HBM4EU science and policy for a healthy future

HORIZON2020 Programme
Contract No. 733032 HBM4EU

Statistical Analysis Plan

Deliverable Report

D 10.2

WP 10 – Data management and analysis

Deadline: December 2017

Upload by Coordinator: 15 December 2017

Entity	Name of person responsible	Short name of institution	Received [Date]
Coordinator	Marike Kolossa-Gehring	UBA	05/12/2017
Grant Signatory	Argelia Castaño	ISCIII	05/12/2017
Pillar Leader	Argelia Castaño	ISCIII	05/12/2017
Work Package Leader	Greet Schoeters	VITO	05/12/2017
Task leader	Martine Vrijheid	ISGLOBAL	05/12/2017

Responsible author	Martine Vrijheid	E-mail	martine.vrijheid@isglobal.org
Short name of institution	ISGlobal	Phone	+34 932 14 73 46
Co-authors	Parisa Montazeri, parisa.montazeri@isglobal.org Loïc Rambaud, Loic.RAMBAUD@santepubliquefrance.fr Nina Vogel, nina.vogel@uba.de Jelle Vlaanderen, J.J.Vlaanderen@uu.nl Sylvie Remy, sylvie.remy@vito.be Eva Govarts, eva.govarts@vito.be Greet Schoeters, greet.schoeters@vito.be		

Table of Contents

1 Authors and Acknowledgements

Lead authors

Lead authors contributed substantially to the development of these statistical analysis plans. Below we list them individually by section of contribution, name, and partner institution.

Section	Name	Partner
General Statistical Analysis Plan, Task 10.3 leader	Martine Vrijheid Parisa Montazeri	ISGlobal
Section 10 – Calculation of EU Reference Values	Loïc Rambaud Nina Vogel	ANSP UBA
Section 11 – Uncertainty Analysis	Jelle Vlaanderen	IRAS
Revision and input of the plan	Sylvie Remy Eva Govarts Greet Schoeters	VITO

Contributors

Contributions were made by members of Deliverable 10.2. Below we list them individually by section of contribution, name, and short name of partner institution.

Section	Name	Partner
Section 5.6.1 - Urinary HBM analysis	Anna-Maria Andersson Hanne Frederiksen	RegionH
Section 7 - Time Trends, non-parametric methods	Jiří Kalina	MU
Section 12 - Substance Specific Analysis		
Phthalates/DINCH	Nina Vogel André Conrad Maria Rüther	UBA
	Anna-Maria Andersson Hanne Frederiksen	RegionH
	Ivars Vanadzins Inese Martinsone Zanna Martinsone	RSU
	Rosa Lange Marike Kolossa	UBA
Bisphenols	Loïc Rambaud	ANSP
	Anna-Maria Andersson Hanne Frederiksen	RegionH
	Tamar Berman	MOH-IL
	Robert Barouki	INSERM
PFASs	Maria Uhl	EAA
	Tina Kold Jensen Louise Dalsager	SDU

Section	Name	Partner
	Helena Garcia Marta Esteban López Susana Pedraza Argelia Castaño	ISCIII
	Hanns Moshammer Eva Schernhammer	MUW
	Lubica Palkovicova Murinova Denisa Richterova	SZU
	Anna Maria Ingelido	ISS
	Thorhallur I Halldorsson	UI
	Maria Uhl Ingrid Hauzenberger	EAA
Flame Retardants	Lisa Melymuk Jiří Kalina Jana Klánová Richard Hulek Jana Boruvkova	MU
	Cathrine Thomsen Enrique Cequier	NIPH
	Marika Berglund Karin Leander	KI
	Kamil Conka	SZU
	Jana Klánová Lisa Melymuk	MU
	Jana Hajslova Jana Pulkrabova	VSCHT
Cadmium(Cd) & Chromium(Cr)	Milena Horvat Janja Snoj Tratnik	JSI (Cd)
	Anna Pino Alessandro Alimonti Elena De Felip	ISS (Cr)
	Loreta Strumylaite Rima Kregzdyte	LSMU
	Beata Janasik Wojciech Wasowicz	NIOM
	Bojana Bazec Agnes Somen Ivan Erzen	NIJZ
	Carla Ancona Francesca Mattei Francesco Forastiere	DEP
	Rebeca Ramis	ISCIII
Occupational Exposure	Tiina Santonen Simo Porras Sanni Uuksulainen	FIOH
	Lode Godderis Radu Corneliu Duca Matteo Creta	KULeuven

Section	Name	Partner
	Anne Saber Marie Frederiksen	NRCWE
	Lisbeth E. Knudsen	UCPH
Air Pollution/PAHs	Denis Sarigiannis Spyros Karakitsios	AUTH
	Sylvie Remy	VITO
	Tamar Berman	MOH-IL
	Denis Sarigiannis	
Mixtures	Jelle Vlaanderen Roel Vermeulen	IRAS
	Eva Govarts	VITO
	Erik Lebret Mirjam Luijten	RIVM

2 Definitions

Confounder	A covariate that is related to both the outcome (dependent variable) and the exposure (independent variable) but is not a mediating factor.
Covariate	A participant variable (e.g. age, sex, ethnicity, etc…) that may or may not be related to the outcome being studied. If the covariate is related to both the exposure and the outcome, then it becomes a confounder or an intermediate.
Directed Acyclic Graph (DAG)	A graphical display of causal relationships among variables in which each variable is assigned a fixed location on the graph (called a node), and in which each direct causal effect of one variable on another is represented by an arrow. All drawn arrows indicate a causal relationship, the absence of a drawn arrow indicates no or unknown relationship. http://www.dagitty.net/ can help to draw DAGs easily.
Human Biomonitoring (HBM)	Assessment of people's exposure to toxic substances in the environment by measuring the substances or their metabolites in human specimens, such as blood, plasma, breast milk or urine.
Multicollinearity	Phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy.
Normality	The assumption of normality is the supposition that the underlying random variable of interest is distributed normally, or approximately so.
Sphericity	The condition where the variances of the differences between all combinations of related groups are equal. Violation of sphericity is when the variances of the differences between all combinations of related groups are not equal.
Stepwise Selection	A method in building multiple regression models that allows dropping or adding variables at the various steps, (i.e. moves in either direction, forward or backward).

3 Introduction and Aims

The aim of the following guide is to set forth a statistical analysis plan (SAP) to answer the exposure related research questions defined in the scoping documents for each of the prioritized chemicals. The general part of the SAP includes statistical plans for the evaluation of time trends, geographic comparisons, evaluation of exposure determinants, a strategy for the calculation of EU reference values, and a plan for conducting uncertainty analysis.

Statistical plans for each specific substance will include the definition and harmonisation of the variables (codebook), the statistical test to be applied, specific exclusion/partitioning criteria for calculating reference values, uncertainty analysis, data descriptions, and visualisations.

4 Statistical issues common to all HBM analysis

4.1 Exclusions

We recommend that HBM biomarkers with more than a certain percentage of missing values from a given data set (not below LOD but no measurement available) be excluded from the analyses (e.g. 70% missing). The exact definition of this cut-off depends on the datasets available and the definition of the study population and should be further discussed and the ultimate decision made by substance specific groups on a case-by-case basis.

4.2 Variable Consistency

To increase efficiency, the desired variables should be defined, preferably with common nomenclature to facilitate data clean-up and avoid confusion. This is to be discussed in substance specific groups before beginning an analysis. This includes covariates for the analysis, as well as, HBM biomarkers variables.

For the HBM biomarker variables and some main accompanying variables (e.g. age, sex…) a data template and harmonised codebook (available at https://www.hbm4eu.eu/data-management/) has been developed by WP10 for the data collections to transfer their data to the HBM4EU repository. For other accompanying variables needed for the specific research question(s), the substance groups should in conjunction with the identified data collections make a proposal for harmonising the information across the different data collections.

4.3 Transformation of HBM variables

Skewed biomarker variables will be transformed to achieve normality for proper interpretation of the output statistic, or categorised (e.g. by percentiles) if no transformation works. The distribution of all transformed variables will be examined to make sure that transformations do not lead to extreme/influential observations. After transformation, check variable normality through use of: histogram, P-P plot (probability–probability plot or percent–percent plot), Q-Q plot (quantile-quantile plot). P-P plots compare the cumulative distribution and are thus more sensitive around the mean, farther deviations from the middle indicate no normality. Q-Q plots compare the quantiles of a data distribution and are more sensitive at the tails; farther deviations at the tails indicate no normality.

Normality is not a requirement of regression models. However, normality of residuals is, and transformations to achieve a more symmetric distribution typically help in providing a more homogenous range of values, and in reducing the number of influential observations.

4.4 HBM concentrations below the LOD/LOQ

Two laboratory quality control limits, limit of detection (LOD) and limit of quantification (LOQ), are commonly utilised to evaluate biomarkers, additionally limit of blank (LOB) is sometimes given. LOD and LOQ parameters are related but have distinct definitions. The intent of both parameters is to define the smallest concentration of an analyte that can be detected with no guarantee about the bias or imprecision of the result by an assay, the concentration at which quantification as defined by bias and precision goals is feasible, and finally the concentration at which the analyte can be quantified.

We will define these terms according to the definitions given by Armbruster(2008).

- LOB can be defined as the highest apparent analyte concentration expected to be found when replicates of a blank sample containing no analyte are tested.
- LOD can be defined as the lowest analyte concentration likely to be reliably distinguished from the LOB and at which detection is feasible.
- LOQ is the lowest concentration at which the analyte can be reliably detected at which some predefined goals for bias and imprecision are met (usually determined by the lab conducting testing, LOQ ≥ LOD). The LOQ may be equivalent to the LOD or it could be a higher concentration.

The way LOD and/or LOQ are determined depends on the laboratory, and different approaches are used for this. As such, it is very important when using biomarker data from different data collections to compare the LODs and/or LOQs applied and, if this information is available, how they were determined.

4.4.1 Treatment of concentrations under the LOD/LOQ

Depending on the laboratory, LOD or LOQ may be given and used as the cut-off for reporting values. Some laboratories may not report values measured below the LOD or LOQ, rather just denote that the value was under the pre-defined limit. This can cause incomplete data. Ways to deal with this include; 1) complete case analysis (CCA), where observations with values below the LOD/LOQ are simply eliminated. This introduces bias by eliminating low values and is not recommended, 2) replacement by fixed value, where every value below the LOD/LOQ is replaced by a constant such as $LOQ/2$ or $LOQ/\sqrt{2}$ (Richardson, 2003; Schisterman, 2006), 3) single imputation, provides the dataset with a specific number (e.g. between 0 and LOD) in place of the missing data by analysing the other responses and looking for the most likely value that corresponds to that individual and then selects one of those possible responses at random and places it in the dataset, or 4) multiple imputation, similar to single imputation but more complex as it imputes more than one data set, setting the imputed values to fall between the interval (0 to LOD), to try to come up with a variance/confidence interval that one can use to better understand the differences between imputed datasets (Den Hond, 2015).

When choosing between single or multiple imputation keep in mind that multiple imputation is much more complex. In cases where very little data is missing, single imputation may be the simpler option and solve the issue without many serious errors. In more complex cases, multiple imputation introduces variability into the data (within the set parameters) and finds a range of possible responses which seeks to reduce error, thus being the preferred option.

Datasets will need to be reviewed on a case-by-case basis, taking into consideration the percentage of missing data, research question, and analysis to be conducted, within substance specific groups to decide the most appropriate method.

For the HBM4EU biomarker analyses, it is recommended to utilise the following treatment of values below LOD/LOQ values:

We recommend that for HBM biomarkers for which more than 80% of observations are below the LOD/LOQ be dichotomised into detected/undetected (Harel, 2014). For the other HBM biomarkers, we recommend values below the LOD/ LOQ are singly imputed using a quantile regression approach for the imputation of left-censored missing data. Note that this kind of imputation refers to substituting a value below the LOD/LOQ by a number between 0 and the LOD/LOQ, usually a very small range. This method of imputation has been shown to produce better results than simply replacing the values by half the LOD/LOQ (replacement by fixed value; Bernhardt, 2015). Skewed

variables should already be transformed so that the distribution can be approximated by a truncated normal distribution. Imputation for these values should be done at each study site, using the same syntax for continuity.

In the case that the laboratory does not give the LOD nor LOQ used for analysis, they should be contacted for the information. If they still cannot provide this information, substance specific groups will need to determine whether this data is of quality and should or should not be included in the analysis.

4.5 Treatment of missing data

To prevent loosing information and introducing potential selection biases, missing values of HBM biomarkers and covariates should be imputed. In cases where only very little HBM biomarker values are missing, single imputation is recommended. However, in more complex cases where both covariates and HBM biomarkers are missing multiple imputation is recommended. This is a commonly used and accepted method to deal with missing data that provides valid inferences under the missing at random (MAR) assumption, which assumes that missing data are associated to observed variables and not to unobserved information (Stuart, 2009). In multiple imputation, missing values are imputed stochastically several times (e.g. M times, leading to M completed datasets). Imputing missing values several times allows the quantification of the uncertainty in results associated with imputation, and to account for this uncertainty in the final standard errors, confidence intervals and p-values. The analysis of M multiply imputed datasets and the combination of results to provide a single final estimate is implemented in most statistical software for regression models. Multiple variables needing to be imputed are defined within the command and systematically imputed by the program taking into consideration the whole dataset.

4.5.1 Multiple imputation of missing data

A different set of imputed datasets will be created for each substance group, based on a common protocol. This protocol should be established and agreed upon by the leaders of the papers in each substance specific group, as they will also be in charge of generating the imputed datasets.

It is recommended that missing values of HBM biomarkers and adjustment variables be imputed using the method of chained equations (mice; White, 2011), using the *mice* package (available in R) or the *ice* package available in Stata (Van Buuren, 2011). It is recommended that the imputation models include all variables that are expected to be associated with the HBM biomarker, including all variables that will be included in the final model, and all variables that are expected to be associated with the probability of having missing values (Stuart, 2009; White, 2011).

Given the potentially large number of variables that may be involved in the analysis, this may cause the imputation models to have too many variables. Thus, it is recommended that imputation models have no more than 15 to 25 variables (Van Buuren, 2011). Depending on the statistical software being used functions are available to do this (e.g. *quickpred* in R).

Predictive mean matching should be used to impute continuous covariates (this is the default option in the *mice/ice* function). This will preclude imputations being outside of the observed range and will preserve skewed distributions. However, ideally continuous variables would be transformed to a symmetric distribution before imputation, and be transformed back to the original scale after imputation, if needed. Logistic or multinomial regression will be used for imputing binary and categorical exposures, respectively. M=20 imputed datasets is recommended to be created for each analysis (White, 2011). With your MI data set, you can now run analysis per normal by using a prefix to identify the dataset as MI (e.g. *mi: command* in Stata, or with (*imp, lm(command))* in R). The software will take into consideration the M=20 datasets when fitting the models.

Complete guides for using *mice/ice* can be found;

For Stata (*ice*) - https://www.jstatsoft.org/article/view/v045i04/v45i04.pdf
For R (*mice*) - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.169.5745&rep=rep1&type=pdf

4.5.1.1 Sensitivity analysis after imputation

Complete-case analyses should also be performed and the results compared to the imputation results as a sensitivity analysis (not relevant in cases where large amount of data are missing, e.g. 60% or more). Comparisons of participants with complete vs. incomplete information should also be carried out to highlight and understand potential differences in results.

4.6 Adjustment for HBM analysis

4.6.1 Urinary HBM analysis: adjustment for dilution level for exposure biomarkers

In studies where substances are measured in urine, the measured concentrations are influenced by the urinary dilution level. As such, this must be taken into account when performing the statistical analyses. Creatinine, osmolality, and specific gravity are all surrogate estimators.

For analysis, we are restricted to the dilution measurements available in the different studies. As such, in the substance specific parts of the statistical analysis plan this should be taken into account by each of the substance groups, who then decide based on the available data which adjustment measurement is best, and available.

Where possible, it is recommended that the investigation and comparability of different dilution methods should be added as a substance-specific research question to further enhance comprehension on the matter.

Creatinine

The u-crea (urinary creatinine) concentration is not a direct measure of urinary concentration/dilution, but as creatinine excretion is considered to be of relatively low variability and constant over time with only slight influences by exogenous factors (e.g. diet). U-crea is commonly used in medicine and biomonitoring to adjust for urinary concentration/dilution. However, several physiological parameters influence creatinine excretion (e.g. renal function and muscle mass). Variation related to sex and age exists and is likely to be related to both sex and developmental differences in muscle mass. Additionally, ethnicity may influence creatinine excretion (see also Barr, 2005, for determinants of creatinine excretion). The unit of u-creatinine is often given as g/L or mmol/L.

Osmolality

Different osmometric method exist (freezing point depression, vapor pressure osmometry, membrane osmometry) but freezing point depression osmometry is the most widely used and has several advantages over the other methods. The theory behind freezing point osmometry to determine the osmolality of an aqueous-based solution is that when a solute (particles) is dissolved in a solvent (water), the freezing point of that solution is lower than that of the solvent alone. As more solute is added, the freezing point decreases further. By precisely measuring the freezing point of the solution, the osmolality, or concentration, can be determined. Osmolality adjustment has been shown to be a more robust adjustment than some of the other methods (Middleton, 2016, Yeh, 2015). Osmolality is the concentration of a solution expressed in osmoles of solute particles per kilogram of soluent and is often presented in the unit Osm/kg or mOsm/kg.

Specific gravity

Specific gravity (SG) is the ratio of the density of a substance (urine) to the density of a reference substance (distilled water). Being a ratio of densities, SG is a dimensionless quantity. As density (or in other words the mass at a given volume) varies with temperature and pressure reference and sample must be compared at the same temperature and pressure or be corrected to a standard reference temperature and pressure.

4.6.1.1 Methods to adjust or normalise for urinary concentration

While most investigators agree that adjustment is beneficial, there is controversy over the best approach.

The ratio model

Traditionally, adjustment has often been done by dividing urinary biomarker concentrations with the parameter used to express the urinary concentration in a simple ratio model (e.g. ng substanceX /g u-crea or ng substanceX/osmolality). However, the ratio model assumes that the parameter used for adjustment only vary due to the urinary concentration/dilution. This is not true for u-crea nor SG and although osmolality seems to be less affected than u-crea by factors other than the urinary concentration, u-osmolality has also been found to be associated with the daily protein intake (Yeh, 2015). Therefore the ratio model may result in some samples being under- or over-adjusted. This approach is particularly problematic in statistical analysis of associations between ratio adjusted biomarker concentration and health outcomes if the parameter used for adjusting also is related to the health outcome (e.g. SG increased in diabetes – not due to urine concentration but due to the presence of sugar molecules).

Normalisation to a standard urinary concentration

U-osmolality and SG have been used to adjust for the urinary concentration by normalising samples to a standard or a population average urinary concentration. The measured biomarker concentration is corrected by multiplying it with the ratio of a standard urinary concentration and the measured urinary concentration.

For normalisation of concentrations by SG the following equation is used (Sauvé, 2015):

$$C_{corr} = \frac{\mathrm{C}_i\ (\mathrm{SG}_{ref} - 1)}{(\mathrm{SG}_{meas} - 1)}$$

where C_{corr} is the corrected concentration, C_i is the measured biomarker concentration, SG_{meas} is the measured specific gravity, and SG_{ref} is the reference SG value (usually between 1.016 – 1.024).

For normalisation of concentration by u-osmolality the following equation is used (Middleton, 2016):

$$C_{corr} = \frac{\mathrm{C}_i\ \times \mathrm{Osm}_{ref}}{Osm_{meas}}$$

where C_{corr} is the corrected concentration, C_i is the measured biomarker concentration, Osm_{meas} is the measured osmolality, and Osm_{ref} is the reference osmolality value (usually between 0.700 – 0.800 Osm/kg).

Both of these two methods aim at normalising a measured biomarker concentration to a standardised (usually population average) urinary concentration and can be useful for a meaningful presentation of descriptive statistics of measured biomarkers and for comparing these between studies.

The covariate model

As the ratio model is not truly valid when the parameter used for adjustment differs significantly among different demographic groups an alternative method has been suggested for adjustment for creatinine in statistical analyses. Barr et al. suggested that in multiple regression analysis of population groups the biomarker concentration should be included unadjusted in the regression analysis with u-crea added as a separate independent variable. This works both in models were the urinary biomarker concentration is the dependent variable and in models were the urinary biomarker concentration is used as an independent variable to predict e.g. a health outcome (Barr, 2005). Similar approaches could be used for u-osmolality and SG.

O'Brien et al. present two covariate-adjusted standardisation approaches for which they showed that their Methods 3 and 7 perform well for chemicals measured in urine.

- Method 3: The goal of this method is to control the covariate-independent, short-term multiplicative effect of hydration on urinary diluteness. The first step is to fit a model for ln(creatinine) as a function of the covariates hypothesised to have a direct effect on it. Next, standardise by calculating,

$$C_{ratio} = \frac{\mathrm{EP}}{(\frac{\mathrm{Cr}}{Cr})}$$

  where, Cr and Cr denote the observed and fitted creatinine, respectively and EP is the urinary biomarker concentration. Finally, standardise C_{ratio} and fit:

$$logit(\Pr[\mathrm{D}]) = \alpha + \beta \times \mathrm{C_{ratioz}} + \delta \times \mathrm{W}$$

- Method 7: The goal of this method is to control for variation due to hydration and reduce confounding by blocking back-door paths between creatinine and risk factors related to both creatinine and the disease. This method is an extension of Method 3, with the inclusion of creatinine as a covariate:

$$logit(\Pr[\mathrm{D}]) = \alpha + \beta \times \mathrm{C_{ratioz}} + \lambda \times \mathrm{creatinine} + \delta \times \mathrm{W}$$

4.6.1.2 Specific issues related to adjustment for urinary concentration/dilution in statistical analyses

Conversion factors between differently adjusted concentrations

We are restricted to the dilution measurements available in the different studies and if different studies used different parameters of urinary concentration it begs the question whether it may be possible to adjust for urinary concentration by converting one measure of urinary concentration to another. The relationship between u-crea and u-osmolality (Anyabolu, 2017, Appendix I), between u-crea and SG (Carrieri, 2001), and between SG and u-osmolality (Voinescu, 2002) have been studied. Not surprisingly u-osmolality, SG, and u-crea are all mutually highly correlated as they all are influenced by the urinary concentration. However, depending on the study, differences in the relationships are observed – presumably related to the compositions of the study population or the study design (timing of urine sampling, spot urine vs. 24 hour etc). As such, it is not feasible to define a standard conversion factor between different parameters for urinary concentrations that can be used between studies unless the studies are very similar in design and study population composition.

4.6.2 Blood HBM analysis: adjustment for blood lipids for fat soluble exposure biomarkers

Similarly, when measuring substances in blood, a correction for the blood lipid content should be considered in case of lipid-soluble contaminants, as individuals with higher lipid concentrations tend to carry proportionally higher concentrations of lipid-soluble contaminants. O'Brien (2016) puts forth two adjustment approaches, Methods 2 and 6, that offer solutions to lipid adjustment issues.

Unlike the case of creatinine, more traditional standardisation techniques involving standardising the biomarker measure by dividing it by the measured serum lipid level are more appropriate.

- Method 2: Compute the ratio of the measured concentration to serum lipid level and then estimate the effect per SD:

$$logit(\Pr[D]) = \alpha + \beta \times \text{ratioz} + \delta \times W$$

- Method 6: This method is an extension of Method 2, with the inclusion of serum lipid level as a covariate with the goal of reducing confounding between lipids and risk factors related to both lipid levels and the disease in question.

$$logit(\Pr[D]) = \alpha + \beta \times \text{ratioz} + \lambda \times \text{lipid} + \delta \times W$$

4.7 Individual and aggregated data definitions

Data definitions presented here are derived from HBM4EU Work Package 10 - Data Management and Analysis, D 10.1 – Data Policy.

4.7.1 Individual Data

- Anonymised single measurement data: Re-identification is completely impossible. All possible de-identification keys have been destroyed; de-identification is not possible by combining variables or by matching with any other data.
- Pseudonymised single measurement data: The dataset does not contain directly identifiable variables. However there is a risk of re-identification: e.g. in combination with an identification key, by combining variables in the dataset, or by combining the dataset with any other data.

In case of filtered or generalised single measurement data, data should be specified as:

- Filtered: By removing the attributes that directly or indirectly violate the privacy, such as specific address information, precise spatial coordinates, the identity of the target population, etc.
- Generalised: By the replacement of the specific location of the samplings with coordinates representing a symbolic place, such as the centroid of the town centre, or by removing the number of digits indicating longitude and latitude coordinates

4.7.2 Aggregated Data

- Spatially aggregated: The summary statistics represent aggregation of measurements at Country level, NUTS 1, 2, 3 Level, City Level, etc.
    - Temporally aggregated: The summary statistics represent measurements of a sampling aggregated by months, years, etc.
    - Spatially/temporally aggregated: The combination of spatially aggregated and temporally aggregated.
    - Semantically aggregated: The summary statistics refer to groups of class of targeted population (humans/biota).

The statistical analysis methods to follow depend to some extent on the data available and what kind of data is available (individual or aggregate). Some common cases of data availability and their appropriate method for analysis have been outlined below by section. The 2018 update of the SAP will include further updates where appropriate to adapt the plans to the actual data available for each substance.

4.8 Statistical Issues References

1. Armbruster DA, Pry T. Limit of blank, limit of detection and limit of quantification. Clinical Biochemist Reviews. 2008;29(Suppl 1): S49-52.

2. Barber TE, Wallis G. Correction of urinary mercury concentration by specific gravity, osmolality, and creatinine. Journal of Occupational Medicine. 1986;28(5): 354-9.

Barr DB et al. Urinary creatinine concentrations in the U.S. population: implications for urinary biological monitoring measures. Environ. Health Perspect. 2005 113:192-200.

3. Bernhardt PW, Wang HJ, Zhang D. Statistical methods for generalized models with covariates subject to detection limits. Statistics in Biosciences. 2015;7(1): 68-89.

4. Den Hond E, Govarts E, Willems H, Smolders R, Casteleyn L, Kolossa-Gehring, et al. First Steps toward Harmonized Human Biomonitoring in Europe: Demonstration Project to Perform Human Biomonitoring on a European Scale. 2015;123(3): 255-63.

5. Harel O, Perkins N, Schisterman EF. The use of multiple imputation for data subject to limits of detection. Sri Lankan Journal of Applied Statistics. 2014;5(4): 227-46.

6. Middleton DRS, Watts MJ, Lark RM, Milne CJ, Polya DA. Assessing urinary flow rate, creatinine, osmolality and other hydration adjustment methods for urinary biomonitoring using NHANES arsenic, iodine, lead and cadmium data. Environmental Health. 2016;15: 68.

7. O´Brien KM, Upson K, Cook NR, Weinberg CR. Environmental chemicals in urine and blood: Improving methods for creatinine and lipid adjustment. Environmental Health Perspective. 2016;124(2): 220-7.

8. Richardon DB, Ciampi A. Effects of exposure measurement error when an exposure variable is constrained by a lower limit. American Journal of Epidemiology. 2003;157(4): 355-63.

9. Sauvé J-F et al. Creatinine and specific gravity normalization in biological monitoring of occupational exposures. J. Occupational Environ. Hygiene 2015 12:123-129.

10. Schisterman EF, Vexler A, Whitcomb BW, Aiyi Liu. The limitations due to exposure detection limits for regression models. American Journal of Epidemiology. 2006; 163(4): 374-83.

11. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. Journal of statistical software. 2011;45(3).

12. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine. 2011;30(4): 377-399.

13. Yeh H-C et al. Urine osmolality in the US population: Implications for environmental biomonitoring. Environmental Res. 2015 136:482-490.

5 Overview of Statistical Analysis

Overview of SAP		
Recommended statistics for all analysis: indication of data normality, count (n), mean, standard deviation, geometric mean (95%CI), median or p50, and specific percentile points (p25, p75, p90, p95, p99).		
Case	**Aggregate or Individual Data**	**Recommended Analysis**
Time Trends		
Case 1: Evaluating time trends in a country/region given a number of time points (5+).	Aggregate	*Parametric* – t-test, linear least square regression *Non-parametric* - Mann-Kendall test and Theil-Sen regression
	Individual	*Parametric* - linear least square regression *Non-parametric* – convert to aggregate statistic, follow aggregate plan
Case 2: Assessing a temporal trend with a limited number of time points (2-4). E.g. before and after a regulation	Aggregate	*Parametric* – t-test *Non-parametric* - Mann-Whitney U test
	Individual	*Parametric* – ANOVA *Non-parametric* – Mann-Whitney U test, Kruskall Wallis test
Case 3: Evaluating time trends by other variables (e.g. SES, lifestyle, nutrition…) and between countries.	Aggregate (Individual data recommended)	*Parametric* - measures analysis of variance (MANOVA) *Non-parametric* – Friedman test
	Individual	*Parametric* - measures analysis of variance (MANOVA), mixed models, GEE *Non-parametric* – Friedman test
Geographical Comparison		
Case 1: Comparison between multiple countries or multiple sites within one country.	Aggregate	*Parametric* – t-test, ANOVA *Non-parametric* – Mann-Whitney U test, Kruskall Wallis test
	Individual	*Parametric* – t-test, ANOVA *Non-parametric* – Mann-Whitney U test, Kruskall Wallis test *Both* – Multi-level models
Exposure Determinants		
Case 1: Examine one or more determinants that may influence biomarker concentrations of defined substances.	Individual (aggregate data not recommended)	Multiple regression models
Case 2: Comparison of clusters (or groups) of countries.	Individual (aggregate data not recommended)	Cluster analysis

6 Time Trends

Time trends are an important result from longitudinal and repeated cross-sectional studies that can provide a dynamic look into a population´s exposure status. Statistical time-trend analysis reveals the direction and significance of temporal changes and allows researchers to draw conclusions about the rate and size of an exposure(s) on different populations. This can be used to test and generate hypotheses for future research.

Time trend analysis, like any statistical inference, needs a minimum count of observations for the parameter of interest (in this case; time). At least five time points of exposure data within specific area are needed to identify some kind of trend. This minimum time sample can become limiting, so that time trend analysis may have to be replaced by a standard group of comparisons (e.g. ante-post regulation; Becker, 2013, Göen, 2011, Hoffman, 2017, LaKind, 2015).

In addition, researchers should keep in mind the timing of sample collection when analysing a time trend. A consistent and well established trend typically includes data taken at evenly spaced intervals over a period of time. If samples are taken for several years in a row (e.g. annual from 2006 to 2010) and then no sample taken again until 2017 (i.e. 7 years after the most recent sample), it may be best to consider the initial measurements as a trend, and instead compare the measurement in 2017 to the previous.

6.1 Time Trends Analysis Plan, by case

6.1.1 Case 1: Evaluating time trends in a country/region given a number of time points (5+).

- Are there different time trends for unregulated and regulated *substance(s)*?
- What does the time trend look like for *substance 1, 2 and 3...*and are there differences between them?
- Do time trends of *substance(s)* differ by country or European region?

For these cases both aggregate and individual data can be used.

Analysis Plan using Aggregate Data

Within task 10.1 on Data Management, a script is developed to calculate the aggregated statistics from study individual data. Obligatory statistics for time trend analysis are the geometric mean and median (p50), the median is usually preferred at it is more robust for both normally and log-normally distributed data. Aggregated data should include the geometric mean with 95%CI, and specific percentile points (p25, p50, p75, p90, p95, p99), arithmetic mean, standard deviation, count (n), indication of data normality or not. Additionally, data should be comparable (i.e. if one study has data from men aged 50-75 and another on children aged 6-9, this is not comparable). When combing the data, clear and specific research questions should be defined including details about the population to be compared (e.g. age, sex, working status, socio-demographic, etc…).

Giving a detailed description of the data when using aggregate data is of great importance to clarify comparability of the data, and describe the basic statistics (e.g. mean levels, do any exceed already existing guidelines, etc…).

Normal Data (parametric approach)

The use of the geometric mean is for log-normal distributed data (from transformation). Geometric means can be compared to one another by use of t-test statistic, as well as to the reference value(s) defined in Section 10, Calculation of EU Reference Values. To use the t-test statistic you

will also need the standard deviation and n of the sample size. In addition, a description of the data distribution is recommended.

Classic parametric approaches such as linear least square regression are suggested here as they serve for both the identification (assessment of a statistical significance of the Pearson correlation) and quantification of the trend (slope of the regression line) when enough data points are available (at the very minimum 5 points).

Non-normal Data (non-parametric approach)

For data that is non-normal, we can utilise the non-parametric analogue of the central tendency statistic, the median. This non-parametric statistic allows for the comparison of data aggregated by season or annually using median aggregation with the Mann-Whitney U test. This test is not prone to extreme values and/or outliers and does not demand any concrete distribution of data. The power of this test is slightly lower compared to t-test (preferred in the case of normal or log-normal distributed data). You will also need the standard deviation and n of the sample size. In addition, a description of the data distribution is recommended.

In the case of violated assumptions of exposure data normality, we suggest to use the Mann-Kendall test which can be used in both the basic and seasonal form to identify whether a trend is present (i.e. assess the significance of the temporal change; Khambhammettu, 2005), and the Theil-Sen regression to quantify it (i.e. estimate the linear - or exponential in case of transformed data trend; Fernandes, 2005). Both these methods are robust enough not to be influenced by any outlying values. For these tests, at least 5 points are necessary to provide a reliable test result.

Visualisation

Differences between groups (countries, sites, etc…) can be shown by tabulated aggregate statistics and supplementary diagnostic plots; bar graphs, and/or distribution curves.

Analysis Plan using Individual Data

Statistical analysis of individual data follows them same plan as aggregate data (see above).

Data will initially be explored through univariate descriptive statistics; measures of central tendency (arithmetic mean, geometric mean, median), dispersion (range, variance, percentiles), and estimates´ errors.

As with aggregate data, normal or log-normal data should be analysed using linear least square regression. For non-normal data there is no simple non-parametric method for time trend analysis, thus it is recommended to use Theil-Sen regression after performing annual aggregation. More complex methods can be explored as needed on a case-by-case basis.

Visualisation

Descriptive statistics from the univariate analysis can be shown through graphs or tables, frequency distribution tables, bar graphs, histograms, scatter plots and/or box and whiskers plot.

6.1.2 Case 2: Assessing a temporal trend with a limited number of time points (2-4).

- Has there been a decrease of the regulated category A, B *substance* levels in the population (general/children) from year 20XX to 20XX?
- Did a regulation of substance have an impact on the time trend of *substance* exposure in Europe?
- Did the restriction placed due to the REACH Regulation (or other regulation) have a favourable impact; i.e. a reduction of GM/median concentration of the already regulated (before 20XX) *substance(s)*?

These research questions typically have do with assessing the role a regulation played in the time trend of a certain substance or in a country where different sampling campaign have been done, thus a smaller number of time points may be used. However, given the lower number of data points, a linear trend is not appropriate here, but rather the use of a statistical test to determine a significant increase or decrease in level of HBM biomarker.

Aggregate or individual data may be used.

Analysis Plan using Aggregate Data

Please see 7.1.1 Case 1 for aggregated statistics to consider.

Normal Data (parametric approach)

The use of the geometric mean is for log-normal distributed data (from transformation). Geometric means can be compared to one another by use of t-test statistic, as well as to the reference value(s) defined in Section 10, Calculation of EU Reference Values. To use the t-test statistic you will also need the standard deviation and n of the sample size. In addition, a description of the data distribution is recommended.

Non-normal Data (non-parametric approach)

For data that is non-normal, again we will utilise the non-parametric analogue of the central tendency statistic, the median. For this we recommend the Mann-Whitney U test.

Analysis Plan using Individual Data

Normal Data (parametric approach)

For normal or log-normal distributed data, we recommend the t-test and analysis of variance (ANOVA) to compare group means. While the t-test is limited to comparing means of two groups (i.e. two time points, pre and post-regulation), the one-way ANOVA can compare more than two groups (i.e. up to 5 time points, including points before and after a regulation was put into place). With at least 5 time points, you could additionally consider following the plan for mapping a time trend given in 7.1.1 Case 1.

Non-normal Data (non-parametric approach)

For non-normal data we recommend using the Mann-Whitney U test for comparing 2 time points, and the Kruskall-Wallis test for comparing 3-4 time points.

Visualisation

Descriptive statistics from the univariate analysis can be shown through graphs or tables, frequency distribution tables, bar graphs, histograms, and for individual data you may also consider scatter plots.

6.1.3 Case 3: Evaluating time trends by other variables (e.g. socio-demographic, lifestyle, nutrition…) and between countries.

- Describe the time trend of measurements taken of *substance(s)* annually for X amount of years by country, age, sex, socio-economic status (SES), etc…

Analysis Plan using Aggregate Data

Due to the nature of using several confounders and covariates, following this case is for individual data only. However, aggregate data may be used in cases where individual data is unavailable and sub-groups are well-defined and stratified by parameter of interest (e.g. sex, age, SES). Follow the same aggregate plan as defined in 7.1.1 Case 1. Well defined groups (e.g. age) can be described and compared to one another, as well as plotted linearly for a better visual comparison).

Analysis Plan using Individual Data

Confounding variables should be defined a priori as variables that are known to be related to the biomarker through the use of directed acyclic graphs (DAGs). Covariates should also be defined as possible determinants and their relationship with the substance biomarker.

Data will initially be explored through univariate descriptive statistics; measures of central tendency (arithmetic mean, geometric mean, median), dispersion (range, variance, percentiles), estimates´ errors, and frequencies for categorical variables.

Next, we recommend using multivariate analysis of variance (MANOVA). MANOVA compares means across one or more variables that are based on repeated observations allowing us to analyse how the chosen covariates influence the slope over time (time trend), and not simply the biomarker level. With MANOVA, standard assumptions apply; normality, sphericity, and randomness. As with other analysis of variance tests, the basic idea behind MANOVA is the same as the paired t-test.

If the assumption of normality is not met, use of a non-parametric version of MANOVA called the Friedman test should be considered. This test statistic is based on ranks; the mean rankings (averaged over all subjects) at each time-point are compared to each other. The Friedman test will not give information about the direction of the trend, rather from the mean ranking you will be able to see a decrease or increase from measurement to measurement.

To create more complex models with inclusion of other variables we recommend using either mixed models (Laird, 1982) or generalised estimating equation (GEE) models (Zeger, 1986). Variations of these models have been developed for both categorical (percentile or tertile) and continuous outcomes. The primary difference between the two approaches is that mixed models are full-likelihood methods and GEE models are partial-likelihood methods. The advantage of a partial likelihood method is that they are simpler to compute, and they generalise easily to a wide variety of outcome measures with different distributional forms. Full-likelihood methods provide estimates of person-specific trend lines that can be useful in understanding inter-individual variability (Gibbons, 2010).

A mixed model will contain variables for fixed effects and random effects, allowing for the inclusion of fixed factors and effects (e.g. sex, age) and random factors and effects (e.g. HBM biomarker samples). This type of model is ideal when using multiple covariates, and/or with a lower sample size.

With GEE the relationships between the variables of the model at different time-points are analysed simultaneously. Before carrying out a GEE analysis, the within-subject correlation structure must be chosen. As there is not one answer for this correlation structure should be based on the correlation of the measured data. Data can be considered correlated with a correlation coefficient (r) >0.59. The following are options for correlation structure.

- Dependent correlation structure: for data with more or less equivalent differences in correlation
- Autoregressive correlation structure: data with a steep decrease in correlation over time
- Unstructured correlation: all correlations are assumed to be different

After examining the data and choosing the structure, the GEE model can be fit with predictor variables. Unlike MANOVA, time can be included as a continuous variables or categorical (2 time points) for situations of before and after a regulation (Twisk, 2003).

For time trends between countries linear regression models could also be used as described in the above paragraph.

Visualisation

With multiple explanatory variables showing the descriptive statistics for the variables in a table will give a good overview of the data. Use bar charts or line graphs to show geometric mean ratios (GMRs) with 95%CI´s, and added variable plots in the case of regression modelling to show how much of the variability can be explained by each factor.

6.2 Time Trend References

1. Becker K, Schroeter-Kermani C, Seimert M, Ruther M, Conrad A, Schulz C, Wilhelm M, et al. German health-related environmental monitoring: Assessing time trends of the general population´s exposure to heavy metals. Intl Journal of Hygiene and Environmental Health. 2013;216: 250-4.

2. Fernandes R, Leblanc SG. Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. Remote Sensing of Environment. 2005;95: 303-16.

3. Goen T, Dobler L, Koschorreck J, Muller J, Wiesmuller GA, Drexler H, Kolossa-Gehring, M. Trends of the internal phthalate exposure of young adults in Germany-Follow-up of a retrospective human biomonitoring study. Intl Journal of Hygiene and Environmental Health. 2011;215; 36-45.

4. Hoffman K, Butt CM, Webster TF, Preston EV, Hammel SC, Makey C, et al. Temporal trends in exposure to organophosphate flame retardants in the United States. Environmental Science & Technology Letters. 2017;4: 112-8.

5. Khambhammettu P. Mann-Kendall analysis for the fort ord site. Annual Groundwater Morning Report – Former Fort Ord, California. 2005.

6. LaKind JS, Naiman DQ. Temporal trends in bisphenol A exposure in the United States from 2003-2012 and factors associated with BPA exposure: Spot samples and urine dilution complicate data interpretation. Environmental Research. 2015;142: 84-95.

7. Twisk J. Applied longitudinal data analysis for epidemiology (A practical guide). Cambridge: Cambridge University Press, 2003. Online edition.

7 Geographical Comparison

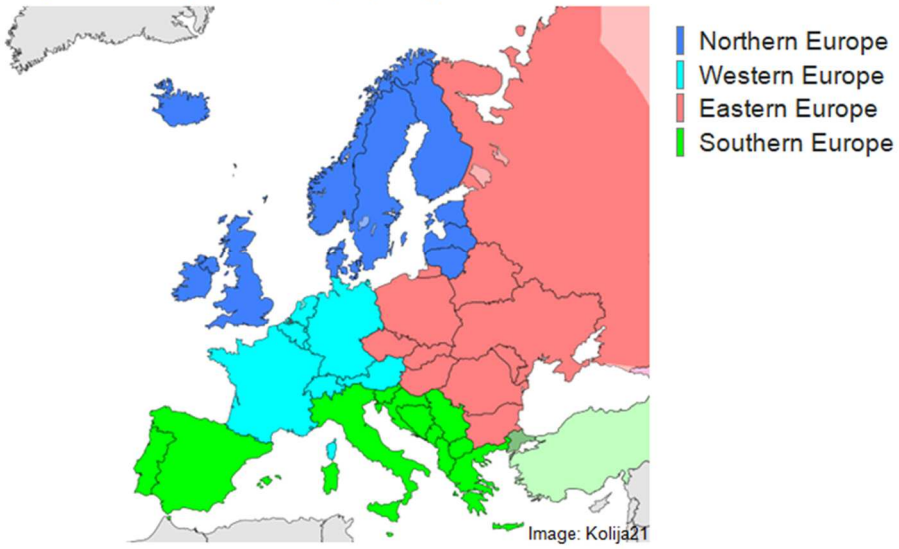
Geographical comparisons in the sense of exposure biomarkers are concerned with the study of the geographic variation of levels of those biomarkers. For the purposes of this study that would mean differences between countries, or differences between multiple sites within one country. This is done through the use of statistical and visual techniques that demonstrate observed differences or similarities.

An important issue is the consideration of time differences in data collection between countries or regions (e.g. one country collected data in 2010 and another in 2016 – how do we compare?). The comparability of datasets will be assessed once it is clear which datasets are available for each analysis. This will be the first step of the data analysis. Procedures to deal with these issues will be described in the substance-specific analysis plans.

For the HBM4EU project, the Sub-regions of Europe geoscheme, as defined by the United Nations (UN) will be used for defining regional areas of comparison (map below). The image can be found for use and modification for inclusion in future works at:

https://commons.wikimedia.org/wiki/File:Europe_subregion_map_UN_geoscheme.svg

Graphic: Subregions of Europe (UN geoscheme)



7.1 Geographical Comparison Analysis Plan, by case

7.1.1 Case 1: Comparison between multiple countries or multiple sites within one country.

- Do exposure levels of *substance(s)* differ significantly between countries or European region?
- Do different regulations in different countries influence exposure to *substance(s)*?
- Do exposure levels of *substance(s)* differ significantly between countries or European region and individual characteristics?

For these cases both aggregate and individual data can be used. However, aggregate data will reveal statistical significance of difference (e.g. among countries or years with respect to ref. values) only with appropriate statistics. Individual data, in turn, will allow comprehensive analysis that includes hypothesised confounders and additional covariates.

Analysis Plan using Aggregate Data

The required aggregate statistics for the different cohorts should be stipulated and communicated well in advance. The statistical software, and possibly the instructions (and/or syntax) used for the calculation should be reported. Recommended statistics for analysis are; indications of normality, geometric mean with 95%CI, (p25, p50, p75, p90, p95, p99), arithmetic mean, standard deviation, and count (n). Aggregate data used in the analysis should meet the requirements defined by the research question (e.g. age, sex, or occupation requirements). For example, if the research question involves children or high-risk occupation groups then using males aged 50-75 would not be appropriate.

From here descriptive comparisons can be made using the geometric mean (standard deviation), sample size, median, and/or percentages, visualisation tools, as well as test statistics t-test, ANOVA for parametric data, and Mann-Whitney U test, Kruskal-Wallis for non-parametric data, to show differences.

Visualisation

Tables of aggregated statistics can be constructed to show the data mentioned previously by country/site, as appropriate. Additionally, a European map can be shown using different colours for varying levels of analysed substance biomarker. Further, bar graphs and distribution curves can be shown to represent the geometric mean levels between countries.

Analysis Plan using Individual Data

Data will initially be explored through univariate descriptive statistics; measures of central tendency (arithmetic mean, geometric mean, median), dispersion (range, variance, percentiles), and estimates´ errors.

If the only interest is to compare substance levels across different countries, then carrying out mean comparison tests as defined above is sufficient for comparison. However, if the aim is to evaluate by further explanatory variables then continue with univariate and multiple regression models.

Conduct univariate and bivariate models for the selected confounding variables of interest. Next, build multiple regression models including those confounders selected a-priori that reached a, p=<0.20 significance level in the bivariate analysis. Using a backward stepwise selection process, drop confounding variables one-by-one starting with those considered to be the least significant and then re-considering all dropped variables for re-introduction into the model. Continue this until final model is achieved. With the completed models, geometric means ratios (GMR) will be given.

Using Multi-level Models (hierarchical or nested models)

Multi-level models are statistical models of parameters that vary at more than one level making them apt for examining variations between countries and variations between individuals at the same time. These models are appropriate for research designs where data of subjects are organised at more than one level (i.e. nested data). The units of analysis are usually individuals (lower level) who are nested within contextual/aggregate units (higher level).

For example, using a multi-level model we can analyse the association between individual levels of a given substance and population density in a certain number of countries. For this, a model with 2 levels would be constructed where individual factors in level 1 are nested in country factors at level 2. The outcome variable of interest is always in level 1.

Data for this analysis will initially be explored through univariate descriptive (see above).

In the first step of multi-level modelling separate analyses are conducted for high level data with the help of level 1 data. In the second step, the regression parameters obtained in step 1 become the outcome variables of interest. (Software programs available, instructions differ) (Woltman, 2012).

Regression diagnostics should be performed to assess the model assumptions (linearity, normality of residuals, multicollinearity, and independence of the error terms). Where assumptions of linearity and normality of residuals are violated, variables should be log-transformed.

Visualisation

Descriptive statistics from the univariate analysis can be shown through graphs or tables, frequency distribution tables, bar graphs, histograms, and/or scatter plots. Final GMR´s can be displayed in a table with 95%CI and/or shown visually using a standard error bar chart.

7.2 Geographical Comparison References

1. Woltman H, Feldstain A, MacKay C, Rocchi M. An introduction to hierarchical linear modeling. Tutorials in Quantitative Methods for Psychology. 2012;8(1): 52-69.

8 Exposure Determinants

Determinants in this study refer to a wide range of personal, social, economic, and environmental factors that may influence the exposure biomarker level of a certain individual.

Determinants to consider, where of interest: lifestyle, nutritional status, genetic background, sex, age, postmenopausal status, elderly, specific regions with elevated levels, and occupation.

8.1 Exposure Determinant Analysis Plan, by case

8.1.1 Case 1: Examine one or more determinants that may influence biomarker concentrations of defined substances.

- What are the determinants of exposure for substance(s) concentration levels?
- Who are the high exposure groups for substance(s); i.e. occupational exposed and not exposed adults, adults and children...?
- Which sub-groups of the population (e.g. age, sex, SES) have exposure levels exceeding HBM guidance values, if available?
- Are there differences in the level of exposure by sub-groups of the population?

Due to the nature of the data required, following this case is recommended for individual data only, no exceptions.

Analysis Plan using Individual Data

Data will initially be explored through univariate descriptive statistics; measures of central tendency (arithmetic mean, geometric mean, median), dispersion (range, variance, interquartile range, percentiles), and estimates´ errors. For categorical data, analyse frequencies.

Conduct univariate and bivariate models for the selected confounding variables of interest. Next, build multiple regression models including those confounders selected a-priori that reached a, p=<0.20 significance level in the bivariate analysis. Using a backward stepwise selection process, drop confounding variables one-by-one starting with those considered to be the least significant and then re-considering all dropped variables for re-introduction into the model.

Due to having multiple explanatory variables in the model we may need to address the problem of multicollinearity; the existence of a high degree of linear correlation amongst two or more explanatory variables in a regression model. We can do this by calculating the variance inflation factor (VIF). If the VIF is >10 then multicollinearity is high. Highly correlated variables should not be included in the same model. Final models will give geometric mean ratios (GMR).

Visualisation

Descriptive statistics from the univariate analysis can be shown through frequency tables. Final GMR´s can be displayed in a table with 95%CI and/or shown visually using a standard error bar chart.

8.1.2 Case 2: Comparison of individual determinants between countries.

- How do individual determinants differ between countries/regions in regards to levels of *substance(s)*?

Due to the nature of the data required, following this case is recommended for individual data only.

Analysis Plan using Individual Data

For this type of study using a cluster analysis is recommended. “A cluster analysis is a multivariate technique that helps regrouping countries in a way that minimises the distance of the clustering variables between countries belonging to the same group, while maximising it among groups (Cornia, 2016)“. Major clustering approaches include hierarchical, optimisation, and model-based methods. A limitation to any of these methods is that there is no internal mechanism to distinguish between important and unimportant indicators, thus indicators must be chosen carefully by the research team (Onda, 2015).

- Hierarchical methods connect data points based on a measure of distance between the data points to form clusters.
- Optimisation methods produce a partition of the data into a number of groups k that must be pre-specified by the analyst, by choosing k data points as pre-assigned “cluster centres”, and then assigning data points to those centres in a way that minimises the squared distances between members within that cluster.
- Model-based methods use an expectation-maximisation algorithm that assigns data points to a fixed number of Gaussian distributions.

Visualisation

A dendrogram is commonly used for cluster analysis, it is a “tree like” diagram used to illustrate the arrangement of the clusters produced by hierarchical clustering. A geographical 2D or 3D diagram with individual points (countries) clustered together may also be an option to consider. Finally, when using individual data a geographic heat map to show the distribution of clusters can be used.

8.2 Exposure Determinant References

1. Cornia GA, Scognamillo A. Clusters of least developed countries, their evolution between 1993 and 2013, and policies to expand their productive capacity. CDP Background Paper No. 33. July 2016.

2. Onda K, Crocker J, Kayser GL, Bartram J. Country clustering applied to the water & sanitation sector: a new tool with potential application in research & policy. International Journal of Hygiene and Environmental Health. 2014;217(0): 379-85.

9 Calculation of EU Reference Values

A proposed strategy to derive HBM European Reference Values (ERV) within the HBM4EU initiative.

9.1 Definitions of terms used

General Population	All individuals from a country, a European geographical area or whole Europe.
Individual	Reference individual selected from a biomonitoring survey with defined criteria (for representativeness) among the general population.
Population	Group of all possible reference individuals.
Population sample	Adequate number of reference individuals taken to represent the population. Individuals from this sample must be comparable to reference individuals.
Exposure data	Set of all measurements of one exposure biomarker concentration within a specific matrix and related to one reference individual belonging to the reference population sample.
Distribution	Set of exposure data in ascending order. Its frequency distribution is used to determine the percentile which will serve as a basis for derivation of a reference value.
Reference value (RV)	Value derived from one percentile from the reference distribution. This value corresponds to the definition of the “reference limit” in the IFCC[1] recommendations.
European Reference Value (ERV)	Reference value derived from the 95th percentile by application of the HBM4EU strategy. Ideally, one value is derived for the whole European area (based on availability of data), separately for each substance, and at a specific time point or time frame which is indicated in the name of the ERV (e.g. $ERV_{2014\text{-}2018}$)
Observed value	Value obtained by analytical measurement within an individual belonging to general population in order to be compared with a reference value.

In this chapter, the terms individual, population, population sample and distribution should be respectively understood as “reference individual”, “reference population”, “reference population sample” and “reference distribution”. The qualifier “reference” is not systematically added in order to make the text more easily readable.

9.2 HBM Reference Value Concept

9.2.1 General information

According to Ewers *et al.* in 1999, reference values (RVs) are statistical descriptions of the ranges of concentrations typically seen in a specified population but which have no direct relationship to health effects or risk assessment. So, by construction, RVs must not be considered as health

[1] IFCC : The international Federation of Clinical Chemistry and Laboratory Medicine (http://www.ifcc.org)

related values. As RVs focus only on the level of exposure of a population to a chemical substance, they do not relate to any or provide any criteria for identifying health risks.

The main purpose of a human biomonitoring RV is to provide a comparison basis for interpretation of levels of exposure to chemical substances of individuals or subpopulations (Angerer, 2007). It allows for the detection of unusually exposed individuals compared to the population used for defining the RV. RVs are derived estimates from data collected in human biomonitoring studies and indicate the upper margin to background exposure to a chemical at a given time. A repeated derivation of RV for the same chemical substance in time with comparable data allows for the following of time trends of exposure levels. It is also a way to assess regulatory policies on the limitation of exposure levels.

For a chemical substance of interest, the RV is a particular level of internal exposure that indicates the splitting point of the sample population in two different categories: 1) the individuals which are exposed or non-exposed to the substance and belong to the background exposure and 2) the highly exposed which is the group of particular interest in risk management and analysis for determinant of exposure. Traditionally, in a frequency distribution of exposure levels from a population, percentiles are used to split the distribution in portions regarding the level of exposure (e.g. the value of the exposure level which splits the sample into 95% below or equal to and 5% above the value). Consequently, RVs are frequently derived from a chosen percentile in the distribution. To be able to quantify high levels of exposure within the general population, RVs need to be derived in empirical studies.

9.2.2 Selection of a percentile for deriving RVs

Although there is no absolute consensus across the scientific community on the necessity to derive RVs from human biomonitoring studies, the use of one single RV per chemical substance, in addition to an exposure distribution, brings about several advantages to specify from which level of exposure an individual/subgroup is unusually highly exposed and exceeds the background level of exposure.

The choice of the 95th percentile as a basis to derive a RV is inspired by IUPAC guidelines and IFCC recommendations on the elaboration of reference values (Solberg, 1987, 2004; Poulson, 1997). It is motivated by the convention in hypothesis testing where the highest 5% of values indicate unusually high values. In addition, it is standard to indicate frequency portions in log-normal distributions in units of standard deviations from the mean, including an approximately 95% cut-off at the right end. IUPAC and IFCC defined the reference interval of a biological measurement distribution as the 0.95 central inter-quantile interval, or the interval between the 0.025 and the 0.975 quantiles. As human biomonitoring is traditionally focused on high levels of exposure, the lower 2.5 percentage of the frequency distribution (i.e. between quantiles 0 and 0.025) is not of particular interest, so the reference interval can be moved to the left side of the distribution and, by consequence, the upper limit is moved from 0.975 to 0.950. This approach is based on the idea that in terms of human exposure to chemicals, people at the upper end of exposure distribution are more vulnerable. Exceedance of RVs in individuals or in subgroups indicates the requirement for follow-up to understand key exposure sources and determinants of exposure for those individuals or respective subpopulations (Angerer, 2011).

On the international scale, when a RV is made available from biomonitoring studies, it is most frequently derived from the 95th percentile of the distribution. As is shown in the cases of Germany (Schulz, 2007), Canada (Saravanabhavan, 2017), South Korea (Lee, 2012), Czech Republic (Cerna, 2012), UK (Bevan, 2013), Spain (Pena-Fernandez, 2014), Italy (Alimonti, 2000) and more recently France (not published).

The HBM4EU initiative offers a new opportunity to derive HBM European RVs (ERVs) for the European population and for a broad variety of chemical substances. Thus, the use of a standard and uniform definition of a HBM RV is of real concern. Harmonised data will be gathered from the participating countries (28 by now) in order to generate an ERV for each chemical substance, either for the whole European zone, or for a specific geographical section of Europe where data are available.

9.3 Proposed strategy to elaborate HBM ERVs

Among the existing possibilities to derive a HBM reference value across countries, there is a preference in studies on European samples and a slight leaning in non-European countries towards using the 95th percentile. To ensure a better comparability between ERVs from the HBM4EU initiative and RVs from HBM international studies, the 95th percentile will be the basis for deriving ERVs in the HBM4EU project. From a regulatory perspective of the research community, this choice enables a straightforward interpretation of RVs across studies, countries, chemical substances, time points and subgroups. To take into account that the sample derived percentile underlies statistical errors, the confidence interval (CI) for each percentile should also be computed. For this, the 95% CI has emerged as the standard CI.

However, to maximise the international comparability of ERVs, the use of the 95th percentile as a basis for the derivation of a RV does not appear to be a sufficient criterion. The selection of participants, the methodology of data collection, the outlier's management or the statistical method used for calculation are also important parameters to take into account.

From a statistical perspective, to provide valid information, ERVs need to be derived from a sufficient amount of HBM data, from an adequate sample size, collected within the time period considered, and be representative of the population. The 95th percentile used as a basis to derive the ERV should be taken from a distribution built with exposure data. Depending on the chemical substance of interest, exposure data are selected from available data by the application of exclusion criteria (see later) that keeps only relevant data to elaborate a valid and interpretable ERV. Another approach is to use data from HBM studies specifically designed for the derivation of ERVs on which all collected data could be considered as reference data (cf. Decision memo on strategy for deriving representative HBM data of the EU population).

So, from HBM data collected from the fieldwork, we suggest a two-step approach consisting in the application of two different kinds of criteria, independent of the design of the HBM survey:

- **Exclusion criteria** to select reference data (individuals) from globally available HBM data in order to build a reference distribution.
- **Authorisation criteria** that applies to reference data and assesses the capability to derive a valid RV from them.

9.3.1 Exclusion criteria/construction of reference sample

The establishment of reliable RVs is a major undertaking consisting of the selection of suitable individuals. Applying exclusion criteria could be regarded as an *a posteriori* selection of individuals in order to build an adapted population sample for deriving RVs (Solberg, 1984). As said previously, if HBM studies are originally designed to do so (by applying same criteria while recruitment of individuals), it may not be useful to apply exclusion criteria *a posteriori*.

The aim of applying exclusion criteria is to exclude individuals or sub-groups of populations, which might confound concentration levels in general or substance specifically. Individuals for which the interpretation of collected data could lead to a suspicion that their behaviour or practices prior to

the biological collection may have had an impact on the exposure biomarker should be excluded from the sample. These exclusion criteria are defined based on the knowledge of confounding factors for the measure of levels of exposure.

For example, individuals that have consumed fish or seafood during the last three days before sampling will be excluded from the sample for calculating ERVs for Mercury or Arsenic. The same applies to smoking in the last days if the aim is to derive a RV for cadmium, nickel, cobalt or antimony. Depending on the population for which the RV is derived (e.g. non-smokers, pregnant women, general population, etc.) different exclusion criteria should be applied. The list with possible exclusion criteria is long and their application depends on the availability of information for the selection of individuals to build the population sample (e.g. sometimes information is not collected by questionnaires and exclusion criteria cannot be applied). We propose a list of general exclusion criteria, which can be considered as partitioning criteria:

- Specific food intakes prior to the biological matrix collection (e.g. fish or seafood) and fasting status;
- Specific physiological status that may have an impact on the measured concentration, for example individuals with creatinine urinary concentration below 0.3 g/L or above 3.0 g/L (WHO, 1996), and pregnancy;
- Drug intake;
- Smoking status;
- Time and date of biological matrix collection, avoid the use of old samples or samples not corresponding to the period of interest.

For specific purposes, exclusion criteria may be considered as partitioning criteria to obtain reference values from specific subgroups of the population sample. This categorisation of criteria has to be decided for each chemical substance. As an example, for the computation of a RV for cadmium, smoking status can be used as an exclusion criteria and the RV will only relate to the exposure of non-smokers, or, smoking status can be considered as a partitioning criteria and three RVs will be derived, one for the general population (including smokers and non-smokers), one RV for non-smokers, and one RV for smokers. It is important to emphasise that these exclusion criteria must be seen as a characterisation of a specific behaviour that can modify a biomarker concentration prior to matrix collection, more than a general condition/behaviour of an individual which can be seen as partitioning criteria rather than exclusion criteria.

Within HBM4EU, for each of the groups of prioritised substances, a list of specific exclusion and/or partitioning criteria is being developed (see further in the part of the SAP corresponding to specific considerations by substance).

9.3.2 Outlier management

One important question about the data preparation is the management of outliers. IFCC recommends to not automatically discard values identified as possible outliers but to handle them according to best judgment (Solberg, 1988). This is done as to not manipulate the outcome by removing outliers. Therefore, we propose to not systematically identify and exclude extreme values of the reference distribution as the impact on the final RV is difficult to assess. However, this does not prevent us from investigating the reference data to identify an aberrant value, which could arise from an abnormal analytical result (e.g. an analytical error in the laboratory or the inclusion of an individual that should have been excluded). It is, however, unlikely that this kind of data would reach this step in the process as it would have been discarded earlier.

9.3.3 Authorisation criteria

This second step aims to apply authorisation criteria used to validate the possibility of computing a 95th percentile from the reference distribution in order to derive a valid and interpretable RV.

9.3.3.1 General population representativeness

Deriving a RV from HBM data which were not collected from a representative population is not advisable as it will lead to a non-valid and non-interpretable RV. Typically, RVs are derived from biomonitoring data from nationally representative surveys which are considered appropriate. In the case that the collected data do not seem representative of the general population, applying the exclusion criteria may lead to representativeness (thus it can be compared to an *a posteriori* selection of individuals).

If the population sample used to build a reference distribution of levels of exposure cannot be characterised as representative of the population for whom we want to derive the RV, then the distribution (which is by consequence not a reference distribution) must not be used to compute a 95th percentile in order to derive a RV.

In practice, within the HBM4EU project, if the aim is to derive a global RV relevant for the whole European population, the population sample must include all age groups of interest, both male and female individuals, individuals from different socioeconomic status, etc. It must also include reference data coming from the four different geographical areas of Europe (North, South, East, and West) to achieve EU wide coverage. If the reference subgroup does not include all ages it must be stated that the given metrics are only representative for those specific age groups.

Depending on the availability of data from different participating countries, it may not be feasible to derive an ERV for the entire European population during the first years of the HBM4EU project. This could constrain to focus on the available data from participating countries and to the derivation of ERVs for only those countries as a first step.

9.3.3.2 Population size

The size of the population sample has a major impact on the determination on the 95% CI of the 95th percentile. In order to be able to compute a narrow and reliable CI of the 95th percentile of the distribution, the population sample should contain at least 120 valid measurements (Solberg H.E., 1987, 1988, 2004). This number must be set as the *minimum* for deriving a RV. It is important to keep in mind that the size of the reference population subgroup should be large enough to include different subgroups to increase the possibility of deriving specific RVs for those subgroups (age, sex, BMI).

9.3.3.3 Analytical method

Exposure data coming from different studies could be a major issue in deriving general RVs because of potentially different conditions of data collection (Solberg, 1991). To ensure that results are comparable and to reduce/control the impact of inaccuracy or imprecision on the data, analytical methods used for measuring concentrations values must be adequately described. According to Saravanabhavan et al. (2016) it is essential to assess the overall quality of the biomonitoring data in terms of the specificity and the sensitivity of the analytical methods used, and the quality assurance or control procedures followed during sampling, sample pre-treatment, and instrumental analysis. Within HBM4EU, analytical protocols, SOPs and criteria will be established within WP9 to meet the QA/QC standards. Data used to derive ERVs need to meet these criteria.

Once the population sample has been built by application (or not) of exclusion criteria and once these authorisation criteria are met, the reference distribution can be drawn and the 95th percentile and its 95% CI can be computed and be used as basis for deriving ERV.

9.4 ERV Calculation

9.4.1 One ERV for the European general population

9.4.1.1 Statistical method

Percentiles of the reference distribution may be estimated by parametric or non-parametric methods. Parametric estimation requires that the reference distribution fits in a specific distribution type (e.g. Gaussian distribution or that the distribution can be transformed to be approximately normally distributed). On the other hand, non-parametric method techniques make no assumptions about type of distribution. Because of its simplicity, IFCC recommends to use non-parametric methods to calculate the 95th percentile of the reference distribution. The non-parametric method selected for the recommendation is the simple rank-based procedure described by Reed et al. (1971). Since the publication of this recommendation, the bootstrap non parametric method has been the most significant development that can be used to estimate 95th percentile and 95% CI. According to Solberg (1984), it's currently the best method available. In addition, since the algorithm used to calculate percentiles differs by statistical programs and settings, one specific recommendation will be suggested for calculating RVs within the HBM4EU project.

The HBM4EU initiative allows gathering of existing HBM data from participating countries in order to elaborate a RV at European scale. As ERV, we propose to round the 95th percentile of the reference distribution within its 95% CI. In order to set a general criterion for rounding the exact value of RV, it is proposed to round the 95th percentile at the value immediately below, while taking into account the number of digits which is set, depending on the scaling and the precision of the analytical method used. Exceptionally, if the rounded value of the 95th percentile falls below the lower bound of the 95% CI, the lower bound itself should be used. This process aims at being precautionary within the interpretations of exposure levels by comparison with the ERV.

To enable comparison of ERV with already published RVs derived from a different percentile, additional typical percentiles (e.g. 75, 90 and 99) which are not defined as RV should be reported. Reporting percentiles in addition to the RV would thus have two roles: to describe the distribution of the sample and to enhance comparability to past studies. The frequency distribution curve of exposure could also be presented to provide further information visually.

9.4.1.2 Dealing with the limit of quantification (LOQ)

In the Canadian Health Measures Survey (CHMS), RVs are derived only if more than 60% of reference data are over the limit of detection (LOD). It is based on the idea that there is no specific need to derive a RV for a chemical substance that is not widely detected in the Canadian population. However, it could be of a certain relevance to supply the information that a chemical substance is barely detected in a population (60% below LOD). As the value of the 95th percentile is not modified in the reference distribution depending on the number of reference data below LOQ or LOD, RVs seems to be an appropriate tool for giving this information. At this point we do not propose to set a minimal amount of quantified reference exposure data to derive a RV.

Nevertheless, in order to be able to compute a 95% CI of the 95th percentile, we suggest using the recommendations provided in section 5.4.1 (see above) for the imputation of values below LOD/LOQ.

9.4.1.3 Unit of ERV

RVs should be expressed in the same unit as the reference data they are derived from. That means generally in µg/L or µg/mL for urine, blood, serum or plasma matrices and in µg/g of hair for hair matrix. There is no consensus anymore in the scientific community about the necessity to express levels of exposure in µg/g of creatinine for urine matrix to adjust for the dilution level of the urine. Other methods dealing with the dilution level are specific gravity, and osmolality. However, in order to ensure comparability with RVs published in µg/g of creatinine, it is relevant to translate the new ERV in this unit if raw data allows it. In some studies (e.g. COPHES), creatinine concentrations <0.3 g/L or >3.0 g/L were excluded (WHO 1996) to calculate reference values.

9.4.1.4 Naming of ERV

As these RVs are derived for the European area within the HBM4EU project, we have proposed to call these values ERVs for European Reference Values.

ERV could be derived for different time periods. As such, it seems necessary to add an identifier to indicate the period to which the ERV refers and facilitate future comparability of ERVs. Therefore, we suggest to add the time period to each reported ERV. For example, an ERV based on data collected from 2014-2018 would be called $ERV_{14\text{-}18}$.

ERVs are typically derived from the 95th percentile. As such, one could indicate from which percentile the ERV was derived as well (i.e. ERV_{95}). However, we assume that the 95th percentile will be the commonly used basis for deriving RVs and thus refrain from adding this number to the ERV.

9.4.2 ERVs for subpopulations

In case a sufficient amount of reference data exists, it is recommended to derive ERVs for different subgroups of the population (sub-ERV). This can be done through the use of partitioning criteria that stratifies the population according to specific characteristics. It is important to remember that in order to derive these sub-ERVs each subpopulation sample has to meet the specific authorisation criteria (e.g. include at least 120 individuals). Nevertheless, this process should only be carried out if sub-ERVs from subpopulations are significantly different from each other, otherwise one global ERV is sufficient for the whole population (e.g. if there is no difference between sub-ERV for children and sub-ERV for adults then there is no need to derive two sub-ERVs).

There could be several partitioning criteria applied. We propose that for deriving sub-ERVs in the HBM4EU framework, those criteria are limited to age, sex and European geographical area. In contrast to North American HBM studies, race or ethnicity cannot be considered as potential partitioning criteria within the scope of the HBM4EU project because the inclusion of these criteria would violate the legislation in many participating countries.

9.4.2.1 Age

As a partitioning criterion, age shouldn't necessarily be categorised by equal intervals but rather focus on particular ranges that represent relevant characteristics of life phases. According to the sampling strategy, there are 7 age groups considered in HBM4EU: babies (0-2y); toddlers (3-5y); children (6-11y); adolescents (12-19y); young adults (20-39y); adults (40-59y) and elderly (60-79y).

9.4.2.2 Sex

Sex seems to be the easiest partitioning criteria to apply on the reference population as it contains only two modalities (male/female).

9.4.2.3 European geographical areas

The geographical splitting of Europe has been set in the sampling strategy according to the United Nations geoscheme for Europe into four regions: Northern Europe, Eastern Europe, Western Europe and Southern Europe. This clustering can be used as partitioning criteria to derive a sub-ERV from the 95th percentile for each European area and enable comparability of levels of exposure between geographic regions.

9.4.2.4 Combination by age, sex and geographical region

The sampling strategy framework forecasts as a minimum the inclusion of 150 males and 150 females for each of the selected age groups in three countries in each of the four European geographical areas. By having more than 120 individuals for each group (the minimum sample size to derive ERVs), this sample size, depending on missing values, theoretically allows deriving RVs partitioned by sex, age group and geographical region. During the first years of the project, while using only existing data, the population size requirement will likely not be met for each geographical area which could lead to the derivation of ERV only for some areas or none at all.

9.4.2.5 Other relevant partitioning criteria

For some particular chemical substances, additional partitioning criteria may be of certain interest regarding main confounder variables that are known to have a great impact on the biomarker's concentration. Instead of using this confounder as an exclusion criteria to focus on a reference population that is not impacted by it, this constraint can be taken into account in order to derive two different RVs, one including a reference population sample which is impacted by the confounding factor, and another which is not. For example, for Arsenic RVs, after deriving one RV for the reference population sample, it has been decided and is feasible to derive one RV for high fish consumers and one for low fish consumers. In this case, it may be more relevant not to publish one global RV but rather to focus on the two partitioned RVs. With a similar hypothesis, two RVs could be derived for Cadmium, one for smokers and one for non-smokers.

9.5 Conclusion on the strategy to derive European reference values within HBM4EU

The definition of a European strategy to derive HBM reference values within the HBM4EU initiative is an opportunity to unify the derivation and the use of ERV in Europe and other countries with human biomonitoring programs. It sets a standardised framework which is of particular interest as HBM ERV are used to compare exposure levels to chemical substances, follow the evolution of time-trends, or to assess regulatory policies on the limitation of exposure levels. The European strategy presented here suggests to use one single reference value based on the 95th percentile and its 95% CI for deriving an ERV. To facilitate comparison with older HBM findings, additional percentiles are reported..

The derivation process is based on the successive application of the following steps (cf. following figure):

- Exclusion criteria to build a reference population or subpopulation sample for which the ERV is derived;
- Authorisation criteria to ensure that statistical analysis will produce a valid and interpretable ERV (e.g. a minimum of 120 individuals to derive one ERV);
- Partitioning criteria in order to produce sub-ERVs that take into account characteristics of subpopulation groups (e.g. age, sex, geographic area).

This strategy should be carried out for each chemical substance within the HBM4EU project. We believe that this will help to build a foundation for the inter-comparability of HBM data between different HBM initiatives in the future.

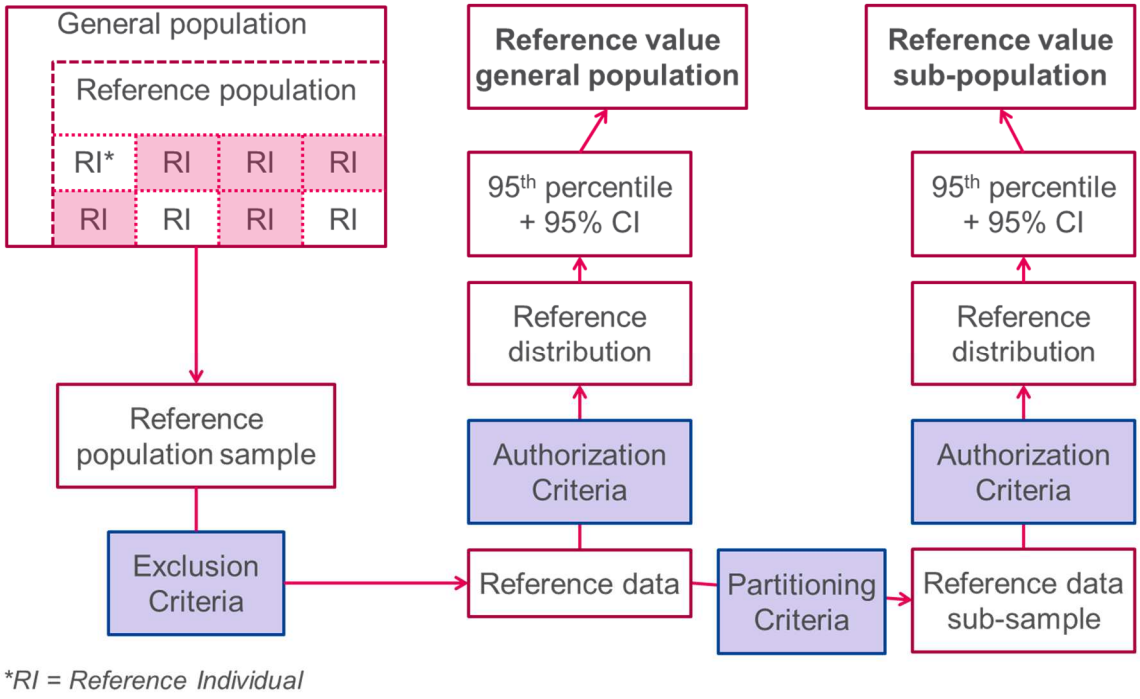


Figure 1: European strategy to derive exposure reference values from human biomonitoring studies within the HBM4EU project

9.6 Section References

1. Ewers U, Krause C, Schulz C, Wilhelm M. Reference values and human biological monitoring values for environmental toxins. Report on the work and recommendations of the Commission on Human Biological Monitoring of the German Federal Environmental Agency. Int Arch Occup Environ Health. 1999;72(4):255-60.

2. Angerer J, Ewers U, Wilhelm M. Human biomonitoring: state of the art. Int J Hyg Environ Health. 2007;210(3-4):201-28.

3. Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section. Approved Recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. J Clin Chem Clin Biochem. 1987;25(5):337-42.

4. Angerer J, Aylward LL, Hays SM, Heinzow B, Wilhelm M. Human biomonitoring assessment values: approaches and data requirements. Int J Hyg Environ Health. 2011;214(5):348-60.

5. Schulz C, Angerer J, Ewers U, Kolossa-Gehring M. The German Human Biomonitoring Commission. Int J Hyg Environ Health. 2007;210(3-4):373-82.

6. Saravanabhavan G, Werry K, Walker M, Haines D, Malowany M, Khoury C. Human biomonitoring reference values for metals and trace elements in blood and urine derived from the Canadian Health Measures Survey 2007–2013. International Journal of Hygiene and Environmental Health. 2017;220(2, Part A):189-200.

7. Lee JW, Lee CK, Moon CS, Choi IJ, Lee KJ, Yi SM, et al. Korea National Survey for Environmental Pollutants in the Human Body 2008: heavy metals in the blood or urine of the Korean population. Int J Hyg Environ Health. 2012;215(4):449-57.

8. Cerna M, Krskova A, Cejchanova M, Spevackova V. Human biomonitoring in the Czech Republic: an overview. Int J Hyg Environ Health. 2012;215(2):109-19.

9. Bevan R, Jones K, Cocker J, Assem FL, Levy LS. Reference ranges for key biomarkers of chemical exposure within the UK population. Int J Hyg Environ Health. 2013;216(2):170-4.

10. Alimonti A, Petrucci F, Krachler M, Bocca B, Caroli S. Reference values for chromium, nickel and vanadium in urine of youngsters from the urban area of Rome. J Environ Monit. 2000;2(4):351-4.

11. Pena-Fernandez A, Gonzalez-Munoz MJ, Lobo-Bedmar MC. "Reference values" of trace elements in the hair of a sample group of Spanish children (aged 6-9 years) - are urban topsoils a source of contamination? Environ Toxicol Pharmacol. 2014:38(1):141-152.

12. Poulson OM, Holst E, Christensen JM. Calculation and application of coverage intervals for biological reference values (Technical Report). Pure and Applied Chemistry. 1997. p. 1601.

13. Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section. The theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta. 1984;137(1):97f-114f.

14. Solberg HE, PetitClerc C. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. J Clin Chem Clin Biochem. 1988;26(9):593-8.

15. Solberg HE, PetitClerc C. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. Clin Chim Acta. 1984;139(2):205f-13f.

16. Solberg HE, Stamm D. International Federation of Clinical Chemistry (IFCC), Scientific Division. Approved recommendation on the theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. Eur J Clin Chem Clin Biochem. 1991;29(8):531-5.

17. Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal Program. Clinical Chemistry and Laboratory Medicine2004. p. 710.

18. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. Clin Chem. 1971;17(4):275-84.

19. WHO. Trace Elements in HUman Nutrition and Health. Geneva: World Health Organisation, 1996.

10 Uncertainty Analysis

10.1 Qualitative description of uncertainty in parameters relevant for biomonitoring data

The starting point of interpreting human biomonitoring data is an assessment of their quality [1]. The first stage of quality assessment involves a qualitative description of the quality of the generated biomonitoring data. We propose to collect this information for each relevant dataset in a table with the goal to contain each of the major aspects involved in generation of the biomonitoring data and, if possible, also to include a statement of the estimated magnitude and direction of the uncertainty.

The table below contains an overview of the minimal set of information that should be reported for a dataset to be able to assess the quality of the biomonitoring data.

Aspect	Description
Biomonitoring method used	Description of how method compares to state of the art
Coefficient of variation (CV)	Report of the CV (%) of the assay for each biomarker
Recovery	Percent recovery by the assay of known added amount of analyte in the sample or as the difference between the mean and the accepted true value together with the confidence intervals
LOD/LOQ	Description of how LOD and LOQ were determined
Design of the survey	Overview of the sampling design that was used, including oversampling in strata (age, sex, race, …), etc., ideally including sampling weights
Response rate in the survey	Information on the response rate in the survey
*** Other aspects that are identified by the consortium as relevant*	

10.2 Quantitative description of uncertainty in exemplary cases described in Chapter 5

As mentioned in chapter 4.5 (p. 11), to allow the assessment of the level of uncertainty in estimated statistics either a standard error or confidence intervals should be reported. Specific instructions for each of the cases described in Chapter 5 are provided below.

Overview of uncertainty analyses for cases defined in chapter 5		
Case	**Aggregate or Individual Data**	**Recommended description of uncertainty**
Time Trends		
Case 1: Evaluating time trends in a country/region	Aggregate	*Normal data* – parametric standard error or 95% CI of the trend estimate *Non-normal data* – bootstrapped standard error or 95% CI of the trend estimate

given a number of time points (5+).	Individual	*Normal data* – parametric standard error or 95% CI of the trend estimate *Non-normal data* – bootstrapped standard error or 95% CI of the trend estimate
Case 2: Assessing a temporal trend with a limited number of time points (2-4). E.g. before and after a regulation	Aggregate	See Case 1, Aggregate.
	Individual	Provide standard deviation of the calculated means
Case 3: Evaluating time trends by other variables (e.g. SES, lifestyle, nutrition…) and between countries.	Aggregate (Individual data recommended)	*Normal data* – parametric standard error or 95% CI for each of the estimated parameters *Non-normal data* – bootstrapped standard error or 95% CI for each of the estimated parameters
	Individual	*Normal data* – parametric standard error or 95% CI for each of the estimated parameters *Non-normal data* – bootstrapped standard error or 95% CI for each of the estimated parameters
Geographical Comparison		
Case 1: Comparison between multiple countries or multiple sites within one country.	Aggregate	Comparison of means: Provide standard deviation of the calculated means
	Individual	Comparison of means: Provide standard deviation of the calculated means Multi-level hierarchical models – parametric standard error or 95% CI for each of the estimated parameters
Exposure Determinants		
Case 1: Examine one or more determinants that may influence biomarker concentrations of defined substances.	Individual (aggregate data not recommended)	Parametric or bootstrapped standard error or 95% CI for each of the estimated parameters
Case 2: Comparison of clusters (or groups) of countries.	Individual (aggregate data not recommended)	Cluster analysis - cross-validation can be used to determine the best number of clusters from the data

10.3 Regression diagnostics

For all analyses that involve a regression approach. It is suggested to conduct a standard set of procedures that can help to assess the validity of the fitted regression model (and parameter estimates based on this model) in any of a number of different ways. In the table below we provide an overview of six commonly conducted regression diagnostics. Additional procedures are available and are suggested to be applied in specific situations.

Factor influencing validity	Suggested statistic(s) or plot(s)
Influential observations	Leverage plot
Non-normality of residuals	Normal Q-Q plot
Non-constant error variance	Scale-Location plot
Multicollinearity between predictors	Calculate variance inflation factors
Nonlinearity of the estimated association	Partial residual plots
Non-independence of Errors	Durbin–Watson statistic

10.4 Section References

1. Human Biomonitoring for Environmental Chemicals Consensus Study Report (2006).

11 Specific Statistical Analysis by Substance Group

The following sections provide statistical information by substance group. The information is divided into four main research areas; general exposure levels (reference values), time trends, geographic comparisons, and exposure determinants. For each research area substance groups have provided substance specific research questions and the variables necessary to address these. Statistical analysis for these research questions can be found under the research area of the same name in the general statistical analysis plans (i.e. Time Trends for phthalates refers to Time Trends found in Section 6 of the general statistical analysis plans). A more detailed elaboration of these plans for each individual research question to be addressed including background information and exact statistical methods will be developed in Task 10.4.

For all substances, specific chemicals/metabolites to be investigated are only listed once under General exposure levels to reduce repetition. However, these are obligatory for all analyses.

Abbreviations

BMI	Body mass index	L	Litre
LOD	Limit of detection	g	Grams
LOQ	Limit of quantification	Kg	Kilogram
OBL	Obligatory	mg	Milligram
OPT	Optional	µg	Microgram
Osm	Osmolarity	ng	Nanogram
S/m	Siemens per meter	Cm	Centimetre
		Km	Kilometre

11.1 Phthalates/DINCH

11.1.1 General exposure levels

- What is the current exposure of the EU population to the 16 phthalates (Cat A, B and C) and their substitute DINCH?

Variable	Description of variable	Unit / score	OPT / OBL
Phthalates, Cat A:	Metabolites of Phthalate diesters of Cat A:	Urine concentration [µg/L]	OBL
DEHP	MEHP		
	5OH-MEHP/MEHHP		
	5oxo-MEHP / MEOHP		
	5cx-MEPP/MECPP		
BBzP	MBzP		
DnBP	MnBP		
	3OH-MnBP		
DiBP	MiBP		
	2OH-MiBP		
DiNP	OH-MiNP / MHiNP		
	oxo-MiNP / MOiNP		
	cx-MiNP / MCiOP		
DEP	MEP		
Phthalates, Cat B:	Metabolites of Phthalate diesters of Cat B:	Urine concentration [µg/L]	OBL
DiDP/DPHP (with sum method)	[OH-MiDP/OH-MPHP] , [MHiDP/OH-MPHP]		
	[oxo-MiDP/ oxo-MPHP] , [MOiDP/oxo-MPHP]		
	[cx-MiDP/cx-MPHP] , [MCiNP/cx-MPHP]		
DPHP (with specific method)	OH-MPHP		
	oxo-MPHP		
	cx-MPHP / MPHxP		
DnOP	MnOP		
	3cx-MBP / MCPP		

Variable	Description of variable	Unit / score	OPT / OBL
DMP	MMP		
DnPeP	MnPeP		
DCHP	MCHP		
DINCH	OH-MINCH / MHiNCH		
	oxo-MINCH / MONCH		
	cx-MINCH/ MCOCH		
Phthalates, Cat C:	Metabolites of Phthalate diesters of Cat C:	Urine concentration [µg/L]	OBL
DnHP	MnHP		

Variable	Description of variable	Unit / score	OPT / OBL
LOD	Limit of detection for each metabolite	Urine concentration [µg/L]	OBL
LOQ	Limit of quantification for each metabolite	Urine concentration [µg/L]	OBL
Type of method for LOD determination	Method used to determine LOD		OPT
Type of method for LOQ determination	Method used to determine LOQ		OPT
Urine density		g/L	OPT
Urine volume		L	OPT
Creatinine level	Concentration of creatinine in urine	Urine concentration [µg/L]	OPT
Osmolarity	Osmotic concentration	Osm/L	OPT
Urine specific gravity	Ratio of urine density compared with water density		OPT
Conductivity	Electrolytic conductivity	S/m	OPT
Type of participant		Background population // occupational exposed // pregnant women // etc...	OBL
Sample type		spot- // morning- // 24 h-urine	OBL
Laboratory	Analysing laboratory	Laboratory number	OPT

11.1.2 Time Trends

- Are there different time trends for unregulated and regulated phthalates and DINCH? (Starting with Cat. A substances)
- Is there a sufficient decrease of the regulated Cat. A substance levels (GM/median) in the population (general/children?) from year 2007 until today (2017)? (DEHP, DnBP, DiBP, BBzP)
- Do time trends for phthalates/DINCH differ between countries[2]?
- Which determinants contribute to time trends?

Variable	Description of variable	Unit / score	OPT / OBL
Year	Year of sampling	Year	OBL
Month	Month of sampling	Month	OPT
Day	Day of sampling	Day	OPT

11.1.3 Geographical Comparisons

- Do the exposure levels of the general population and subpopulations differ significantly ($p < 0.05$) between the countries?
- Do different regulations in different countries influence exposure to phthalates and DINCH?
- Had the restriction under REACH the favourable impact, that is a reduction of GM/median concentrations of the already regulated (before 2015) phthalates (DEHP, BBP, DnBP, DiNP, DiDP, DnOP), especially for children?
- Can countries be grouped according to similarity in concentration levels (e.g., northern, southern…)?

Variable	Description of variable	Unit / score	OPT / OBL
Country	Country of participant/study	Leaning on HBM4EU participant number (Germany = UBA = 1 // Austria = EAA = 2, etc.)	OBL
Region in Europe	United Nations geoscheme for Europe	1=North // 2=South // 3=West // 4=East	OBL
Type of regulation for the metabolite/ substance	Extent of regulation or restriction of a metabolite in its respective country	% (of the substance e.g., in a lotion)	OBL
Year of regulation for the metabolite/ substance	Year when regulation became effective for the metabolite/substance	Year	OBL
Sunset date / potential phase outs for the metabolite/ substance	Date of sunset or potential phase out	Date	OPT

[2] See section 11.1.3 for country codes

11.1.4 Exposure Determinants

- What are determinants for phthalates/DINCH concentration levels?
- What are the high exposure groups? (Is there a statistical significant difference in mean concentration between adults and children / occupational exposed and non-exposed adults / between men and women/ …?
- What portion of the population has exposure levels exceeding the HBM guidance values - if existing- or TDI)?
- What portion of subgroups of the population (e.g., in age, sex, socio-economic status) has exposure levels exceeding the HBM guidance values - if existing- or TDI)?

Variable	Description of variable	Unit / score	OPT / OBL
Personal factors:			
Sex	Sex of participant	0 = male // 1 = female	OBL
Age	Age of the participant at time of sampling	Completed years	OBL
Height	Height of participant	Cm	OBL
Weight	Weight of participant	Kg	OBL
BMI[3]	BMI of participant	kg/m^2	OPT
Socio-economic status	Socio-economic status of participant	Low // Medium // High	OBL
Education	Education of participant	Low // Medium // High; ISCED scale	OBL
Occupation status	E.g. What is your current labour status?	Current labour status, categories[4]	OPT
Occupation	E.g. To which industrial sector does this occupation belong?	Industrial sectors[5]	OPT
Main job task with plastic devices or material	E.g. Do you have contact with plastic devices, material or substances (at least once per week)?	Yes // No	OPT
Present position	E.g. Which position do you presently hold?	ISCO-88 - International Standard Classification of Occupations	OPT
Personal care and health care factors:			

[3] Needed if height and weight are not provided
[4] Example. Carrying out a job or profession, including unpaid work for a family business or holding, including an apprenticeship or paid traineeship, etc. // full time // part time // unemployed // pupil, student, further training, unpaid work experience // in retirement or early retirement or has given up business // permanently disabled // in compulsory military or community service // fulfilling domestic tasks // other inactive person
[5] Phthalate/DINCH-relevant professions. E.g. painter

Variable	Description of variable	Unit / score	OPT / OBL
Use of sunscreen	Last 4 weeks	Daily // several times a week // 1x a week // 1-3x a month // almost never // never	OPT
Use of shampoo		Times per week	OPT
Use of other hair products		Times per week	OPT
Use of deodorant		Times per week	OPT
Use of body lotion	Body lotion, creams	Times per week	OPT
Use of make-up		Times per week	OPT
Use of nail polish		Times per week	OPT
Fragrances	Fragrances (perfume, eau de toilette etc.)	Times per week	OPT
Medicine intake		In groups e.g. painkillers, asthma medication, etc.	OPT
Life style and food factors:			
Main source of drinking water	What is your main source of water for drinking?	1 = public water & supply // 2= commercial producers // 3= own well/ private water supply	OPT
Fish consumption	How often did you eat fish/fish products in the last 4 weeks	daily // several times a week // 1x a week // 1-3x a month // almost never	OPT
Consumption of meat/ cold meat (ham, cut)	In the last 4 weeks	As above	OPT
Convenience food/ fast food	In the last 4 weeks	As above	OPT
Dishes served in a canteen	In the last 4 weeks	As above	OPT
Milk	In the last 4 weeks	As above	OPT
Chocolate	In the last 4 weeks	As above	OPT
Cereals	In the last 4 weeks	As above	OPT
Cheese	In the last 4 weeks	As above	OPT
Ice cream consumption	In the last 4 weeks	As above	OPT
Hazelnut spread	In the last 4 weeks	As above	OPT

Variable	Description of variable	Unit / score	OPT / OBL
local food	Home-grown fruit/ vegetables in the last 4 weeks	As above	OPT
Use of chewing gum	Frequency of chewing gum	Times per week	OPT
Hand washing	Frequency of washing hands	Times per day	OPT
Diseases	Which of the following diseases did you experience?	Categories of diseases	OPT
Use of plastic toys	E.g. How often did your child [name] play with toys consisting of or containing rubber-like plastic in the last week?	Daily // less than daily // never	OPT
Use of plastic gloves	E.g. How often did you wear rubber-like plastic gloves (not latex) in the last week?	Daily // less than daily // never	OPT
Use of pet toys		Daily // less than daily // never	OPT
Smoker		Yes // No	OPT
Smoker in household		0 // 1 // more than 1	OPT
Age when began smoking		Age in years	OPT
PVC on floor	E.g. Do you have PVC flooring in your flat/house?	Yes // No	OPT
PVC on floor		M^2	OPT
PVC on walls	E.g. Do you have PVC wall paper in your flat/house?	Yes // No	OPT
PVC on walls		M^2	OPT
Rooms with carpets?	E.g. Do you live in rooms with carpets?	Yes // No	
Mainly carpet in the home	Does a carpet cover more than half of the floors in the home?	Yes // No	OPT
Pet at home	E.g. Do you have a pet at home?	Yes // No	OPT
Degree of urbanization		High // middle // low or Urban // rural // urbanized	OPT
Season	Month of sampling	1-12	OPT
Time of day	Time of day of sampling		OPT

11.2 Bisphenols

11.2.1 General Exposure Levels

- What is the current (level of) exposure of the EU population (including foetus) to bisphenols compounds (category A, B, C)?
- What are the reference values for EU population to bisphenols compounds (especially to cat A and B)?
- Are their differences in the level of exposure by some sub-population groups (age, gender, occupational activities)?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Bisphenol A	Urinary BPA levels	Urine concentration [µg/L]	OBL	Urine is the common biological matrix used in previous studies in European countries among several others and should be a starting point for a harmonised approach. We need information about free and/or conjugate forms.
Bisphenol F and Bisphenol S	Urinary BPS/BPF levels	Urine concentration [µg/L]	OBL	There is a gap of data on BPS and BPF. We need information about free and/or conjugate forms.
Other Bisphenols (Cat. C)	Urinary concentration for :BPB, BPAF, BPAP, BPBP, BPCI2, BPE, BPPH, BPM, BPP, BIS2, DHDPE, BPFL, BPZ, BP4'4	Urine concentration [µg/L]	OPT	There is need of prioritisation among different Cat. C Bisphenol if not possible to assess all of them. We need information about free and/or conjugate forms.
Creatinine	Creatinine concentration in urine	Urine concentration [µg/L]	OBL	Standard of use is to be determined in the statistical analysis plan
Occupational status				
Occupational activity		List of activities	OBL	In relation with bisphenol exposure (cashiers, plastic industry,

11.2.2 Time Trends

- What is the time-trend for Bisphenol A exposure?
- Are there differences in time-trend for bisphenols B and F exposure (if possible to trace)?
- Did the European regulation on BPA have an impact on the time-trend of BPA exposure in Europe?
- Is there an effect of the different regulations in different countries on the exposure levels or time-trends for BPA or substituents?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Year	Year of sampling	Year	OBL	
Month	Month of sampling	Month	OPT	
Country	Country of sampling	Name of country	OBL	
Regulation in the country, for each compounds	Type of regulation or extent of restriction of a substance in its respective country		OBL	
Year of regulation for the compounds	Year when regulation became effective for the compounds	Year	OBL	
Type of used concerned by regulation			OBL	In relation with the determinants of exposure
Type of population concerned by regulation			OBL	In relation with the determinants of exposure
Country-level variables	Variables on country level which are expected to contribute to differences in exp.		OPT	

11.2.3 Geographical Comparisons

- What is the current (level of) exposure in each 4 European geographical area to bisphenols compounds (category A, B, C) based on the measurement in urine matrix?
- Which are the reference values for each geographical area for BPA?
- Do the exposure levels, mixture composition differ significantly between countries?
- Can countries be grouped according to similarity in concentration levels?
- Do the high exposure population groups varies between European countries and why?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of participant/study	Name of Country	OBL	
Time of sampling	Time of sample collection	Morning / afternoon / evening	OBL	
Type of sampling	Sampling method	Morning void sampling/spot urine sampling/24 hour pool sampling	OBL	
Analytical method	Name of method		OBL	Define for which bisphenol compounds method is available and operational
Type of tube		Material	OBL	Should allow to appreciate the risk of contamination
Urine dilution assessment method	Type of adjustment method for taking into account urine dilution	creatinine, osmolarity, specific gravity	OBL	
LOD	Level of detection for each compound (and form)	Urine concentration [µg/L]	OBL	Standardised procedure for reporting concentrations lower than LOD
LOQ	Level of quantification for each compound (and form)	Urine concentration [µg/L]	OBL	Standardised procedure for reporting concentrations lower than LOQ

11.2.4 Exposure Determinants

- Which determinants contribute to the exposure to BPA and their main substituents (BPS, BPF)?
- What are the high exposure groups and what are their sources of exposure and/or routes of contamination to bisphenols compounds?
- What is the portion of the population with exposure level exceeding the HBM value for BPA?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Height	Height of participant	cm	OBL	Use for BMI calculation
Weight	Weight of participant	kg	OBL	Use for BMI calculation
Age	Age of the participant at time of sampling	Completed Years	OBL	
Sex	Sex of participant	1 = male 0 = female	OBL	
Education	Education of participant, ISCED scale	Low Medium high	OBL	For children, parent or household; same operationalisation as in task 7.2.
Socio-economic status	Socio-economic status of participant	Low Medium high	OBL	For children, parent or household; same operationalisation as in task 7.2.
Pregnancy status	Pregnant women	Yes // No	OBL	Only for women
Smoking status	Recent exposure to tobacco (<24h) active or passive	Yes // No	OBL	A focus on the recent exposure to the tobacco, e.g. in the last 24 hours
Place of residence	Living place	Urban // rural	OBL	
Place of birth	Origin	Urban // rural	OPT	If ethnicity is not possible
Year of residence	Starting year for residence in actual place	Year	OPT	
Plastic floor covering	Spending time on plastic floor covering	In house Yes // No Other (kindergarten) Yes // No	OPT	
Time spent inside house/ use of electronic devices in house		Number of hours in the household	OPT	Type of activities inside, type of electronic devices used

Variable	Description of variable	Unit score	OPT / OBL	Comment
Paint	Recently painted house (indoor)		OPT	
Consumption of plastic bottle water	Type of drinking water consumed and frequency	Bottle water Tap water Water coolers Others	OBL	A focus on a plastic bottle water consumption
Use of plastic crockery/dishes		Ordinal variable (Never, rarely, sometimes, often, very often, every day)	OBL	
Use of plastic electric kettles		Ordinal variable (Never, rarely, sometimes, often, very often, every day)	OBL	
Use of food mixers		Ordinal variable (Never, rarely, sometimes, often, very often, every day)	OBL	
Use of containers for heating foods in microwave		Ordinal variable (Never, rarely, sometimes, often, very often, every day)	OBL	
Use of plastic (polycarbonate) baby bottles		Yes // No	OBL	Only relevant for young children
Food consumption	Type of food consumed and frequency of consumption (packed food and/or ready meals, canned food, canned beverage)	Modality to be defined according to the completeness of the data from the questionnaire (type of food and frequency of consumption)	OPT	Focus on food likely to be more contaminated to the bisphenols (offal, seafood, dried vegetables, mushrooms…) This could be based on the detailed food questionnaire and frequencies of consumption set up from WP7.
Occupation	Occupations in relation with a potential high risk of exposure to bisphenols	Yes // No	OBL	Cashiers, plastic industry workers, health care workers, food service workers,

Variable	Description of variable	Unit score	OPT / OBL	Comment
Use of thermal papers	Contact with different type of thermal papers	Modality to be defined according to the completeness of the data from the questionnaire (type of thermal papers contacts and frequencies)	OPT	Examples of thermal papers products: receipts from cash registers, transportation tickets such as for trains and airlines, entertainment tickets, parking tickets, lottery tickets, self-adhesive labels and tags, fax paper
Use of plastic medical devices	Transfusion/infusion bag, hemodialysis, newborn incubators	Yes // No (for each variable taken individually)	OPT	As a patient
Have a dental sealant		Yes // No Year of application	OPT	
Use of glasses and/or contact eye lenses		Ordinal variable (Never, rarely, sometimes, often, very often, every day) for glasses and contact eye lenses	OPT	
Use of cosmetics		Yes // No	OPT	The question (unit scores) should be extend to the frequency of use and modes

11.3 Per-/Polyfluorinated Compounds

11.3.1 General Exposure Levels

- What is the current exposure of the EU population to PFASs and do they exceed Guidance values (reference and HBM values), where available?

Variable	Description of variable	Unit score	OPT / OBL	Comment
PFCA, PFSA Cat. A[b]	PFOS,PFDS *PFBA, PFPeA, PFHxA, PFHpA*, PFOA, PFNA, PFDA, PFUnDA, PFDoDA, PFTeDA, *PFBS, PFHxS, PFHpS*	Serum concentration [µg/L]	OBL	Breast milk (exposure to young children through breast-feeding) and urine (short-chain PFCAs and PFSAs) may also be a feasible matrix[a].

Variable	Description of variable	Unit score	OPT / OBL	Comment
FASAs Cat. B	FOSA, N-EtFOSA, EtFOSE, N-MeFOSA, EtFOSAA	Serum concentration [µg/L]	OBL	[a]
PAP, PFPAs Cat.C	6:2diPAP,8:2 diPAP, 6:2 mono PAP, 8:2 mono PAP; PFHxPA, PFOPA, PFDPA, ADONA	Serum concentration [µg/L]	OBL	Breast milk may be the desired matrix and sometimes urine may also be a feasible matrix[a].
PFAS (all)	PFCA, PFSA, FASAs, PAP, PFPAs	Serum concentration [µg/L]	OBL	Mixture effects should be considered. Uncertainty regarding the total PFASs exposure has to be considered. TOF or oxidizable fractions methods should be planned and performed, in order to be able to quantify also the so far unidentified compounds. LC-MS-MS is the most common technique used for determination of PFASs in any kind of matrix. Sample preparation with a protein precipitation step followed by either on-line or off-line clean-up is also the most commonly used.
LOQ		Serum concentration [µg/L]	OBL	A low limit of quantification (LOQ) is necessary for gen pop, well below the LOQ for occupational exposure.
Type of participant		Background population // occupational exposed // pregnant women // etc...	OBL	
Sampling	Serum sampling conditions (time of collection etc.)		OBL	
Sample type		spot- // morning- // 24 h-urine	OBL	

[a] Inter-laboratory comparisons and certified reference materials are so far only available for the most abundant PFCAs and PFSAs.
[b] compounds in *italic* are presumably Cat B

11.3.2 Time Trends

- Has restriction of PFOS according to the POP Regulation led to a reduction in exposure, especially for children?
- What is the impact of the EC restriction of manufacturing, marketing and use of PFOA under REACH?
- Can differences in PFASs profiles be observed in different population groups and time periods?
- (How do PFAS pattern change after regulatory action?)

Variable	Description of variable	Unit score	OPT / OBL	Comment
Year of regulation for PFOS	Year when regulation became effective for PFOS	Year	OBL	
Year of regulation for PFOA	Year when regulation became effective for PFOA	Year	OBL	REACH restriction published in 2017; no country specific regulation in place
Type of uses concerned by regulation	Determinants of exposure		OBL	In relation with the determinants of exposure
PFAS pattern	Pattern of PFAS exposure	Year	OBL	

11.3.3 Geographical Comparisons

- Do the exposure levels differ significantly between the countries?
- Can countries be grouped according to similarity in concentration levels (e.g. northern, southern…)?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of participant/study	Name of Country	OBL	
PFAS exposure	Exposure pattern	Name of Country	OBL	

11.3.4 Exposure Determinants

- Is exposure driven by diet, consumer exposure, occupation or environmental contamination?
- What are the PFASs levels and health effects in vulnerable population groups?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Personal:				
Sex	Sex of participant	0=male 1=female	OBL	
Number of children	Number of breastfed children including child birth date		OBL	
Breastfeeding history	Duration of breastfeeding for each child including, no of months exclusive breastfeeding and		OBL	

Variable	Description of variable	Unit score	OPT / OBL	Comment
	no. of months partial breastfeeding			
Menopausal status		Postmenopausal = yes //no. If yes, 1=natural 2=hormone therapy 3=surgical 4=unknown	OPT	
If yes, age of menopause		Age in years	OPT	
Menstrual Cycle	Length (in days) including information on contraceptives (spiral) that affect menstruation.		OPT	
Blood loss	History of blood donation and blood loss.		OPT	
Age	Age of the participant at time of sampling	Completed years	OBL	
Ethnicity	Ethnicity of participant in respective country		OPT	Also possible to look at migration status
Height	Height of participant	cm	OBL	
Weight	Weight of participant	kg	OBL	
BMI	BMI of participant	kg/m^2	OPT	Only needed if height and weight cannot be provided
Socio-economic status	Socio-economic status of participant	1=low 2=medium 3=high	OBL	(for children, parent or household)
Education	Education of participant, ISCED scale	1=low 2=medium 3=high	OBL	(for children, parent or household)
Occupation	Occupation in specific workplaces: e.g. Fluoropolymer/ elastomer production; chrome plating, AFFF in firefighting foams, production and use of lubricants, printing inks, pesticides…		OPT	
Personal care and health care factors:				
Use of cosmetics	Specific type e.g. water resistant make up, nail polish,..	Frequency	OPT	
Medicine/medical equipment			OPT	
Lifestyle and food factors:				
Drinking water	Origin	Amount/day	OBL	

Variable	Description of variable	Unit score	OPT / OBL	Comment
Diet	Fish, seafood	Frequency	OPT	
Take away/ Fast food		Frequency	OPT	
Baking paper, specific food packaging		Frequency	OPT	
Microwave popcorn		Frequency	OPT	
Housing/ environment factors:				
Type of flooring			OPT	
Use of furniture polish/ specific cleaning agents/ impregnation/ coating agents/ paints		Frequency	OPT	
Use of ski wax		Frequency	OPT	

11.3.4.1 Vulnerable Groups General Exposure Levels

- Vulnerable population groups (depending on available data):
  - Newborns/Children …. Available
  - Pregnant/breastfeeding women ….. Available
  - Elderly ….. Limited
- What are PFASs levels in vulnerable populations?
- Do subgroups of vulnerable populations belong among highly exposed population groups?
- What are determinants of PFASs exposure in vulnerable populations and are they different from general population? (in cooperation with RQ on exposures)

Data needed – newborns

Variable	Description of variable	Unit score	OPT / OBL	Comment
PFASs	Levels of PFASs	ng/mL	OBL	Levels in cord blood
LODs	Levels of detection	ng/mL	OBL	
Gender	Male/Female		OBL	
Gestational age at birth		Weeks	OBL	
Birth weight		Grams	OPT	
Maternal age	Age at delivery	Years	OBL	
Ethnicity	Categories		OPT	Depending on country – to be specified
Education of mother/parents	Categorised education	Levels of education	OBL	Depending on country – to be specified
Parity		Primiparous // multiparous	OPT	
BMI of mother	Maternal BMI before pregnancy	kg/m^2	OPT	

Data needed – children

Variable	Description of variable	Unit score	OPT / OBL	Comment
PFASs	Levels of PFASs	ng/mL	OBL	Levels in blood serum
LODs	Levels of detection	ng/mL	OBL	
Gender	Male/Female		OBL	
Age	Age of a child at sampling	Months/years	OBL	Months in infants, years in older children
Breastfeeding	Was child breastfed?	Yes // No	OBL/OPT?	
Ethnicity	Categories		OPT	Depending on country – to be specified
Education of parents	Categorised education	Levels of education	OBL	Depending on country – to be specified

Data needed – pregnant women/mothers

Variable	Description of variable	Unit score	OPT / OBL	Comment
PFASs	Levels of PFASs	ng/mL	OBL	Levels in breast milk/blood serum
LODs	Levels of detection	ng/mL	OBL	
Age	Age at sampling	Years	OBL	
Ethnicity	Categories		OBL	Depending on country – to be specified
Education	Categorised education	Levels of education	OBL	Depending on country – to be specified
Parity		Primiparous // multiparous	OPT	
BMI of mother	Maternal BMI before pregnancy	kg/m^2	OPT	

11.3.4.2 Contaminated areas

- Is exposure driven by diet, consumer exposure, occupation or environmental contamination?
- Which areas and environmental media in Europe are contaminated with PFASs?
- What is the cause of contamination? Which factors contribute most to elevated body burdens?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Place of birth		Rural // Urban	OPT	
Place of residence		Rural // Urban	OBL	
Place of residence near a fluorochemical industrial facility		Yes // No	OBL	
Place of residence near civilian airports, military bases, wastewater treatment facilities, or fire fighting training facilities		Yes // No	OBL	To be specified the type of exposure source
Place of residence near agricultural areas characterised by the use of soil conditioners		Yes // No	OBL	To be specified the type of soil conditioner
Years of residence	Years of residence in actual place	Year	OBL	
Consumption of tap water		L/day	OBL	
Use/Consumption of groundwater or surface water	Use of groundwater or surface water for drinking or cooking	Yes // No	OBL	
Consumption of locally produced food		Yes // No	OBL	To be specified the type of consumed food
Consumption of own grown vegetables and own raised livestock		Yes // No	OBL	To be specified the type of consumed food
Consumption of fish and seafood from a local body of water		Yes // No	OBL	

11.4 Flame Retardants (FR)

Fundamental variables: these variables are necessary for interpretation of multiple research questions and are not repeated within individual sections below

Variable	Description of variable	Unit score	OPT / OBL	Comment
Levels of Cat. A FRs	Conc. of Cat. A FRs in human matrices	Serum or milk or other concentration (ng/g lipid weight)	OBL	Information from other WPs necessary to identify relevant matrices
Levels of Cat. B FRs	Conc. of Cat. B FRs in human matrices	Serum or milk or other concentration (ng/g lipid weight)	OBL	Information from other WPs necessary to identify relevant matrices
Levels of Cat. B metabolites	Conc. of Cat. B FR metabolites in human matrices	Urine concentration (ng/ml adjusted urine)	OBL	Information from other WPs necessary to identify relevant metabolites
Levels of Cat. C FRs	Conc. of Cat. C FRs in human matrices	Serum or milk or other concentration (ng/g lipid weight)	OBL	Information from other WPs necessary to identify relevant matrices
Levels of Cat. C metabolites	Conc. of Cat. C FR metabolites in human matrices	Urine concentration (ng/ml adjusted urine)	OBL	Information from other WPs necessary to identify relevant metabolites
Year	Year of sampling	Year	OBL	
Country	Country of participant/study	Name of Country	OBL	
Regulation for the substance	Type of regulation or extent of restriction of a substance in its respective country		OBL	Ideas on quality/type/extent of regulation? CGL: Distinguish between annex XIV, SVHC candidates for authorisation, restrictions
Year of regulation for the metabolite/ substance	Year when regulation became effective for the metabolite/ substance	Year	OBL	
Type of participant		Background population // occupational exposed	OBL	”

11.4.1 General Exposure Levels

- What are average levels of FR exposure for European population?

Variable	Description of variable	Unit score	OPT / OBL	Comment
LOD	Limit of detection for each FR	Serum or milk or other concentration (ng/g lipid weight)	OBL	
LOQ	Limit of quantification for each FR	Serum or milk or other concentration (ng/g lipid weight)	OBL	
LOD	Limit of detection for each FR metabolite	Urine concentration (ng/ml adjusted urine)	OBL	
LOQ	Limit of quantification for each FR metabolite	Urine concentration (ng/ml adjusted urine)	OBL	
Urine adjustment method	Creatinine or specific gravity or other adjustment		OBL	
Levels of Cat. A metabolites	Conc. of Cat. A FR metabolites in human matrices	Urine concentration (ng/ml adjusted urine)	OPT	Information from other WPs necessary to identify relevant metabolites
Detection of Cat. D or E FRs or metabolites	Detection frequency or presence/absence of Cat. D and E FRs	1=detected 2=not detected	OPT	

11.4.2 Time Trends

- Do we see a shift in temporal trends of exposure in conjunction with regulatory measures? E.g. EU restrictions on PBDEs, voluntary industry phase-out of PBDEs, Stockholm Convention listing of HBCDD, PBDEs.
- Are different time trends observed for different PBDE congeners?
- Do time trends differ between countries?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Month	Month of sampling	Month	OPT	
Day	Day of sampling	Day	OPT	

11.4.3 Geographical Comparisons

- Do FR exposure levels differ between countries?
- Do different regulations in different countries influence exposure levels to FRs?
- Can we generalise regional groupings of geographic differences, e.g., Northern Europe/Scandinavia, Eastern Europe, etc.?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of participant/study	Name of Country	OBL	
Country-level variables	Variables on country level which are expected to contribute to differences in exp.		OPT	

11.4.4 Exposure Determinants

- What are the determinants of exposure for FRs?
- Identification of highly exposed groups – infants/toddlers? People with certain dietary patterns? Identify if occupationally exposed groups differ significantly from general population.
- Is there a significant difference between exposure levels in adults and children?
- Is there a significant occupationally exposure population?
- How do exposure levels relate to HBM guidance values or other intake thresholds?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Age	Age of the participant at time of sampling	Years	OBL	
Sex	Sex of participant	1=male 0=female	OBL	
Degree of urbanisation	Urban/suburban/rural /remote	1=urban 2=suburban 3=rural 4=remote	OPT	
Height	Height of participant	cm	OBL	
Weight	Weight of participant	kg	OBL	
BMI	Body mass index	kg/m^2	OPT	Only if height and weight cannot be provided
General nutrition information			OPT	Vegetarian, high fish consumption, local food, etc..

Variable	Description of variable	Unit score	OPT / OBL	Comment
Maternal transfer			OPT	Linked information if HBM from mother-child cohorts
Breastfeeding			OPT	
Occupation			OPT	
Socio-economic status	Socio-economic status of participant	1=low 2=medium 3=high	OPT	(for children, parent or household)
Indoor environment				Electronics, furnishings, renovations, age of residence
Time spent inside house/ use of electronic devices		Number of hours	OPT	Type of activities inside, type of electronic devices used

11.5 Cadmium (Cd) & Chromium (Cr)

11.5.1 General Exposure Levels

- What is the current (last 5 years) exposure to Cd and Cr(VI) of the European population?
- What is the level of exposure, environmentally and occupationally relevant to Cr(VI) in the EU population?
- Are the overall exposure levels (in different population groups) above any health-relevant assessment levels (HBM guidance values, TDI)?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Blood Cd	Level of Cd in blood	µg/L	OBL	
Urinary Cd	Level of Cd in urine	µg/L	OBL	
Blood Cr	Level of Cr in whole blood	µg/L	OBL	
Urinary Cr	Level of Cr in urine	µg/L	OPT	
Red cells Cr VI	Level of Cr VI in red cells	µg/g	OBL	
Exhaled Breath Condensate (EBC)	Level of Cr VI in EBC	µg/L	OPT	
Breast milk Cd	Level of Cd in breast milk	µg/L	OPT	
Urine sampling	Type of urine sampling	1=spot 2=24-h	OBL	
Creatinine	Level of creatinine in urine	µmol/L g/L	OBL	Obligatory in case of spot urine samples

Variable	Description of variable	Unit score	OPT / OBL	Comment
Specific gravity	Specific gravity of urine	g/L	OPT	Alternative to creatinine
LOD blood Cd	Level of Cd detection in blood	µg/L	OBL	
LOQ blood Cd	Level of Cd quantification in blood	µg/L	OBL	
LOD urine Cd	Level of Cd detection in urine	µg/L	OBL	
LOQ urine Cd	Level of Cd quantification in urine	µg/L	OBL	
LOD blood Cr	Level of Cr detection in blood	µg/L	OBL	
LOQ blood Cr	Level of Cr quantification in blood	µg/L	OBL	
LOD urine Cr	Level of Cr detection in urine	µg/L	OBL	
LOQ urine Cr	Level of Cr quantification in urine	µg/L	OBL	
LOD red blood cells Cr	Level of Cr detection in red blood cells	µg/g	OBL	
LOQ red blood cells Cr	Level of Cr quantification in red blood cells	µg/g	OBL	
LOD EBC	Level of Cr VI quantification in EBC	µg/L	OBL	
LOQ EBC	Level of Cr VI quantification in EBC	µg/L	OBL	
LOD breast milk Cd	Level of Cd detection in breast milk	µg/L	OPT	
LOQ breast milk Cd	Level of Cd quantification in breast milk	µg/L	OPT	

11.5.2 Time Trends

- Is there a significant time trend of Cd and Cr(VI) levels in existing population studies?
- Has the regulation under REACH had the favourable impact like a reduction of GM/median concentrations?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Year	Year of sampling	Year	OBL	

11.5.3 Geographical Comparisons

- Do the exposure to Cd and Cr(VI) differ significantly between countries and population groups?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of sampling	Name of country	OBL	
Region	Geographical region	1=North 2=South 3=West 4=East	OBL	Variable to be generated based on the country of sampling
Place of residence	Living place of participant	Town or GPS code	OPT	

11.5.4 Exposure Determinants

- What are the groups at risk? (Considering: life-style, nutritional status and genetic background, gender, age; postmenopausal women, elderly, regions with elevated levels in the environment, occupational settings, co-exposure to chemical mixtures).
- Provide information on the quantitative relationship between HBM Cr(VI) concentration and dermatological risks and cancer risks.
- Understand the role of factors as food and beverage intake, smoking, exercise, habits on HBM data of Cr(VI).

Variable	Description of variable	Unit score	OPT / OBL	Comment
Population	Type of population	1=newborns 2=children 3=adolescents 4=adults 5=occupational	OBL	
Age	Age of participant at time of sampling	Years	OBL	
Gender	Gender of participant	1=male 2=female	OBL	
SES	Socio-economic status of participant	1=low 2=medium 3=high	OBL	In case of a study population of children: parent or household

Variable	Description of variable	Unit score	OPT / OBL	Comment
Education	Education of participant	1=primary or lower 2=secondary 3=university or higher	OBL	In case of a study population of children: parent or household
Area	Type of area of residence	1=city/town 2=suburban 3=countryside	OBL	Can be extracted from GPS code or town of residence if provided
Traffic	Density of traffic in the residential area	1=no traffic 2=light traffic 3=intense traffic	OBL	o Obtained from GIS databases OR o As perceived by the subjects
Height	Height of participant	cm	OBL	
Weight	Weight of participant	kg	OBL	
Smoking	Smoking habit of participant	1=non smoker 2=smoker 3=former smoker	OBL	
Passive smoking	Passive smoking of participant	1=no 2=yes 3=former	OBL	
Cotinine	Cotinine level in urine	µg/L	OPT	To better assess smoking
Local food	Consumption of local food	1=no 2=yes	OBL	
Seafood	Consumption of seafood	Decide based on available data	OBL	
Tattoo	Presence of tattoo	1=no 2=yes (more than 2 tattoos)	OPT	
Dietary items				Select from what is available
Jewellery	Use of jewellery	Decide based on available data	OPT	
Level of essential elements	Level of essential elements (Zn, Se, Cu, Fe,…) in biological matrices (preferably blood)	µg/L	OPT	Select from what is available

11.6 PAHs and Air Pollution

11.6.1 Time Trends

- Is there a significant change of the regulated data-rich substance levels (GM/median) in the population (both in terms of general population and in terms of susceptible population sub-groups such as children) over the last ten years?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Year	Year of sampling	Year	OBL	Changes in PAHs emissions will be reflected
Month	Month of sampling	Month	OBL	This is important to identify seasonal differences
Regulation for the substance	Type of regulation or extent of restriction of a substance in its respective country	Year of implementation	OBL	Changes in the regulation will be reflected in environmental PAHs levels

11.6.2 Geographical Comparisons

- Do the exposure levels of data-rich substances differ significantly between countries? Do spatial and temporal analyses of available data reveal hot spots or time patterns of exposure? What are the main reasons for differences in exposure?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of participant/study	Name of Country	OBL	Different emission patterns and meteorological conditions affect PAHs exposure levels
Traffic fleet	Traffic fleet composition (diesel, gasoline, LPG, electric	Percentage of fleet components	OPT	Traffic fleet composition affects PAHs presence in ambient air
Biomass use	Biomass burning for space heating	Percentage of biomass burning for space heating	OPT	Use of biomass burning increases levels of PAHs in ambient air
Hot spots	Hot spots associated with significant PAHs emissions	Yes // No	OPT	Presence of specific hot spots increases environmental PAHs occurrence

11.6.3 Exposure Determinants

- What are the most important determinants of aggregate exposure (e.g. are PAH and benzene exposure primarily driven by lifestyle factors, by environmental factors or by workplace environments?)
- What are the high exposure groups? Do available HBM data reveal differences in sub-groups that depend on gender, age group, socio-economic status, etc.?
- Are the overall exposure levels in the general population, children, and pregnant women above any health-relevant assessment levels (reference dose or HBM guidance values)?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Smoking	If someone is smoking	Yes // No	OBL	Smoking affects the levels of PAHs metabolites measured in urine
Consumption of smoked food	Estimate of the amount of smoked food intake on a daily or weekly basis	Frequency in the week, amount of smoked food	OPT	Consumption of smoked food is a key determinant of the levels of PAHs metabolites measured in urine
Proximity to heavily trafficked road	Describes the distance from a heavily traffic road	Km	OPT	Proximity to heavily trafficked road increases the levels of PAHs metabolites measured in urine
Proximity to incineration	Describes the distance from an incineration	Km	OPT	Proximity incinerators increases the levels of PAHs metabolites measured in urine
Proximity to petrochemical processing facilities	Describes the distance from petrochemical processing facilities	Km	OPT	Proximity to petrochemical processing facilities increases the levels of PAHs metabolites measured in urine
Indoor use of coal/biomass burning	Describes the use of biomass indoors	Yes // No	OPT	Indoor use of coal/biomass burning increases the levels of PAHs metabolites measured in urine
Occupational activities	Activities that contribute to PAHs exposure (road construction, petrochemical industry, cooking)	Yes // No	OPT	
Age	Completed years	Years	OBL	Considerations about urinary volumes and creatinine excretion are reflected in the levels of PAHs metabolites measured in urine

11.6.4 General Exposure Levels

- How high is the current (year 2012 or more recent) exposure (both external and internal) of the EU population to data-rich substances?

Variable	Description of variable	Unit score	OPT / OBL	Comment
1-hydroxypyrene	Metabolite of pyrene, indicator of exposure to PAHs mixture	Urine concentration [µg/L]	OBL	This is the metabolite usually measured in urine for PAHs assessment
3-hydroxybenzo[a]pyrene	Metabolite of benzo[a]pyrene, indicator of the most toxic and only one regulated PAH	Urine concentration [µg/L]	OPT	This is the metabolite usually measured in urine for benzo[a]pyrene, the most toxic and the only regulated PAH in ambient air
LOD	Level of detection for each metabolite	Urine concentration [µg/L]	OBL	Standardised procedure for reporting concentrations lower than LOQ or lower than LOD (e.g. replaced by LOQ/2 and LOD/2, respectively) is currently developed in the statistical analysis plan
LOQ	Level of quantification for each metabolite	Urine concentration [µg/L]	OBL	Standardised procedure for reporting concentrations lower than LOQ or lower than LOD (e.g. replaced by LOQ/2 and LOD/2, respectively) is currently developed in the statistical analysis plan
Creatinine level	Level of creatinine in urine	Urine concentration [g/L]	OBL	Standard of use is to be determined in the statistical analysis plan
Sampling	Urine sampling conditions (time of collection etc.)	Time of the day	OPT	There is variability within the day that is important to be captured
Sample type	Indicates the time regime of sampling	spot- // morning- // 24 h-urine	OPT	There is variability within the day that is important to know to what extent has been captured

11.7 Occupational Exposure / ANILIN Family including MOCA

11.7.1 General Exposure Levels

- What are the background levels of general (adult) population to aniline, paracetamol, o- and p-toluidine? Identify reference values.
- Levels of diisocyanate metabolites (like MDA, TDA) in the urine of general population? Assumption is that levels are <LOD when the reference value can be set equal to the detection limits.
- Effect of smoking on urinary levels of anilines in general population?

Variable	Description of variable	Unit score	OPT / OBL	Comment
aniline	U-aniline U-N-Acetyl-4-aminophenol (Hb-adducts)	urine concentration	OBL	also U-asetanilide has been as a marker for aniline, good option but less used.
o-toluidine	total urinary o-toluidine (free and conjugated) (Hb-adducts)	urine concentration	OBL	
p-toluidine	total urinary p-toluidine (free and conjugated) (Hb-adducts)	urine concentration	OBL	
diisocyanates	U-MDA U-TDA	urine concentration	OBL	these are likely to be below detection limits in general population
smoking	smoking	yes/no	OBL	

11.7.2 Time Trends

- Effect of planned restriction of diisocyanates on the levels of diisocyanate metabolites (like MDA, TDA) in the urine of workers.

Variable	Description of variable	Unit score	OPT / OBL	Comment
time	time before 2019 and after 2019	year of sampling	OBL	Restriction is only under preparation but is estimated to become enforced in 2019. Baseline data should be analysed now.

11.7.3 Geographical Comparisons

- Differences between different countries in the general population urinary aniline, N-Acetyl-4-aminophenol, o-, and p-toluidine levels.

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country of the study	name of the Country	OPT	

11.7.4 Exposure Determinants

- What is the exposure of workers in different occupations to aniline, diisocyanates, o- and p-toluidine?
- Effects of smoking on the aniline compound levels in urine?
- What is the exposure of different age groups to paracetamol (N-Acetyl-4-aminophenol)?

Variable	Description of variable	Unit score	OPT / OBL	Comment
Occupation	Occupation	Occupation according to ISCO08 system or similar	OBL	
smoking	smoking	yes/no	OBL	
sex	M/F	M/F	OPT	
age (in the case of paracetamol)	Age of the participant at time of sampling	Completed Years	OPT	

11.8 Chemical Mixtures

The statistical analysis of existing HBM data of chemical mixtures is explorative in nature, with the primary question “What are the HBM mixture levels in the European population?”.

An extensive analytical plan has been developed within HBM4EU work package 15 “Mixtures HBM and human health risk”, described in Annual Deliverable AD15.3.

Specific (research) questions that are addressed in task 15.1 are:

- How can we rank/order individuals on the basis of low-high body burdens to mixtures?
- What patterns can we observe amongst body burdens of different substances within individuals? (e.g. are people with high levels for some substances more likely to be high on others as well?)
- Are such patterns indicative for specific sources or pathways of exposures?
- Can we identify hotspots or risk groups with high body burdens of mixtures?
- Can we develop aggregate/hybrid indicators that encapsulate toxicity of the mixture in a meaningful way? (e.g. hazard index approach or CAGs.)
- 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 AD15.3. 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 and OMICS studies.

Main determinants that will be included in these analyses are listed below. The list of determinants will be further expanded based on which mixture data becomes available in the HBM4EU repository, incorporating the determinants listed for each priority group in this deliverable.

Variable	Description of variable	Unit score	OPT / OBL	Comment
Country	Country name		OBL	
Region	United Nations geoscheme for Europe	1=North 2=South 3=West 4=East	OBL	
Year	Year of sampling	Year	OBL	
Month	Month of sampling	Month	OPT	
Day	Day of sampling	Day	OPT	
Sex	Sex of participant	Male=0 Female=1	OBL	
BMI	Body mass index	kg/m^2	OBL	
Age	Age of the participant at time of sampling	Completed years	OBL	

12 Appendix

Correlation between urinary osmolality and creatinine

Two datasets with existing data on the osmolality and creatinine concentrations in the same urine samples were used:

Dataset 1: 24 hour pools from 129 children and adolescents aged 6-20 years (65 males)

Dataset 2: first morning voids, spot urine and 24 hour pools from 5 adult persons > 40 y (2 males), in total 41 urine samples.

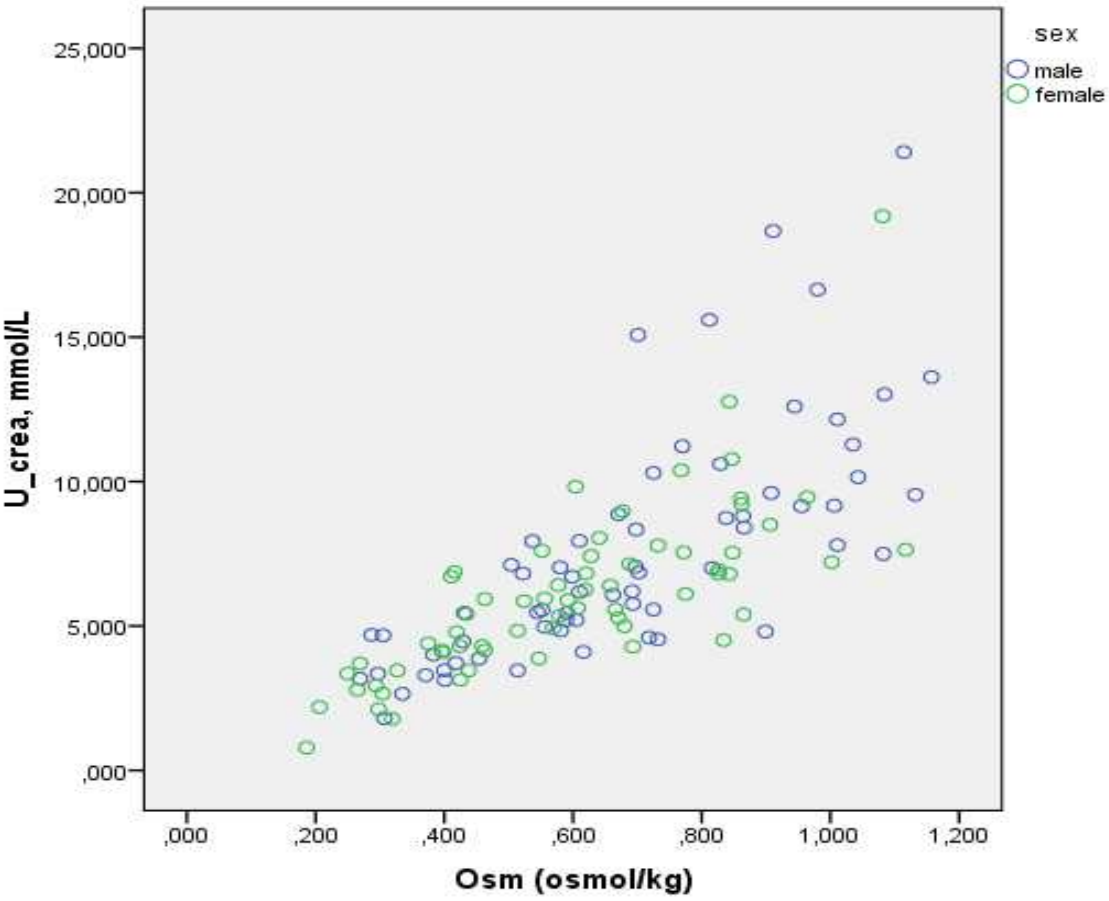


Figure 1: Osmolality and U-creatinine in 24-hour urine pools from 129 children and adolescents (data set 1)

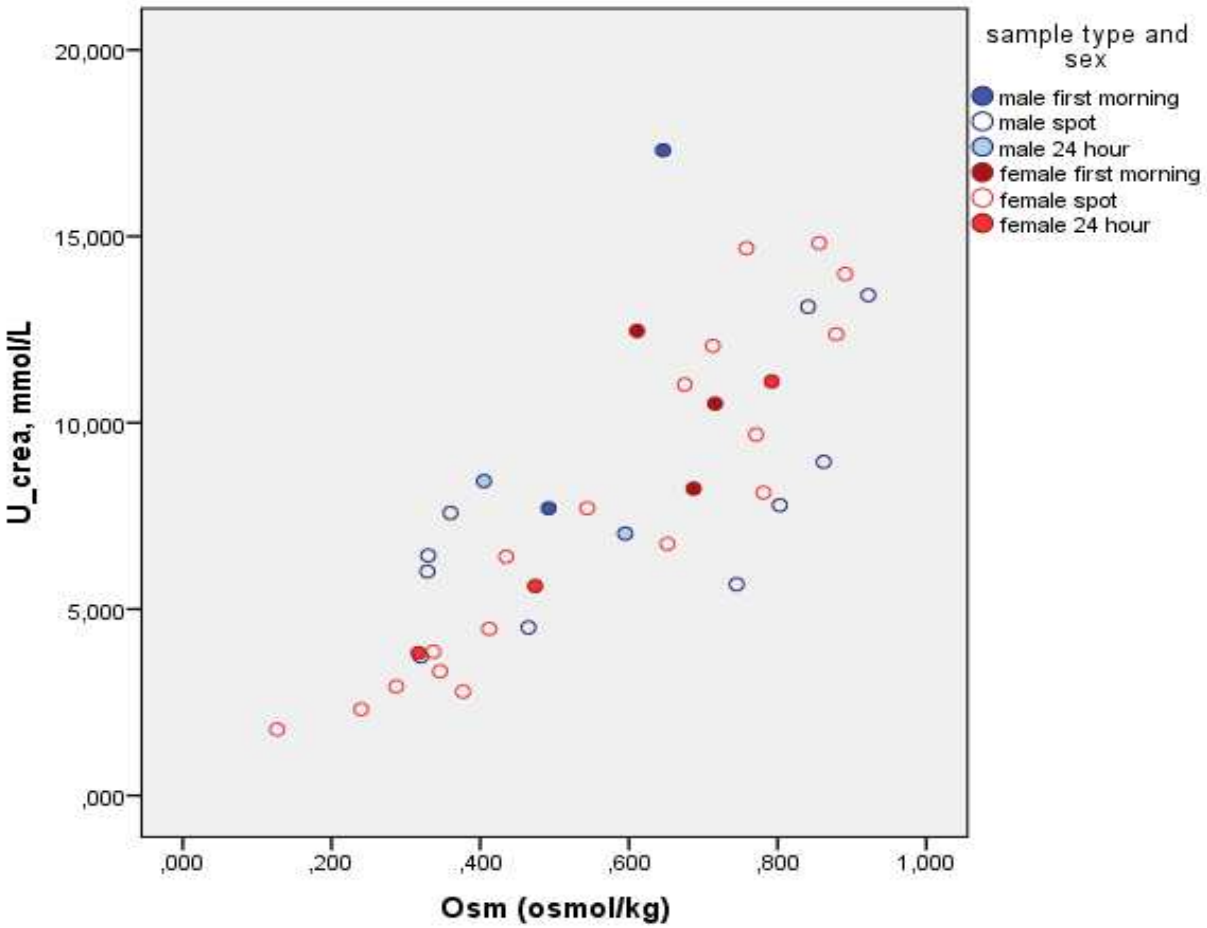


Figure 2: Osmolality and U-creatinine in mixed urine samples (spot, first morning, 24 hour pool) from five adults (data set 2)

The descriptive statistics of u-crea concentrations and urinary osmolality for the two data sets are given in Table 1a-b, respectively. As some laboratories/studies use the unit mmol/L and others use g/L for u-crea both are presented. Likewise osmolality is presented as both Osm/kg and mOsm/kg.

Table 1a
Descriptive statistics: Data set 1 - 24-h pools 129 children and adolescents

	Minimum	Maximum	**Mean**	Std. Deviation
Urine volume, mL (g)	2863	4214	**1157.8**	719.8
U-creatinine, mmol/L	0.79	21.40	**6.83**	3.57
U-creatinine, g/L	0.09	2.42	**0.77**	0.40
Osmolality, Osm/kg)	0.186	1.157	**0.643**	0.237
Osmolality mOsm/kg	186	1157	**643**	237

Table 1b
Descriptive Statistics: Data set 2 - mixed urine samples from 5 adults

	N	Minimum	Maximum	**Mean**	Std. Deviation
U-creatinine, mmol/L	38	1.77	17.31	**8.12**	4.03
U-creatinine, g/L	38	0.20	1.96	**0.92**	0.45
Osmolality, Osm/kg)	41	0.127	0.922	**0.552**	0.227
Osmolality mOsm/kg	41	127	922	**552**	227

Linear regression between u-crea (mmol/L) and u-osmolality (Osm/kg) were explored using SPSS with and without adjustment for sex and age (dataset 1). A significant association between urinary u-crea and u-osmolality were observed as expected as both parameters are reflecting the urinary concentration/dilution. No significant association of sex with u-crea was observed but age was significantly positively associated with u-crea in dataset 1. In order to determining the empirical relationship between u-osmolality and u-crea we performed bivariate regression (Table 3). The relationship between u-osmolality and u-crea differed slightly between the two dataset, which may be due to the difference in age (children vs adult). Nevertheless we also merged the two data sets in order to obtain the overall associations. As data set 1 was bigger than data set 2 the obtained overall associations are likely biased towards the younger age.

Based on the observed bivariate relationship between the u-osmolality and u-crea concentration we converted the measured u-osmolality (Osm/kg) to an estimated u-crea (mmol/L) using the following equations:

Data set 1: calculated u-crea (mmol/L) = -0.4 + 11.2 x measured u-osmolality (Osm/kg)

Data set 2: calculated u-crea (mmol/L) = -0.18 + 14.5 x measured u-osmolality (Osm/kg)

Data set 1+2: calculated u-crea (mmol/L) = -0.04 + 11.4 x measured u-osmolality (Osm/kg)

Subsequently, we compared this calculated u-crea to the actual measured u-crea to explore the feasibility of using a conversion factor to transform a measured u-osmolality into a meaningful calculated u-crea.

Table 2: Linear relationship between u-crea (mmol/L) and u-osmolality (Osm/kg)

	intercept	B	95% CI	p-value
Data set 1:				
Adjusted for age	-6.12	13.4	11.7 - 15.1	<0.001
Bivariate regression	-0.40	11.2	9.5 – 13.0	<0.001
Data set 2:				
Bivariate regression	-0.18	14.5	10.7 – 18.2	<0.001
Data set 1 +2:				
Bivariate regression	-0.04	11.4	9.7 – 13.1	<0.001

Table 3: Comparison of measured and calculated U-creatinine

	Minimum	Maximum	**Mean**	Std. Deviation
Data set 1:				
Measured U-creatinine, mmol/L	0.79	21.40	**6.83**	3.57
Calculated* U-creatinine, mmol/L	1.69	14.10	**7.87**	2.68
Data set 2:				
Measured U-creatinine, mmol/L	1.77	17.31	**8.12**	4.03
Calculated# U-creatinine, mmol/L	1.66	13.16	**7.81**	3.29
Data set 1 +2:				
Measured U-creatinine, mmol/L	0.79	21.40	**7.12**	3.70
Calculated** U-creatinine, mmol/L	1.41	13.15	**7.04**	2.70

Calculated U –crea (mmol/L) was calculated from measured Osm/kg based on the regression line: *U-crea = -0.40 + 11.2*Osm/kg, #U-crea =-0.18+ 14.5*Osm/kg, **U-crea=-0.04+11.4*Osm/kg

In data set 1 the calculated u-crea concentration slightly overestimated the actual u-crea concentration while in data set 2 the calculated u-crea concentration slightly underestimated the actual u-crea concentration. Although we obtained similar means between measured and calculated u-creatinine when combining the two data sets this is probably masking differences in the associations in different age groups.

Exploring the ratio between measured and calculated u-crea as a function of the measured concentration of u-crea showed a clear decrease in the ratio with increasing measured u-crea indicating that the relationship between u-osmolality and u-crea are not linear over the range of u-crea concentrations; at low concentrations of u-crea a calculated u-crea based on the u-osmolality overestimate the u-crea concentration and at high concentrations it is underestimated (Figure 3).

Nevertheless, when we used respectively the measured and the calculated u-crea concentrations to adjust measured monoethylphthalate (MEP) and bisphenol A (BPA) concentrations we obtained a fairly good agreement and correlation (Figure 4) and population means that were not statistically significantly different by a paired T test (Table 4).

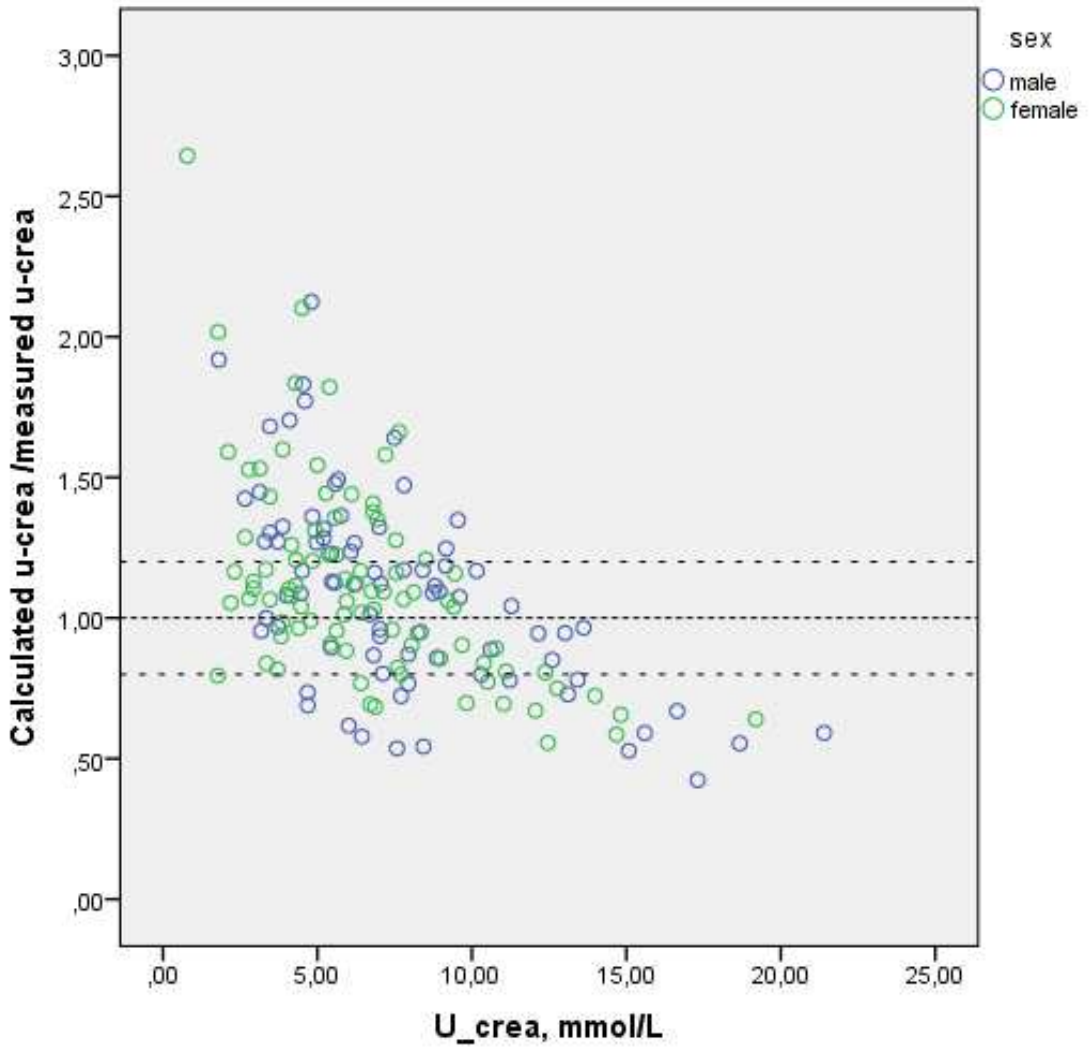


Figure 3: The ratio between calculated and measured u-crea (mmol/L) plotted against the measured u-crea concentration (Data set 1+2 combined). Dashed lines at 0.8, 1, and 1.20

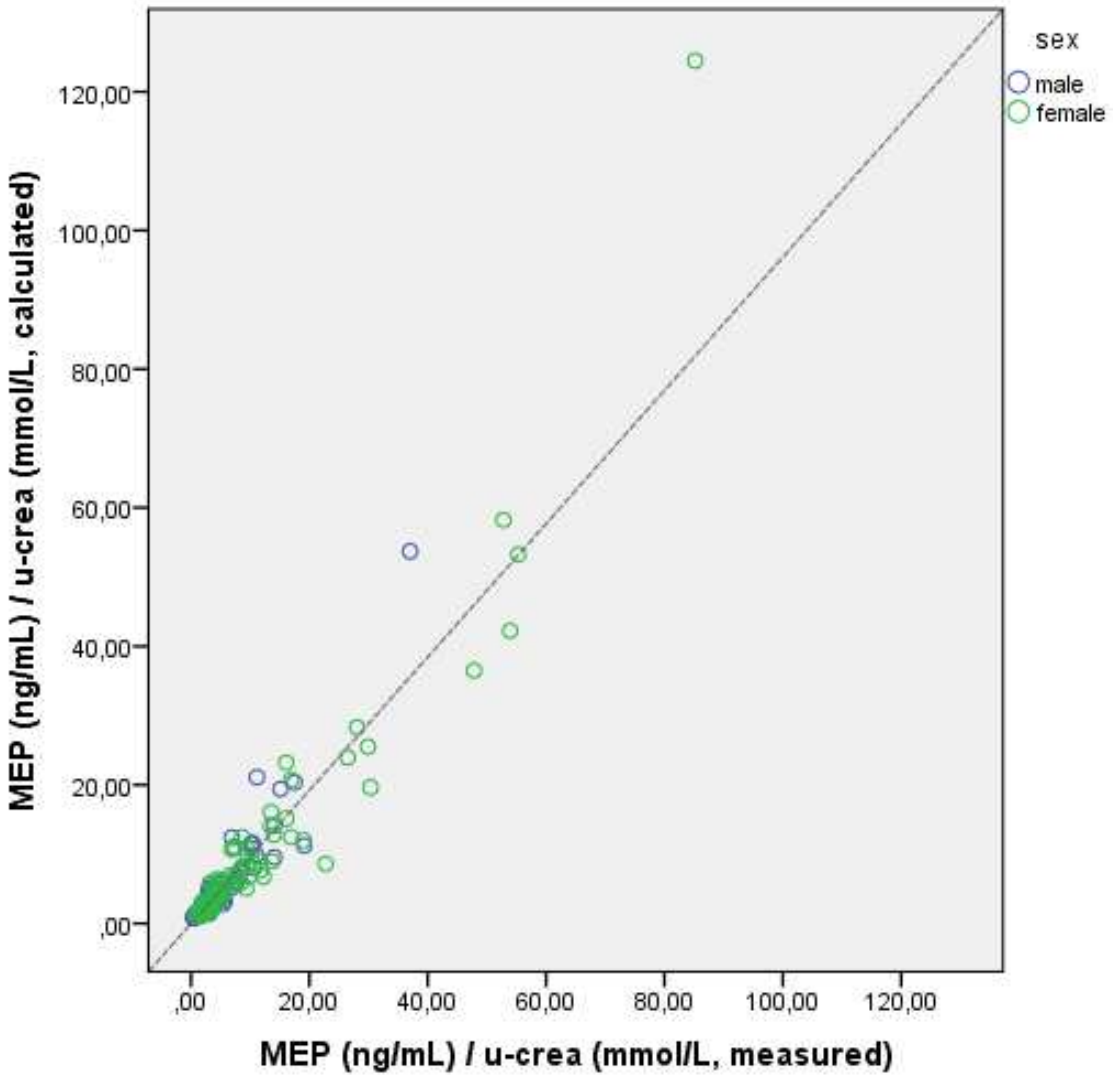


Figure 4: Measured urinary MEP concentration (ng/L) adjusted for calculated (by sex) u-creatinine concentration (mmol/L) plotted against measured urinary MBP concentration (ng/L) adjusted for measured u-creatinine concentration (mmol/L) (Data set 1+2). The dashed line is the identity line.

Table 4: Comparison of measured MEP and BPA concentrations adjusted by measured or calculated* u-creatinine (Data set 1+2 combined)

	Minimum	Maximum	Mean	Std. Deviation
ng MEP/mmol u-crea (measured)	0.42	85.20	**7.89**	11.11
ng MEP/mmol u-crea (calculated)	0.79	124.47	**7.68**	12.79
ng BPA/mmol u-crea (measured)	0.00	34.34	**0.74**	2.846
ng BPA/mmol u-crea (calculated)	0.00	37.56	**0.80**	3.13

* calculated as described for the combined data sets in footnote of Table 3

Paired T-test of difference between adjustment with measured versus calculated u-crea:

MEP: mean difference = 0.14 (95% CI: -0.53 – 0.80), p-value=0.682
BPA: mean difference = -0.05 (95% CI: -0.11 – 0.00), p-value=0.065