DERIVED VARIABLES TOOL BACKGROUND DOCUMENTATION

VITO

Table of contents

Introduction 2
Imputation of biomarker levels below LOD/LOQ 2
Sum parameters 2
Urinary biomarkers: adjustment for urinary dilution 3
Lipid soluble biomarkers analyzed in blood: adjustment for blood lipids 3
Lipid soluble biomarkers analyzed in breast milk: adjustment for lipids in breast milk 4
References 4
ANNEX 1: Overview of sum parameters' calculation implemented in the derived variables tool 6
ANNEX 2: Derived variables codebook 8

Introduction

The derived variables tool adds derived variables, based on the individual level data, to your data template. This document provides an overview of which variables are derived and the methodology on how they are derived.

Imputation of biomarker levels below LOD/LOQ

Values below the limit of detection (LOD) or limit of quantification (LOQ) or between LOD and LOQ are imputed using two single imputation methods:
1) Fixed value imputation using medium bound imputation, and
2) random single imputation using a censored lognormal distribution.

1) The medium bound imputation is performed as follows: (i) values lower than LOD are imputed by LOD/2; (ii) values below LOQ are imputed by LOQ/2, and (iii) values between LOD and LOQ are imputed by (LOD+LOQ)/2. The derived imputed variable using medium bound imputation is provided as an additional variable with suffix (**_MEB**).

2) The distribution based random single imputation technique is performed as follows:
   a. Assume the data is log-normally distributed and left-censored.
   b. Use maximum likelihood estimation (MLE) to fit a lognormal distribution to the censored data.
   c. For each censored data point, sample a value from the fitted lognormal distribution. Sampling will be done in a range that is determined by a lower and upper bound.

   In those cases where the censored value is indicated to be either lower than the LOD or lower than the LOQ, the upper bound is equal to the respective LOD or LOQ value and the lower bound is equal to 0. In case the censored value is indicated to be between LOD and LOQ, the lower bound is equal to LOD and the upper bound is equal to LOQ.

   Due to the requirement of sufficient data to estimate the distribution, the imputation using a censored lognormal distribution is only done if: (i) at least 30% of the values are above LOD/LOQ and (ii) at least 10 unique values are observed above the LOD/LOQ for a specific biomarker per data collection. If less than 30% of the values are detected in a specific data collection, the imputed variable is empty for that data collection. The new derived variable is named with the suffix (**_IMP**). Due to the nature of the random imputation, if the user decides to repeat the imputation process, the imputed values will differ each time the tool is used.

For biomarkers with more than 70% of the observations below LOD/LOQ, it is recommended to dichotomize the observations into <LOD/LOQ versus ≥ LOD/LOQ. In the derived variables tool, this derived binary variable is computed for all biomarkers regardless of percent below LOD/LOQ and it is provided with a suffix (**_BIN**).

Depending on the research question, the research team might decide to use the data resulting from either of the above recommended techniques or a different imputation technique using the observed values. If the team opts to use the imputed dataset, it's recommended to use the imputed values using censored lognormal distribution (**_imp**) as it provides less biased estimates compared with medium bound imputation (**_meb**). Single imputation using the censored lognormal distribution procedure has been shown to produce better results than simply replacing the values by half the LOD/LOQ (replacement by fixed value). However, it is important to mention that the random imputation method implemented in the tool does not include predictor variables for better estimation of the location and variance of the distribution, like for example correlated exposure biomarkers.

Sum parameters

For each substance group a whole range of biomarkers might be measured. Some biomarkers are metabolites from one and the same parent compound. For these markers it may also be useful to

investigate sum-parameters. Another reason to calculate specific sum parameters can be the availability of a health-based HBM guidance value for this sum of biomarkers.
Within the derived variables tool, sum-parameters are only calculated if the following conditions are fulfilled:

1. all the compounds forming the sum are analyzed.
2. at least one of the compounds has at least 60% of the values detected.

Since the sum possibly also incorporates compounds with more than 70% of the values below LOD/LOQ, and for these compounds the single random imputed biomarkers are not created, the sum parameter is taken as the sum of the medium bound imputed compounds.
Furthermore, additional sum parameters are calculated based on their molecular weight if the conditions (1) and (2) are fulfilled. This is done by dividing the metabolites by their molecular weight and calculating then the sum parameters. The created variable is named by adding "**_MOL**" to the sum parameter name.
Sum parameters implemented in the derived variables tool, if conditions above are fulfilled, are indicated in **Annex 1.** If other/additional sum parameters are considered relevant, please contact: PEHDataPlatform@vito.be

Urinary biomarkers: adjustment for urinary dilution

In the derived variables tool, urinary biomarker data standardized for creatinine and normalized for specific gravity (SG) are calculated according to the following formulas:

Creatinine standardization
Creatinine is used as a parameter to standardize the urinary biomarker concentration, i.e., biomarker concentration divided by gram creatinine (e.g., µg/g creatinine).

$$biomarker_{crt} = 100\frac{biomarker_{val}}{crt}$$

where the unit for marker value $biomarker_{val}$ in (µg/L) and creatinine crt in (mg/dL), and biomarkers $biomarker_{crt}$ to be reported in µg/g of crt.

Specific Gravity normalization
For normalization of urinary concentrations by SG the following equation is used(Sauvé et al., 2015):

$$biomarker_{corr} = \frac{biomarker_{val}\,(SG_{ref} - 1)}{(SG_{meas} - 1)}$$

where $biomarker_{corr}$ is the normalized concentration, $biomarker_{val}$is the measured biomarker concentration, SG_{meas} is the measured specific gravity, and SG_{ref} is the reference SG value of 1.024.

The new derived variables are named with the suffix (**_CRT and _SG**). The standardization and normalization are performed on all urinary markers and are applied on the imputed variables using the medium bound imputation and the random single imputation. The standardization and normalization are also performed on sum parameters. For creatinine standardization, the imputed creatinine variable (CRT_IMP) will be used if any value was below LOD or LOQ for creatinine.

Lipid soluble biomarkers analyzed in blood: adjustment for blood lipids

Within the derived variables tool, total lipid is calculated using an enzymatic summation method applying one harmonized formula. The total blood lipid content was calculated using the formula proposed by (Bernert et al., 2007; Phillips et al., 1989) as advised by the analytical experts of HBM4EU:

$$Lipid_{enz_{harm}}(\text{mg/dL}) = 2.27 * (\text{chol}_{\text{tot}}) + \text{trigl} + 62.3\ \text{mg/dL}$$

This harmonized lipid variable is calculated with the imputed values of cholesterol and triglycerides and is named "**lipid_enz_harm**". This harmonized lipid value is the preferred value to use for lipid standardization. When total cholesterol and triglyceride are not available, i.e., lipid_enz_harm variable cannot be produced, the imputed variable for lipid_enz (which is the enzymatic summation lipid value provided by the study owner), is used. If the lipid_enz is not provided in the dataset, gravimetric determined lipids are used for standardization, the imputed variable lipid (lipid_imp).
In the derived variables tool, lipid soluble biomarkers analyzed in blood are standardized for lipid using the following formula:

$$biomarker_{lip} = 100\frac{biomarker_{val}}{lip}$$

where $biomarker_{lip}$ is the lipid standardized variable (µg/g lipid), $biomarker_{val}$ is the marker's original value (in µg/L) and lip is the variable's corresponding lipid value (in mg/dL).
The biomarker data standardized for lipids are named with the suffix (**_LIP**). The standardization is performed on all lipid-soluble markers measured in blood (currently, these include biomarkers belonging to Flame retardants, Dioxins, Organochlorine compounds (OCs), Pesticides, and Furans) and it is applied on the imputed variables using the medium bound imputation and the random single imputation. The standardization is also performed on sum parameters.

Lipid soluble biomarkers analyzed in breast milk: adjustment for lipids in breast milk

Within the derived variables tool, lipid-soluble biomarker data analyzed in breast milk standardized for total lipid are calculated. The formula is the same as the correction used in blood samples.

$$biomarker_{lip} = 100\frac{biomarker_{val}}{lip}$$

where $biomarker_{lip}$ is the lipid standardized variable (µg/g lipid), $biomarker_{val}$ is the marker's original value (in µg/L) and lip is the variable's corresponding lipid value (in mg/dL).

The new derived variables are named with the suffix (**_LIP**). The standardization is performed on all biomarkers measured in a breast milk sample that are lipid soluble, using the imputed variables with the medium bound imputation and the random single imputation. For the standardization the imputed lipid variable (lip_imp) is used.

References

Bernert, J.T., Turner, W.E., Patterson, D.G., Needham, L.L., 2007. Calculation of serum "total lipid" concentrations for the adjustment of persistent organohalogen toxicant measurements in human samples. Chemosphere 68, 824–831. https://doi.org/10.1016/j.chemosphere.2007.02.043

Phillips, D.L., Pirkle, J.L., Burse, V.W., Bernert, J.T., Henderson, L.O., Needham, L.L., 1989. Chlorinated hydrocarbon levels in human serum: effects of fasting and feeding. Arch Environ Contam Toxicol 18, 495–500. https://doi.org/10.1007/BF01055015

Sauvé, J.-F., Lévesque, M., Huard, M., Drolet, D., Lavoué, J., Tardif, R., Truchon, G., 2015. Creatinine and Specific Gravity Normalization in Biological Monitoring of Occupational Exposures. Journal of Occupational and Environmental Hygiene 12, 123–129. https://doi.org/10.1080/15459624.2014.955179

ANNEX 1: Overview of sum parameters' calculation implemented in the derived variables tool.

Compound Group	PARC Sum Parameter Name	Compounds conforming the sum parameter	PARC Label
DINCH	summinch	Cyclohexane-1,2-dicarboxylate-mono-(7-carboxylate-4-methyl) heptyl ester + Cyclohexane-1,2-dicarboxylate-mono-(7-hydroxy-4-methyl) octyl ester	cx-MINCH + OH-MINCH
Flame retardants	sumdp	Syn-dechlorane plus + Anti-dechlorane plus	Syn-DP + Anti-DP
	sumhbcdag	Hexabromocyclododecane alpha + Hexabromocyclododecane gamma	HBCDα + HBCDγ
Metals	sumastfmma	Arsenic(III) acid + Arsenic(V) acid + Monomethylarsonic	As(III) + As(V) + MMA
	sumtas	Arsenobetaine + Arsenic(III) acid + Arsenic(V) acid + Dimethylarsinic + Monomethylarsonic	AsB + As(III) + As(V) + DMA + MMA
	sumtra	Arsenic(III) acid + Arsenic(V) acid + Dimethylarsinic + Monomethylarsonic	As(III) + As(V) + DMA + MMA
	ias	Arsenic(III) acid + Arsenic(V) acid	As(III) + As(V)
Per-/poly-fluorinated compounds (PFASs)	sumpfhxspfos	Linear Perfluorohexane sulfonic acid + Linear Perfluorooctane sulfonic acid	L-PFHxS + L-PFOS
	sumpfoapfna	Linear Perfluorooctanoic acid + Linear Perfluorononanoic acid	L-PFOA + PFNA
	sumfourpfas	Linear Perfluorooctanoic acid + Linear Perfluorononanoic acid + Linear Perfluorohexane sulfonic acid + Linear Perfluorooctane sulfonic acid	L-PFOA + PFNA + L-PFHxS + L-PFOS
Pesticides (Glyphosate)	sumglyphampa	Glyphosate + Aminomethylyphosphonic acid	GLY + AMPA
Pesticides (pyrethroids)	sumdcca	cis-DCCA + trans-DCCA	cis-DCCA + trans-DCCA
	sumpba	3-phenoxybenzoic acid + 4-fluoro-3-phenoxybenzoic acid	3PBA + F-3-PBA
Phthalates	sumeightphthal	Mono(2-ethyl-5-oxo-hexyl) phthalate + Mono(2-ethyl-5-hydroxy-hexyl) phthalate + Mono(2-ethylhexyl) phthalate + Mono(2-ethyl-5-carboxy-pentyl) phthalate + Mono-benzyl phthalate + Mono-isobutyl phthalate + Mono-n-butyl phthalate + Mono-ethyl phthalate	5oxo-MEHP + 5OH-MEHP + MEHP + 5cx-MEPP + MBzP + MiBP + MnBP + MEP

	sumlmwphthal	Mono-isobutyl phthalate + Mono-n-butyl phthalate + Mono-ethyl phthalate	MiBP + MnBP + MEP
	summinp	7-OH-(Mono-methyl-octyl) phthalate + 7-Carboxy-(mono-methyl- heptyl) phthalate	OH-MiNP + cx-MiNP
	sumohcxmehp	Mono(2-ethyl-5-hydroxy-hexyl) phthalate + Mono(2-ethyl-5-carboxy-pentyl) phthalate	5OH-MEHP + 5cx-MEPP
	sumohoxocxmehp	Mono(2-ethyl-5-oxo-hexyl) phthalate + Mono(2-ethyl-5-hydroxy-hexyl) phthalate + Mono(2-ethylhexyl) phthalate + Mono(2-ethyl-5-carboxy-pentyl) phthalate	5oxo-MEHP + 5OH-MEHP + MEHP + 5cx-MEPP
	sumohoxomehp	Mono(2-ethyl-5-oxo-hexyl) phthalate + Mono(2-ethyl-5-hydroxy-hexyl) phthalate	5oxo-MEHP + 5OH-MEHP
	sumohoxomehptwo	Mono(2-ethyl-5-oxo-hexyl) phthalate + Mono(2-ethyl-5-hydroxy-hexyl) phthalate + Mono(2-ethylhexyl) phthalate	5oxo-MEHP + 5OH-MEHP + MEHP
	summidp	6-OH-Mono-propyl-heptyl phthalate + Mono(2,7-methyl-7- carboxy-heptyl) phthalate	OH-MiDP + cx-MiDP
	sumhmwphthal	Mono(2-ethyl-5-oxo-hexyl) phthalate + Mono(2-ethyl-5-hydroxy-hexyl) phthalate + Mono(2-ethylhexyl) phthalate + Mono(2-ethyl-5-carboxy-pentyl) phthalate + Mono-benzyl phthalate	5oxo-MEHP + 5OH-MEHP + MEHP + 5cx-MEPP + MBzP
Polycyclic Aromatic Hydrocarbons (PAHs)	sumdiohnap	1-hydroxynaphthalene + 2-hydroxynaphthalene + 1-,2-dihydroxynaphthalene	1-NAPH + 2-NAPH + 1,2-NAPH
	sumohnap	1-hydroxynaphthalene + 2-hydroxynaphthalene	1-NAPH + 2-NAPH
	sumoneninehphe	1-hydroxyphenanthrene + 9-hydroxyphenanthrene	1-OH-Phenanthrene + 9-OH-Phenanthrene
	sumtwothreehfluo	2-hydroxyfluorene + 3-hydroxyfluorene	2-OH-Fluorene + 3-OH-Fluorene
	shphe	1-hydroxyphenanthrene + 2-hydroxyphenanthrene + 3-hydroxyphenanthrene + 4-hydroxyphenanthrene + 9-hydroxyphenanthrene	1-PHEN + 2-PHEN + 3-PHEN + 4-PHEN + 9-PHEN
	sumhfluo	2-hydroxyfluorene + 3-hydroxyfluorene + 9-hydroxyfluorene	2-FLUO + 3-FLUO + 9-FLUO
	sumhphetwo	1-hydroxyphenanthrene + 2-hydroxyphenanthrene + 3-hydroxyphenanthrene + 4-hydroxyphenanthrene	1-PHEN + 2-PHEN + 3-PHEN + 4-PHEN

	sumhpheone	1-hydroxyphenanthrene + 2-hydroxyphenanthrene + 4-hydroxyphenanthrene	1-PHEN + 2-PHEN + 4-PHEN
	stwoninehphe	2-hydroxyphenanthrene + 9-hydroxyphenanthrene	2-PHEN + 9-PHEN
	stwothreehphe	2-hydroxyphenanthrene + 3-hydroxyphenanthrene	2-PHEN + 3-PHEN

ANNEX 2: Derived variables codebook

VARNAME	VARIABLE	TYPE	NUMDIGITS	UNIT
Biomarker_bin	Dichotomized biomarker value into <LOD/LOQ or ≥ LOD/LOQ	varchar		
biomarker_imp	Imputed biomarker values (distribution based random single imputation method)	decimal	(10/4)	µg/L
biomarker_imp_crt	Imputed biomarker values (distribution based random single imputation method) standardized for creatinine	decimal	(10/5)	µg/g crt
biomarker_imp_sg	Imputed biomarker values (distribution based random single imputation method) normalized for specific gravity	decimal	(10/5)	µg/L
biomarker_imp_lip	Imputed biomarker values (distribution based random single imputation method) standardized for lipid	decimal	(10/4)	µg/g lipid
biomarker_meb	Imputed biomarker values (medium bound method)	decimal	(10/4)	µg/L
biomarker_meb_crt	Imputed biomarker values (medium bound method) standardized for creatinine	decimal	(10/5)	µg/g crt
biomarker_meb_sg	Imputed biomarker values (medium bound method) normalized for specific gravity	decimal	(10/5)	µg/L
biomarker_meb_lip	Imputed biomarker values (medium bound method) standardized for lipid	decimal	(10/4)	µg/g lipid
sumparameter	For some biomarkers sum parameters were calculated. See tab "SUMPAR-ANNEX"	decimal	(10/4)	µg/L
sumparameter_crt	Sum parameters values standardized for creatinine	decimal	(10/5)	µg/g crt
sumparameter_sg	Sum parameters values normalized for specific gravity	decimal	(10/5)	µg/L
sumparameter_lip	Sum parameters values standardized for lipid	decimal	(10/4)	µg/g lipid
sumparameter_mol	sum parameter in molar unit. See in sheet SUMPAR - ANNEX the molar units used for each of the biomarkers used in the sum parameters	decimal	(10/4)	µmol/L
sumparameter_mol_ crt	Sum parameters values standardized for creatinine, in molar units	decimal	(10/5)	µmol/g crt
sumparameter_mol_sg	Sum parameters values normalized for specific gravity, in molar units	decimal	(10/5)	µmol/L
sumparameter_mol_lip	Sum parameters values standardized for lipid, in molar units	decimal	(10/4)	µmol/g lipid