



List of criteria for selecting cohort studies

Deliverable Report 13.1

WP 13 - Establishing exposure-health relationships

Deadline: June 2017

Upload by Coordinator: 27 June 2017

Entity	Name of person responsible	Short name of institution	Received
Coordinator	Marika Kolossa-Gehring	UBA	16.06.2017
Grant Signatory	Jana Klanova	MU	16.06.2017
Pillar Leader	Robert Barouki	INSERM	16.06.2017
Work Package Leader	Jana Klanova	MU	16.06.2017
Task leader	Thorhallur Ingi Halldorsson	UI	16.06.2017

Responsible author	Thorhallur Ingi Halldorsson	E-mail	tih@hi.is
Short name of institution	UI	Phone	+354 5438407
Co-authors in alphabetical order	Anna-Maria Andersson (REGIONH), Marika Berglund (KI), Anna Bergstrom (KI), Ludek Blaha (MU), Eve Bonnenfeldt Jørgensen (AU), Marie Aline Charles (INSERM), Hubert Dirven (NIPH), Tommaso Filippini (UNIMORE), Fiona Fordyce (NERC), Aleksandra Fucic (IMROH), Eva Goverts (VITO), Thorhallur Ingi Halldorsson (UI), Tina Kold Jensen (SDU), Jana Klanova (MU), Nathalie Lambrechts (VITO), Karin Leander (KI), Marcel Mengelers (RIVM), Sophie Ndaw (INRS), Markku Sallmén (FIOH), Dimosthenis Sarigiannis (AUTH), Greet Schoeters (VITO), Marco Vinceti (UNIMORE), Martine Vrijheid (ISGlobal), Agneta Åkesson (KI)		

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 2

Table of contents

1	Authors and Acknowledgements.....	3
2	Background.....	4
2.1	Scenario 1: Selection of studies when collection of new data is needed	5
2.2	Scenario 2: Selection of existing studies when no additional data collection is needed.....	9

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 3

1 Authors and Acknowledgements

Authors in alphabetical order

Anna-Maria Andersson, REGIONH
Marika Berglund, KI
Anna Bergstrom, KI
Ludek Blaha, MU
Eve Bonnenfeldt Jørgensen, AU
Marie Aline Charles, INSERM
Hubert Dirven, NIPH
Tommaso Filippini, UNIMORE
Fiona Fordyce, NERC
Aleksandra Fucic, IMROH
Eva Goverts, VITO
Thorhallur Ingi Halldorsson, UI
Tina Kold Jensen, SDU
Jana Klanova, MU
Nathalie Lambrechts, VITO
Karin Leander, KI
Marcel Mengelers RIVM
Sophie Ndaw, INRS
Markku Sallmén, FIOH
Dimosthenis Sarigiannis, AUTH
Greet Schoeters, VITO
Marco Vinceti, UNIMORE
Martine Vrijheid, ISGlobal
Agneta Åkesson, KI

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 4

2 Background

The overall context for deliverable D13.1 is given in the five year work plan:

*“The aim of our work is to establish causal links between exposures and health by developing criteria (D13.1) for selecting data and samples from established cohort studies like HELIX or HEALS as well as obtaining data for new studies **in coordination with WP8 and WP9**. Depending on the substance (exposure) and health outcome of interest and based on the criteria developed, we will carefully consider what study design and covariate information from questionnaires or registries are needed to draw robust conclusion on the relationship between exposure and health for each of the substances under consideration.”*

*“On the basis of the criteria (D13.1) developed for selecting studies, and based on identified data gaps for each of the priority substances (D13.3) we will, in coordination with WP7 and WP8, identify and initiate research collaboration with selected cohorts that are appropriately designed and **can provide 1) necessary samples for measurement or 2) provide existing data that will allow us to conduct new targeted studies**. These targeted studies will allow us to draw more robust conclusions on exposure health relationships for each of the priority substances (D13.7).”*

Deliverable D13.1, is then defined in more detail in the 2017 annual work plan:

*“Under coordination of UI, ISGlobal, SZU, VITO, AUTH and **in collaboration with task 7.1 and task 8.2** a set of criteria (qualitative and quantitative) for selecting existing studies for further exploration of the exposure-health relationships will be developed (D13.1, M6). The selection criteria will include assessment of cohort size and power analysis.”*

As formulated in the 5 year work plan, the study under consideration should address data gaps on human health that have been identified within the HBM4EU program for one or more of the priority substances. The purpose of this deliverable is therefore to set the stage for how we make decisions on which studies/cohorts to choose once we have, in collaboration with other relevant WPs, identified

- a) what research questions need to be specifically addressed within WP13
- b) which cohorts/databases may be suitable for selection
- c) which biomarkers are in place or is there need for new analyses (link with WP14)

When developing the selection criteria it was considered necessary to divide them into two different scenarios that covered *selection of studies where analyzes of new samples or other data collection is needed*; and *selection of studies where all necessary data has been collected*. This latter scenario primarily addresses joint analyses of more than one study by merging individual participant data or producing summary estimates through meta-analyses. Below each of these criteria is formulated followed by further justification and explanations.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 5

2.1 Scenario 1: Selection of studies when collection of new data is needed

Broadly speaking the first scenario deals with selection of existing studies when chemical analyses or other collection of data is needed. With respect to the exposure, analyses of new samples would most likely be needed for those priority substances that have not been frequently measured in existing European studies. New samples may also need to be analyzed, even if there are existing studies with quantified exposures as that on its own is no guarantee that the study is suitable to address the given research question. This scenario also covers cases when suitable studies exist where exposure has been quantified but information on the outcome or covariates is missing (see explanation below in the text). More formally this scenario is formulated as follows:

SCENARIO 1: Initiating new projects based on selecting existing studies where collection of additional data is needed. This may involve:

- a. New chemical measurements from archived samples or samples that must be collected.
- b. Collection of additional information on the health outcome(s) of interest such as through existing registries, biomarker analyses or additional follow-up.
- c. Both **a** and **b**.

The selection of studies should be guided by choosing the most suitable study that can answer the research question under consideration in the most cost-effective manner. This may make collection of new samples unfeasible but this option is kept open, in case it may be needed. Although option “c” may seem unlikely it may come relevant if new studies in Europe are being initiated or existing studies are conducting follow-up during HBM4EU program period and data collection can be incorporated at low (or no) cost. However, unless it can be properly justified use of existing samples and data is prioritized.

When having to select studies that fall under scenario 1 the following criteria should be evaluated to judge if the study is eligible and suitable:

Eligibility criteria for scenario 1:

- a) Co-funding is a central part of the HBM4EU EJP. Thus, only those studies are eligible for selection where co-funding is guaranteed. This includes co-funding in terms of work (person months) and co-funding for chemical analyses if new analyses are performed.
- b) Ethical permissions also needs to be documented and the participants had to have given informed consent that covers any new chemical analyses and other use of data for the study (either open or specific informed consent).
- c) Permission or intent to submit chemical data to IPCHEM is also needed. A minimum requirement is that the selected study provides aggregated data and meta-data of the study.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 6

Selection criteria for scenario 1:

1. Quality of the exposure assessment

- a) *Type of sample*: Does the study have appropriate bio-material (blood, urine or other material) to allow for reliable quantification of exposure? If not, can such samples be collected at low cost? Collection of new samples can only be considered if studies with existing samples are not suitable and sample collection can be achieved at low cost without compromising the time constraint of the study.
- b) *Sample collection and storage*: When and how were the samples collected? Here a range of study specific factors will have to be considered on a case by case basis.

Examples include:

- Was the preparation and storage of the sample appropriate for the substance or biomarker under consideration? Is it possible that quality of the measurement may be compromised due to time and conditions between sampling and treatment, duration of storage and temperature, number of freeze-thaw cycles and other factors?
 - Is cross-contamination likely? Here the types of tubes used, availability of field blank on other factors must be considered.
 - Specific criteria related to physiology - e.g. the appropriateness of fasting versus non-fasting blood samples must be considered. Same considerations are needed for the type of urine sample and other types of samples.
 - Were samples taken at a similar time during the day or during the year (seasonal changes)? If relevant, can variation in timing of sample collection be adjusted for?
 - Are relevant adjustment factors available or can they be quantified (such as urine creatinine)?
 - Are there other factors that may influence the interpretation of the exposure-health relationship?
- c) *Time frame*: What time frame does a single measure of exposure reflect?
- Does it cover the most relevant time window?
 - Would use of repeated samples be an advantage? If yes then studies that have collected samples at multiple time points may be ranked higher. This issue will be of importance for substances with short elimination half-life.
- d) *Exposure level*: What is the expected exposure level and range in the study population?
- Is it anticipated that exposure levels in the samples are sufficiently high to influence the health outcome under consideration?
 - If many studies with similar exposure have already addressed the same research question then the added value of an additional study should be justified. If exposure has been quantified this would be known and could be directly evaluated.
- e) *Sources of exposure*: What would be the main source of the exposure for the study population and is such information recorded? This would perhaps be obvious in the case of occupational studies. However, among studies in the general population having

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 7

some information or at minimum some idea on main sources of exposure is relevant. For example, if main sources of exposure are unusual/rare then the generalizability of the study has to be considered carefully.

2. Quality of the Outcome

- a) *Assessment*: Is the outcome reliably assessed and validated?
- If the outcome of interest is a direct measure of health status we may want to favor objective measures based on clinical examination, health records or registry based data over self-reported outcomes. Validated outcomes are preferred.
 - If the outcome of interest is based on quantification of one or several biomarkers (surrogate markers) then the validity of those measures need to be assessed. Same criteria as stated in 1b) above (under exposure assessment) should be considered.

3. Quality of the study design:

- a) *Temporal separation*: Prospective studies are generally preferred. In cases (some biomarker and omics studies) where the exposure is expected to immediately affect the outcome, use of cross sectional studies could be justified. Even if prospective studies are available it should be assessed whether exposure is measured at a relevant and sensitive time window and that the temporal separation between the exposure measure and recording of the outcome is sufficient. In other words, is it biologically plausible that exposure at the time the bio-material was collected could have influenced the outcome under consideration at the time it was assessed?
- b) *Confounder control*: Does the study have information on relevant confounders that need to be taken into consideration? Are these confounders recorded or quantified in sufficient detail?
- c) *Statistical power*: Is the proposed study sufficiently powered to address the exposure health relationship under consideration? Power calculations need to be performed and underlying assumptions must be clearly stated.

4. Other practical requirements that need to be considered

- a) What are the rules for access to samples and data? That is, is it likely that access to samples and data will be granted so the time constraints of the study are not compromised?
- b) Is there an access fee? If yes, that cost should be weighed in with other expenses.

Justification for these criteria:

For the priority substances analyses of new samples will always be expensive, particularly when linking exposure to a health outcome that requires samples size of several hundreds of participants. Funding such studies may however be needed as some of the priority chemicals (Cr (VI) and DINCH) have only been minimally addressed with respect to human health.

It may also happen that a study exists where the exposure of interest has been quantified but the outcome is not available but can be quantified. For example, if the outcome of interest is a biomarker of inflammation and samples are available then it may be feasible to invest in those analyses to do the study.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 8

It is also important to remember that new studies will be initiated based on identified data gaps. Data gap per definition means that limited information exists. It is by no means given that it will always be the exposure that needs to be quantified. Therefore, the criteria above cover cases where either exposure, outcome or both are missing; as well as more complex situations where samples have to be collected.

How to apply these criteria:

All studies need to fulfill the eligibility criteria. The cost of the study is a major factor, which is why studies with existing samples are favored and cost-effectiveness ranks high. There is no perfect scoring system for the other criteria given above but the selected study must be considered appropriate after taking into consideration the points raised with respect to study design and quality of the exposure and outcome assessment.

What studies should we focus on:

Although classical epidemiological studies are likely to be considered, the selection should not only restrict to such studies. In the 5-year work plan investigation of “omics approaches” is mentioned, which is more related to understanding biological pathways that may or may not turn out to have a direct causal link to disease. This will allow for natural linking of the work to task 13.1, which addresses adverse outcome pathways (AOPs). If feasible we could initiate some studies in tasks 13.2 that would be designed to support or refute some of work on AOPs for the priority chemicals. These types of studies may potentially be less costly and could be used as a platform to do more targeted epidemiological studies on certain health outcomes.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 9

2.2 Scenario 2: Selection of existing studies when no additional data collection is needed

This scenario is intended to cover selection of studies when the aim is to do combined analyses of several studies either through direct merging of individual participant data or by producing summary estimates through meta-analyses. The added benefit of doing joint analyses of several comparable studies compared relying on single studies include:

- a) More precise risk estimates.
- b) Being able to examine generalizability of associations across studies where incidence of the outcome and covariate constellation may differ.
- c) Examination of common health outcomes in a harmonized way between studies (broader exposure range) resulting in more robust dose-response estimates than can be obtained from single studies.
- d) The possibility of examining relatively rare health outcomes.
- e) Addressing different susceptibility, for example for vulnerable sub-groups, as well as being sufficiently powered to test for effect modifications

Points a-b) would be the advantage when producing summary estimates through meta-analyses, while points a-e) would be the added benefits of directly merging individual participant data. This type of work has previously been done when evaluating the associations between prenatal exposures to PCBs and birth weight through meta-analyses (Govarts E et al EHP 2012) and by merging individual participant data in a combined analyses (Casas M et al. Environ Int. 2015). These types of studies are an excellent example of how to use existing data to address policy relevant questions on environmental health.

Scenario 2 may also cover selection of individual studies when no additional data collection is needed. This would be an option in cases where we can identify 1) unexplored data for a single study 2) which can answer research questions relevant for HBM4EU and 3) those who own the data are willing to collaborate.

The HBM4EU contribution for scenario 2 would primarily be person months to do the relevant analyses. In addition, on a case by case basis there may also be costs associated with getting access to data that need to be covered.

More formally this scenario can be described as follows:

SCENARIO 2: Initiating new projects that are solely based on existing data. This could involve:

- a. New analyses of existing data from multiple but comparable studies by merging individual participant data or by producing summary estimates through meta-analyses.
- b. New analyses of existing data from a single study where exposure and health outcome already exist but have not previously been explored.

Eligibility criteria for scenario 2:

- a) For the proposed analyses, co-funding in terms of person months must be guaranteed.
- b) For each study included, ethical and other relevant permissions needed to carry out the new analyses need to be documented.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 10

Selection criteria for scenario 2:

1. **Study population:** Are the study populations in each individual study comparable enough to justify either merging individual participant data or producing summary estimates through meta-analyses? Relevant factors to consider may be age, disease status, pregnancy and lactation and others.
2. **Study design:** Are the designs of the studies under consideration comparable enough to justify either merging of individual participant data or generation of summary estimates through meta-analyses? Prospective studies will generally be preferred over other study designs. Aspects that may need to be considered include:
 - a) *Comparable designs:* Merging pro- and retrospective studies may for example lead to spurious results that would be hard to interpret.
 - b) *Biological plausibility* also needs to be considered: That is for the joint analyses, is it possible that exposure at the time the bio-material was collected in these studies could have influenced the outcome under consideration at the time they were assessed? Is the temporal separation comparable in all studies?
 - c) *Confounder control:* Do all the studies have information on relevant confounders that need to be taken into consideration? Are these confounders recorded or quantified in sufficient detail and can they be combined?
 - d) *Sample size considerations:* When merging individual participant data for several studies, estimating the minimum effect size that the analyses can detect would be appropriate (that is, estimating the minimum RR, mean difference or slope or other effect estimates using the anticipated number of participants and assuming $\alpha=0.8$ and $\beta=0.05$). Similar considerations on the precision of effect estimates would be appropriate when doing meta-analyses.
3. **Outcome:** Are the health outcomes in each of the study under consideration comparable enough to justify merging of individual participant data or producing summary estimates through meta-analyses? Restricting the work to fewer studies would in many cases be more suitable than trying to increase number of participants.
4. **Exposure:** Are the types of samples used for exposure assessment in each of the studies comparable enough to allow for merging of individual data or producing summary estimates through meta-analyses? Here several factors need to be considered:
 - Did all the studies use the same type of sample (blood, urine, breast milk)? For obvious reasons care should be taken when comparing different types of samples (such as blood *versus* breast milk).
 - When and how were the samples collected? Some considerations and comparison between studies should be made.
 - Can each study document the quality of the exposure assessment? This may include results from inter-laboratory comparisons and other relevant quality control materials.
 - What would be the main source of the exposure in each of the studies? Some reflections on this may be justifiable.
5. **Other practical recruitments that need to be considered**
 - What are the rules for access to data for each study?
 - Is there an access fee? If yes, that cost should be weighed in with other expenses.

D 13.1 List of criteria for selecting cohort studies	Security: Public
WP 13 - Establishing exposure-health relationships	Version: 1.1
Authors: Task 13.2 partners	Page: 11

For scenario 2b (i.e. new analyses of existing data from a single study) same considerations apply but then comparison between studies is for obvious reasons not needed.

How to apply these criteria: and what should we focus on

All studies need to fulfill the eligibility criteria. Merging of individual participant data should be aimed for if possible as it would allow for more in-depth analyses (such as exploring effect modifications and associations among vulnerable sub-groups) than simple meta-analyses can achieve. The problem with such an approach is getting direct access to data and the fact that such analyses are time consuming. It may also be difficult to find several comparable cohorts to merge. Still many pregnancy cohorts would easily fulfill the criteria above (there are several pregnancy cohorts that included analyses of PFAS and phthalates). Once data has been merged several health endpoints can be examined. In other cases, meta-analyses may be considered more practical and this has to be decided on case by case basis.

Since all data collection has taken place the focus of the criteria for scenario 2 is on evaluating the suitability of combining studies and making sure that there is added value in doing joint analyses.