How to use PARC HBM data in MCRA

Hilko van der Voet, WR-BIOM, 20-12-2022

Aim of this note

To describe how human biomonitoring (HBM) data should be organised and how a simple run in the MCRA web system can be made.

Introduction

In PARC project Real-life mixtures, data from HBM studies will be analysed for co-exposures that might be relevant for mixture risk assessment. This will be done by multivariate analyses methods (e.g., SNMU, cluster analysis, graphical network analysis) which have been implemented in the MCRA system for probabilistic risk assessment (https://mcra.rivm.nl, see also van der Voet et al., 2020). Another purpose is to compare measured HBM exposures with modelled exposures. When these are reasonably close, the model can then be further used to identify risk drivers, e.g., substances in certain foods combinations, or in certain non-dietary sources.

In PARC, WP 7 is coordinating FAIR data organization. For HBM data, a data format is being developed by VITO, in collaboration with WP 4. The MCRA system aims to follow this data format development. In this note we describe how HBM data organised according to the first version of the PARC data format (as of September 2022) can be uploaded and used in MCRA to obtain summary statistics, boxplot representations, and a maximum cumulative ratio (MCR) plot (Price & Han, 2011). In another note, the possibilities for multivariate analysis are described.

Procedure to upload and inspect HBM data in MCRA

It is assumed that you have an MCRA account. If not, register on https://mcra.rivm.nl. Registrations have to be approved, so this may take some time.

Annex to this document, two data files are distributed:

1. Example study population.xlsx
2. ExampleData_BasicCodebook_v2.0.xlsx

It is assumed that these files are available to you.

Steps

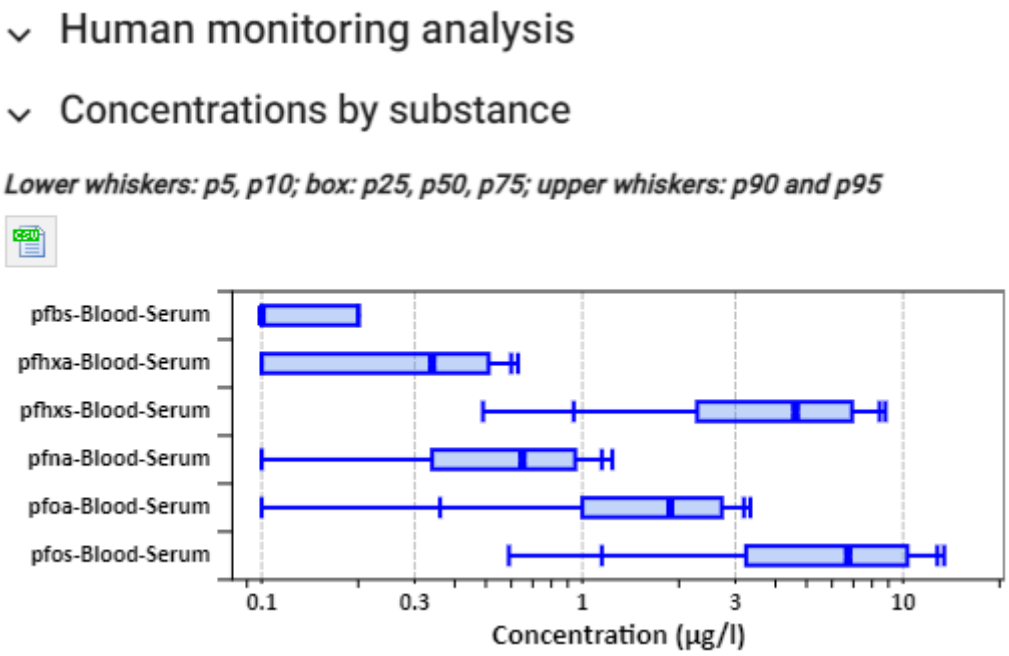
1. The data files are prepared for the case study of the FLEHS study and contain several tables as suggested by VITO. For your own data, please prepare these in the same format in connection with VITO:
    a. The population file specifies the target population. The IdPopulation is used to link to the HBM data, specified as 'Study ID' in the STUDYINFO tab of the example HBM file.
    b. The example HBM file contains these worksheets:
        i. STUDYINFO
        ii. SAMPLE
        iii. GROUP
        iv. SUBJECTUNIQUE
        v. SUBJECTREPEATED
        vi. DATA_US
        vii. DATA_BS.

See Appendix for more information. The exact formats may be adapted still (In PARC, VITO is in charge of this).

2. Upload the data to MCRA:
    a. Login to MCRA at https://mcra.rivm.nl
    b. Go to Data. Click on your personal folder (labelled with your MCRA user name).
    c. Press the green '+' button in the lower right corner of the screen. Choose 'Upload new file(s)'. Select the two data files above, and upload them to MCRA.
3. Create an MCRA Human monitoring analysis action:
    a. Go to Workspaces (Click All workspaces button top right in the blue bar). Press the green '+' button in the lower right corner to create a Workspace.
    b. In the workspace, press the green '+' button in the lower right corner to create an action
    c. Click 'Show all action types', and then scroll down and select Human monitoring analysis
    d. Choose an appropriate name for your action (e.g. 'PARC HBM') and click Next
    e. Choose Risk type. 'Acute' will evaluate exposures per individual-day. In the current version, from HBM data all measurements from each individual will be used separately. 'Chronic' will calculate averages per individual person over all days per individual. For the example file, we choose 'Acute'.
    f. Select multiple substance analysis.
    g. Click Create. Now the action is created but still has to linked to the input data.
4. Link data and choose action settings:
    a. Click on Human monitoring data, in the list shown in the left panel (green section) or as Input in the main Human monitoring analysis screen (works the same). Click on the pencil icon to select a Human monitoring data source. Browse to the folder with your uploaded data and select 'ExampleData_BasicCodebook_v2.0.xlsx' (or your own adapted version). At the bottom of the pop-up, choose 'Toggle all'. This will load all available data tables in the file. Then click Select.
    b. Click on Populations (purple section). Select 'Use data' and click on the pencil icon to select a Populations data source. Browse to the folder with your uploaded data and select 'Example study population' (or your own adapted version). Then click Select.
    c. In the Human monitoring data settings, choose the survey/study (PARC HBM sample data in the example file), and a sampling method (e.g., Blood (Serum) in the example file). Then click Save changes (red button).
    d. [Optional] Click on Substances (purple section). Click 'set filter to select the substances of interest. Click Save. Then click Save changes (red button).
    e. In the Human monitoring analysis settings, choose how to handle censored values (e.g., as 0.5 * LOR, where the LOR (limit of reporting) means LOQ if that is provided, or else LOD). Also choose how to impute any missing values. Impute from data means that for each missing value a random value from the available measurements per substance is selected.
5. Run the action:
    a. Click the Run button, triangle in the grey horizontal action bar at the top of the screen.
    b. You are transferred to the Results screen. Wait for the run to finish. Meanwhile, you can change the name of the output if you want by clicking the pencil symbol.
6. Inspect the results:
    a. When ran to completion, open the output by clicking the name of the output.
    b. Click on 'Concentrations by substance' to show the results of the processed HBM data as boxplots and in a table with statistics.
    c. Click on 'Details >> MCR co-exposure' to display the Maximum Cumulative Ratio against the total exposure for each individual. Note that in this example each point in the plot corresponds to an individual day. Each point will correspond to an individual if 'Risk type' in 'Human monitoring analysis' is 'Chronic' instead of 'Acute'.
    d. Details can be found under the main tab Sub-action results. For example, Human monitoring data will show boxplots and statistics of the data before censored value and

missing value imputation. Note that output tables such as 'Human monitoring samples per sampling method and substance' can be saved by pressing the 'csv' icon right above the table. Figures can be stored by right clicking them.

Example: blood serum concentrations of substances in an example population



In this example, censored values were imputed with 0.5 times the LOR and missing values (if any) were imputed from the data.

References

Price, P.S and Han. X.L.. Maximum cumulative ratio (mcr) as a tool for assessing the value of performing a cumulative risk assessment. International journal of environmental research and public health, 8(6):2212–2225, 2011. https://doi.org/10.3390/ijerph8062212

van der Voet, H., Kruisselbrink, J.W., de Boer, W.J., van Lenthe, M.S., van den Heuvel, J.J.B., Crépet, A., Kennedy, M.C., Zilliacus, J., Beronius, A., Rorije, E., Sprong, C., and van Klaveren, J.D. The EuroMix model toolbox MCRA 9. 2019. URL: https://zenodo.org/record/3462181.

Appendix 1. Example of HBM data

Notes:

- The example is simulated data in file ExampleData_BasicCodebook_v2.0.xlsx.

STUDYINFO

Currently, only the Study ID is relevant, and should match the Population id in the Populations data source.

SAMPLE

This table lists all samples, and identifies the sampling type / matrix (US or BS in the example)

GROUP

Groups identify repeated measurements of the individuals. In the example, subjects were resampled after one year.

SUBJECTUNIQUE

This table specifies characteristics of the individuals such as cohort, age of the mother at birth, birth weight etc.

SUBJECTREPEATED

This table specifies characteristics and questionnaire answers of the individuals at the second sampling moment. Not used in this example.

DATA_US and DATA_BS

These tables specify the actual data per matrix (sampling type). For example, in worksheet DATA_BS, the actual measurements in spot urine are included with a column for each substance. Codes -1, -2, and -3 are used for non-detects (<LOD), measurements known to be between LOD and LOQ or non-quantifications (<LOQ), respectively. For each substance, there are two more optional columns to specify the LOD and/or LOQ values.

id_sample	pfhxa	pfhxa_lod	pfhxa_loq	pfoa	pfoa_lod	pfoa_loq	pfna	pfna_lod	pfna_loq
301	-3		0.2	-3		0.2	0.68		0.2
302	-3		0.2	0.97		0.2	1		0.2
303	-3		0.2	1.21		0.2	0.45		0.2
304	0.25		0.2	2.49		0.2	1.09		0.2
305	0.27		0.2	1.71		0.2	0.59		0.2