

EUCCONET-CHICOS Workshop

“Combining Birth Cohort Data – Lessons from Past and Ongoing Studies”

Centre of Research in Environmental Epidemiology (CREAL)

Barcelona, Spain

Summary

European birth and child cohort research has resulted in an extensive amount of data on child health outcomes and risk factors. This has led to efforts to combine data in order to increase statistical power when outcomes or exposure are rare, or when new study designs requiring large sample sizes are applied (e.g. in the field of genetics). Major challenges in combined studies result from difficulties in the identification of eligible cohorts, in administrative obstacles, willingness and restrictions in sharing data, and in the harmonisation and standardisation of analysis data, amongst others. The objective of this meeting was to discuss past and current methodology, opportunities and benefits of data combination. During the plenary session, speakers with experience in combining data as part the European projects ENRIECO, EAGLE and CHICOS presented issues encountered in their studies.

Parallel sessions focussed on case studies which are currently evaluating issues around data pooling in Fish consumption in pregnancy and fetal growth; Alcohol consumption during pregnancy and birth weight; Selected maternal occupations and fetal health; Social inequalities in preterm delivery; and Prenatal environmental exposures (POPs) and birth outcomes. These topics were chosen based on the current stage of various available case studies, their common issues in harmonisation of outcome data, and their relevance to public health issues in children. Sessions were led by participants who are currently conducting pooled studies as part of ongoing EC-funded and other projects.

The meeting was deemed very beneficial to get to the core of issues through face-to-face discussions with a variety of cohorts. It has led to a better understanding of data availability and to increased data contributions, and has stimulated future collaborations. Heterogeneity in data between cohorts varies greatly due to differences in questions, follow-up, country specific variables, etc. It was clear that for variables such as birth outcomes, social class, education, occupation, smoking, and dietary factors, harmonisation between cohorts is possible but requires detailed evaluation of differences in data collection methods and definitions used. In particular, in defining the outcome variables of common interest for the 5 case studies that were presented, the lack of consensus in definitions for preterm birth and small for gestational age was apparent.

Discussion ensued over the benefits of pooled over meta-analysis, with the former requiring a higher level of harmonisation between variables but giving more flexibility in the analyses afterwards, and the latter being applicable also when cohorts cannot transfer individual data for central analysis. It is recommended that case studies compare both approaches as part of their statistical analysis protocols.

Combining studies also depend on the willingness of cohorts to do so, which has seen a positive development over the years, with the increase of international collaborations. An important step has been made in further collaboration between cohorts to develop specific protocols for data combination and analysis.

Description of the scientific content of and discussion at the event

The workshop started with a plenary session presenting experience in data preparation and management, issues in formulating birth outcome definitions, co-variate harmonisation, and meta- versus pooled analysis. Following this, several case studies were discussed in detail during parallel sessions.

The case study **“Fish consumption during pregnancy and birth outcomes”** evaluates the effect of fish intake during pregnancy on birth outcomes in European birth cohorts. Representatives from nine participating birth cohorts were present, and five new cohorts showed interest to participate and are now preparing their data. Issues discussed concerned methodological aspects of data combination on exposure/outcome variables, as well as the statistical analysis plan. The discussion focused on the difficulties of harmonising data on fish intake and gestational age across European birth cohorts, the definition of exposure categories, the treatment of extreme variables, and the difficulties of pooled analysis. As a first step, the participants agreed on the implementation of meta-analysis on cohort specific models, with the use of a priori defined confounders. Future steps for the analysis were discussed during the session and were presented on the second day of the meeting (March 30th) during the plenary session. First results will be circulated among the participants by June 2012, and the time plan is to complete the analysis by October 2012.

Previous studies have shown that women with light alcohol use during pregnancy have decreased risks of preterm birth and low birth weight compared to abstainers. The purpose of the study **“Alcohol consumption during pregnancy and birth weight”** is to examine if these effects can be explained as artefacts arising from a “healthy drinker” effect. During the session, preliminary descriptive results were presented on gestational age, birth weight, reproductive history and alcohol intake during pregnancy, following discussion of some of the problems in data cleaning and categorisation of the variables.

Some of the problems discussed were how to clean data on birth weight and gestational age and how to choose a reference for classifying infants as being small for gestational age. Different methods for estimating gestational age and how timing of ultra sound scans may influence precision on estimates were also discussed. Questions about these problems had been sent to the participants in advance to allow them to find this information before the workshop. Another central problem is how to create a uniform classification of alcohol intake. For most cohorts, alcohol use during pregnancy was assessed through questions with intake in categories; however, the choices of categories differ widely between cohorts. Discussed were various ways to estimate and impute the underlying continuous mean alcohol intake, and also how varying policies and alcohol culture influenced reported alcohol use during pregnancy and discussed whether timing of questions (first, second or third trimester of pregnancy) could influence results.

The **“Selected maternal occupations and fetal health”** study addressed the different issues including:

- Occupational coding: classification of occupational groups was based on three-digit ISCO88 codes, but the meeting decided that it should be based on 4-digits ISCO88 codes in order to make a more detailed classification. It would be beneficial to create a new group for “shop assistants”. Definitions of occupations will be sent to cohorts in order to see if they can be comparable between cohorts; Also, the distributions of each occupational group will be sent to cohorts to verify if

they fit with the expected percentages in each country. A new working variable for based on maternity leave will be created.

- Outcome data: Birth weight will be used as a continuous exposure. In the case of small for gestational it was decided that a document should be written on how to create this variable to be used for all the case studies on birth outcomes (issue to be addressed: use external or internal references, or create one European reference giving weights to the different cohorts).
- Harmonization of covariates: Different models will be constructed: first including maternal age, parity, child sex, maternal BMI, gestational age and gestational age²; and secondly including maternal education, alcohol, smoking, and ethnicity. A model will be constructed using a common variable for all cohorts and then, a sensitivity analysis will be done using cohort-specific variable.
- An analysis will be developed for comparison of 1-9 occupational groups (ej. health, day-care, cleaning) with “unexposed” occupational groups (ej. teachers, clerks). All analysis will be restricted to women who worked during pregnancy.
- Common covariates for all cohorts will be included in the models.
- Control groups: a general one could be “white collar workers”; for hairdressers could be one or several occupational groups from the same socio-economic status (shop assistants).
- Sensitivity analysis: using cohort-specific variables, exclude cohort by cohort, exclude outliers of birth weight, and exclude diabetic mothers.
- Stratification: by start of enrolment, geographical differences and type of delivery.

The objective of the **“Social inequalities in preterm delivery”** study is to gain a better understanding of why women with longer education have lower risk of preterm birth by comparing data between cohorts and examining contextual effects. The session was structured in two parts; the first dealt with gestational age and preterm birth rates, the second with availability and harmonisation of socio-economic markers.

In the first part, data on gestational age at study was presented which highlighted how study designs influence the possibilities for studying preterm birth. Also discussed were methods of estimating gestational age and how this was done in different cohorts. This discussion helped to clarify some of the problems we had been unable to solve by email. In particular, it helped clarify which methods were used to determine gestational age and whether the cohorts had already performed a clean-up of the data. In the second part, data on socio-economic and demographic variables were presented and their interpretations were discussed, in particular in the context of pregnant women. Discussed was which socio-economic data are available in each cohort and it was found that more information is available in the cohort than was received by the study leaders so far.

The variables considered were income, education, occupation-based socio-economic status, labour market status, ethnicity/country of origin and marital status. The only variable that is available in all cohorts is maternal education, but the grouping of this variable varies a lot between cohorts. This is partly due to different education systems but also to different questionnaires. Representatives from the ABCD and Generation R cohorts, which both includes a high number of immigrants, made the point that education is difficult to interpret for immigrant women, and that it may be necessary to restrict analysis sample to non-immigrants.

Persistent organic pollutants (POPs) are measured in different matrices, which provides complications because of variations in concentration. Data analysis showed slight differences in coefficients between those published by the ENRIECO project and those found by CHICOS. During the study session **“Prenatal environmental exposures (POPs) and birth outcomes”**, a solution was proposed to plot the

generic conversion rate and apply this to all cohorts. Several other issues were discussed:

- Lipid adjustment: a first analysis without lipid adjustment will be performed and then a sensitivity analysis using lipid-adjusted values will be done;
- It was found that co-variables are identified in cohorts in the similar ways apart from social class. For this, cohort-specific definitions are used, but with a common format/principle;
- Multiple imputation of missing values has been performed generating ten datasets for each cohort. Adjusted results from imputed datasets are half from those determined by ENRIECO?! It could be that the data were not the same, but most likely to be affected by statistical method used, and the random effect of PCB;
- PCB153 (a polychlorinated biphenyls found in the environment and electrical equipment) shows a linear decline in birth weight;
- A general document will be drafted on how to define "small for gestational age" variable in order to be used for all the case studies on birth outcomes.

Assessment of the results and impact of the event on the future directions of the field

The birth cohort workshop helped to stimulate contacts between the birth cohorts and participation in other ongoing case studies. A great benefit of the face-to-face meetings has been that case study leaders were able to identify common problems much more easily. A strong experience with all ongoing data combining studies is that face-to-face contact is extremely important to improve the involvement of the cohorts and leads to increased willingness to share data.

Specifically for the alcohol case-study, it meant that study leads were able to get in touch with three cohorts that had not been reached because of errors in contact information. The discussion of the alcohol variables emphasised that development of similar questionnaires is a necessary part of being able to combine data from different cohorts in the future. For POPs, problems identified include the use of common versus cohort specific variables and imputation.

The formulation of definitions is one of the main points of attention of the workshop, which will be taken forward in various study areas. For occupation studies, cohorts will continue the discussion on comparability of occupation definitions internationally, and develop new categories to harmonise data collection. The discussion on outcome measurements is to be continued to see whether common ground exists for SGA (small for gestational age) and GA (gestational age).

Overall, the workshop was very well received by the delegates and organisers alike, and has benefited in-depth discussion that the case studies needed. The cohorts were actively involved in these sessions: these sessions were not just about presenting results, but about discussing the data with them and agreeing analysis protocols. It has contributed to establishing new collaborative links between cohorts, new data contribution to the current studies, and a push for developing guidelines addressing definitions. A lot of input is being received and need to be considered before further decisions to be made, including the development of guidelines for international use.

Many studies were founded on data obtained through the ENRIECO coordination action, and the currently running CHICOS action addressing child health provides another major platform for birth cohorts. This workshop has stimulated new involvement of cohorts in this project, who will continue the discussions during workshops of CHICOS in Turin (May 2012) and EUCCONET in Paris (October 2012). The meeting in Barcelona has also revived the discussion on how a continuation of the cohort interaction is envisioned after the round-up of activities associated with initiatives such as EUCCONET and CHICOS. To begin with, documentation of the current case studies will be published to prevent future studies reinventing the wheel.

Annex I – Workshop Programme

Day 1 Thursday 29th March

13:00—14:30 Plenary (Room Xipre)

Welcome and Introduction to CHICOS case studies – Martine Vrijheid

Lessons from combined studies in ENRIECO and other European projects

1. Data preparation and management – experience from ENRIECO and EAGLE meta-analyses – Marie Standl
2. Outcome definition - Anne-Marie Nybo Andersen
3. Covariates – Maribel Casas
4. Analysis issues – experience from ENRIECO - Mark Nieuwenhuijsen

14:30 – 15:30 **Parallel Sessions for discussion of ongoing case studies**

Presentation of progress/first results, discussion of protocols and harmonisation of variables

1. Fish consumption in pregnancy and fetal growth (Room 150.7 – Rodes)
2. Alcohol consumption during pregnancy and birth weight (Room 182 – Sardenya i Corsega)
3. Selected maternal occupations and fetal health (Room 173.06 – Xipre)

15:30-16:00 **Break**

16:00-18:00 **Parallel Sessions 1,2,3 continued**

DINNER

Day 2 Friday 30th March

9:00-10:30 Parallel Sessions 4, 5

Presentation of progress/first results, discussion of protocols and harmonisation of variables

4. Social inequalities in preterm delivery (Room 182 – Sardenya i Corsega)
5. Prenatal environmental exposures (POPs) and birth outcomes (Room 173.06 – Xipre)

10:30-11:00 *Break*

11:00-12:30 Parallel Sessions 4,5 continued

12:30-13:00 Plenary (Room Xipre)

Presentation of studies – short feedback from each study

13:00-14:00 *Lunch*

14:00-15:30 Plenary (Room Xipre)

Presentation of studies – continued

Summary of experiences and overall issues in data pooling (conclusions on issues from day 1)

Guidelines for future combined studies

15:30 **end meeting**

Parallel session 1, 29th March:

Fish intake during pregnancy and birth outcomes

14:30 – 15:30 Part 1

- Description of exposure and outcome variables

15:30-16:00 *Break*

16:00 - 18:00 Part 2

- Description of confounding variables
- First results: Univariate analysis
- First results: Multivariable analysis
- Discussion: Harmonization of exposure- data, selection of confounders, cohort-specific effects

Parallel session 2, 29th March:

Alcohol consumption during pregnancy and birth weight and preterm birth

14:30-15:30 Part 1

- Presentation of participants
- Background for study
- Identification of cohorts
- Presentation of available pregnancy and delivery data for each cohort:
 - Gestational age at enrolment
 - Gestational age
 - Birth weight and definition of small for gestational age
 - Data on reproductive experience
- Discussion: How can we operationalise the pregnancy data?

15:30-16:00 *Break*

16:00-18:00 Part 2

- Presentation of data on alcohol use for each cohort:
 - Alcohol use during pregnancy
 - Alcohol use before pregnancy
 - Paternal alcohol use
- Discussion: How can we operationalise the alcohol data?
- How we plan to move forward with the analyses
- Discussion: Input for the analysis plan

Parallel session 3, 29th March: Selected maternal occupations and fetal health

14.30-15:30 Part 1

- Presentation of the case study participants
- Progress of the case study (Maribel Casas):
 - Participating cohorts, datasets received, progress up to now
- Occupational coding
 - Description of data received
 - Pre-specified analysis groups
 - Unclassified jobs
 - Control population
 - JEM on physical load (Rotterdam)
- Outcome data
 - Description of data received
 - Exclusions
 - Gestational age definition

15:30-16:00 *Break*

16:00-18:00 Part 2

- Harmonization of covariates
 - Description of data received
 - Selection of confounders for main analysis
- Analysis protocol
 - First results comparing birth weight in different occupational groups
 - Definition of occupational groups and control groups
- Next steps and time line
 - Completion current analysis
 - Other JEMs

Parallel session 4, 30th March: Social inequalities in preterm delivery

9:00-10:30 Part 1

- Presentation of participants
- Background for study
- Identification of cohorts
- Description of included cohorts
- Presentation of available pregnancy and delivery data for each cohort:
 - Gestational age at enrolment
 - Gestational age
 - Other data on pregnancy
- Discussion: How can we operationalise the pregnancy data?
- Presentation of available socio-economic data and potential mediators for each cohort:
 - Socio-economic markers
 - Lifestyle factors
- Discussion: How can we operationalise the socio-economic data?

10:30-11:00 *Break*

11:00-12:30 Part 2

- Presentation of analysis plan
 - Comparative study
 - Mediation study
- Presentation of preliminary results
- Discussion of analysis plan

Parallel session 5, 30th March: Persistent organic pollutants and birth outcomes

9:00-10:30

- Presentation of the case study participants
- Progress of the case study:
 - Participating cohorts, datasets received, progress up to now
- ENRIECO datasets:
 - Descriptive analysis
 - Coefficients obtained vs Govarts et al. paper
- CHICOS datasets:
 - Imputation
 - Descriptive analysis
 - Cohort specific conversion factors

10:30-11:00 Break

11:00-12:30

- Analysis protocol:
 - First results with birth weight
 - Further analyses
- Next steps and time line:
 - Completion current analysis
- Progress of other POPs case studies:
 - CHICOS POPs case study and respiratory effects (Mireia Gascon)
 - OBELIX: POPs and growth (Nina Iszatt)