science and policy
for a healthy future

HBM4EU project

Calculating European Reference Values (ERVs) and data analysis of HBM data aligned studies in HBM4EU

Eva Govarts - VITO

2nd HBM4EU Training School 2018

HBM4EU data analysis

Aim: data analysis aligned studies at EU level

Needs:

- Harmonized variables → **pooled dataset**
- Central data analysis

→ harmonized codebooks developed by WP10

→ research protocols developed by T10.4 partners

→ exposure distributions / European reference values calculated by statistical working group (ANSP, UBA, VITO)

Existing HBM data

= heterogeneous data

- Different matrices
- Different time periods
- No EU coverage
- Not always access to individual data
- Different study populations
- No analytical quality assurance

HBM data aligned studies

= more homogeneous data

- Same matrix (only check alignment of type of urine sample (morning, spot, 24h), blood sample (fasting/non-fasting, serum/plasma))
- Time period 2014-2018
- EU coverage
- Access to individual data
- Selected study populations
- Quality assurance labs

Different strategy depending on homogeneity of the data!!

Different strategy

Heterogeneous HBM data ⟷ Homogeneous HBM data

DATA

- Aggregated data calculated with the HBM4EU R-script
- Individual data → pooled dataset

STATISTICAL ANALYSES

- Exposure distributions across data collections (comparison over time, region, …)
- Reference value calculations by applying exclusion criteria (specific food intake) and/or partitioning criteria (age, sex, EU geographical region)

Exposure distributions ⟷ European Reference Values (ERVs)

Exposure distributions

R-script developed by WP10 data management team

→ harmonized aggregated descriptive statistics

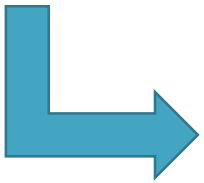
- Whole population + stratified by some parameters (sex, age, educational level, season, smoking status, sampling year)
- Values below LOD/LOQ imputed assuming normal/log-normal distribution
- Descriptive statistics
  - N
  - % below LOD-LOQ
  - Percentiles (P5, P10, P25, P50, P75, P90, P95) and 95% CI
  - Mean, standard deviation/error, 95% CI
  - Geometric mean and its 95% CI

Three steps approach

- Step 1: Inventory of available HBM data collections

- Step 2: Construction of a suitable pooled dataset

- Step 3: Authorization to derive a valid ERV



valid ERV

Inventory of available HBM data collections

→ Selection of homogeneous HBM data

- Type of substance / biomarker / matrix
- Type of population
- Age range
- Time period
- Sampling
- Representativeness (national, regional, etc.)
- Quality assurance and comparability analytical methods

Construction of a suitable pooled dataset

- Constitution of one dataset by pooling all data coming from the selected data collections (reference sample)
- Apply splitting criteria to focus on a dataset for which it's relevant (and possible) to derive an ERV
  - Exclusion criteria: exclude participants that are not considered appropriate to be derived an ERV from (e.g. pregnant status)
  - Partitioning criteria: create sub-groups to derive sub-ERVs when relevant (when sub-ERVs are significantly different for sub-groups of the population) (e.g. sex, age)

Authorization to derive a valid ERV

- Sufficient sample size: 120 individuals to calculate 95% CI for the 95th percentile
- Representativeness: is the dataset representative of the population for which the ERV is derived?

 If authorization criteria are not met → **return to previous step to modify dataset**

Once dataset is built:

→ percentiles of the reference distribution may be estimated by non-parametric methods

→ UBA, ANPS, VITO will work on a R-script to calculate ERVs in a harmonized and uniform way

Research protocols

Within T10.4 research protocols will be developed by expert partners for the statistical analyses at EU level.

For the aligned studies, protocols will focus on:

- Geographical comparisons
- Determinants of exposure
- Time trends (in collaboration with T8.2)

Time trends (in collaboration with task 8.2)

T 1 (Literature)

- ✓ 2006-2010
- ✓ Published data

T 2 (Task 8.2)

- ✓ 2011-2013
- ✓ New Analysis in DEMOCOPHES samples
  (children 6-11 yrs)
  (Mothers 20-45 yrs)
- ✓ DINCH
- ✓ phthalates, FR, bisphenols and PFASs. (depending on sample availability)

T 3 (Task 8.1)

- ✓ 2014-2019
- ✓ New analysis
- ✓ DINCH

Research questions: issues

Determinants of exposure:

what if important variables/information is missing for a certain data collection?

→ list of mandatory – optional variables

mandatory missing → data collection cannot participate

optional missing → sensitivity analysis without data collection

State the research question

Background

Objective

Methods

- Study Design: Studies needed
- Variables needed
  - Exposure variables (specify which biomarkers)
  - Accompanying variables

    → **indication of mandatory and optional variables!**
- Statistical analyses
- Substance specific issues

Organization, Time schedule, Publication

Organization-publication

- Each participating data collection will be included in the study protocol working group and is as such informed about the data analyses at EU level

- Each participating data collection is entitled to request to include 2 co-authors in scientific publication of results considering provided data

(see publication policy at https://www.hbm4eu.eu/private/work-package-webpage/scientific-and-administrative-management/ under work package 2)

Aligned studies

Each data collection is free to do data analysis on their own data collection's data, e.g.

- Determinants of exposure at level of the data collection
- Looking into the association with health effects possibly included in the data collection's questionnaires, biomarkers of effect?
- …

Thank you for your attention

If you have any questions about data analysis

Don’t hesitate to contact the HBM4EU WP10 statistical working group!

VITO: eva.govarts@vito.be

ISGlobal: martine.vrijheid@isglobal.org

UBA: nina.vogel@uba.de

ANSP: loic.rambaud@santepubliquefrance.fr

Contacts

Eva Govarts: eva.govarts@vito.be
Sylvie Remy: sylvie.remy@vito.be
Liese Gilles: liese.gilles@vito.be
Greet Schoeters: greet.schoeters@vito.be

Speaker's information

Eva Govarts works as researcher – biostatistician at the Flemish Institute for Technological Research (VITO), Mol, Belgium. She received training in biomedical sciences, applied and biostatistics. In HBM4EU she is task leader of task 10.4 on the data analysis and generation of European reference values (RVs) and together with Greet Schoeters she is co-leading WP10.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733032.