF CONTENTS

ABSTRACT

Background: Exposome appears as a very promising tool for better understanding the complexity of interactions between genome and epigenome and environment, especially when investigating large population studies.

Objectives: The HEALS project aims at identifying the complex links among genes, epigenetic mechanisms, environment and human disease focusing on allergies and asthma, metabolic and neurodevelopmental/neurodegenerative disorders based on individual exposome characterization and how should this be implemented in large cohorts.

Methods: HEALS relies on the re-analysis of existing cohort studies and the deployment of a pilot European Exposure and Health Examination Survey (EXHES), the two approaches including both singletons and twins. Although the analysis will start from the collection of biomonitoring data, a wide array of omics technologies (completed by confirmatory *in vitro* testing) will be employed and related to environmental stressors. Lifetime exposure assessment will involve novel technologies such as sensors and agent based modelling. Mapping the different omics responses onto regulatory networks and disease pathways in relation to exposures will allow understanding the intermediate developmental stages from exposure to disease at individual as well as population level. This objective will be attained through the refinement of an integrated methodology and the application of the corresponding analytical and computational tools for performing environment-wide association studies (EWAS) in support of EU-wide environment and health assessments.

Discussion: HEALS is expected to provide additional insights into the way to synthesize different data and methodological tools for assessing internal and external exposome aand their complexity, overall aiming to a better understanding the potential mechanisms and the origin of disease. This includes i) how different environmental factors contribute cumulatively to disease and ii) the common nodes of exposure and molecular events resulting in phenomenally different health outcomes

Conclusions: HEALS is a comprehensive methodological advance aiming to provide the way of linking interdisciplinary research towards the understanding of genome, epigenetic mechanisms and lifetime environmental interaction at individual and population level.

INTRODUCTION

The exposome (Wild 2005) represents the totality of exposures from conception onwards, simultaneously identifying, characterizing and quantifying the exogenous and endogenous exposures and modifiable risk factors that predispose to and predict diseases throughout a person's life span.

The exposome came as a complement to the human genome; although decoding of human genome (Schmutz et al. 2004) increased our understanding of the underlying causes of disease, genome explains only a percentage of population burden. Indeed according to Rappaport et al. (2014) two thirds of all people worldwide die of chronic disease (mostly heart disease and cancer) which are caused by combinations of the genome and exposome (representing all exposure – internal and external). However, disease risks attributed to genome alone are modest representing less than 15% suggesting that more than 85% of risks results from exposome and interaction of genome and exposome. Thus, it is evident that environmental factors are equally or eventually more important and what is actually critical is the interaction of environmental factors with the biological systems. Towards a better understanding of the causal links among genome, environment and disease, unraveling the exposome implies that both environmental exposures and genetic variation are reliably measured simultaneously.

Unraveling the exposome is daunting, particularly in the light of the enormous amount of information that needs to be integrated. As a result of dedicated actions and projects following the European Commission's (EC) Environment & Health Action Plan 2004-2010, various harmonization efforts have occurred. Projects such as COPHES (harmonization of HBM), EHES (harmonization of Health Surveys), EU-menu (harmonization of data collection on food consumption) or CHICOS (harmonization of child cohort studies) or U-BIOPRED (unbiased biomarkers in prediction of respiratory disease outcomes) all aim at providing common ground for the often disparate information which was scattered across Europe. In addition, European twin registries have collected biological material and longitudinal phenotypic and exposure data on tens of thousands of twins providing a valuable resource for investigating the development of complex phenotypes and their underlying biology. The HEALS project is a logical progression from many of the achievements of the Environment and Health Action Plan (EHAP) 2004-2010. Making optimal use of the availability of harmonized data across Europe, HEALS introduces significant advances to environmental and health data fusion, including assimilation of data from satellite remote sensing for direct measurement of environmental exposure to airborne pollutants such as particulate matter (PM) and for providing accurate spatially-resolved estimates of population exposed to environmental pollutants.

Two other projects comprise the EU exposome initiative, namely Exposomics (led by Imperial College, in London, UK) and HELIX (led by CREAL in Barcelona, Spain). Exposomics focuses on the development of a systematic way to measure the influence of environmental exposures on health. Towards this aim, Exposomics deals with the development of a personal exposure monitoring (PEM) system (including sensors, smartphones, geo-referencing, satellites) for collecting data on the individual external

exposome as well as on analyzing biological samples (internal markers of external exposures) using multiple “omic” technologies. Relationships between external exposures (as measured by PEM) and global profiles of molecular features (as measured by omics) in the same individuals constitutes the overall methodology (Vineis et al. 2013), opening the way to ‘exposome-wide association studies’ (EWAS). The ultimate goal is to use the new tools in risk assessment and in the estimation of the burden of environmental disease with a special focus on the molecular epidemiology of cancer (Chadeau-Hyam et al. 2013; Vineis and Wild 2014).

The Human Early-Life Exposome (HELIX) project (Vrijheid et al. 2014) is a collaborative research project that aims to implement novel exposure assessment and biomarker methods to characterize early-life exposure to multiple environmental factors and associate these with omics biomarkers and child health outcomes, characterizing thus the early-life exposome. HELIX uses a multilevel study design, drawing on nested study populations on four different levels of data collection, including data from existing, as well as from an originally designed sub-cohort for HELIX that includes 1,200 mother–child pairs (Vrijheid et al. 2014). Similarly to Exposomics, a wide array of external and internal exposome assessment strategies will be applied.

A fourth project, CROME (Cross-Mediterranean Environment and Health Network) is also part of the EU-funded exposome projects albeit with a regional focus on the Mediterranean basin. CROME focuses on integrating human biomonitoring measurements into the exposome construction process as a means to shed more light into the links between environmental and dietary exposures to toxic metals and persistent organic chemicals and adverse health outcomes with emphasis on cancer (Sarigiannis et al. 2015) and neurotoxicity.

In the USA, two are the main exposome-related initiatives funded by NIEHS, namely the HERCULES center at Emory University in Georgia and the Exposome Reseach Center at the University of California at Berkeley. HERCULES takes a multi-faceted approach to the exposome, attempting to estimate the allostatic load from environmental exposures over an individual’s lifecourse. The UC Berkeley Center instead focuses on refining untargeted metabolomics as the main technological means towards unraveling the individual exposome and linking it to human disease adopting a top-down approach (Rappaport and Smith 2010).

HEALS (Health and Environment-wide Associations based on Large population Surveys) brings together a comprehensive array of novel technologies, data analysis and modeling tools that support the efficient design and execution of large-scale exposome studies. The HEALS approach brings together and organizes environmental, socio-economic, exposure, biomarker and health effect data; in addition, it includes all the procedures and computational sequences necessary for applying advanced bioinformatics coupling advanced data mining, biological and exposure modeling so as to ensure that environmental exposure-health associations are studied comprehensively. The overall approach will be verified in a series of population studies across Europe, tackling various levels of environmental exposure, age windows and gender differentiation of exposure, and socio-economic and genetic variability. **The main objective of HEALS is the refinement of**

an integrated methodology and the application of the corresponding analytical and computational tools for performing environment-wide association studies in support of EU-wide environment and health assessments. The HEALS approach will be refined on the basis of **pre-existing European population data** and then it will be applied in a **pilot environment and health examination survey** covering ten EU Member States. The lessons learned will be translated into scientific advice towards the development of protocols and guidelines for the setting up of a European environment and health examination survey.

Exposome studies will require novel tools to address the complexity of emerging environmental health issues. Critical for success will be the ability to bring together existing geospatial, environmental, health and socioeconomic data, and to collect new high resolution data using innovative environmental micro-sensors, remote sensing or other community and omics/systems biology based approaches to describe the exposome for e.g. endocrine disruption-related syndromes and sex-related changes (menopause), neurodegenerative or respiratory diseases. Focus will be on susceptibility windows during growth (including pregnancy) and development, and on the unequal distribution of the burden of epigenetically active food and environment-related disease to vulnerable populations such as the young, elderly, socio-economic disadvantaged, gender and ethnic minorities.

The individual exposome has to be intended as dynamic and continually changing. Indeed, all exposures and their determinants and modifiers can vary over the course of a day, not to mention over the weeks, months, and years that make up a lifetime, as our bodies, diets, risk factors and lifestyles change. Because sources and levels of exposure change over time, and because capturing all these changes verges on the impossible in the impracticality of “high-resolution real-time” monitoring of all the exposures for the entire lifetime, the exposome has to be constructed by assessing the exposures at critical life periods through representative snapshots that act as demonstrative measures of these critical periods. Indeed, mapping the entire lifecycle of an individual may not be necessary if critical lifetime events where an individual’s geospatial lifeline crosses a noteworthy environmental event (Sabel et al. 2009) are recognized and understood. Thus, one major challenge consists in identifying critical life stages that are informative at most as well as the snapshots reflecting the exposures and the downstream consequences at the individual level.

Hence, the most relevant exposure episodes in an individual’s life could be reconstructed and linked to socio-economic conditions at critical life stages such as prenatal exposure, puberty, or the reproductively active period. Whereas exposure during all life stages may entail adverse effects, **children, pregnant women** and the **elderly** are particularly susceptible; thus these population sub-groups will be the focus of this project. Modelling the mobility patterns of the population at risk at the individual level is challenging. There are considerable conceptual and computational difficulties involved in intersecting data on the distributions of pollutants, and/or the patterns of movements of recipient individuals or groups, reflecting the limitations of available data on environmental conditions and human distributions. With the advent of geographic information systems (GIS), GPS to

track individuals, and personal environmental monitoring, undertaking such analyses throughout an individual's lifetime is now possible.

For the first time, HEALS will try to **reverse the paradigm of "nature versus nurture"** and adopt one defined by **complex and dynamic interactions** between DNA sequence, epigenetic DNA modifications, gene expression and environmental factors that all combine to influence disease phenotypes. HEALS will start from analysis of data collected in on-going epidemiological EU studies involving mother/infant pairs, children, or adults including the elderly to evidence relevant environmental exposure/health outcome associations. These associations will aid in designing pilot surveys using an integrated approach, where the selection of biomarkers of exposure, effects and individual susceptibility results in integrated risk assessment. In the context of this new paradigm, a relevant contribution for a better understanding of the diseases comes also from twin studies. Reviewed evidence makes it clear that substantial phenotypic differences in monozygotic (MZ) twins, that are genetically identical, arise from both non-shared environment and epigenetic DNA modification patterns responsible for incomplete penetrance of genes. Data obtained recently demonstrated that, in young MZ twins, epigenetic changes such as locus-specific inter-individual DNA methylation differences arise both *in utero* and after birth (Feil and Fraga 2012; Flintoft 2005). Environmental conditions that can affect the epigenome of an individual include both external and internal factors. Individual behaviours such as smoking and alcohol consumption, physical activity, dietary intake, temperature changes and stress are external factors that have been proposed to have a long-term influence on epigenetic modifications. However, it is possible that small defects in transmitting epigenetic information through successive cell divisions, or maintaining it in differentiated cells, accumulate in a process that could be considered as an ''epigenetic drift'' associated with aging. Indeed, external and internal environmental factors may have long-lasting effects on metabolism and health, sometimes even in subsequent generations. Knowledge of epigenetic mechanisms (differential DNA methylation in promoter and intragenic CpG islands as well as in repeated sequences, miRNA expression, skewed X-inactivation, imprinting, chromatin modification) and underlying causes provides a new model for understanding MZ twin discordance and discovering mechanisms affecting disease susceptibility. Studies in dizygotic (DZ) twins also have revealed epigenetic changes that could in part be environmentally induced. This is why HEALS uses twins studies as an additional advantageous model to comprehend the development of disease by showing the proper roles of epigenome, environment and their interaction in the expression of complex phenotypes and their underlying biology. Two different types of twin studies will be considered in HEALS: 1) twin cohorts or registries already existing in many European countries including the elderly for which bio-banks and data on environmental exposures and health are available; 2) a new birth cohort survey including newborns with an oversampling on twins matched by singletons in which harmonized and standardised data on -omics, environmental stressors and health will be collected. In the coming years, longitudinal phenotypic information coupled with biological material collected by worldwide twin registries and birth cohorts will constitute an important resource for large-scale Environment-wide Associations based on Large Population Surveys (EWAS) studies. The

following key issues will be addressed to characterize the individual exposome:

- **Human biomonitoring (HBM)** and **biobanking** are seen as a central theme. Sources and levels of exposure change over time and it is envisioned that individual exposomes can be (re)constructed by analyzing toxicants in (preferentially non-invasively collected) human matrices obtained during critical life stages.
- **Understanding of the interaction between HBM and exposure modeling (EM) or estimation** is another key factor for elucidating the exposome. HBM only has a limited ability to identify specific exposure sources, and may not be applicable to all relevant environmental stressors (e.g. PMx or noise). EM is therefore an essential part of any strategy aimed at unraveling the exposome. Integrating HBM data, EM and environmental monitoring data will lead to a more conclusive picture of the exposome and related health outcomes and will be of use in future large-scale population studies.
- **Lifestyle/behaviour patterns** (such as time-activity-location, food consumption, use of consumer products, etc.) are needed to understand individual and population-based geospatial lifelines.
    - Spatial information and initiatives to harmonize their collection (INSPIRE, GMES) have the ability to transform the way scientists and policy makers think about exposure to environmental stressors.
    - At the same time, behavioural information functions as the most accessible and direct way for policy makers and risk assessors to understand and manage an individual's exposure patterns.
    - These exposure factors can be used to derive aggregate and cumulative exposure models, leading to probabilistic exposure assessments. In this context, data from the European Social Survey (ESS) a biannual multi-country survey started in 2002 covering over 30 nations (http://ess.nsd.uib.no/) will provide contextual variables on change and continuity in a wide range of social variables including well-being, health and security, human values, demographics and socio-economics.
- **Innovations in sensor technology** create possibilities to collect environmental data at unprecedented depth and breadth. Finding the right balance between limited amounts of high quality data from standardized environmental monitoring campaigns (e.g. EIONET, AIRBase) and large amounts of moderate quality data by sensor networks will transform the way we understand and interact with our environment.
- Due to the substantial technical and ethical hurdles involved in collecting real individual space-time movement data for whole populations, we propose simulating movement and interaction behaviour using **agent-based models** (ABM) informed by sensor technologies. ABM aim to simulate and organise social behaviours in order to understand the dynamics of real-world systems. A style of computational modelling, ABM, focuses on simulating individuals (agents) and their interactions with other agents and their environments, as part of a larger complex system. The use of ABM will enable us to better understand the behaviour of individuals and populations in social and evolutionary settings, and to 'fill-in' the gaps in the exposome currently not available from real-world monitoring and sensor data.
- Current toxicological state of the art couples estimations of **biologically effective dose**

(BED) with early biological events to derive dose-effect models, which can be used in combination with the probabilistic exposure estimates (Georgopoulos et al. 2008) to derive biomarkers of exposure and/or effect. Combined use of epidemiological, clinical and genetic/epigenetic analysis data will shed light on the effect of risk modifying factors such as lifestyle choices and DNA polymorphisms and methylation (D Sarigiannis et al. 2009**)**. Distinguishing exposures of limited duration or periods of specific relevance (e.g., in utero) from chronic exposures may render new biomarkers more specific. Exposure assessed prospectively and tightly linked to proposed periods of vulnerability of the epigenome (e.g., periods of placental invasion or sex specification in utero) would be ideal. Observation of real clinical data and/or results of biomonitoring, coupled with exposure/effect biomarker discovery systems, will produce predictive biomarkers allowing estimations of individual response to toxic insults. **Metabolomics and adductomics** are key to this analytical and data interpretation process. They will be functionally integrated with **transcriptomics** and **proteomics** to provide the mechanistic underpinning for establishing causality in the association between health status and exposure to environmental stressors.

METHODS

Overall methodological concept

HEALS does not follow a top-down or bottom-up approach to building the exposome. Its take is to start from the available biological signals (biomonitoring data or other, more sensitive biological markers) and draw in parallel functional links with both (a) phenotypes or confirmed endotypes of disease and (b) external exposures over an individual's lifecourse. We call this the "middle-out" approach (Figure 1).

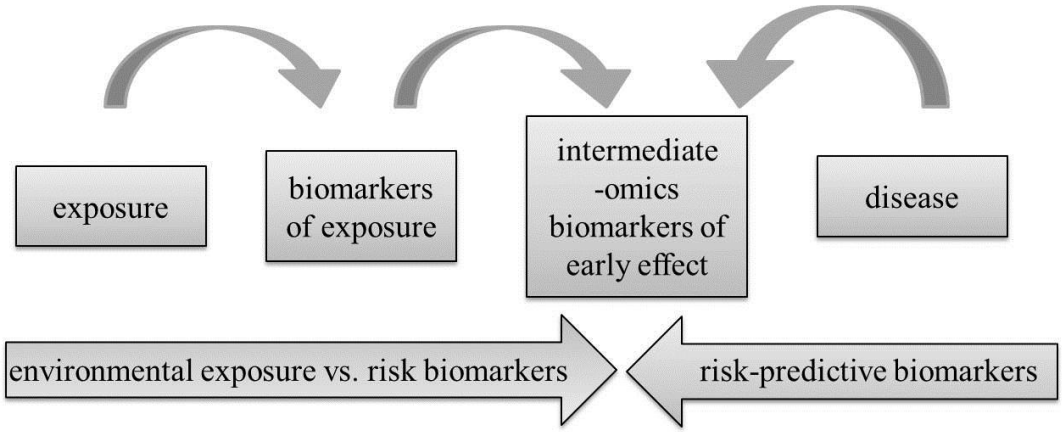

Figure 1: The "middle-out" HEALS approach

It builds upon the concept of connectivity between different biological systems from gene regulatory networks to cell and intracellular function, to tissue and organ response to human physiology. This is a high dimensional biology approach to the linkage between human exposure to exogenous and endogenous stressors and adverse health outcomes that allows us to develop credible adverse outcome pathways. A key departure of the

HEALS approach compared to environmental health paradigms thus far can be summed up to the recognition that “there are no confounders”. A truly comprehensive exploration of the multi-causal associations between environmental exposures and state of human health should not be constrained by the *a priori* determination of confounders, especially when the true mechanism of action underlying the phenotype of disease onset is not thoroughly known. On the contrary, it should embrace multi-causality and take stock of the interplay between exogenous and endogenous stressors to define the allostatic load, i.e. the departure from homeostasis that may be associated to disease phenotypes.

Such as a comprehensive depiction of the high dimensional biology that bridges the gap between exposure assessment and clinical diagnosis of disease requires the use of a complex network of exposure and biological information generated through numerous sensing and assay batteries. The overall methodological concept of HEALS and the different assays involved is graphically illustrated in Figure 2. This includes a wide array of state of the art technologies across all major disciplines of the environmental exposure, biochemistry, molecular biology, toxicology, bioinformatics and epidemiology arena. We see the exposome as a key methodological tool to functionally integrate the different types of information and channel them through a specific workflow in order to shed light on the environmental causes of diseases taking into account gene-environment interactions.

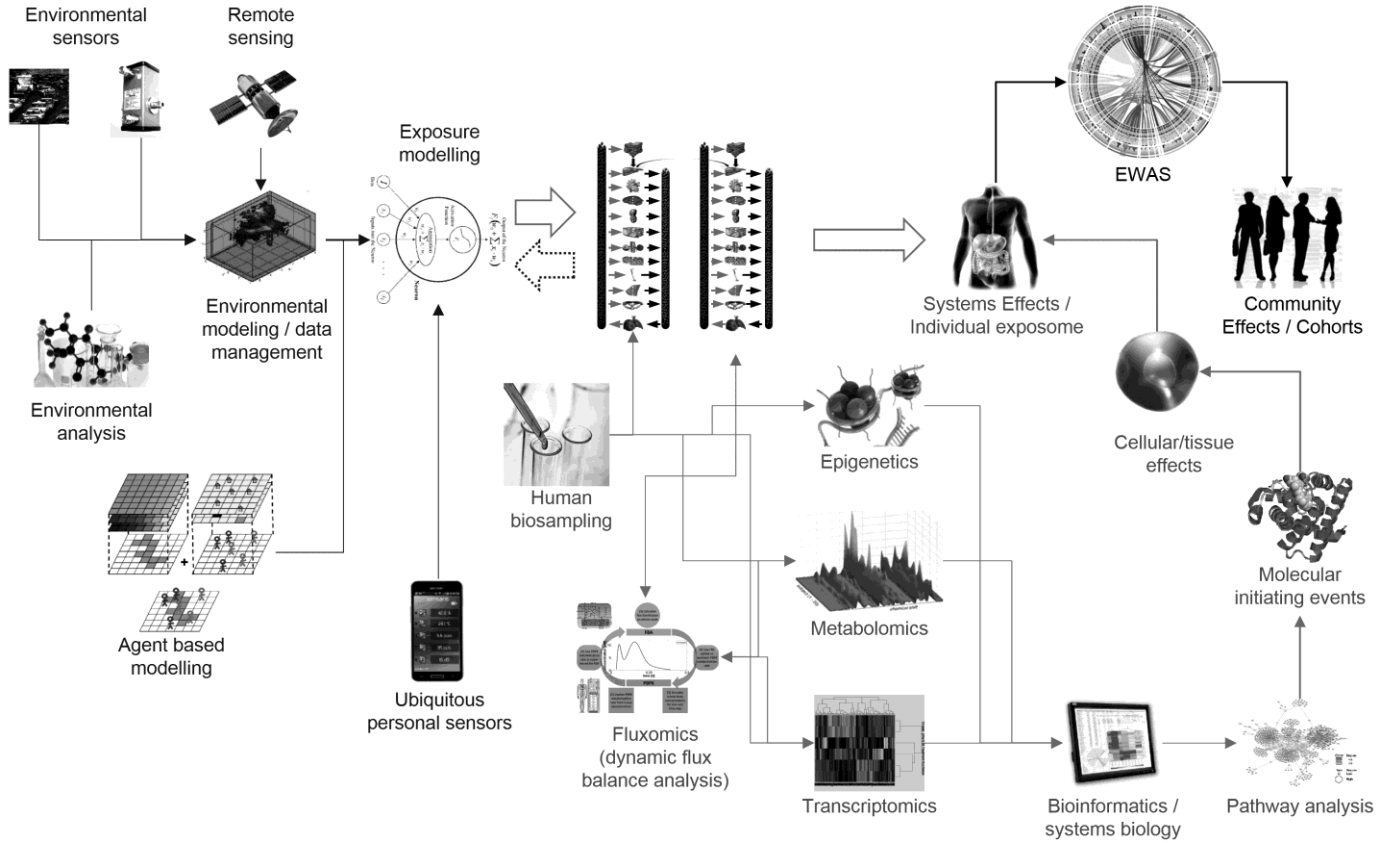

Figure 2: Graphical representation of the HEALS workflow and methodology

Assessment of the external exposome

Environmental and microenvironmental assessment – integration of data and models

Current external exposure assessment methodologies lack detailed resolution in time and space and generally omit consideration of the whole life-course of individuals. In addition, large-scale exposure assessments of the European population to environmental stressors rarely exist. A major objective of HEALS is to integrate the existing dataset and to fill data gaps in order to unravel the external exposome of individuals and population subgroups to multiple stressors via different pathways. Towards this aim, an extensive data collection / data mining scheme will be employed. The overall workflow of external exposome assessment in HEALS is graphically illustrated in Figure 3.

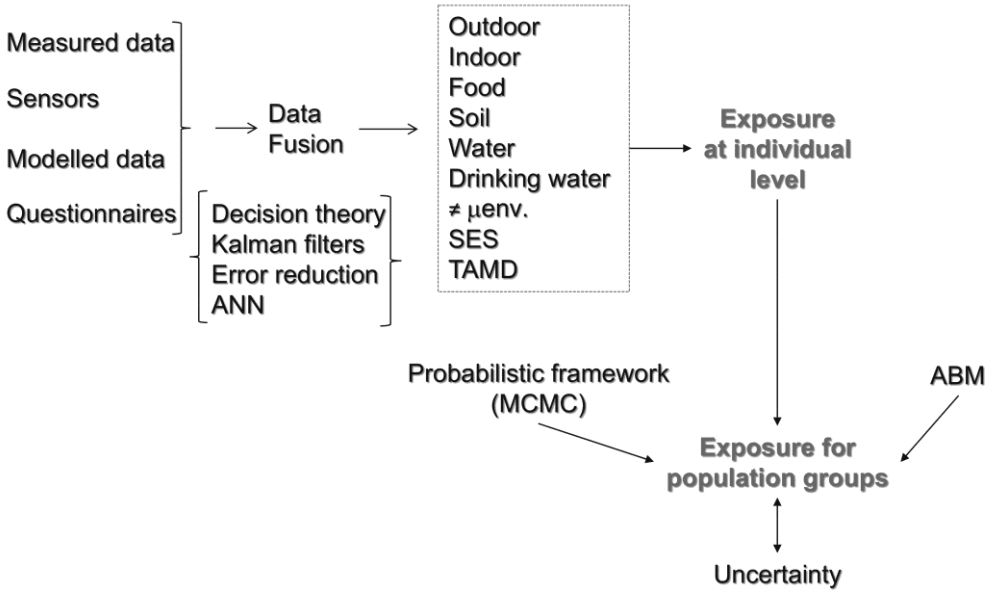

Figure 3: Workflow of external exposome assessment according to the HEALS paradigm

Data will be mined from different information sources including past and ongoing research and survey projects at the national and European levels and EU-wide monitoring systems specifically in areas of interest to the population studies addressed in the project. In this regard a critical step is the harmonization of this data to ensure that, for example, PMx data is directly comparable across Europe. Environmental data, combined with exposure models, will provide the basis for the development of a methodological and computational framework for estimating the external exposures of selected population groups to multiple stressors via different exposure routes (i.e. inhalation, ingestion and dermal). Depending on data availability, geospatial analysis and multimedia modelling will be used to estimate concentrations of the analysed toxic substances in 'microenvironments' and food. Methods for data integration as well as for handling missing data will be exploited by using several data fusion techniques (e.g. Kalman filter and ANN) aiming at maximizing the information

available through “intelligent” merging of the disparate environmental data available and the agent-based and other exposure modelling techniques such as personal sensor technologies. Probabilistic exposure modelling methodologies based on Markov-Chain Monte Carlo techniques will be used to estimate external exposure for selected population groups across time and space and to integrate uncertainty associated with the measurement and modelling of the various agent-specific (individual) exposure doses. Thus, based on “individual” exposure estimates, estimations for wider and coherent population subgroups will be derived. Additional detail on the implementation of novel technologies such as personal and remote sensors and agent-based modeling is given below.

Sensors

Technological advances in the recent years have produced sophisticated monitoring devices which can be carried or worn by a person during his/her regular daily routine allowing for personal exposure to be monitored explicitly. Smartphone apps, wireless devices and the downsizing of monitoring technologies and costs make it possible for various environmental stressors and exposure factors to be measured more easily and frequently, thus providing a more reliable “time–geography of exposure” shifting the current paradigm from a population to an individual level.

To unravel the individual external exposome in HEALS we will use several types of personal sensors which can be grouped according to the type of data they can provide: passive pollution measuring sensors to measure the pollution levels and climatic variables (temperature and humidity) encountered in the different locations where users spend their days, tracking location and physical activity sensors to gather information about the spatial patterns of user location and physical activity. Direct reading monitors will help us to identify whether peak exposures are more important than average exposure values, identify specific exposure pathways that dominate in critical time windows over an individual’s lifetime, and finally build individual exposure profiles. Combining information on individual position with spatially resolved pollution levels allows us to assign pollutant concentrations to a person as they move through different microenvironments. Moreover, information on individual physical activity as tracked by personal sensors allows the estimation of the breathing rates during different activities which in turns translated into inhaled dose. We are also studying the possibility to use personal sensors able to provide real-time data on air pollution exposure (NO_2, O_3, PM, temperature, humidity, sounds and location), which if reliable will constitute an added value.

This highly novel and promising approach will give us access to an unprecedented amount of “individualized exposure data,” which could greatly improve our understanding of exposure and health associations but which are worthless without interpretation (e.g. human behaviour recognition). This requires statistical advances, sophisticated data mining techniques, computing power as well as a careful sharing of data sources while also maintaining privacy protections for personal data. Big data is difficult to be used with classical relational databases, desktop statistics and traditional visualization packages. What is common for big data treatment is that it is not just about storing huge amounts of data; it is

the ability to mine and integrate data, extracting new knowledge from it. Appling this innovative framework to construct the individual exposome in the pilot EU-wide Exposure and Health survey (EXHES) as well as in the existing cohorts, HEALS will bring advances in this area to overcome the current limiting factors related to the analysis and the interpretation of the enormous wealth of data generated necessary to move the current approach from a population to a personalized level.

Agent-based modeling

Using data fusion techniques, traditional health and exposure data derived from fixed monitoring networks will be supplemented by a range of emerging novel techniques and technologies such as ABM, mobile phone apps, environmental sensor-webs, micro-sensors and satellite remote sensing. In addition we will considerably improve exposure modelling and phenotype identification using deterministic and probabilistic approaches, and applying new epidemiological and statistical methods to relate modelled exposure to health outcomes. ABM will be informed by data relating to an individual's behaviour within his/her environment (such as movement data within specific micro-environments) and between individuals exploring interactions around health related behaviours and risk determinants such as low SES. Using these parameters and the evolution of agents, simulations will produce detailed information relating to the emulated systems, data that can be used to fill in the gaps that exist in traditional datasets. This holistic approach is highly novel, taking the best from existing monitoring and sensor technology, but supplementing it with computational modeling simulations where real-world data is unavailable at the spatial and temporal scales that the individual exposome requires. Although commonly used elsewhere, ABM and fusion methods have not been applied to our knowledge in environmental epidemiology. This array of novel technologies, coupled with state-of-the-art environmental monitoring of chemical health stressors will provide a complete and dynamic picture of external exposure to environmental chemicals.

Socio-economic status

Individual health and well-being are influenced by past and present behaviour, health care provision and 'wider determinants' including social, cultural and environmental factors. It can be argued that socio-economic factors are as important as the physical environment in determining health impacts on human populations, since a disproportionate share of the burden of environmental exposure falls on vulnerable groups of society (defined as low SES, ethnic minorities, the elderly and young) due partly to issues of environmental (in)justice. In addition, SES can explain differences in external exposure because of different prevalence of specific behaviours in some groups, e.g. differences in diet between SES groups.

In HEALS we shall explore social and cultural factors at play, and evaluate how socio-economic status (SES) should be taken into account when modelling external exposure. Indeed, health outcomes are related to an individuals' environment, including factors such as water, soil and air content, exposure to hazardous materials, tobacco smoke, occupation, marital status, social support, characteristics of the home, in addition to the composition of the local built environment. Although social deprivation, or SES, is commonly treated as one of the confounders most widely adjusted for in ecological

epidemiological studies, in HEALS we will consider it as covariate to explain the myriad (albeit indirect) causation routes between social factors and health outcome.

Using existing national population censuses and surveys, we will use geospatial analysis methods to distribute our exposure estimates across all sectors of society at a local neighbourhood scale, for all of Europe. Gaps in censuses and the surveys will be modelled using agent-based modelling (ABM) to inform the division of exposure across different SES groups. Moreover working with the prospective twin cohort in the EXHES study, we will directly survey the twin cohort's parents for socio-economic status at the household level, acting as a check on the ecological level estimates.

Assessment of internal exposome

HBM

It is generally agreed that our understanding of the links between environmental exposure and human health can be greatly improved by the development of biological indicators of exposure, early toxicological effect, and genetic susceptibility. Human biomonitoring can be defined as "the method for assessing human exposure or their effect to chemicals by measuring these chemicals, their metabolites or reaction products in human specimens, such as blood or urine" (CDC 2005). It includes (1) biomarkers that allow assessment of exposure to a chemical on the basis of its measurement in a biological matrix (biomarker of exposure), (2) changes that have occurred in the biochemical or physiological makeup of an individual because of this exposure (biomarker of effect), or (3) biomarkers that assess a person's susceptibility to alter the progression along the exposure-effect continuum (biomarker of susceptibility) (NRC 2006). HBM data provide an integrated overview of the pollutant load any subject is exposed to, and hence may serve as good aggregate exposure proxy. The internal dose of a xenobiotic, has a much greater value for environmental health impact assessment as the internal body concentration is much more relevant than mere exposure data.

Human Biomonitoring (HBM) data will support and allow for a new approach to exposure assessment even when the quantity and quality of external exposures are unknown or ambiguous. Biomonitoring data can be used to compare exposures of the general population with special subpopulations and with toxicological animal data. They can also be used in risk assessment and risk management. For risk assessment, biomarker measurements are used to estimate dose, which can then be compared with toxicological parameters normally obtained from animal studies. A key task in human biomonitoring is their interpretation so as to put them in perspective exposure data with presumed toxic doses. Typically, large-scale human biomonitoring surveys gather biomarker measurements from single time points. These data are then used to make inferences about longer periods of toxicant intake, assuming that biomarker values are representative of steady-state conditions. However, this may not be a justified assumption, and may require additional investigation. By repeated sampling of individuals with particularly high and particularly low biomarker values (e.g. > P_{95} and < P_5), more insight in the toxicant exposure can be

obtained. The time-lag between sampling depends on the chemical half-life and the matrix used for biomarker determination.

Quantitative interpretation of biomonitoring data can best be accomplished by linking PBBK modeling with exposure pathway modeling in a probabilistic framework. This is carried out by combining PBBK models and HBM data to make inferences about environmental exposure scenarios for biomarker data collected in population-based studies (Mosquin et al. 2009). PBBK models when used in combination with MCMC techniques may allow disentangling the contribution of different environmental compartments through exposure reconstruction within a probabilistic framework. Ultimately, as biomarker data also reflect individual accumulation, distribution, metabolism and excretion (ADME) characteristics of chemicals, HBM data offer an excellent opportunity for the validation of the PBBK models.

Multi-omics, in vitro confirmatory analysis and exposure biology workflow

The HEALS approach to the internal exposome relies on evaluating the maximum available information from multiple omics data. Towards this aim, a well-structured exposure biology workflow has to be followed, as graphically illustrated in Figure 4.

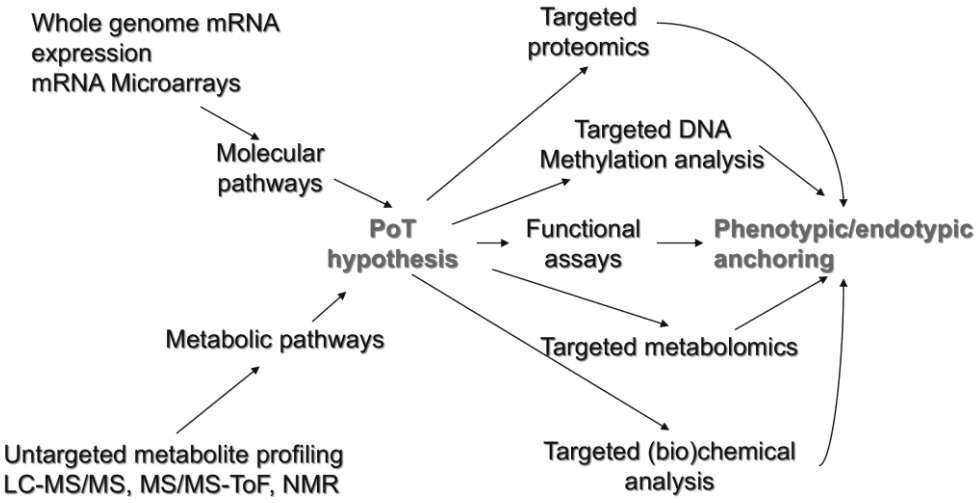

Figure 4: Analytical exposure biology workflow according to the HEALS paradigm

This workflow is specifically designed to elucidate the causal links of genome, environment and disease, entailing a stepwise process:

1. The assessment is initiated from an agnostic point of view; no former hypothesis on associations between environmental stressors and adverse health outcomes has to be formed unless driven by some evidence of perturbed pathways. Untargeted –omics (transcriptomics and metabolomics/adductomics) analysis is performed on human biological matrices using state of the art microarray and liquid chromatography / mass spectrometry and nuclear magnetic resonance techniques.

2. The results of the coupled –omics analyses performed in step (1) are analysed jointly with advanced bioinformatics and consultation of annotated libraries to identify gene regulatory and metabolic pathways that are modulated jointly. These pathways would allow us to construct putative hypotheses on Pathways of Toxicity (or of Adverse Health Outcome).
3. The hypothesized pathways of toxicity or of adverse health outcome can be mechanistically explored through the use of targeted –omics (including proteomics and metabolomics), functional assays (e.g. Comet or micronuclei assays for genotoxicity), DNA methylation for epigenetic analysis and biochemical biomonitoring. These targeted analyses would allow us to identify the very few pathways that may mechanistically interpret the observed associations between exposure and health outcomes. Our analyses at this step would be further supported by in vitro assays for mechanistic confirmation and biology-inspired modeling and bioinformatics for data analysis.

The final result of the above staggered analytical exposure biology framework would be:

(a) Identification of endotypes (early biological event markers) that can be associated causally with phenotypes of adverse health outcomes.

(b) Determination of allostatic status of the analysed individuals

(c) Association of allostatic status with individual risk of disease and eventually burden of disease on the population level

Technically, the different arrays of analytical exposure biology will include: (a) transcriptomics, meaning the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using high-throughput methods, such as microarray analysis (b) Metabolomics (untargeted and targeted); entails the global analysis of metabolites in easily accessible human body fluids. Established and robust sample preparation protocols (e.g. for plasma, urine) in combination with the latest generation of NMR and ultra-high-pressure liquid chromatography (UPLC)-mass spectrometers (MS), high-pressure liquid chromatography-electrochemical detection (HPLC-ED), and LC-ECA (electrochemical coulometric arrays) will be applied. For novel matrices (e.g. meconium) ad hoc protocols have been developed. These approaches will be applied to samples from selected human cohorts to determine the impact of exposure to chemicals (e.g. pesticides, industrial solvents etc.) on the human exposome. (c) Adductomics; deals with the measurement of adducts of electrophiles with DNA, blood proteins (haemoglobin, albumin), and glutathione. HPLC-ED and MS will be used to evaluate levels of 8-hydroxyguanine and O6-alkylguanine in blood samples from human populations, in relation to genetic susceptibility (DNA repair genotype). MS based technology (GS-MS) will be employed to quantify blood protein adducts. Also, LC-MS/MS for measuring protein adducts will be further developed for application in HBM settings to contribute to the definition of the exposome. With respect to SNP profiling in toxicologically relevant genes, genotyping assays will be developed and optimized using Taqman technology on the existing Biomark Fluidigm platform. In addition, functional assays of specific DNA repair proteins will be

performed. Genome-wide DNA methylation profiling and miRNA expression are key to investigating the importance of epigenetic effects on environmental health. Array- and sequencing-based DNA methylation profiling technologies will be used to meet this goal. In particular, Agilent microarray miRNA profiling and genome-wide bisulfite sequencing will be used to analyse miRNA expression and DNA methylation respectively, on the biosamples of the cohorts according to the following procedure:

i) Apply SNP analysis, miRNA analysis and next-generation sequencing to define genetic susceptibility for chemicals at population level (e.g. DNA repair phenotypes, Phase II reaction genotypes)
ii) Identify differences between epigenetically influenced and independent SNPs
iii) Develop sample handling and metabolomics workflows
iv) Generate metabolite expression data; identify biomarkers from cohorts and *in vitro* models (glutathione, s-adenosyl methionine, bisphenol A)
v) Develop methodologies (LC-MS/MS) and generate DNA and protein adduct data from cohorts.
vi) Determine methyl/hydroxymethylcytosine at specific genomic sequences (promoters, CpG islands, repeated sequences)
vii) Identify Illumina methylation at 20 genes, and apply Methylator phenotypic analysis for this set of genes.

The causal link between exposure and disease endpoint as revealed by the identification of –omics markers of effect requires mechanistic confirmation, which will be based on in vitro assays assessing complex toxic endpoints relevant to disease endpoints on cell models derived from liver (hepG2/ HuH7, HepaRG), adipocytes (hMADS) or cells differentiating into neurons (C11) to establish for the different technologies omics compound specific hazard signatures. This will also support the system biology model of causation developed using advanced bioinformatics. Next to confirming the systems biology exposome model, these mechanistic studies provide kinetic constants and compound receptor-binding data useful to further develop the model. In this context, it is possible to use these data to design in vitro human cell-based assays and systems biology experiments, to mechanistically anchor omics observations from cohort studies in light of the linkage between exposure and health outcome data. The hypothesis is put forward that the correspondence of omics signatures in cohort biomaterials with those obtained from experimental in vitro models can demonstrate the validity of chosen biomarkers. In fact this approach can be used to corroborate if: (a) signatures are more closely related to the actual compound exposure (e.g. biotransformation) and thus represent a marker for exposure; (b) more closely related to mechanisms of disease, and consequently represent a marker for health impact. Overall, empirical perturbation detected from HBM data (including omics markers) can be integrated into in vitro testing to help verify system biology-level or “real-world” human toxicity of environmental stressors (Pleil 2012).

Internal dose modelling

The use of internal dose modelling aims at integrating exposure data and modelling output with HBM data. Its goals are to (a) provide the time history of the internal exposure

profile, focusing on susceptible developmental stages; (b) assimilate the biomonitoring data related to the cohorts to estimate the individual exposome in quantitative terms; and (c) derive reliable Biologically Effective Dose (BED) values for the compounds of interest so that they can be associated to observed health outcomes. The key component of the above is the development of a lifetime (including gestation and breastfeeding) generic PBBK model (Sarigiannis and Karakitsios 2012) incorporating mixtures interaction (Sarigiannis and Gotti 2008) and a framework for biomonitoring data assimilation (Georgopoulos et al. 2008). Aiming to expand the applicability of the generic PBBK model to cover as much as possible the chemical space, parameterization of the model for known and new chemicals with limited information is done through the development of QSAR models. The generic PBBK model will also be used to reconstruct exposure from HBM data (Andra et al. 2015). A tiered approach will be followed as a function of data availability (periodicity and size of sampling, specimen type) and requirements of the exposure reconstruction analysis (temporal analysis of exposure, contribution from different routes), ranging from Exposure Conversion Factors (ECFs) (Tan et al. 2006), up to Markov Chain Monte Carlo analysis. Inputs involve spatial and temporal information on micro-environmental media concentrations of xenobiotics and corresponding information on human activities, food intake patterns or consumer product use that result in intakes; outputs are the observed biomarkers; and the error metric can be defined in terms of population variation (the latter has to be lower than the intra-individual variation, which may be associated to measurement or other random error source). On the individual level, PBBK will be combined with multimedia models and survey questionnaires to identify exposure sources. PBBK modelling will also be used to estimate the internal doses of xenobiotics that exceed levels associated with biological pathway alterations (Judson et al. 2011) and, eventually, health risk. The latter can involve the use of specific omics results (metabolomics analysis) and associations of BED to early biological responses. In addition, BED would be used to quantify the effect of compound-induced extracellular perturbations on metabolic states, so as to directly couple the PBBK model with metabolic regulatory networks. Direct coupling defines a feedback loop that connects clearance and metabolite production rates to metabolism regulation (Eissing et al. 2011) via dynamic flux balance analysis (FBA) (Krauss et al. 2012), thus linking to the regulatory networks identified in the bioinformatics analysis carried out in HEALS.

Bioinformatics

Another key aspect of the HEALS methodological approach is the development of innovative bioinformatics strategies for biomarker prediction. The bioinformatics tools currently available for biomarker detection and analysis range from statistical approaches to data mining. The latter can be grouped as descriptive or predictive. The first describes the dataset in a concise and brief manner and presents general properties of the data. The second constructs one or a set of models, performs inference on the available dataset, and tries to predict the outcome on new datasets. Data mining has been carried out on different types of data, including clinical (Exarchos et al. 2006), biological (Exarchos et al. 2009) and environmental (DA Sarigiannis et al. 2009), exhibiting constantly excellent results.

Association rule mining is a data mining technique commonly used for local-pattern discovery in unsupervised learning systems (Han and Kamber 2011). In particular, it is the

discovery of association relationships or correlations among a set of items. Many efficient algorithms have been proposed in the literature for association rule mining, with the most common being the Apriori and the FPgrowth algorithms. Data associations can also be detected through *cluster analysis* where similar objects are grouped together based on certain attributes. Several algorithms have also been developed for clustering with the most commonly used ones being based on partitioning (K-means), connectivity (graph clustering) and neural networks (self-organizing maps, SOMs).

When prediction is the focus of the analysis, it can be reached through *classification* or *regression*. The available techniques for both tasks are numerous, with decision trees, neural networks and support vectors being more widely spread among the research community. Each of the above technique has its own strong features and limitations, hence in many multifactorial problems, as is the case of exposome a *model combination* (i.e. meta-model) approach has been adopted to improve the prediction performance (Exarchos et al. 2007; Papaloukas et al. 2008) as single model usually fails to learn and generalize efficiently on the entire training set. Building a meta-model offers the ability to each one of the component models to specialize on certain types or even subsets of data. By post-processing the architecture of the derived meta-model as well as by interpreting its inference mechanism, significant *multivariate profiles* can be revealed that best describe the available data. The systematic examination of these profiles can indicate the most important traits of the problem under study, which in HEALS means robust candidates for predictive biomarkers.

In HEALS, bioinformatics further advance the concept of exposome, by integrating the omics responses of different omics levels, through mapping of regulatory networks and disease pathways. Functional integration of different omics results using bioinformatics tools aiming at development of adverse outcome pathways (AOPs) (Gutsell and Russell 2013) for the endpoints addressed in HEALS. Different omics data that will be derived both from existing on-line databases and from *in vitro* laboratory experimentation and will be mapped onto regulatory pathways. They will be analysed using network visualization environments such as Agilent GeneSpring, Thompson-Reuters MetaCoreTM and Reactome/Functional Interaction network plug-in for Cytoscape, so as to create systems toxicology hypotheses from human data, related to the endpoints addressed in HEALS, as well as to identify common nodes across several pathways of exposure to different compounds of relevance. This is especially important under the prism of cumulative exposure; metabolomic data, in combination with the transcriptome data, show that the toxicologic mechanism for the combined toxicity of several compounds exerting interactions beyond additivity occur at a network level (Roede et al. 2014).

Cohort studies

Use of big data and re-analysis of previous studies

The HEALS approach and tools will be put to test through their application in a number of population studies (incl. twins studies) across different exposure settings tackling key health endpoints of the SCALE initiative and the Parma Declaration for both children and the elderly. The overall population size involved in these studies is up to ca. 335,000 individuals

covering different age, gender and socio-economic status groups. The cohorts involved are dispersed across Europe to provide sufficient geographic coverage so as to facilitate drawing conclusions at the EU-wide scale. These cohorts have been carefully selected to comprise twins. The classical twin methods combined with novel technologies represent a powerful approach towards identifying and understanding the mechanisms and pathways that underlie complex traits. In this way, epigenetic modification of both exposure to and effect of co-exposure to chemical, physical and biological environmental stressors throughout a person's lifetime will be easier to take into account when deducing environmental exposure and health associations.

HEALS will make use of environmental and health data including bio-banks from pre-existing twin registries from:

- Denmark (Danish Twin Registry: 14,000+ twins born 1931-1969),
- Finland (Finnish Twin Cohort Study: 12,966 MZ and DZ twin pairs (25,932 individuals) with both members currently alive),
- Italy (Italian Twin Register: 25,000 twins),
- Netherlands (Netherlands Twin Register: 87,000 twins),
- Norway (Norwegian Twin Registry: 31,440 twins, MoBa Study: 1,900 twins),
- Sweden (Swedish Twin Registry: 20,000 MZ pairs, 25,000 same-sex DZ pairs, and 30,000 opposite-sex DZ pairs, respectively) and the
- UK (EpiTwins: 5,000 twins followed-up 20 years).

EXHES study

The technological and computational integration proposed in HEALS will be tested through a pilot EXposure and Health Examination Survey (EXHES) that will be organized in 10 EU countries to test the applicability, technical feasibility and cost-effectiveness of the HEALS approach for EU-wide large population surveys and to disentangle the complex interactions among factors underlying the development of asthma/allergies, metabolomics diseases (obesity/diabetes) and neurodevelopmental/degenerative troubles. Taking stock of the results of the already existing population studies, EXHES will combine a longitudinal and a nested case-control phase to allow for better definition of environmental exposures and better characterization of disease and risk phenotypes/endotypes and the underlying mechanisms (through omics) over the limited duration of the project, whilst setting the foundation for post-project follow-up.. The lessons drawn from the EXHES pilot survey will provide the basis for drafting scientific advice, protocols and, eventually, guidance for the setting up of a European Health and Exposure Survey, paving the way to EU-wide assessments.

The objectives of the HEALS pilot European Exposure and Health Examination Survey (EXHES) are:

1. To collect **new harmonized and standardised exposure and health data in European twins and matched singletons**, describe for the first time the normal and pathological development of twins and matched singletons in different European centres and to make comparisons within and between countries;

2. To obtain **baseline measures for assessment of future trends** in environmental exposures and major chronic diseases and to provide a framework for further aetiological research into lifestyle, environmental, epigenetic, genetic and medical care factors affecting health;
3. To determine **the contribution of the exposome to later health and development compared to other factors** operating during childhood. In particular, to confirm that the plasticity of the foetal and early life period may lead to programming of body functions by environmental conditions as early as *in-utero* including low level exposures to environmental toxics;
4. To **provide insight into the potential mechanisms at the origin of diseases**.

The EXHES survey focuses on children health and development: growth and development of the adipose tissue, neurodevelopment, as well as immune, respiratory, and metabolic functions that constitute intermediate phenotypes/endotypes. However, it is also expected that the most frequent conditions, such as respiratory diseases, asthma, allergies, overweight (disease phenotypes) will be observed in the first years of life. Risk and protective markers and factors considered include mother-and-child diet, pre- and post-natal environmental exposure, social factors, and –omics markers. EXHES comprises 2 phases:

- EXHES PHASE I is a **3-year longitudinal** study including up to 300 pairs of twins and 300 matched singletons (on day of birth, sex, social class, region) in each country and their parents, recruited over 18 months. Ca. up to 4,000 mothers, 4,000 fathers and 6000 children, 4,000 of whom are twins will be included in Europe (Croatia, France, Germany, Greece, Italy, Portugal, Slovenia, Spain, UK, Poland).
- EXHES PHASE II is a **nested case-control** study (cases based on the targeted diseases and healthy controls) conducted in a sub-sample of 210 individuals (140 twins and 70 singletons) and 140 mothers in each country for a total of ca. 2,000 individuals in which -omics and geo-referenced exposure will be analysed.

Overall methodology for assessing environmental burden of disease – lifetime risks for individuals

HEALS introduces a novel approach towards defining causal associations between health status and environmental stressors through the integrated use of advanced statistical tools for environment-wide association studies (EWAS). In our approach, environmental factors that are correlated are not considered confounders; rather they are co-variates, which are in "**linkage disequilibrium**" with each other. EWAS findings could then be used to identify further factors that may be in "disequilibrium", for further detailed measurement and causal identification. Internal doses will be coupled to health impacts on the local population through advanced statistical methods to derive the dose–response functions which account for differences in exposure patterns, susceptibility differences and inter-individual variation in health response. The approach starts from the biomarker values measured in different biological matrices (urine and peripheral blood) to estimate through the application of the lifetime generic PBBK model the biological effective dose in the target tissue, which is consistent with the biomarker level measured. To estimate the health impact we will use a

statistical approach based on survey-weighted logistic multivariate regression adjusted for different covariates (age, sex, socio-economic status etc.) linking internal doses with health effects or intermediate biological events that can be associated to health perturbations through pathway analysis considering the interdependence of the covariates (using as metric an analogy of the "linkage disequilibrium" metric used in genome-wide association studies).

The general formulation of the approach is based on the mathematical linkage of health end points (expressed in terms of odds ratio, *p*) with different covariates (age sex, SES, lifestyle choices such as smoking, etc.) and the internal dose in the target tissue (X_{Factor}) such as the expression below:

$$logit\left({}^{p}/_{1-p}\right) = \alpha + \beta_0 \cdot cov_1 + \beta_1 . cov_1 + \beta_2 \cdot X_{factor} \ldots + \beta_n . cov_n$$

where *cov* represents the different covariates used in the model and α and β are the regression coefficients which take into account the interdependence between covariates.

The European dimension will be taken into account through hierarchical multi-level models using logic regression embedded in a linear regression framework. Multilevel models are statistical models applied on parameters that vary at more than one level. Logic regression, an adaptive classification and regression procedure initially developed to reveal interacting single nucleotide polymorphisms (SNPs) in genetic association studies, will be used in any setting with binary predictors, when the interaction of these covariates is of primary interest. Logic regression searches for Boolean combinations of binary variables that best explain the variability in the outcome variable will determine variables and interactions that are associated with the response and/or have predictive capabilities. Considering that adjustment-variable selection in epidemiology can be broadly grouped into background knowledge-based and statistics-based approaches, the use of **Directed Acyclic Graphs** (DAGs) has come to be a core tool in the background-knowledge approach. DAGs present assumed relationships between variables graphically and, based on these assumptions, identify variables to adjust for confounding and other biases. The enhancement in DAGs construction in epidemiology that includes arrow-on-arrow representations for effect modification proposed by Weinberg (2007), boosted their applicability, including the identification of the direction of unmeasured confounding bias (VanderWeele et al. 2008), adjustment for socioeconomic status in occupational cancer studies (Fleischer and Diez Roux 2008), identifying variables that need to be adjusted for high resolution spatial epidemiology, as well as metabolic syndrome confounders (Shahar 2010). Recent applications include the assessment of complex gene-environment interactions (Geneletti et al. 2011), as well as asserting that the role of ozone in studies of temperature and mortality is a causal intermediate that is affected by temperature and that can also affect mortality, rather than a confounder. From the methodological point of view, Evans et al. (2012) combined directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. **Bayesian inference** based modeling has recently been used for the minimization of covariates and determinants in epidemiology, considering that such models offer a more scientifically defensible framework

for epidemiologic analysis than the fixed-effects models now prevalent in epidemiology (Greenland 2000). Their applicability domain includes Bayesian regression models for analyzing longitudinal binary process data with emphasis on dealing with missing data (Su and Hogan 2008), Bayesian adjustment for covariate measurement errors (Hossain and Gustafson 2009) and more specifically for health association studies as Bayesian sensitivity analysis for mismeasured and unobserved confounders (Gustafson et al. 2010). In HEALS, the tools mentioned above are combined to provide an integrative framework for identifying the proper covariates that affect the health endpoints of relevance under the light of the health surveys, as well as to constrain the magnitude and direction of bias parameters. In practice, HEALS ushers in "**enviromics**", the study of a wide array of environmental factors in relation to health and biology

DISCUSSION

Shortcomings of existing approaches

Despite the worldwide awareness about the relevance of introducing exposome in environmental health studies, there is still lack of understanding the opportunities for interdisciplinary research, aiming at understanding the complex gene and environment interaction. Although significant advances are proposed by molecular epidemiology and statistical approached based on gene-environment wide associations, a more in depth and synthetic approach is required to understand the causality of environmental burden of disease.

Advances proposed in HEALS

External exposome

Firstly, it is important to capture individual lifetime exposure and to quantify differences in exposure among individuals living in phenomenally similar environments. This will be greatly facilitated by using personal sensors and developing the algorithms for translating the acquired information into personalized activity patterns. This will allow us to draw a more accurate timeline of environments and microenvironments encountered, as well as to the physiological conditions (inhalation and cardiac rate) occurring to the contact medium. Considering that it is impossible to have the detailed full picture of large population groups, extrapolation from the detailed individual profiles to population groups have to be wisely executed, accounting for the different parameters resulting in different behavioral patterns such as age, gender and socioeconomic status; agent based modelling seems a promising tool towards this direction.

Internal exposome

Translating external exposure into internal exposure and biological responses is the next crucial step. Firstly, we need to take into account that internal dose resulting from the same external exposure might vary significantly among individuals; this is the result of differences in physiology (developmental stage, bodyweight, inhalation rate, obesity status), co-

exposure to other compounds (environmental or pharmaceutical) and also to polymorphisms related to exposure to xenobiotics. Before moving forward to genome and environment wide associations, all these parameters have to be taken into account for, utilizing PBBK modelling and the SNPs profiling related to genes regulating xenobiotics metabolism. Thus, association between exposure and response have to be initiated after “adjusting” external exposure into individual internal dose profiles.

In depth analysis of the connectivity approach and the integrated high dimensional biology

HEALS functionally integrates the broadest array of technologies involved in exposome studies. Although collection of HBM data is the starting point of the assessment, bi-directional mechanistic links (external exposure to HBM and from HBM to disease) should be in equally investigated in depth; misclassification of exposure or its translation into actual biologically effective dose results in loss of valuable information. From the other side, analysis of internal exposome should be more comprehensive, explaining how the exogenous molecules identified result in changes in endogenous molecules, both found in HBM samples (Rappaport et al. 2014). Moving from associations to disease causality results the mechanistic understanding of all intermediate events related to human homeostasis unbalancing. Identifying these perturbations requires data acquisition (multi-omics, with a special focus on transcriptomics and metabolomics) and data interpretation and this can be effectively described by regulatory and pathway analysis (Roede et al. 2014). It is also very important to highlight the importance of cumulative exposure to the environmental disease burden. The results of the agnostic transcriptome and metabolome search on biological samples with regard to the identification of potential genomic signatures that can serve as exposure biomarkers. Even when exposure to single compounds only shows intermediate results, it has been found that co-exposure to a real life mixture may have more than additive effects on gene expression modulation (D Sarigiannis et al. 2009). After the agnostic tier it is possible to identify not only single genes that have shown significant modulation in expression levels, but also determine the biological pathways that are regulated by gene networks that were significantly modulated with regard to their induction levels from exposure to xenobiotics. Combined with the changes in metabolomics profile, pathway analysis can reveal the key pathways involved in each perturbation, (e.g. p53 or oxidative stress) and how these are differentially modulated from specific chemical families, while specific genes or gene sequences as well as combination of other small molecules could be characterized as molecular markers of exposure. By further proceeding with the targeted analysis described in the exposure biology workflow described above, causalities among genome, environment and specific endotypes or disease phenotypes can be deciphered.

What will be the benefits from HEALS in terms of public health protection

HEALS foresees the execution of a pilot exposure and health assessment at the European level (EXHES) focusing on families of neonates including twins and singletons in a longitudinal study design in order to account properly for epigenetic effects on the link between internal and external exposome and health outcomes. The study will set the stage

for harmonized and standardized exposure and health data collection and interpretation in Europe; it will set the baseline for assessment of future trends in environmental exposures and chronic disease and will establish the contribution of the exposome to child development and later life health vs. other factors operating during childhood. Addressing exposure and health in population aged in susceptibility windows during growth (including pregnancy and early life) and elderly, HEALS will focus on the environment of vulnerable populations including children which is one of the priorities of the Parma declaration. Overall, HEALS will provide new information of the potential role of environmental chemicals to disease etiology and search for predictive markers which may contribute to improve prevention management strategies.

CONCLUSIONS

HEALS introduces a new exposome based paradigm for interdisciplinary scientific work in the area of environment and health. This denotes an approach that builds on the exploration of the interconnections between the co-existence of multiple stressors and the different scales of biological organization that together produce the final adverse health effect. This marks a clear departure from the conventional paradigm, which seeks to shed light on the identification of singular cause-effect relationships between stressors and health outcomes. It entails creating a new way of combining health-relevant information coming from different disciplines, including (but not limited to) environmental science, epidemiology, toxicology, physiology, molecular biology, biochemistry, mathematics and computer science. According to the HEALS paradigm, all factors affecting internal and external exposome are treated as co-variates, rather than as confounders. The functional integration of these different information classes into a unique framework will enhance our understanding of the complex interaction between genome and exposure to environmental factors. This will be verified through the re-evaluation of existing cohorts (including re-analysis of samples) and the deployment of a new cohort, which in turn will act as pilot for the design of future relevant studies.

In this regard the estimated impacts of the HEALS concept and methodology are expected to be extensive leading to potentially highly significant improvements in public health by:

(a) reducing the uncertainty in exposure assessment and consequently the assessment of risk from environmental exposures;

(b) reducing the fragmentation of exposure data and improving the harmonization and comparability of data; and

(c) elucidating the mechanistic basis of the links between environmental exposure and health outcomes making use of both existing and new data collected in EU-wide study (EXHES).

All the above would be expected to put HEALS at the forefront of the worldwide research in environment-wide association studies paving the way for a new paradigm in unraveling the complex interactions between environment and health.

REFERENCES

Andra SS, Charisiadis P, Karakitsios S, Sarigiannis DA, Makris KC. 2015. Passive exposures of children to volatile trihalomethanes during domestic cleaning activities of their parents. Environmental Research 136:187-195.

CDC. 2005. Third national report on human exposure to environmental chemicals. Atlanta, Georgia, USA.

Chadeau-Hyam M, Campanella G, Jombart T, Bottolo L, Portengen L, Vineis P, Liquet B, Vermeulen RCH. 2013. Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. Environmental and Molecular Mutagenesis 54:542-557.

Eissing T, Kuepfer L, Becker C, Block M, Coboeken K, Gaub T, Goerlitz L, Jaeger J, Loosen R, Ludewig B, Meyer M, Niederalt C, Sevestre M, Siegmund HU, Solodenko J, Thelen K, Telle U, Weiss W, Wendl T, Willmann S, Lippert J. 2011. A computational systems biology software platform for multiscale modeling and simulation: Integrating whole-body physiology, disease biology, and molecular reaction networks. Frontiers in Physiology FEB.

Evans D, Chaix B, Lobbedez T, Verger C, Flahault A. 2012. Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. BMC Medical Research Methodology:156.

Exarchos TP, Papaloukas C, Fotiadis DI, Michalis LK. 2006. An association rule mining-based methodology for automated detection of ischemic ECG beats. IEEE Transactions on Biomedical Engineering 53:1531-1540.

Exarchos TP, Tsipouras MG, Exarchos CP, Papaloukas C, Fotiadis DI, Michalis LK. 2007. A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. Artificial Intelligence in Medicine 40:187-200.

Exarchos TP, Tsipouras MG, Papaloukas C, Fotiadis DI. 2009. An optimized sequential pattern matching methodology for sequence classification. Knowledge and Information Systems 19:249-264.

Feil R, Fraga MF. 2012. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13:97-109.

Fleischer NL, Diez Roux AV. 2008. Using directed acyclic graphs to guide analyses of neighbourhood health effects: An introduction. Journal of Epidemiology and Community Health 62:842-846.

Flintoft L. 2005. Identical twins: epigenetics makes the difference. Nat Rev Genet 6:667-667.

Geneletti S, Gallo V, Porta M, Khoury MJ, Vineis P. 2011. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. Emerging Themes in Epidemiology 8.

Georgopoulos PG, Sasso AF, Isukapalli SS, Lioy PJ, Vallero DA, Okino M, Reiter L. 2008. Reconstructing population exposures to environmental chemicals from biomarkers: challenges and opportunities. Journal of Exposure Science and Environmental Epidemiology 19:149-171.

Greenland S. 2000. When should epidemiologic regressions use random coefficients? Biometrics 56:915-921.

Gustafson P, McCandless LC, Levy AR, Richardson S. 2010. Simplified Bayesian Sensitivity Analysis for Mismeasured and Unobserved Confounders. Biometrics 66:1129-1137.

Gutsell S, Russell P. 2013. The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment. Toxicology Research 2:299-307.

Han J, Kamber M. 2011. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems):Morgan Kaufmann.

Hossain S, Gustafson P. 2009. Bayesian adjustment for covariate measurement errors: A flexible parametric approach. Statistics in Medicine 28:1580-1600.

Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB, Houck KA, Thomas RS, Wetmore BA, Dix DJ. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chemical Research in Toxicology 24:451-462.

Krauss M, Schaller S, Borchers S, Findeisen R, Lippert J, Kuepfer L. 2012. Integrating Cellular Metabolism into a Multiscale Whole-Body Model. PLoS Computational Biology 8.

Mosquin PL, Licata AC, Liu B, Sumner SCJ, Okino MS. 2009. Reconstructing exposures from small samples using physiologically based pharmacokinetic models and multiple biomarkers. Journal of Exposure Science and Environmental Epidemiology 19:284-297.

NRC. 2006. Human Biomonitoring for Environmental Chemicals, Committee on Human Biomonitoring for Environmental Toxicants, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies. Washington DC:National Research Council.

Papaloukas C, Granseth E, Viklund H, Elofsson A. 2008. Estimating the length of transmembrane helices using Z-coordinate predictions. Protein Science 17:271-278.

Pleil JD. 2012. Categorizing biomarkers of the human exposome and developing metrics for assessing environmental sustainability. Journal of toxicology and environmental health Part B, Critical reviews 15:264-280.

Rappaport SM, Smith MT. 2010. Environment and disease risks. Science 330:460-461.

Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. 2014. The Blood Exposome and Its Role in Discovering Causes of Disease. Environmental Health Perspectives 122:769-774.

Roede JR, Uppal K, Park Y, Tran V, Jones DP. 2014. Transcriptome–metabolome wide association study (TMWAS) of maneb and paraquat neurotoxicity reveals network level interactions in toxicologic mechanism. Toxicology Reports 1:435-444.

Sabel CE, Boyle P, Raab G, Löytönen M, Maasilta P. 2009. Modelling individual space-time exposure opportunities: A novel approach to unravelling the genetic or environment disease causation debate. Spatial and Spatio-temporal Epidemiology 1:85-94.

Sarigiannis D, Marafante E, Gotti A, Reale GC. 2009. Reflections on new directions for risk assessment of environmental chemical mixtures. International Journal of Risk Assessment and Management 13:216-241.

Sarigiannis DA, Gotti A. 2008. Biology-based dose-response models for health risk assessment of chemical mixtures. Fresenius Environmental Bulletin 17:1439-1451.

Sarigiannis DA, Karakitsios SP, Gotti A, Papaloukas CL, Kassomenos PA, Pilidis GA. 2009. Bayesian algorithm implementation in a real time exposure assessment model on benzene with calculation of associated cancer risks. Sensors 9:731-755.

Sarigiannis DA, Karakitsios SP. A dynamic physiology based pharmacokinetic model for assessing lifelong internal dose. In: Proceedings of the AIChE 2012, October 28 - November 2, 2012 2012. Pittsburgh, PA.

Sarigiannis DA, Karakitsios SP, Zikopoulos D, Nikolaki S, Kermenidou M. 2015. Lung cancer risk from PAHs emitted from biomass combustion. Environmental Research 137:147-156.

Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, Escobar J, Flowers D, Fotopulos D, Garcia C, Gomez M, Gonzales E, Haydu L, Lopez F, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salazar A, Tsai M, Myers RM. 2004. Quality assessment of the human genome sequence. Nature 429:365-368.

Shahar E. 2010. Metabolic syndrome? A critical look from the viewpoints of causal diagrams and statistics. Journal of Cardiovascular Medicine 11:772-779.

Su L, Hogan JW. 2008. Bayesian semiparametric regression for longitudinal binary processes with missing data. Statistics in Medicine 27:3247-3268.

Tan YM, Liao K, Conolly R, Blount B, Mason A, Clewell H. 2006. Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. Journal of Toxicology and Environmental Health - Part A: Current Issues 69:1727-1756.

VanderWeele TJ, Hernán MA, Robins JM. 2008. Causal directed acyclic graphs and the direction of unmeasured confounding bias. Epidemiology 19:720-728.

Vineis P, van Veldhoven K, Chadeau-Hyam M, Athersuch TJ. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ Mol Mutagen 54:461-467.

Vineis P, Wild CP. 2014. Global cancer patterns: causes and prevention. Lancet 383:549-557.

Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, Thomsen C, Wright J, Athersuch TJ, Avellana N, Basagaña X, Brochot C, Bucchini L, Bustamante M, Carracedo A, Casas M, Estivill X, Fairley L, van Gent D, Gonzalez JR, Granum B, Gražuleviciene˙ R, Gutzkow KB, Julvez J, Keun HC, Kogevinas M, McEachan RRC, Meltzer HM, Sabidó E, Schwarze PE, Siroux V, Sunyer J, Want EJ, Zeman F, Nieuwenhuijsen MJ. 2014. The Human Early-Life Exposome (HELIX): Project Rationale and Design. Environmental Health Perspectives 122:535-544.

Weinberg CR. 2007. Can DAGs clarify effect modification? Epidemiology 18:569-572.

Wild CP. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiology Biomarkers and Prevention 14:1847-1850.