HORIZON2020 Programme
Contract No. 733032 HBM4EU

# HBM mixture database description and proposal statistical analysis plan

## Deliverable Report

## AD 15.3

## WP 15 - Mixtures, HBM and human health risk

## Deadline: December 2017

## Upload by Coordinator: 12 December 2017

| Entity | Name of person responsible | Short name of institution | Received [Date] |
|---|---|---|---|
| Coordinator | Marike Kolossa-Gehring | UBA | 04/12/2017 |
| Grant Signatory | Erik Lebret | RIVM | 25/11/2017 |
| Pillar Leader | Robert Barouki | INSERM | 25/11/2017 |
| Work Package Leader | Erik Lebret | RIVM | 25/11/2017 |
| Task leader | Erik Lebret | RIVM | 25/11/2017 |



| Responsible author | Jelle Vlaanderen | E-mail | j.j.vlaanderen@uu.nl |
|---|---|---|---|
| Short name of institution | IRAS | Phone | +31-30-2535947 |
| Co-authors | Ilse Ottenbros (IRAS, RIVM), Roel Vermeulen (IRAS), Eva Govarts (VITO), Denis Sarigiannis (Thessaloniki), Spyros Karakitsios (Thessaloniki), Chris Roth (ANSES), Amélie Crépet (ANSES), Nina Vogel (UBA), Marco Vinceti (UNIMORE), Tommaso Filippini (UNIMORE), Miranda Loh (IOM), Erik Lebret (RIVM), Nicole Probst-Hensch (Swiss TPH), Pavel Čupr (MU), Maria Silva (INSA), Greet Schoeters (VITO), Sylvie Remy (VITO), Anne Marie Vinggaard (TU Denmark) | | |

# Table of contents

# 1 Authors and Acknowledgements

**Lead authors**

Jelle Vlaanderen, Roel Vermeulen – IRAS

Ilse Ottenbros – RIVM/IRAS

Erik Lebret - RIVM

Denis Sarigiannis, Spyros Karakitsios - AUTH

Chris Roth, Amélie Crépet - ANSES

Miranda Loh - IOM

**Contributors**

Eva Govarts, Sylvie Remy, Greet Schoeters – VITO

Nina Vogel – UBA

Marco Vinceti, Tommaso Fillippini – UNIMORE

Nicole Probst-Hensch - Swiss TPH

Pavel Čupr – MU

Maria Silva – INSA

Anne Marie Vinggaard – TU Denmark

We gratefully acknowledge the input of the participants to the Work Package 15 (WP15) Workshop on November 7-8, 2017, Bilthoven, during which this deliverable was discussed. We also acknowledge the input of the WP15 partners on drafts of the manuscript.

# 2 Introduction and background

## 2.1 Definition of mixtures for the purpose of this deliverable

The phenomenon of mixtures (in the context of HBM) refers to the common occurrence at the level of the individual of chemical xenobiotic substances. There is no broadly accepted operational definition of mixtures. The European Commission communication on "*The combination effects of chemicals – Chemical mixtures*"[1] was published in response to a request from the European Parliament for the Commission to consider the extent to which the existing legislation *"adequately addresses risks from exposure to multiple chemicals from different sources and pathways, and on this basis considers appropriate modifications, guidelines and assessment methods".* In the communication mixtures are differentiated as follows:

- **Intentional mixtures:** these are manufactured formulated products that are marketed as such. The composition of such mixtures and the hazardous properties and classification of the constituents is (generally) known (e.g. pharmaceuticals, plant protection products);
- **Mixtures originating from a single source:** also known as 'unintentional mixtures'[2] these are the result of discharges to the environment during the production, transport, use or disposal of goods, often contain a mixture of chemical substances. The composition can either be known (for example an effluent) or it can be unknown (e.g. waste related); and
- **Mixtures of chemicals originating from multiple sources and through multiple pathways**: also known as "coincidental mixtures" these relate to multiple substances from multiple and varying sources. Their composition is unknown and can vary in both space and time (e.g. exposure to humans to multiple chemicals from food and drinking water).

Intentional, unintentional and coincidental mixtures can arise from combinations of ambient environments and indoor sources, food products or contamination, consumer products, cosmetics, occupational exposures, medication and medical implants and lifestyle (e.g. smoking, recreational drugs, tattoo ink). In principle, every single substance, once it enters the body, will exhibit its health effects in interaction with a person's genetic makeup and acquired characteristics, and in concert with all other (xenobiotic) substances from previous and simultaneous exposures. These mixtures thus form a challenge to (experimental and observational) science, to mechanistic and casual assessment of risks and to regulation of substances and general risk management policies[3,4].

In this document the term mixture is used to describe any combination of exposure or internal dose biomarkers that has been measured in one or more biological matrices of a person during a single time point. These biomarkers include the chemical substances themselves, and also their metabolites. The mixture might include compounds from multiple substance groups (e.g. perfluorinated compounds, phthalates, etc.), but also multiple compounds from the same substance group (e.g. pesticides, mycotoxins, etc.). For the HBM4EU project, in a first stage data on the first set of prioritised substances (that is phthalates/DINCH, bisphenols, per-/polyfluorinated compounds, flame retardants, cadmium and chromium, PAHs and air pollutants, aniline family) is gathered both from existing as well as ongoing studies in the HBM4EU repository (see section 3.1). For the statistical analysis of mixture data we are limited to existing data for these substances, the focus will be on all quantified substances in one individual at a given point of time.

## 2.2 Overall objectives and tasks of WP15

The **overall aim** of "WP15 Mixtures, HBM and human health risks" is to improve the precision and efficacy of HBM to inform science, policy and regulatory actions with respect to dealing with mixtures. More specifically, WP15 has four main objectives:

1. To develop summary indicators to describe the exposure and body burdens of mixtures with an emphasis on defining priority mixtures and drivers of mixture toxicity;
2. To re-evaluate existing HBM mixture data and the collection of new HBM mixture data, to identify real-life exposure patterns to mixtures;
3. To further develop and apply practical approaches to identify and assess the potential health risks and impacts of mixtures;
4. To inform policy makers, stakeholders and the public at large about mixture exposures and associated health risks.

The work in WP15 is broken down in three tasks:

**15.1** Re-analysis of existing data on mixtures from earlier HBM studies using a key set of summary indicators, which consists of:

- statistical analyses on existing data with focus on exposure profiles that take into account coincidental exposures.
- toxicity-based approaches, e.g. develop cumulative assessment groups (CAG) based on common working mechanisms
- hybrid approaches, using Hazard Index (HI) and/or Biomonitoring Equivalents (BE)

**15.2** Joint survey on HBM mixtures in 3-5 countries

**15.3** Identification of mixture health effects

This deliverable AD15.3 "HBM mixture database description and proposal statistical analysis plan" focuses on preparatory ground work associated with the work in task 15.1. Activities include the description of a functional database structure, together with WP10, development of a statistical analysis plan for existing data, the testing of this analysis plan on an available subset of data, while in parallel compiling existing HBM mixture data in the data repository developed in WP10.

## 2.3 Questions underlying the tasks in 15.1

The statistical analysis of existing HBM data is explorative in nature, with the primary question "What are the HBM mixture levels in the European population?". More specific (research) questions that are addressed in task 15.1 are:

1. How can we rank/order individuals on the basis of low-high body burdens to mixtures?
2. What patterns can we observe amongst body burdens of different substances within individuals? I.e. are people with high levels for some substances more likely to be high on others as well?
3. Are such patterns indicative for specific sources or pathways of exposures?
4. Can we identify hotspots or risk groups with high body burdens of mixtures?
5. Can we develop aggregate/hybrid indicators that encapsulate toxicity of the mixture in a meaningful way? E.g. hazard index approach or CAGs.
6. Can we define mixture levels of excess risk, based on toxicity based aggregation of HBM mixture data?

These questions are the starting point for the statistical analysis and are further addressed in more detail in section 4. While we are charting unexplored territory in HBM mixture data, for each analysis, we will explore one or more methods, either from the literature on HBM, from environmental epidemiology or from exposome[5] and OMICS studies[6].

# 3 HBM mixture database description

In WP7 an inventory on existing HBM data has been compiled in June 2017, including some 120+ studies. These included studies within the framework of national HBM programmes as well as targeted (research) studies. Many of these studies included multiple pollutants, although only 14 studies were identified as studies on mixtures. As of November 2017, the database is not yet complete and WP7 will make available a tool to further develop the inventory.

It is currently not yet clear which data can be made available at the individual level. Individual level data on mixtures for the same sampling time point are essential for the tasks in 15.1. From the inventory and the data stored in the repository (see section 3.1), task 15.1 will identify available HBM mixture data at the individual level for which clearance for use is or can be obtained. The repository will provide data sets that are harmonised and quality controlled, but since the repository does not provide the functionality for statistical analysis, selected datasets will be requested according to the protocols described in the HBM4EU data plan and data policy. The data, upon signing user agreement forms, can then be downloaded to the platform of IRAS, RIVM, and other partners specified in the data use proposal to perform the necessary statistical analyses.

## 3.1 HBM4EU repository

Mixture data to be used in WP15 will be identified from and stored in the HBM4EU repository, using the inventory of WP7 as point of departure. Also, we will approach other Horizon2020 mixture projects through the 'Mixture project cooperation' started in 2017, to explore the availability of individual HBM mixture data. In this exercise we will include data from Exposome projects.

Within WP10 („Data Management Plan" D10.1) a repository is developed to store the HBM data becoming available for the consortium. Upon in-house pseudo-anonymisation of the data, the data are to be transferred to the HBM4EU repository by data providers/owners. Data shall be transferred using a harmonised template and following the HBM4EU codebook. To obtain permission to use the data by consortium members, the lead data user has to submit a proposal, containing among others the purpose of data use, named lead data user and other data users, the required studies, variables, sampling time frame and a start and end date to perform the analyses. The WP leader verifies the fit of the purpose with HBM4EU objectives and checks whether the same analyses are not yet performed by other consortium partners. Depending on the conditions set out by the data provider in the HBM4EU data transfer form, the data provider is consulted to either approve or refuse the proposal. Upon approval of the proposal, the lead data user and each other data user shall complete and sign the HBM4EU data access and use agreement. Upon signed agreement, the lead data user and other data users obtain the necessary permissions to access the data via a dedicated section on the HBM4EU repository and to perform the data analyses for the purpose specified in the approved proposal.

Data analyses should adhere to good epidemiological practice and the presentation of papers should follow the STROBE guidelines. The data users can choose their software for data analysis while respecting the conditions outlined in the data access and use agreement. Signing the data access and use agreement, the data user agrees to

1) not abide the rules set out in the agreement,
2) accepts the terms and conditions of HBM4EU repository, and
3) accepts the HBM4EU Data Policy.

This includes, but is not limited to, protection of the study participants identity, safeguard to not share the data with unauthorised people (anyone not signed a data use agreement linked to the approved proposal), commitment to upload all derived variables, analysis pipelines, intermediate results, and/or results to the repository upon analysis to a section of the repository dedicated to this

agreement, commitment to not store the data on any place that is accessible by others, commitment to securely destroy all data related to the proposal stored outside the repository - including all possible back-ups as well as (intermediate) results from the analyses - at the latest at the end-date of the approved proposal, commitment to follow the procedures regarding publication and/or dissemination of results. All details are available in the HBM4EU data management plan (DMP) and data policy (https://www.hbm4eu.eu/data-management/).

## 3.2 Required data structure

WP15 has identified the required data structure as depicted below in Figure 1. This involves not only the individual HBM mixture data per se, but also associated meta-data about the study under which the data were collected, analytical details, etc. Also, to be able to perform toxicity-based approaches, auxiliary information needs to be added to enrich the original HBM mixture data. Information from the literature and from other work packages needs to be included, such as Hazard Quotients (to develop the Hazard Index), Mode of Action/Adverse Outcome Pathway (MoA/AOP). While developing the database, we will work according to the WP10 Data Management Plan and the ethical requirements with respect to data protection.

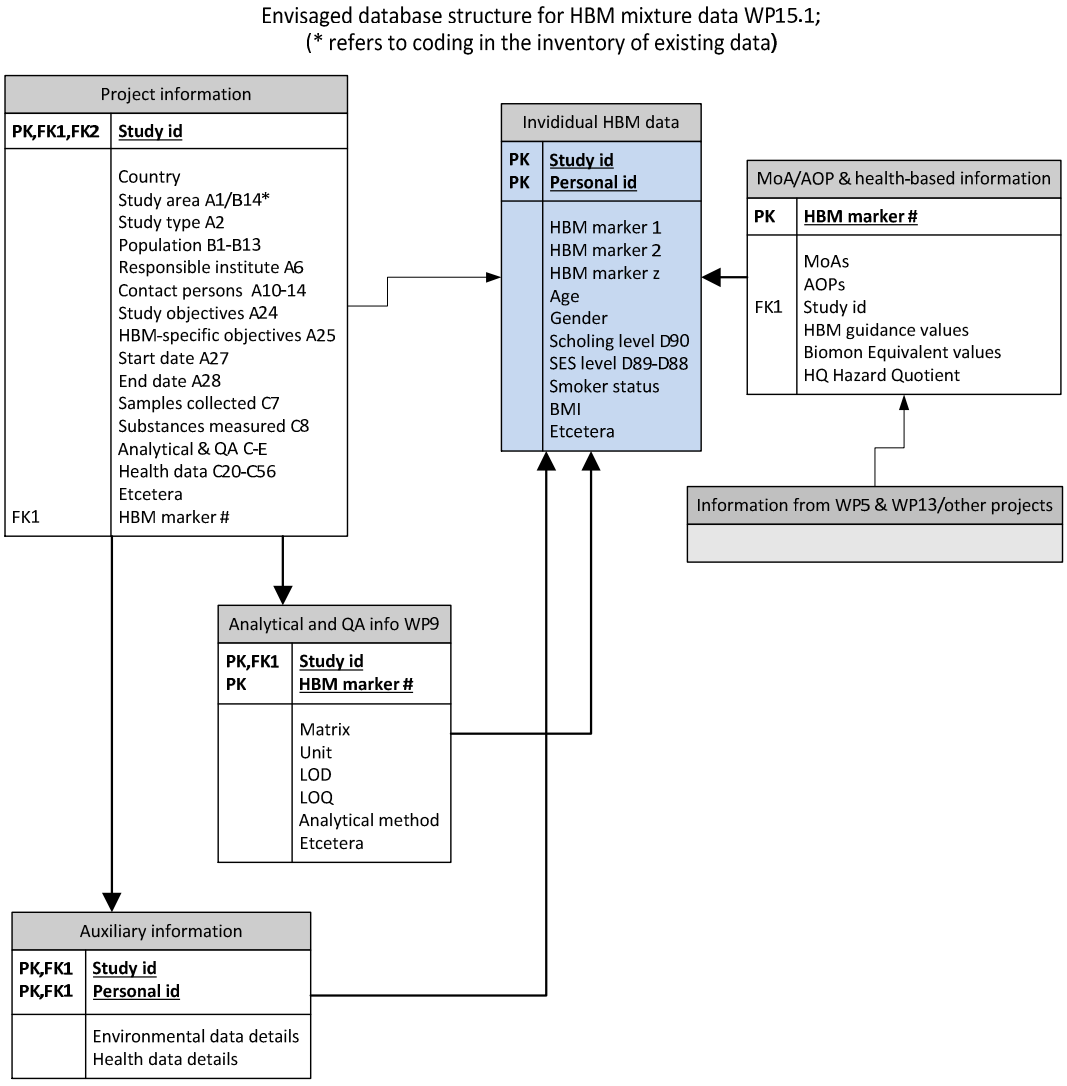


**Figure 1: The required database structure for HBM mixture data in WP15**

# 4 Proposal statistical analysis plan

This statistical analysis plan describes the suggested analysis strategy to address the main questions related to exposure to chemical mixtures, as raised in the HBM4EU scoping document. The questions pertain to the description of the distribution of (patterns in) biomarker mixtures (section 4.2), the identification of determinants that explain observed variation of patterns in biomarker mixtures (section 4.3), and the ability to rank European individuals based on their cumulative biomarker mixture burden (section 4.4). The proposed work in sections 4.2 and 4.3 is purely focused on (determinants of) exposure distributions and is closely related to the activities conducted for other HBM4EU priority chemicals conducted in work package 10. The work in section 4.4 incorporates a toxicity component and therefore links with task 15.3 within work package 15 and work package 5.

## 4.1 Descriptive analyses and data transformation

Descriptive analysis, assessment of data quality and data transformation will follow the general statistical analysis plan (D10.2) developed within WP10, task 10.3. It is likely that many biomarkers will follow a more or less log linear distribution or, more generally, right-skewed distributions. Thus non-parametric statistical procedures or transformations (like Z-scores, rank order, or log transformations) will be in order. Protocols for dealing with data below the limit of detection (LOD) or the limit of quantification (LOQ) in the specific studies also need to be taken into account, as well as varying LODs/LOQs between studies; following procedures laid out in D10.2. Alternative approaches, based on existing data, for rank ordering individuals on HBM mixture data will be performed and compared.

## 4.2 Description of correlation patterns within biomarker mixture data

After descriptive analyses and decisions about appropriate data transformations with the partners in task 15.1, so called Circos plots will be generated for each dataset in the HBM4EU repository for which mixture data (as described in section 2) are available. Circos plots are an efficient way to represent the dependence between multiple biomarkers in a dataset. An example[7] of a Circos plot (here also including non-biomarker exposures) is provided in Figure 2.

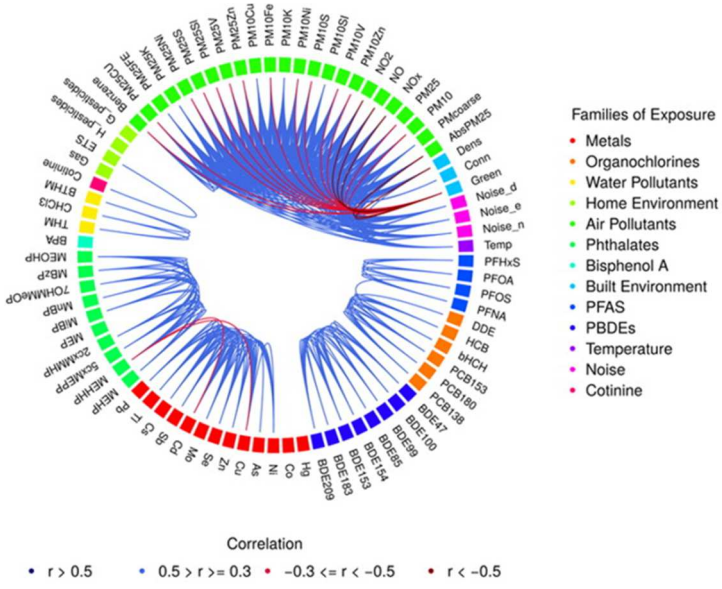


**Figure 2: Example of a Circos plot developed within Horizon 2020 project HELIX[7] providing an overview of dependencies between biomarkers within a biomonitoring dataset.**

The Circos plots will be used to:

1. Describe the mixtures present in datasets part of the HBM4EU repository.
2. Assess the overlap in mixture data across datasets available in the HBM4EU repository.
3. Assess visually differences in correlation patterns between datasets, studies, countries, etc.

### 4.2.1 Differential network analysis

To further quantitatively explore differences in correlation networks across stratifications by determinants such as by country, region, sampling period, socio-economic status, gender, BMI, and age (see for a full description of potential determinants section 4.3) differential network analysis[8,9] will be used. Differential network analysis is a form of weighted correlation network analysis where network nodes correspond to biomarkers and connection strengths are determined by the pairwise correlations between biomarkers.

Within differential network analysis biomarkers are grouped based on the correlation patterns between them. Outcomes from these analyses will indicate:

1. whether the overall structure of two networks are different;
2. whether the connectivity of a particular set of “interesting” biomarkers (e.g. matched by known toxicity effects) has changed between two networks;
3. whether the connectivity of a given single biomarker has changed between the two networks.

### 4.2.2 The sparse non-negative matrix under-approximation (SNMU) method

The approach based on the combination of sparse non-negative matrix under-approximation (SNMU) and clustering algorithms will be applied to biomonitoring data to identify and characterise individual clusters of the sampled population and their associated mixtures. The SNMU[10] is a modified version of the Non-negative matrix factorisation (NMF) method[11] which consists of factorising a non-negative matrix into the product of two low rank nonnegative matrices. For example, applied to a biomonitoring data matrix E (composed of P substances x N individuals), the SNMU method decomposes E as $E=WxH+\varepsilon$ where W and H are matrices and $\varepsilon$, a minimised residual matrix term. The matrices W and H provide very different and important information: H provides information about the exposure profiles (the dominating mixtures in individual exposures) while W provides information about mixture profiles (the dominating substances in mixtures). Each column of the matrix W corresponds to a specific mixture and each element w of a column corresponds to the contribution of the substance p to the mixture (see Figure 3), the highest values of W therefore correspond to the most important substances within the mixture. At the same time, each column of the matrix H corresponds to an individual and each element H of a column corresponds to the contribution of a certain mixture to an individual's overall exposure. To calculate W and H, the SNMU method is performed using an optimisation method with a non-negativity constraint which consists of minimising the residual term $\varepsilon$.

An important feature of the SNMU method also is that it takes into account the dependencies between biomarkers and of the biomonitoring levels. This method is therefore perfectly adapted to identify the main mixtures in biomonitoring databases and to classify them in decreasing order of importance in the population. Moreover, within each mixture the method quantifies the contribution of the different substances to the mixture.

The results of the SNMU method will be combined with those obtained with the Circos plot to provide an overall picture of the biomonitoring mixture data and exposure of the European population (Figure 3).

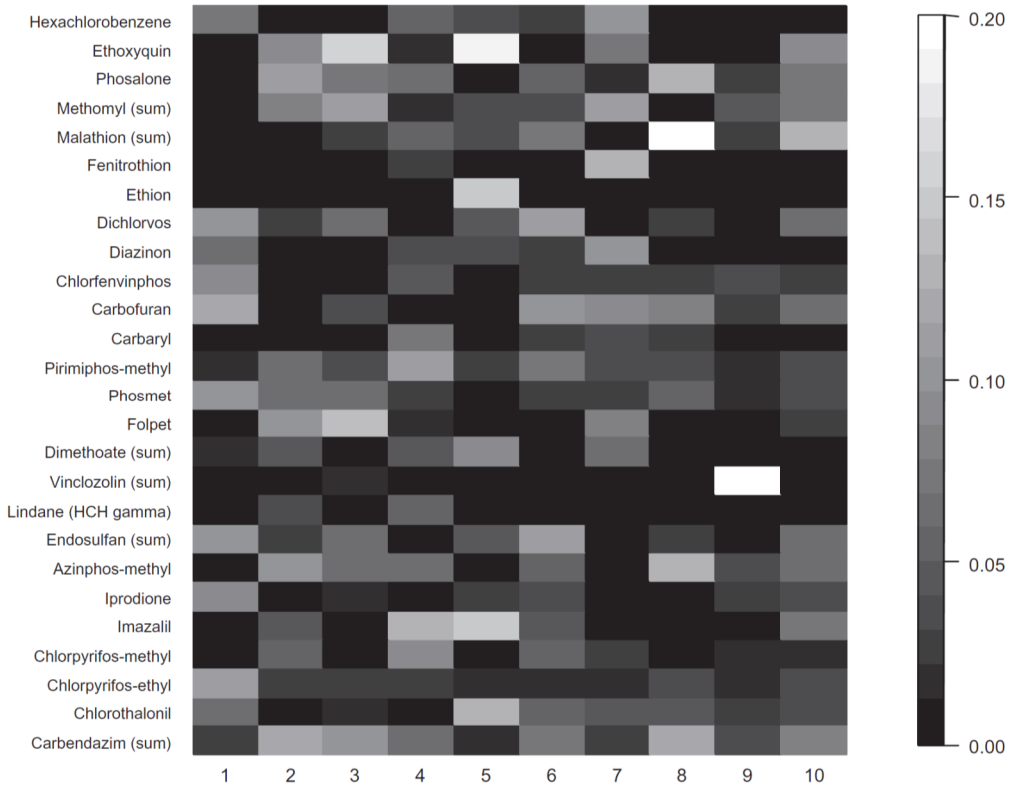


**Figure 3: Example of a W matrix obtained from exposure to 26 pesticides. 10 mixtures were defined, the different shades of grey corresponding to the contribution of the substances in each mixture.**

## 4.3 Identification of determinants of mixture composition

Determinants of mixture composition will be identified using a principal component regression framework.

### 4.3.1 Determinants considered

The minimal set of determinants that will be included in the analysis will include country, region (including rural and urban), sampling period, socio-economic status, gender, BMI, and age (or year born to look at cohort effects). Additional determinants will be included based on their availability (as indicated in the inventory in WP7 and priority group specific inventories that have been conducted in WP10, Task 10.3-10.4).

### 4.3.2 Principal component regression

Principal component regression (PCR) is an analysis technique that is based on principal component analysis (PCA), which is a method to convert the set of observations of possibly correlated biomarker measurements into a set of values of linearly uncorrelated variables called principal components. The rationale behind PCA is to represent as much of the variation observed in the original biomonitoring dataset with a (much) smaller set of principal components. For example, metabolites in the original dataset would be bundled in a statistically meaningful way to their mother substance. Principal components reflect correlation structures that exist in the data and can therefore sometimes be labelled based on an aspect that is shared between the variables contributing to it, such as chemical group. Using PCR, principal components are regressed on a set of determinants based on a standard linear regression model to estimate regression coefficients.

### 4.3.3 Cluster analysis of the SNMU approach results

The results of the SNMU method, in particular the H matrix (see section 4.2.2) will be analysed to study the determinants of the mixture compositions, hierarchical clustering being performed to classify individuals with similar exposure profiles, that is individuals exposed to the same type of mixture. A hierarchical agglomerative clustering will be applied on the matrix H containing the contributions of each mixture to the individual exposures. The clusters of individuals thus obtained will be described with regard to different selected variables (country, region, sampling time period, sex, BMI and age). Statistical analyses will be made to identify the specific determinants for the different clusters previously identified. The results of this treatment will be combined with those obtained through PCR to obtain the most coherent results on terms of determinants of mixture composition.

## 4.4 Table to calculate toxic potency of mixture data

One of the goals of WP15 is to assess whether hotspots or groups of individuals with high body burdens of mixtures can be identified. Several publications have provided theoretical and applied frameworks for such an exercise[12–15], though approaches were often restricted to a single chemical group.

As a first step towards the estimation of a cumulative biomarker mixture burden, a database will be created that contains toxic potency information for each relevant exposure. This database on toxic potency and risk indicators will be developed in dialogue with the task leader of 15.3 and through interaction with WP5, WP13 and WP14. Milestone for the toxic potency database is foreseen in January 2019. The main methodologies available today to estimate the human health risk of environmental chemical mixtures include (a not complete selection is presented here):

- Health-based human biomonitoring guidance values
- Toxic equivalence quotient (expressing congeners in terms of its most toxic form)
- Hazard quotient assessment (the ratio of the potential exposure to a substance and the level at which no adverse effects are expected) for the Hazard Index (HI) approach[12] (estimated contributions to total HI in one exposure pattern).

The database will include a reference table which curates available health-based human biomonitoring guidance values from different sources, including both workplace and general population values. For example, it will include workplace reference values such as the Biological Limit Values and Guideline Values (BLVs, BGVs) recommended by the Scientific Committee on Occupational Exposure Limits, Biological Exposure Indices (BEIs) from the ACGIH, the biological tolerance levels (BATS) set by the German MAK commission and any others available. Additionally non-workplace biological equivalence values (BEs), which are derived from reference values such as tolerable daily intakes (TDIs) or HBM values used by the German Biomonitoring Commission and those from other countries, will be included. These will serve as reference values from which a mixture indicator (such as a hazard index) may be calculated. The hazard index (HI) is the sum of the ratios of a population measured value to a biological guidance value (these ratios are also referred to as the hazard quotient, HQ):

$$HI = \sum_{i=1}^{n} \frac{C_i}{R_i}$$

Where $C_i$ = measured concentration of chemical $i$ and $R_i$ = biological reference value for chemical $i$. If the HI is greater than 1, that indicates that mixture is potentially of concern. The HI generally assumes a similar toxic action of the chemical mixture.

Some classes of chemicals, such as dioxins/dioxin-like compounds, and PAH's may be treated as mixtures where a set of congeners are weighted relative to a reference congener (toxic equivalence factor, TEF). This provides a single weighted concentration for the mixture, which assumes a common outcome or mechanism of action, and a certain relative toxicity for each compound in a class. TEFs have been developed for exposure concentrations (e.g. through ingestion), not for concentrations in biological matrices, although there have been efforts to derive biological toxic equivalency weighted serum concentrations of dioxins[16].

The proposed database will contain a table (illustrated in Figure 1 and below as Table 1) type of biomonitoring guidance value (e.g. BEI, BAT, BE, etc.), biomonitoring guidance value, country where biomonitoring guidance value was issued, and, when available, TEF and Hazard quotient will be included. This information will then be used to aggregate information at the level of the mixture (e.g. cumulative assessment groups) yielding toxic equivalence quotients and hazard indices.

**Table 1: Illustration of toxic potency database at the level of single compounds**

| Compound name | Compound group | Health-based human biomonitoring guidance value type | Bio monitoring guidance value | (Country of) origin of BGV | TEF (if available) | TEQ (if available) | Hazard quotient (if available) | Hazard Index (if available) |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Perfluorooctane sulfonate | Perfluorinated compounds | | | | | | | |
| Perfluorooctanoate | Perfluorinated compounds | | | | | | | |
| Diethyl Phthalate | Phtalates | | | | | | | |
| | | | | | | | | |
| Etc. | | | | | | | | |

### 4.4.1 Strategies to generate missing toxic potency data using internal dosimetry models

The expectation is that toxic potency information will not be complete for all compounds relevant for HBM4EU, with no information available for some of the compounds. In the case that there are no direct biomonitoring reference values (health based or not), we identified three possible routes through which data can be augmented.

The first approach involves using the Danish QSAR database[17], which is currently under development and available at http://qsar.food.dtu.dk . This database includes estimates from more than 200 (Q)SARs from free and commercial platforms and related to physicochemical properties, ecotoxicity, environmental fate, ADME and toxicity. (Q)SAR predictions for more than 600,000 chemical substances can be searched. The database is developed by the National Food Institute, Technical University of Denmark, with support from the Danish Environmental Protection Agency, the Nordic Council of Ministers and the European Chemicals Agency. Direct collaboration with this initiative is possible through the participation of Technical University of Denmark (Vinggaard) in this work package.

In addition, internal dosimetry modelling through Physiology Based Toxicokinetic (PBTK) models, offers two options: One option is to reconstruct exposure starting from the existing biomonitoring data[18,19], employing the so-called "reverse dosimetry". Reverse dosimetry can be performed to

estimate the external exposure that is consistent with the measured biomonitoring data through the backward application of PBTK models. In this case the PBTK model is geared with reverse modelling algorithms in order to reconstruct exposure from human biomonitoring (HBM) data. Assimilation of human biomonitoring data and their translation into intake distribution amounts to a computational inversion problem, where the objective is to identify the specific input distributions that best explain the observed outputs while minimising the residual error. 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 results in intakes; outputs are the observed biomonitored levels. 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). After estimating the distribution of the intake that corresponds to the measured biomonitoring data, the Hazard Index is applied[16]; The hazard index (HI) of a mixture of chemicals is the sum of the compound-specific hazard quotients (HQi), which are calculated as the ratio of the exposure (e.g. the daily intake of a substance) to the dose of no concern (i.e. the exposure above which adverse effects on human health can be expected).

A second option, as proposed by WP12, is to use the existing external exposure threshold for each of the compounds of interest and to run the respective PBTK models in the forward mode, so as to derive the predicted biomonitoring equivalents (BE) in a consistent manner for all the compounds of interest. This is a slightly different approach compared to the one adopted by Hays et al.[20], where BEs are usually derived from benchmark dose using simple TK considerations; the approach proposed is to start from the human relevant external reference exposure (e.g. RfD, TDI) and to derive a BE using a human PBTK model. This is of particular importance for rapidly metabolised substances/compounds, because we can capture the time course of the expected biomarker (the substance itself or the appearance of its metabolites in biological fluids) and to associate this with the expected time of sampling (e.g. what is the expected concentration in urine of BPA at 7:00 am, under a typical dietary exposure scenario that accounts for an intake of 4 μg/kg_bw/day). After estimating the BEs for all the compounds of the mixture, we may use a concept similar to hazard index, where the respective Hazard Quotients will be estimated by dividing the biomarker level with the calculated BE for each component of the mixture.

# 5 Work plan and first output of the statistical analysis

As mentioned earlier, the statistical analysis of existing HBM mixture data is novel. There is currently little or no insight in actual data availability at the individual level, in within-study temporal differences, nor in between-study differences, nor about the extent of missing data for individuals. Standard statistical approaches are not yet available. Moreover, it is expected that the data repository for individual HBM mixture data will emerge relatively slow, with limited availability of data in 2018. Therefore, the proposed approaches will be piloted on a limited set of data available to partners in task 15.1. Candidates are the three successive campaigns of the Flemish Environment and Health Study and the data collected within the HELIX project[21], other potential candidates are considered.

Statistical scripts will be developed for each of the in Section 4 described analysis. These scripts will be made available to the other partners for testing on their own data, where desired. Results of the pilot analysis on the smaller test data sets will be reported in D15.3 (November 2018), together with recommendations for further and broader application on the growing data set in the repository.

Parallel to the pilot test of statistical scripts on the test data sets, a database on toxic potency will be developed in dialogue with the task leader of 15.3 and with WP5, WP13 and WP14. Milestone for the toxic potency database is foreseen in January 2019.

# 6 References

1. 'European Commission'. Communication from the Commission to the Council: The combination effects of chemicals – Chemical mixtures, COM(2012) 252 final, Brussels, 31.5.2012. (2012). Available at: http://ec.europa.eu/environment/chemicals/effects/effects_en.htm. (Accessed: 25th November 2017)
2. Kienzler, A., Bopp, S. K., van der Linden, S., Berggren, E. & Worth, A. Regulatory assessment of chemical mixtures: Requirements, current approaches and future perspectives. *Regulatory Toxicology and Pharmacology* **80,** 321–334 (2016).
3. Barrera-Gómez, J. *et al.* A systematic comparison of statistical methods to detect interactions in exposome-health associations. *Environ. Heal. A Glob. Access Sci. Source* **16,** (2017).
4. Agier, L. *et al.* A Systematic Comparison of Linear Regression-Based Statistical Methods to Assess Exposome-Health Associations. *Environ. Health Perspect.* (2016). doi:10.1289/EHP172
5. Rappaport, S. M. Implications of the exposome for exposure science. *J. Expo. Sci. Environ. Epidemiol.* **21,** 5–9 (2011).
6. Vlaanderen, J. *et al.* Application of OMICS technologies in occupational and environmental health research; current status and projections. *Occup. Environ. Med.* **67,** 136–43 (2010).
7. Robinson, O. *et al.* The Pregnancy Exposome: Multiple Environmental Exposures in the INMA-Sabadell Birth Cohort. (2015).
8. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
9. Gill, R., Datta, S. & Datta, S. A statistical framework for differential network analysis from microarray data. *BMC Bioinformatics* **11,** 95 (2010).
10. Gillis, N. & Glineur, F. Using underapproximations for sparse nonnegative matrix factorization. *Pattern Recognit.* **43,** 1676–1687 (2010).
11. Lee, D. D. & Seung, H. S. Algorithms for Non-negative Matrix Factorization. 556–562 (2001).
12. Martin, O. V, Evans, R. M., Faust, M. & Kortenkamp, A. A Human Mixture Risk Assessment for Neurodevelopmental Toxicity Associated with Polybrominated Diphenyl Ethers Used as Flame Retardants.
13. Meek, M. E. B. *et al.* Risk assessment of combined exposure to multiple chemicals: A WHO/IPCS framework. *Regul. Toxicol. Pharmacol.* **60,** S1–S14 (2011).
14. Silins, I. & Högberg, J. Combined toxic exposures and human health: biomarkers of exposure and effect. *Int. J. Environ. Res. Public Health* **8,** 629–47 (2011).
15. Ryan, P. B., Burke, T. A., Cohen Hubal, E. A., Cura, J. J. & McKone, T. E. Using biomarkers to inform cumulative risk assessment. *Environ. Health Perspect.* **115,** 833–40 (2007).
16. Aylward, L. L., LaKind, J. S. & Hays, S. M. Derivation of Biomonitoring Equivalent (BE) Values for 2,3,7,8-Tetrachlorodibenzo- *p* -Dioxin (TCDD) and Related Compounds: A Screening Tool for Interpretation of Biomonitoring Data in a Risk Assessment Context. *J. Toxicol. Environ. Heal. Part A* **71,** 1499–1508 (2008).
17. Danish National Food Institute. *Danish (Q)SAR Database*. (2017).
18. Bányiová, K. *et al.* Long-term time trends in human intake of POPs in the Czech Republic indicate a need for continuous monitoring. *Environ. Int.* **108,** 1–10 (2017).
19. Sarigiannis, D. A. *et al.* Integrated exposure and risk characterization of bisphenol-A in Europe. *Food Chem. Toxicol.* **98,** 134–147 (2016).
20. Hays, S. M., Aylward, L. L. & Lakind, J. S. Introduction to the Biomonitoring Equivalents Pilot Project: development of guidelines for the derivation and communication of Biomonitoring Equivalents. *Regul. Toxicol. Pharmacol.* **51,** S1-2 (2008).
21. Vrijheid, M. *et al.* The human early-life exposome (HELIX): project rationale and design. *Environ. Health Perspect.* **122,** 535–44 (2014).