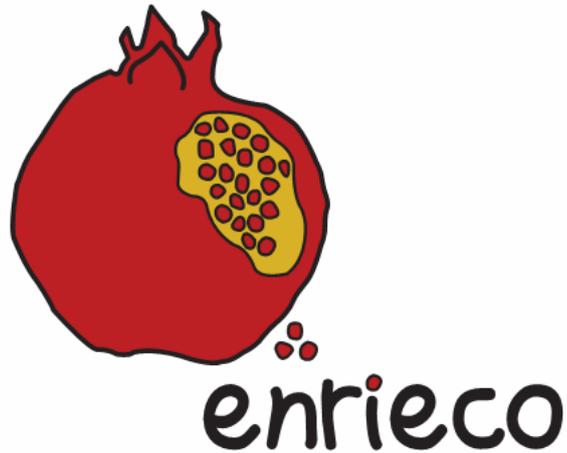


# ENRIECO EU project



## Work package 5 Deliverable 6

Thomas Keil  
Cynthia Hohmann

Institute for Social Medicine, Epidemiology and Health Economics  
Charité University Medical Center Berlin, Germany

31 August 2009



## Content

1	Introduction .....	3
1.1	Objectives.....	3
1.2	Issues related to current ENRIECO Project .....	3
2	Key Responsibilities.....	4
2.1	Organisational Issues related to current ENRIECO Project.....	4
2.2	Trust Building and Agreement .....	5
3	Procedure of Database Building.....	5
3.1	Step 1: Willingness to Participate .....	6
3.2	Step 2: Eligibility of Cohorts .....	6
3.3	Step 3: Collection of Individual Participant Data .....	6
3.4	Step 4: Combined Analysis .....	6
3.5	Issues of Database Building related to current ENRIECO Project .....	7

## ENRIECO WP5 - Deliverable Number 6

### Protocol for Database Building

Thomas Keil and Cynthia Hohmann  
cynthia.hohmann@charite.de

## 1 Introduction

Work package 5 (WP5) combines individual participant data from ongoing birth cohorts with similar outcome and exposure assessments. Some of the birth cohorts are rather small, with a few hundred participants. However, for analysing certain outcome and exposure assessments the sample sizes of many cohorts may not be sufficient on their own. Combining their data will increase statistical power to examine rare exposures. Furthermore, after harmonising the data across the cohorts, one will be able to investigate regional differences among European cohorts.

### 1.1 Objectives

The objectives of WP5 are to obtain data access, build databases, and conduct analysis including:

1. Preparation of protocols for data access, database building and analyses
2. Preparation of protocols for exposure response analyses
3. Conduct of specific analyses on exposure and health data to obtain exposure response functions (through case studies)
4. Conduct of specific meta/pooled analyses to obtain exposure response functions (through case studies).

### 1.2 Issues related to current ENRIECO Project

The ENRIECO working groups aim to examine associations between indoor environmental exposures (working group 1: dampness/mould; working group 2 and 3: tobacco smoke) and asthma and allergic diseases in children. Individual participant data from European birth cohorts focusing on asthma and allergies will be used for analyses.

⇒ Similar experience was gathered from the Global Allergy and Asthma European Network (GA<sup>2</sup>LEN). Pet ownership and association with asthma, allergic rhinitis and allergic sensitisation was investigated with combined analysis of 11 cohorts with over 22,000 children. Experience with this case

study of the GA<sup>2</sup>LEN network of excellence showed that combining and analysing individual participant data from birth cohorts is a laborious and time consuming process in terms of data collection, data management and harmonising variables of interest across the cohorts.

## 2 Key Responsibilities

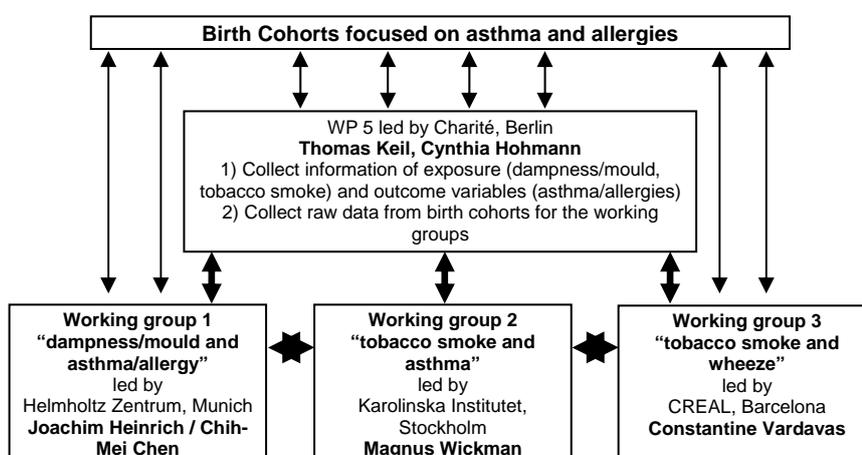
For the complex task of coordinating and combining different birth cohorts a clear definition of key responsibilities is necessary. While the work package leader is in charge of coordination processes, the working groups (WG) have the key responsibility for final analysis.

Interpretation of data is the joint responsibility of both research groups. All researchers will participate at different stages in the process of manuscript preparation and revision.

### 2.1 Organisational Issues related to current ENRIECO Project

For the current project, the work package leader, Charité University Medical Center (Berlin, Germany), is responsible for the coordination and communication process between birth cohorts, as well as for the preparation of the datasets and actual pooling of data. The Helmholtz Zentrum (Munich, Germany) (WG1 on dampness/mould), the Karolinska Institutet (Stockholm, Sweden) (WG2 on tobacco smoke and asthma), and the University of Crete (Greece) (WG3 on tobacco smoke and wheezing) will be responsible for data analysis.

Figure 1. Organisational Structure of WP5



### 1. Key Responsibility for Coordination Process:

- ⇒ Contacting potential cohorts, providing them with project information
- ⇒ Preparing data collection sheets, sending them to interested cohorts
- ⇒ Preparing overviews on available data for final data analysis decisions of work package leaders
- ⇒ Contacting cohorts with available information for individual participant data
- ⇒ Collection of individual participant data, harmonising of variables
- ⇒ Providing work groups with available and harmonised raw variables to conduct data analysis

### 2. Key Responsibility for Final Analysis:

- ⇒ Preparing analysis and agreement plan
- ⇒ Decision of final changes and adaptations of analysis according to available data/for which variables individual participant data will be ordered for final analysis
- ⇒ Conducting final analysis

## 2.2 Trust Building and Agreement

As participating cohorts are considered to provide valuable, previously collected, longitudinal data, trust building between the cohort's study investigators represents one of the most important parts of the communication process led by the work groups leader. For first contact, introduction and discussion of the project, personal meetings should be arranged.

From the beginning, transparency of the data collection and analysis process is essential. There should be a clear analysis plan that everyone agrees on. Decision of leadership of the specific project and authorship order on publications should also be clear from the start (e.g. two authorships per cohort in addition to those who actively coordinate the project, perform data analyses and write the manuscript). During the process, regular updates of data collection and analysis status are recommended. Preliminary results should be discussed personally (e.g. face-to-face airport meetings or telephone conferences) before performing final analyses.

## 3 Procedure of Database Building

To achieve combined analysis of individual participant data, several steps are required which will be described in the following.

### **3.1 Step 1: Willingness to Participate**

After the principle investigator of the cohorts were contacted by the work package leader and informed about the study, a personal meeting with all interested cohorts is advisable for general discussion of the project's content and possible analyses. This helps to substantiate the analyses strategy and to elaborate the analysis plan and a Memorandum of Understanding. These documents are collected and centrally circulated by the coordinator among interested cohorts for final decision on participation.

### **3.2 Step 2: Eligibility of Cohorts**

Interested cohorts receive a collection sheet for available data by the work package leader. They have to precisely fill in the definition of their birth cohort variables, answering categories, and follow-ups at which data was collected. After receiving this information from all cohorts, the coordinator conducts overviews. These are sent to the PIs of each cohort for verification of correctness and completeness and to the working groups performing the analysis. The working groups check which cohorts can participate in planned analyses, and which analyses are possible with the available data. Variables for analyses are chosen by the WGs and coordinator. An overview of all selected variables will be sent to the PIs for double checking and final confirmation that these are the correct variables. Analysis plans are adapted if necessary and verified by PIs.

### **3.3 Step 3: Collection of Individual Participant Data**

Individual participant data, chosen for analyses, is collected by the coordinator. Participating cohorts receive a detailed agreement form, regulating data access, storage, management, analyses, and publication policy including authorship issues (see deliverable 7, ENRIECO WP5).

Variables are harmonised for each cohort separately. Afterwards, a database is built consisting of the harmonised variables (or raw variables if applicable) of all participating cohorts. This database is sent to the WGs performing analysis.

### **3.4 Step 4: Combined Analysis**

Analyses are conducted with data and procedures according to the analysis plan. Preliminary results are personally discussed in a common meeting with all participating PIs. If necessary, the work plan will be adapted and reasons must be documented. Final analyses are conducted after verification by PIs, work group leader and the coordinator.

### 3.5 Issues of Database Building related to current ENRIECO Project

Step 1. In March 2009, the GA<sup>2</sup>LEN birth cohorts BAMSE, PIAMA-NHS, GINI, LISA, DARC, C.O.N.E.R., Isle of Wight, AMICS-Menorca, KOALA and Leicester were contacted for first collection of detailed variable definitions for indoor dampness/mould exposure. Overviews were presented at the ENRIECO workshop of WP5 on 29 May. Representatives from several GA<sup>2</sup>LEN birth cohorts and the additional cohorts from ALSPAC (Bristol), BIB (Bradford), RHEA (Crete), Generation XXI (Porto), NINFEA (Turin), INMA (Spain) and Generation R (Rotterdam) were present at the ENRIECO meeting.

Step 2. In June and July 2009, the Charité as the coordinator of WP5 provided interested GA<sup>2</sup>LEN cohorts and cohorts from the ENRIECO meeting (named above) with analysis and agreement plans of the working groups. They received data sheets for the collection of the detailed definition of variables (see table 2).

These sheets have to be filled out by the end of August with a detailed description of available variable assessments in each cohort.

Step 3. From September to the end of October 2009 the collection of individual participants will take place. Harmonising of variables will be conducted until the end of December 2009 by the work coordinator.

Step 4. Finalised datasets will be sent to the working groups. From January to July 2010, preliminary and final analysis will be conducted by the working groups according to the specific analysis and agreement protocols.

#### **Figure 1. Gant Chart of Database Building ENRIECO, Time Schedule 2009/2010**

WP5 Protocol for database building

	March 2009	April 2009	May 2009	June 2009	July 2009	Aug 2009	Sep 2009	Oct 2009	Nov 2009	Dec 2009	Jan-July 2010
Step 1	contacting potential cohorts										
				providing cohorts with analysis and agreement plans							
Step 2	collection of detailed variable definition										
						variable overviews reconfirmed by PIs and WGs					
Step 3	collection of individual participant data										
									database building, harmonising of variables		
Step 4											conduct of analysis



**Table 2. Collection of Detailed Variable Definitions for Tobacco Smoke Exposure and Asthma**

**First Row: Wording of required variables.**

*Please use suggested variables as a guideline and adjust the wording according to your own assessments and/or add additional variables.*

**Following rows: Required follow-ups.**

*Please tick the appropriate boxes according to the time of assessment.*

<b>1. Second Hand Smoke Exposure (SHSE) in pregnancy and up to 6 months after birth of child</b>								
	1st trimester	2nd trimester	3rd trimester	entire pregnancy	at birth	1 month	2-3 mths	4-6 mths
<b>Maternal Smoking</b>								
maternal smoking								
cig/day								
changes of smoking habits								
what is smoked (e.g. filter-tip, cigarette...)?								
child exposed to ETS?								
<i>other variables?</i>								
<b>Paternal Smoking</b>								
paternal smoking								
cig/day								
changes of smoking habits								
what is smoked (e.g. filter-tip, cigarette, cigar, small cigar, pipe)?								
child exposed to ETS?								
<i>other variables?</i>								
<b>Other Second Hand Smoke Exposure (SHSE)</b>								
smoking of others at home?								
maternal smoke exposure in public								



<b>maternal smoke exposure at work</b>									
<i>other variables?</i>									
<b>2. Wheezing assessment age 0-2 years and 4-6 years; asthma assessment 4-6 years</b>									
	0-3 mths	4-6 mths	7-12 mths	11-18 mths	19-24 mths	-	4-5 yrs	5-6 yrs	6-7 yrs
<b>wheezing</b> <i>please enter precise wording for wheezing variable(s), e.g. current, ever, last 12 months</i>						-			
							4-5 yrs	5-6 yrs	6-7 yrs
<b>asthma medication</b>	-	-	-	-	-	-			
<b>doctor diagnosed asthma</b>	-	-	-	-	-	-			
<b>3. Confounders (please tick the according box)</b>									
	0-1 yr	2-3 yrs	3-4 yrs	4-5 yrs	5-6 yrs	6-7 yrs			
<b>ever had atopic eczema?</b> <i>(please enter precise wording, e.g. doctor diagnosed, parent reported)</i>									
<b>ever had allergic rhinitis?</b> <i>(please enter precise wording, e.g. doctor diagnosed, parent reported)</i>	-	-							
<b>birth weight</b>									
<b>gestational age</b>									
<b>mode of delivery</b>									
<b>family history of asthma and AR – parents</b>									
<b>family history of asthma and AR – siblings</b>									
<b>siblings at birth of child</b>									
<b>educational level of mother at birth of child</b>									
<b>educational level of father at birth of child</b>									
<b>breast feeding duration (in months)</b>									
<b>dampness in housing during 1st year of life</b>									