

**EUROlinkCAT WP3 - Mortality associated with congenital anomalies**

**Statistical analysis plan**

**WP3 leads:** Judith Rankin (UNEW), Anna Pierini (CNR-IFC)

**Researchers:** Svetlana Glinianaia (UNEW), Michele Santoro (CNR-IFC), Alessio Coi (CNR-IFC)

In collaboration with WP2 team

**WP2 lead:** Maria Loane (Ulster University, UU), James Densem (BIOMEDical Computing Ltd), Joachim Tan (Queen Mary University of London (QMUL))

Statistical expertise and advice will be provided by Joan Morris (QMUL).

**The aim** of this analysis plan is to facilitate the work of the WP2 team on the common syntax scripts which will guide the registries in their pre-specified local analyses and to provide each participating WP3 registry with the detailed stepwise instructions for these analyses and the dummy tables for completion.

**DOCUMENTATION**

Please ensure you keep a detailed log of all the steps carried out as part of this project. It is likely we will identify differences in the findings between the registries/locations and we may need to go back to this log to determine whether they are true differences or whether they are artefacts of the data.

**preparing the rEgistry dataset**

1. Each registry needs to extract a file of all cases with a major congenital anomaly irrespective of the outcome from their local EDMP (use the EDMP Version 6.10, 11 January 2016) from **1st January 1995** (or the earliest available year) **until 31st December 2014** (Appendix 1 for data availability by time period for participating registries) – see EDMP Step-by-Step Instructions written by WP2 team.

Include all pregnancies resulting in a:

1. fetal death at ≥ 20 weeks of gestation (FD);
2. termination of pregnancy for fetal anomaly (TOPFA) at any gestational age;
3. live birth (LB)
4. This file (named ‘*Registry number*\_cases\_95-14”) will include all EUROCAT variables listed in Table 1 (and also on page 21 in ‘EDMP Step by Step Instructions’ document produced by WP2 team) and must be stored and archived locally, preferably on another server/pc.

**Linkage of live birth data WITH DEATH DATA (as specified in the WP3 Protocol)**

Each registry will organise the **linkage of all live born cases** included in the dataset specified above to **deaths** that occurredfrom **1st January 1995** (or earliest available year if later than 1995) **to 31st December 2015.**

**STage 1. Assessing Quality of EUROCAT and Mortality data** - see a separate document produced by WP2 team on EDMP Step-by-Step Instructions and cross tabulations.

**STAGE 2. Statistical Analysis**

**Frequency tables (descriptive statistics) for linked live births**

Participating registries will run pre-specified analysesusing individual-based data which will follow this analysis plan and the WP2 team syntax scripts. This will include cross-tabulations for outcomes of interest listed below. The dummy output tables for completion by the registries are drafted in the Excel spreadsheets (see Excel file ‘**WP3\_stage 2\_stat analysis\_final’)**.

1. **For all linked live births and those resulting in deaths the tables for the following EUROCAT variables will be produced:**
2. ‘**T 1 – LBs’ (Linked live births with congenital anomalies and those resulting in deaths by age groups’)**
* by number of babies delivered - ‘NBRBABY’

The following tables will be produced for a) singletons and b) all linked LBs (singletons and multiples):

**EUROCAT recoded variables (see Table 2 on WP3 Common Data Model):**

* by year of delivery – ‘Yeargp’ recoded variable;
* by maternal age categories - ‘matage\_gr’ variable;
* by sex of child - ‘L\_CH\_SEX’ variable (created for linked cases);
* by birth weight categories - ‘BW\_gp’ variable;
* by gestational age categories - ‘GA\_disc\_gp’ variable;
* by maternal BMI categories- ‘BMI\_gr’ variable;

**Non**- **EUROCAT SES variables:**

* by SES variables – L\_MAT\_EDUC (maternal education), L\_MAT\_OCC (Maternal occupation), L\_MATDEPR\_IND (Deprivation index at maternal residence in quintiles), L\_MATMAR\_STA (Maternal marital status), L\_MAT\_CTRY\_B (Maternal country of birth/ place of birth/ country of origin) – see Table 2.
1. **T 2 – Prenatal DS’** – The table stratifies linked deaths for age at death groups up to 1 year of age (0-6 days, 7-27 days and 28-364 days) and those at risk at death (i.e. alive at certain age) by four subgroups of isolated congenital anomaly, i.e. spina bifida, transposition of great arteries, diaphragmatic hernia and gastroschisis (al6, al19, al48 and al50) and by prenatal diagnosis (‘WHENDISC’ variable=6 (prenatally diagnosed for LBs) stratifying by gestational age at prenatal diagnosis (<22, 22-31 and 32+ weeks – ‘GA\_disc\_gp’ recoded variable) vs postnatally diagnosed - ‘WHENDISC’=1-5 or 10).

Separate tables will be completed for singleton LBs and for all LBs (singletons and multiples) for the time periods 1995-2004 and 2005-2014 (four tables).

1. **For linked deaths only the following tables will be produced:**
2. ‘**T 3 - linked deaths’ - Deaths by age at death group and year of delivery**

Frequency of deaths by year of birth and child’s age at death group (first 24 hrs; 1 complete day; 2 complete days; 3 complete days; 4 complete days; 5 complete days; 6 complete days; 7-27 days; < 1 month (use if days unavailable); 28 days to < 3 months (28-90 days); 3-5 months (91-182 days); 6-11 months (183-364 days), < 1 year (use if days/months unavailable); 1 complete year (365-729 days); 2 complete years (730-1094 days); 3 complete years (1095-1460 days); 4 complete years (1461-1825 days); 5 complete years (1826-2190 days); 6 complete years (2191-2555 days); 7 complete years (2556-2921 days); 8 complete years (2922-3286 days); 9 complete years (3287-3651 days); alive on 10th birthday or on 31.12.2015 (whichever earlier).

1. Singletons;
2. All linked live births (singletons and multiples).
3. ‘**T 4A – Cause of death (infant)’** – **Distribution of underlying causes of death for selected congenital anomaly subgroups, all MM groups unless otherwise stated**

Data for all linked live births (singletons and multiples) resulting in deaths will be used for Tables 4A, 4B and 4C.

a) ) Neonatal Deaths (0-27 days)

b) Postneonatal Deaths (28-364 days)

‘**T 4B – Cause of death (child)’** – for deaths at 1-9 years: **Distribution of underlying causes of death for selected congenital anomaly subgroups, all MM groups unless otherwise stated**

**T 4C - Cause of death accuracy** - **Analysis of the accuracy of Cause of Death (CoD) matching for specific congenital anomaly subgroups, all MM groups unless otherwise stated.**

**KAPLAN-MEIER Survival analysis**

1. **Kaplan-Meier (K-M) Survival analysis up to аge 10 years by subgroups of congenital anomalies:**

**‘T 5 – Survival analysis’ (T 5A –T 5I)**

Linked live births will be used for these analyses.

Groups for age at death to be used for the K-M survival analyses:

0-6 days of age, 7-27 days, 1-2 months (28-90 days), 3-5 months (91-182 days), 6-11 months (183-364 days), and then for each year after the first year of life up to 10 years (see Excel tables T 5A-T C). For **T 5D –** after the first year the years are combined to 1-4 years and 5-9 years.

Subjects at risk for each age group, number of deaths, K-M survival estimates (% with 95% CI) will be provided. Surviving children with congenital anomalies should be censored at their 10th birthday or at 31 Dec 2015 if the follow-up was shorter than 10 years, or if they have migrated outside the study area before the end of follow-up.

**‘T 5A – Survival analysis’ –** The first series of the K-M survival analysis will be run for live births with an isolated anomaly (coded as N, A, R and I in 'mult\_malf' EDMP derived variable: N: NTD isolated; A: isolated cardiac; R: isolated renal; I: isolated other) by a selected congenital anomaly subgroup (proposed in Table 4 below based on clinical relevance and on the prevalence of about 1/10,000).

1. Singletons;
2. All linked live births (singletons and multiples).

**‘T 5B – Survival analysis’ –** this series of the K-M survival analysis will be run for all structural anomalies (select isolated cases using A, N, R and I 'mult\_malf' codes and potential multiple anomalies using 'M' code).

1. Singletons;
2. All linked live births (singletons and multiples).

**‘T 5C – Survival analysis’**

**All anomalies (all mult\_malf groups included)**

1. Singletons;
2. All linked live births (singletons and multiples).

**‘T 5D – Survival** **analysis’**

**All linked live births (singletons and multiples) - Down syndrome**

Select those with Down syndrome and categorise:

a) any with Down syndrome (al89 - no exclusions);

b) Down syndrome with CHD and gastrointestinal anomaly (al89 + al17 +al40)

c) Down syndrome with CHD (al89 + any CHD (al17) but not with gastrointestinal anomaly (exclude those with al40);

d) Down syndrome with gastrointestinal anomaly (al89 +al40) and not with CHD (exclude those with al17);

f) Down syndrome with neither CHD nor gastrointestinal anomaly (exclude those with codes al17 and al40);

**‘T 5E – Survival** **analysis’**

**All linked live births (singletons and multiples) - Genetic syndromes and other rare congenital anomalies**

No selection by ‘mult\_malf’ code, use the subgroup numbers ‘al’ for existing subgroups and the subgroup numbers ‘aud’ for new subgroups for inclusion at local level.

Three next series of Kaplan-Meier Survival analysis (**T 5E, T 5F, T 5G**) investigate the survival up to 10 years by subgroups of congenital anomalies and by time period (1995-2004 and 2005-2014) for three different groups of congenital anomalies:

 **‘T 5F – Survival** **analysis’**

**Kaplan-Meier Survival analysis up to 10 years by subgroups of congenital anomalies - time trends (1995-2004 and 2005-2014)**

**All linked live births (singletons and multiples) - isolated anomaly** (cases should be selected using the codes N, A, R and I 'mult\_malf' codes: N: NTD isolated; A: isolated cardiac; R: isolated renal; I: isolated other)

1. 1995-2004
2. 2005-2014

**‘T 5G – Survival** **analysis’**

**All linked live births (singletons and multiples) - All structural anomalies** (select isolated cases using A, N, R and I 'mult\_malf' codes and potential multiple anomalies using 'M' code)

1. 1995-2004
2. 2005-2014

**‘T 5H – Survival** **analysis’**

**All linked live births (singletons and multiples) - all anomalies (all mult\_malf groups included)**

1. 1995-2004
2. 2005-2014

**‘T 5I – Survival** **analysis’**

**All linked live births (singletons and multiples) - Down syndrome with or without additional anomalies**

1. 1995-2004
2. 2005-2014

**Cox proportional hazards regression**

Survival analysis to explore the associations between the risk factors listed below and the congenital anomaly subgroup-specific survival by age groups will be run using Cox proportional hazards regression. The adherence to the proportional hazards assumption, i.e. to examine that there was no evidence of violation, will be tested using Schoenfeld residuals.

**‘T 6A – Cox regression’ - the analysis of the effect of prenatal diagnosis on survival during the first year of life;**

**Results of the Cox proportional hazards regression - for survival up to 7 days of age, 28 days and 365 days**

**Model 1:** the analysis of the effect of prenatal diagnosis on survival during the first year of life (isolated anomalies only including those with 'mult\_malf' codes N, A, and I)

For:

1. Singletons;
2. All linked live births (singletons and multiples).

**Prenatal diagnosis:** Yes: WHENDISC=6; No: WHENDISC=1-5, or 10; exclude those with WHENDISC=9; use AGEDISC variable (gest. age at diagnosis, weeks) to categorise gestational age at prenatal diagnosis.

We aim to analyse the effect of prenatal diagnosis [‘No’ (reference) vs ‘Yes’] on survival up to 365 days for cases of isolated spina bifida, transposition of great arteries, diaphragmatic hernia and gastroschisis (al6, al19, al48 and al50).

Gestational age at diagnosis (AGEDISC variable) will be grouped into: at <22 weeks, 22-31 weeks and 32+ weeks, as earlier gestational age at diagnosis is associated with the lower survival (diaphragmatic hernia).

The adjusted analysis of the effect of prenatal diagnosis will include the following factors (see **‘T 6A – Cox regression’)**:

Byear - year of birth [‘1995-2004’ (reference) vs ‘2005-2014’],

SEX - sex of child [male (reference) vs female],

AGEMO - maternal age (20-34 (reference) vs < 20 years and 35+ years],

GESTLENGTH - gestational age at delivery [37+ weeks (reference) vs <28 weeks, 28-31, 32-36 weeks],

SES (the SES variables will be registry-specific, for example, L\_MATDEPR\_IND for the BINOCAR registries, 5th quintile (least deprived) – references vs 1st quintile (most deprived) and 2, 3 and 4 quintiles.

WEIGHT - Birth weight: 2500-3999 (reference) vs <1000, 1000-1499, 1500-2499 and 4000+g – only to be used in the unadjusted analyses, do not include in the multivariate models.

**For all T 6 tables**, the unadjusted and adjusted models will be run for each of the factors listed above (except for ‘prenatal diagnosis’ which relates to T 6A only) reporting crude β coefficients with the standard errors and the adjusted β coefficients with the standard errors respectively. Crude global PH test p-value will be reported for each unadjusted analysis and adjusted model global PH test (p-value) will be reported for the adjusted models for each outcome and anomaly.

We will follow the usual EUROCAT rule to exclude registries from studies on prenatal diagnosis if unknown or missing is more than 20% of the cases.

**‘T 6B – Cox regression’**

**Model 2:** the analysis of the effect of the listed risk factors on survival up to 365 days (isolated anomalies only based on the 'mult\_malf' codes A,N, R and I - except al6, al19, al48 and al50 analysed in T 6A ).

For:

1. Singletons;
2. All linked live births (singletons and multiples)

**‘T 6C – Cox regression’**

**Model 3:** the analysis of the effect of the listed risk factors on survival up to 365 days - include all structural anomalies (select isolated cases using A, N, R and I 'mult\_malf' codes and potential multiple anomalies using 'M' code)

For:

1. Singletons;
2. All linked live births (singletons and multiples).

**‘T 6D – Cox regression’**

**Model 4**: the analysis of the effect of the listed risk factors on survival up to 365 days - include all anomalies (all mult\_malf groups included)

For:

1. Singletons;
2. All linked live births (singletons and multiples).

**‘T 6E – Cox regression’**

**Model 5**: the analysis of the effect of the listed risk factors on survival up to 365 days - Down syndrome with or without additional anomalies – for ‘All linked live births (singletons and multiples)’

**‘T 6F – Cox regression’**

**Model 6:** the analysis of the effect of the listed risk factors on survival during 1-9 years of age - include all cases with isolated anomalies (include only those with 'mult\_malf' codes N, A, R and I) – for ‘All linked live births (singletons and multiples)’

**‘T 6G – Cox regression’**

**Model 7:** the analysis of the effect of the listed risk factors on survival during 1-9 years of age - include all structural anomalies (include isolated cases using A, N, R and I 'mult\_malf' codes and potential multiple anomalies using 'M' code) – for ‘All linked live births (singletons and multiples)’

**‘T 6H – Cox regression’**

**Model 8:** the analysis of the effect of the listed risk factors on survival during 1-9 years of age - include all anomalies (all mult\_malf groups included)

**‘T 6I – Cox regression’**

**Model 9:** the analysis of the effect of the listed risk factors on survival during 1-9 years of age - Down syndrome with or without additional anomalies – for ‘All linked live births (singletons and multiples)’

**Tables for the analysis of the geographic variation in survival across Europe (see the description in the Addendum on page 21)**

**‘T 7A– Population Mortality’** - Resident populations (0-9 years) and deaths in the background registry population, by year of death and age at death

**‘T 7B –** **Pop Neonatal Mortality’**– Neonatal deaths and stillbirths in the background registry population, by year of death**‘**

**T 8A Adjusted survival’ -** Kaplan-Meier survival estimates at 7, 28 and 365 days adjusted by 1 risk factor at a time, for cases of isolated spina bifida, transposition of great arteries, diaphragmatic hernia and gastroschisis (al6, al19, al48 and al50)

**‘T 8B Adjusted survival’ –** 1- and 10-year Kaplan-Meier survival estimates for isolated structural anomalies adjusted by 1 risk factor at a time

**‘T 8C Adjusted survival’ –** 1- and 10-year Kaplan-Meier survival estimates for Down syndrome adjusted by 1 risk factor at a time

The number of children in the background population and the number of deaths by year of death and age at death should be provided using mortality data or regional/national published statistics if available.

**analysis of the aggregated data**

The WP3 team will analyse the aggregated data submitted to the Central Results Repository (CRR) by each of the participating registries (aggregated tables and analytical results from the independent standardised databases). We will combine estimates produced by each individual registry in a meta-analysis using a random effects model.

**Table 1.** Master table of the list of the required variables (in addition to all the subgroups codes provided in EDMP)

|  |  |
| --- | --- |
|  | **EUROCAT core variables, one row of data per case (see also** [**http://www.eurocat-network.eu/aboutus/datacollection/guidelinesforregistration/guide1\_4**](http://www.eurocat-network.eu/aboutus/datacollection/guidelinesforregistration/guide1_4)**)** |
|  | **EDMP Core variables (shaded in blue)** |
| **Baby and Mother – Variables 1 to 18**  |
|  | **Variable** | **Description** |
| 1 | CENTRE | Centre Number |
| 2 | NUMLOC | Local ID of case |
| 3 | BIRTH\_DATE | Date of Birth |
| 4 | SEX | Sex |
| 5\*\* | NBRBABY | Number of babies delivered |
| 6 | SP\_TWIN  | Specify twin type of birth, like or unlike, zygosity  |
| 7 | NBRMALF | Number of malformed in multiple set |
| 8 | TYPE | Type of birth |
| 9 | CIVREG | Civil registration status |
| 10 | WEIGHT | Birth weight |
| 11 | GESTLENGTH | Length of gestation in completed weeks |
| 12 | SURVIVAL | Survival beyond one week of age |
| 13 | DEATH\_DATE | Date of death |
| 14  | DATEMO  | Date of birth of mother  |
| 15 | AGEMO | Age of mother at delivery |
| 16\* | BMI  | Maternal Body Mass Index  |
| 17  | RESIDMO  | Mother’s residence code  |
| **Diagnosis – Variables 19 to 57**  |
| 19\*\* | WHENDISC  | When discovered |
| 20 | CONDISC | Condition at discovery  |
| 21  | AGEDISC  | If prenatally diagnosed, gestational age at discovery |
| 24  | KARYO  | Karyotype of infant/fetus  |
| 25  | SP\_KARYO  | Specify karyotype  |
| 26\* | GENTEST  | Genetic Test  |
| 27\* | SP\_GENTEST  | Specify genetic test  |
| 28 | PM  | Post mortem examination  |
| 29\*\* | SURGERY  | First surgery for malformation performed or planned  |
| 30 | SYNDROME  | Syndrome |
| 31 | SP\_SYNDROME  | Specify Syndrome  |
| 32 | MALFO1  | malformation |
| 33  | SP\_MALFO1  | Specify malformation  |
| 34  | MALFO2  | As MALFO1  |
| 35  | SP\_MALFO2  | Specify malformation  |
| 36  | MALFO3  | As MALFO1  |
| 37  | SP\_MALFO3  | Specify malformation  |
| 38  | MALFO4  | As MALFO1  |
| 39  | SP\_MALFO4  | Specify malformation  |
| 40  | MALFO5  | As MALFO1  |
| 41  | SP\_MALFO5  | Specify malformation  |
| 42 | MALFO6  | As MALFO1  |
| 43  | SP\_MALFO6  | Specify malformation  |
| 44  | MALFO7  | As MALFO1  |
| 45  | SP\_MALFO7  | Specify malformation  |
| 46  | MALFO8  | As MALFO1  |
| 47  | SP\_MALFO8  | Specify malformation  |
| 57# | OMIM  | OMIM code / Type of Mendelian Inheritance  |
| **Exposure – variables 58 to 78** |
| 58\*\* | ASSCONCEPT | Assisted conception (where available) |
| 59##  | OCCUPMO  | Mother’s occupation at time of conception  |
| **Sociodemographic – Variables 91 to 94**  |
| 91 | MATEDU | Maternal education  |
| 92  | SOCM  | Socioeconomic status of mother  |
| 93  | SOCF  | Socioeconomic status of father  |
| 94  | MIGRANT  | Migrant status  |
| **EDMP-derived variables** |
|  | Byear | Year of birth |
|  | birth\_type | Definitions of stillbirths and spontaneous abortions vary between regions. This variable recodes birth type according to EUROCAT’s specifications: cases with gestational age ≥ 20 weeks are re-coded as “stillbirths” (irrespective of the local definition of stillbirth/spontaneous abortion). |
|  | casestatus | Only cases with casestatus = 1 |
|  | al1-al108 | EUROCAT subgroups: (0 = No, 1 = Yes). Based on EUROCAT coding in Guide 1.4 |
|  | mult\_malf  | Algorithm for case classification into isolated and multiples |
|  | MM\_Final\_Verdict | This is the final aetiological classification following review of potential multiple cases by a panel of 3 geneticists |

\* New variable In Guide 1.4

\*\* Variable compatible with Guide 1.3, but coding has been extended/modified

# Variable name change only

## Guide 1.4 use ISCO-08 classifications

aShould be a calculated field using the difference between date of death and death of birth, it should be provided in days for infants who died after the first 24 hours.

bShould be recorded in addition to the exact age at death variables: L\_CH\_AGED\_H, L\_CH\_AGED\_D or L\_CH\_AGED\_M.

Table 2. Common Data Model – WP3

| ***Variable Name*** | ***Variable Definition and Instructions*** | ***Variable Format*** | ***Variable Values*** |
| --- | --- | --- | --- |
| **Core variable required for linkage to vital statistics or mortality databases and for calculating age at death** |
| L\_CH\_ID | Unique identifier of childA unique ID that links child to another database | As recorded locally |  |
| L\_CH\_DATE\_B | Child’s date of birth Needed to calculate age at death. | DDMonYYYY |  |
| L\_CH\_DATE\_D | Child’s date of deathNeeded to calculate age at death. | DDMonYYYY |  |
| **Variables relating to linkage** |
| L\_MATCH\_TYPE | Match with vital statistics database (or local health care databases) |  Numeric | 1 = Linked to national/vital statistics database – match2 = Linked to national/vital statistics database – non-match3 = Linked to mortality database only – match4 = Linked to mortality database only – non-match5 = EUROCAT death only |
| L\_CONFIDENCE | Strength of match with vital statistics or health care database. Use local data provider’s codes for assessing confidence that the case is correctly matched. If local code unavailable, use suggested coding (see Appendix for full details) | Numeric  | 1=Excellent2=Good3=Fair4=Poor9=Not Matched |
| L\_DATE\_LOST | Date lost to follow-up/ linkage (i.e. due to emigration, adoption or other reason) | DDMonYYYY | .=Not recorded by registry or not available for study |
| L\_YEAR\_LOST | Year lost to follow-up/ linkage (i.e. due to emigration, adoption or other reason) | YYYY | .=Not recorded by registry or not available for study |
| L\_AGEL\_D | Age lost to follow-up/ linkage in complete days | Numeric (1-4 digits) | .=Not recorded by registry or not available for study |
| **Standardised variables relating to child** |
| L\_CH\_YEAR\_B | Child’s year of birth  | YYYY |  |
| L\_CH\_SEX | Child’s sex | Numeric | 1 = Male 2 = Female 3=Indeterminate 9 = Not known.=Not recorded by registry or not available for study |
| L\_CH\_REG\_TYPE | Type of birth/ civil registration of baby | Numeric | 1 = Livebirth 2 = Stillbirth 9 = Not known.=Not recorded by registry or not available for study |
| L\_CH\_NON\_EUR | Citizenship/ Nationality of infant /country of origin | Numeric | 1= National2= Other European3 = Non-European4= Non-national (exact nationality not specified)9 = Not known. = Not recorded by registry or not available for study |
| L\_CH\_BW | Child’s birthweight (grams) | Numeric | 9999=Not known.=Not recorded by registry or not available for study |
| L\_CH\_GA\_B | Child’s gestational age at birth (completed weeks) | Numeric | 99=Not known.=Not recorded by registry or not available for study |
| **Variables relating to mortality** |
| L\_CH\_STATUS | Outcome status-Died = child is known to have died before 10th birthday or 31 Dec 2015 (whichever earlier)-Alive at 10th birthday = child was born on or before the 31st Dec 2005 and: (i) is definitively known to be alive on 10th birthday; or (ii) there is no information on death or lost to follow-up- Censored on 31st Dec 2015 = child was born on or after 1st Jan 2006 and:(i) is definitively known to be alive on 31st Dec 2015; or (ii) there is no information on death or lost to follow-up-Lost to follow up = child is lost to follow-up/ linkage (i.e. due to emigration, adoption or other reason) | Numeric | 1 = Died2 = Alive at 10th birthday3 = Censored on 31st Dec 20154 = Lost to follow up |
| L\_EXIT\_DATE | Date of last day in study (censored/lost/died/alive)- Date lost - if child was lost to follow up - Date of death - if child has died - Date of birth plus 3652 days (approx. 10th birthday) if child was born on or before the 31st Dec 2005 and:(i) is definitively known to be alive on 10th birthday; or (ii) there is no information on death or lost to follow-up- Date of last day of the study (31st Dec 2015) if child was born on or after 1st Jan 2006 and:(i) is definitively known to be alive on 31st Dec 2015; or (ii) there is no information on death or lost to follow-up | DDMonYYYY |  |
| L\_EXIT\_DAYS | Number of days child is in studyThis is calculated as the last date child was in the study (L\_EXIT\_DATE) minus the child’s birth date (L\_CH\_DATE\_B).  | Numeric (1-4 digits) |  |
| L\_CH\_YEAR\_D | Year of child’s death as recorded on the death certificate | YYYY |  |
| L\_CH\_AGED\_H | Age at death in complete hours for day 0 (first 24 hours) = applies to infants who died within the first 24 hoursThis variable can be used to check values recorded under variable L\_CH\_AGED\_D. A child who died within 24 hours but across two dates is coded “0” on L\_CH\_AGED\_D below. | Numeric(1-2 digits) | 0 = Died <1 hour after birth1 = Died 1 complete hour after birth 2 = Died 2 complete hours after birth Etc.23 = Died 23 complete hours after birth88 = Alive at 24 hours99 = Died within first 24 hours, but exact time unknown |
| L\_CH\_AGED\_D | Age at death in complete days (up to 10th birthday). A calculated field using the difference between date of death and death of birth i.e. subtract child’s date of birth from child’s date of death. | Numeric (1-4 digits) | 0 = died <24 hours after birth1 = died 1 complete day after birth 2 = died 2 complete days after birth Etc.8888 = Alive on 10th birthday or by end of study period, whichever is sooner9999 = Died before 10th birthday, but exact time unknown |
| L\_CAUSE\_D\_U | Underlying cause of death - diagnosis There can only be one underlying cause of death | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_P | Primary/ immediate cause of death - diagnosis  | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_C | Contributing cause of death - diagnosis | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_O1 | Other causes of death- diagnosis 1 | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_O2 | Other causes of death- diagnosis 2 | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_O3 | Other causes of death- diagnosis 3 | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_O4 | Other causes of death- diagnosis 4 | String | ICD9 or ICD10 code |
| L\_CAUSE\_D\_O5 | Other causes of death- diagnosis 5Additional “Other causes of death” can be recorded in variables L\_CAUSE\_D\_O6 to L\_CAUSE\_D\_15 (not listed) | String | ICD9 or ICD10 code |
| L\_CH\_PLACE\_D | Place of death  | Numeric | 1 = Home 2 = Hospital3 = Other9 = Not known.=Not recorded by registry or not available for study |
| **Variables relating to mother** |
| L\_MAT\_YEAR\_B | Maternal year of birth | YYYY | 99 = Not known. = Not recorded by registry or not available for study |
| L\_MATAGE\_B | Maternal age at infant's birth in completed yearsMay be used to link information held on the child in local health care databases | Numeric | 99=Not known . =Not recorded by registry or not available for study |
| L\_MULT\_BIRTH | Singleton or multiple birth | Numeric | 1=Singleton2=Twins3=Triplets or higher4= Multiple birth, number unknown9= Not known. =Not recorded by registry or not available for study |
| L\_MAT\_CTRY\_B | Maternal country of birth/ place of birth/ country of origin | Numeric | 1= National2= Other European3 = Non-European4= Non-national (exact nationality not specified)9 = Not known.=Not recorded by registry or not available for study |
| L\_MAT\_BMI | Maternal Body Mass Index (BMI) at first antenatal visit/at bookingExpected range 15 – 50 | Numeric(Whole number only) | Exact BMI value 97 = <3098 = >=3099 = Not known.=Not recorded by registry or not available for study |
| L\_MAT\_EDUC | Maternal education(UNESCO’s International Standard Classification of Education (ISCED)) | Numeric | 1 = Pre-primary /Primary 2 = Any secondary3 = Postsecondary (non-tertiary)4 = Tertiary5=No education9 = Not known. = Not recorded by registry or not available for study |
| L\_MAT\_OCC | Maternal occupation(Based on Paris (INSERM) coding classification) | Numeric | 1= Farmer2= Artisan (ex: baker) / shop-owner3= Professional 4= Intermediate 5= Administrative/public service 6= Business employees: shop assistants/ salesperson 7= Household and personal service 8= Skilled manual worker 9=Unskilled manual worker 0= No occupation declared / student 99= Not known . =Not recorded by registry or not available for study |
| L\_MATDEPR\_IND | Deprivation index at maternal residenceMultiple deprivations scores are ranked into quintiles where 1= Least deprived and 5 = Most deprived (coding scheme used in Wales & Basque Country) | Numeric | 1 = First quintile (Least deprived)2 = Second quintile3 = Third quintile4 = Fourth quintile5 = Fifth quintile (Most deprived)9 = Not known. =Not recorded by registry or not available for study |
| L\_PROXY\_SES | Proxy variable for SES This is registry-specific. Use the agreed proxy variable for each registry.Maternal education:* Tertiary/ post-secondary=High
* Any secondary = Middle
* Primary/ pre-primary/ No education = Low

Maternal occupation: * Professional = High
* Intermediate= Middle
* No occupation = Low

Multiple Deprivation Index* Quintile 1 (Least deprived)=High
* Quintiles 2-4= Middle
* Quintile 5 (Most deprived)= Low
 | Numeric | 1 = High2 = Middle 3 = Low9=Not known. = Not recorded by registry or not available for study |
| L\_MATMAR\_STA | Maternal marital status | Numeric | 1 = Single2 = Married/ Living together 3 = Widow4 = Divorced/ Separated9 = Not known. =Not recorded by registry or not available for study |
| **Recoded variables**  |
| Yeargp | Grouped year of birth1995/2004=1 2005/2014 =2 | Numeric | 1 = 1995-2004 2 = 2005-2014 |
| BMI\_gp | Grouped BMIBMI <10, code as unknownBMI >60, code as unknownBlank or missing, code as unknown | Numeric | 1 = <30 2 = 30+9 = unknown |
| BW\_gp | Grouped BirthweightBW <400g, code as unknownBW >7000g, code as unknownBlank or missing, code as unknown | Numeric | 1= <1000g2 = 1000-1499g 3 = 1500-2499g4 = 2500-3999g5 = 4000+ g9 = unknown |
| GA\_gp | Grouped Gestational ageGA <24 weeks, excluded from studyGA >45 weeks, code as unknownBlank or missing, code as unknown | Numeric | 1 = 24-27weeks 2 = 28-31 weeks3= 32-36 weeks4 = 37+ weeks9 = unknown |
| GA\_disc\_gp | Grouped Gestational age at discoveryGA at discovery <8 weeks, code as unknownGA at discovery >42 weeks, code as unknownBlank or missing, code as unknown | Numeric | 1= <22 weeks2 = 22-31 weeks3 = 32+ weeks9 = unknown |
| matage\_gp | Grouped Maternal age at infant’s birthMaternal age range 10-19 years, code =120-34 years, code=235-59 years, code=3All other values, blanks or missing, code=9  | Numeric | 1= <20 years2= 20-34 years3= 35+ years9=Not known |

**Appendix to Table 2:**

Guideline for coding L\_CONFIDENCE variable (for matching cases in different datasets)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All 3 variables available in both datasets for each individual child** | **Only 2 variables present in both datasets** **for each individual child** | **Only 1 variable available in both datasets for each individual child** |
| Unique ID | Matched | Not Matched | Matched | Matched | - | Matched | Matched | - | - | Matched | Not matched | Matched | - | - |
| Child’s date of birth | Matched | Matched | Not Matched | Matched | Matched | - | Matched | Matched | Not matched | - | - | - | Matched | - |
| Maternal age | Matched | Matched | Matched | Not Matched | Matched | Matched | - | Not matched | Matched | Not matched | matched | - | - | Matched |
| **Final code** | **Excellent** | **Fair** | **Fair** | **Good** | **Fair** | **Fair** | **Good** | **Poor** | **Not linked** | **Fair** | **Not linked** | **Fair** | **Poor** | **Not linked** |

The “-“ symbol = Not present

If all 3 variables present and matched, code=EXCELLENT

If 2 variables present and both matched, code=GOOD if the 2 variables are unique ID number and child’s DOB

If 2 variables present and both matched, code=FAIR if unique ID number is not one of the 2 variables

If 2 variables present and only 1 matched, code=FAIR if unique ID number is matched

If 2 variables present and only 1 matched, code=POOR if unique ID number is the unmatched variable

If 1 variable present and matched, code=FAIR if it is unique ID number

**Table 3.** Congenital anomaly subgroups to use in the survival analysis

(based on EUROCAT Subgroups of Congenital Anomalies (August 2016) with exclusions mentioned in doc 3.2 and doc 3.3 in Guide 1.4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EUROCAT Subgroups | ICD10-BPA | ICD9-BPA† | Comments | Subgroup binary variable number (al) |
| All anomalies \* | Q-chapter, D215, D821, D1810^, P350, P351, P371 | 74, 75, 27910, 2281^, 76076, 76280, 7710, 7711, 77121 |  | al1 |
| Structural anomalies |  |  |  |  |
|  Spina Bifida | Q05 | 741 | Exclude if associated with anencephalus or encephalocele subgroups | al6 |
|  Hydrocephalus | Q03 | 7423  | Exclude hydranencephaly 74232. Exclude association with NTD subgroup | al7 |
|  Severe microcephaly | Q02 | 7421 | Exclude association with NTD subgroup | al8 |
|  Congenital cataract | Q120 | 74332 |  | al13 |
| Congenital Heart Defects | Q20-Q26 | 745, 746, 7470-7474 | Exclude PDA with GA <37 weeks Exclude peripheral pulmonary artery stenosis with GA < 37 weeks | al17 |
|  Severe CHD | Q200, Q201, Q203, Q204, Q212, Q213, Q220, Q224, Q225, Q226, Q230, Q232, Q233, Q234, Q251, Q252, Q262 | 74500, 74510, 7452, 7453, 7456, 7461, 7462, 74600, 7463, 7465, 7466, 7467, 7471, 74720, 74742 | ICD9-BPA has no code for HRH and double outlet right ventricle | al97 |
|  Transposition of great vessels | Q203 | 74510 |  | al19 |
|  VSD | Q210 | 7454 |  | al21 |
|  ASD | Q211 | 7455 |  | al22 |
|  AVSD | Q212 | 7456 |  | al23 |
|  Tetralogy of Fallot | Q213 | 7452 |  | al24 |
|  Pulmonary valve stenosis | Q221 | 74601 |  | al27 |
|  Aortic valve atresia/stenosis | Q230 | 7463 | ICD9-BPA has no code for atresia | al29 |
|  Mitral valve anomalies | Q232, Q233 | 7465, 7466 |  | al110 |
|  Hypoplastic left heart | Q234 | 7467 |  | al30 |
|  Coarctation of aorta | Q251 | 7471 |  | al32 |
|  PDA as only CHD in term infants (GA +37 weeks) | Q250 | 7470 | Livebirths only | al100 |
|  Cystic adenomatous malf of lung | Q3380 | No code |  | al36 |
|  Cleft lip with or without cleft  palate | Q36, Q37 | 7491, 7492 |  | al102 |
|  Cleft palate | Q35 | 7490 |  | al103 |
|  Oesophageal atresia with/ without trachea-oesophageal fistula | Q390-Q391 | 75030-75031 |  | al41 |
|  Duodenal atresia or stenosis | Q410 | 75110 |  | al42 |
|  Atresia or stenosis of other parts of small intestine | Q411-Q418 | 75111-75112 |  | al43 |
|  Ano-rectal atresia and stenosis | Q420-Q423 | 75121-75124 |  | al44 |
|  Diaphragmatic hernia | Q790 | 75661 |  | al48 |
|  Gastroschisis | Q793 | 75671 |  | al50 |
|  Omphalocele | Q792 | 75670 |  | al51 |
|  Multicystic renal dysplasia | Q6140, Q6141 | 75316 |  | al54 |
|  Cong hydronephrosis | Q620 | 75320 |  | al55 |
|  Hypospadias | Q54 | 75260 |  | al59 |
|  Limb reduction defects | Q71-Q73 | 7552-7554 |  | al62 |
|  Club foot – talipes equinovarus | Q660 | 75450 | For morbidity | al66 |
|  Hip dislocation and/or dyspasia | Q650-Q652, Q6580, Q6581 | 75430 | For morbidity | al67 |
|  Polydactyly | Q69 | 7550 | For morbidity | al68 |
|  Syndactyly | Q70 | 7551 | For morbidity | al69 |
|  Craniosynostosis | Q750 | 75600 |  | al75 |
|  |  |  |  |  |
| **Chromosomal anomalies** |  |  |  |  |
|  Down syndrome  | Q90  | 7580  | With or without al17 and al40 | Al89 |
|  |  |  |  |  |
| All subgroups below analysed as rare |  |  |  |  |
|  |  |  |  |  |
| **Chromosomal anomalies** |  |  |  |  |
|  Trisomy 13 | Q914-Q917 | 7581 | For mortality | Al90 |
|  Trisomy 18 | Q910-Q913 | 7582 | For mortality | Al91 |
|  Turner syndrome  | Q96  | 75860, 75861, 75862, 75869  |  | Al92 |
|  Klinefelter syndrome  | Q980-Q984  | 7587  |  | Al93 |
|  |  |  |  |  |
| Rare structural anomalies with a EUROCAT subgroup |  |  |  |  |
|  Encephalocele | Q01 | 7420 | Exclude if ass with anencephalus subgroup | al5 |
|  Arhinencephaly / holoprosencephaly | Q041, Q042 | 74226 |  | al9 |
|  Anophthalmos / microphthalmos | Q110, Q111, Q112 | 7430, 7431 |  | al11 |
|  Anophthalmos | Q110, Q111 | 7430 |  | al12 |
|  Congenital glaucoma | Q150 | 74320 |  | al14 |
|  Anotia | Q160 | 74401 |  | al16 |
|  Common arterial truncus | Q200 | 74500 |  | al18 |
|  Double outlet right ventricle | Q201 | No code |  | al109 |
|  Single ventricle | Q204 | 7453 |  | al20 |
|  Triscuspid atresia and stenosis | Q224 | 7461 |  | al25 |
|  Ebstein’s anomaly | Q225 | 7462 |  | al26 |
|  Pulmonary valve atresia | Q220 | 74600 |  | al28 |
|  Hypoplastic right heart | Q226 | No code |  | al31 |
|  Aortic atresia / interrupte aortic arch | Q252 | 74720 |  | al111 |
|  Total anom pulm venous return | Q262 | 74742 |  | al33 |
|  Choanal atresia | Q300 | 7480 |  | al35 |
|  Hirschsprung’s disease | Q431 | 75130-75133 |  | al45 |
|  Atresia of bile ducts | Q442 | 75165 |  | al46 |
|  Annular pancreas | Q451 | 75172 |  | al47 |
|  Indeterminate sex | Q56 | 7527 |  | al60 |
|  Situs inversus | Q893 | 7593 |  | al79 |
|  VATER/VACTERL | Q8726 | 759895 |  | al112 |
|  |  |  |  |  |
| New subgroups for EUROlinkCAT |  |  |  |  |
|  |  |  |  |  |
| Structural anomalies |  |  |  |  |
| Anomalies of corpus callosum | Q040 | 74221 |  | aud1 |
| Anomalies of intestinal fixation | Q433 | 7514 |  | aud3 |
| Unilateral renal agenesis | Q600 | No code |  | aud4 |
| Accessory kidney | Q630 | 75330 |  | aud5 |
| Bladder exstrophy  | Q641 | 7535 |  | aud6 |
| Epispadia | Q640 | 75261 |  | aud7 |
| Posterior urethral valves | Q6420 | 75360 |  | aud8 |
| Prune Belly | Q794 | 75672 |  | aud9 |
| Arthrogryposis multiplex congenita | Q743 | 75580 |  | aud10 |
|  |  |  |  |  |
| Genetic syndromes |  |  |  |  |
| Di George syndrome  | D821 | 27910 |  | aud14 |
| Goldenhar syndrome | Q8704 | 75606 |  | aud15 |
| Cornelia de Lange syndrome | Q8712 | 759821 |  | aud16 |
| Noonan syndrome | Q8714 | 759896 |  | aud17 |
| Prader-Willi | Q8715 | 759872 |  | aud18 |
| Beckwith Wiedeman syndrome | Q8730 | 759874 |  | aud20 |
| Williams syndrome | Q8784 | No code |  | aud21 |
| Angelman syndrome | Q8785 | No code |  | aud22 |
|  |  |  |  |  |
| Chromosomal anomalies |  |  |  |  |
| Wolff-Hirschorn syndrome | Q933 | 75832 |  | aud23 |
| Cri-du chat syndrome | Q934 | 75831 |  | aud24 |
| Karyotype XXX | Q970 | 75885 |  | aud25 |
|  |  |  |  |  |
| Sequences |  |  |  |  |
| Pierre-Robin sequence | Q8708 | 75603 |  | aud27 |

**Footnote:** \*All Anomalies = ALL cases of congenital anomaly, excluding cases with only minor anomalies as defined in Section 3.2 in Guide 1.4 for cases born post-2005. Cases with more than one anomaly are only counted once in the “All Anomalies” subgroup.

†EUROCAT ICD-9 codes are used with the British Paediatric Association (BPA) extension code: <http://www.eurocat-network.eu/content/EUROCAT-ICD9-with-BPA-Extension.pdf>

aud13 subgroup **-** Meckel-Gruber syndrome (Q6190) - has been removed from the list as it is not going to be analysed due to a very small number of live births.

Standardisation committee meeting September 2018:

Decision to remove subgroups for fetal alcohol syndrome, valproate syndrome and maternal infections due to small numbers and too heterogeneous subgroup for maternal infection.

Standardisation committee meeting 12.11.2018: More subgroups have been deleted due to small numbers in the dataset (inclusion criteria is at least 5 livebirths in 3 registries). Excluded are:

Megalencephaly

Ectodermal dysplasia

Alagille syndrome

Holt-Oram

Caudal regression sequence

**Appendix 1.** Data availability by time period for participating registries

Data **for 1995-2014** will be included for the following 12 EUROCAT registries:

Belgium: Antwerp, Croatia: Zagreb, Denmark: Odense, Finland, Germany: Saxony-Anhalt, Italy: Emilia Romagna, Italy: Tuscany, Malta, Netherlands: Northern, Spain: Basque Country, UK: Thames Valley, UK: Wessex.

The remaining registries will include data for the study subjects for the following periods:

**1997-2014** - France: Paris;

**1998-2014** - UK: East Midlands, UK: Wales;

**1999-2014** - Norway;

**2000-2014** - UK: North;

**2002-2014** - France: Ile de la Reunion;

**2005-2014** - UK: South West, Ukraine: West;

**2007-2014** - Spain: Valencia.

**Addendum ‘Analysis of the geographic variation’**

One of the tasks of the WP3 is aimed at investigating whether there are geographic variations in survival across Europe for selected EUROCAT congenital anomaly subgroups.

As individual data are not transmitted to the CRR it will not be possible to calculate risk indicators by Country or European area. For this reason we will evaluate geographic variation using the estimates produced at registry level.

The approaches we are going to use in order to provide information about geographic variations in survival in infants with congenital anomalies across Europe are briefly described below:

1. Evaluate whether there is variability of the survival indicators among the 21 participating registries. For this aim we will consider the between-registry heterogeneity (through the I2 statistic) produced for the pooled estimate in the meta-analysis.

Geographical differences will be evaluated by the analysis of indicators calculated by each registry. Both crude and adjusted risk indicators for risk factors will be used for this aim.

2. Use information on risk factors to interpret the geographic variations. Geographic differences will be explored using the distribution of specific risk factors for the specific anomalies investigated. Models can be fitted using the observed survival in each registry and the distribution of risk factors for the babies with anomalies in that registry (Tables T 8A – T 8C – see pages 7-8).

3. Evaluate the variability of TOTAL mortality given the observed mortality in children with congenital anomalies. In order to contextualise and better interpret the results, as no controls are required for WP3, general information on mortality (e.g. neonatal mortality) will be collected (Tables T 7A – T 7B – see pages 7-8).

Additional and supplementary analyses will be performed:

a. Pooled estimates of the European areas (i.e. Northern, Western, Eastern, Southern) will be calculated by using the estimates calculated at registry level. The WP3 team will combine estimates produced by each individual registry belonging to a specific European area by using a meta-analysis random effects model. The heterogeneity among estimates of the registries belonging to the same European area will be assessed through the I2 statistic.

b. Specific statistical exploratory techniques can be used to identify latent information among the results produced by the WP3. For this purpose, we could:

- use cluster analysis to identify 3-4 groups of registries according to survival indicators related to all anomalies and subgroups. Results can give insights into registry similarities.

- group registries by defining 3-4 geographic areas *a priori;* then we use random forest to characterize the areas in terms of survival indicators related to all anomalies and subgroups. Upon model fitting, random Forest could help in defining indicators that best differentiate the areas.