

**EUROlinkCAT WP5 - Educational achievements and needs of children with congenital anomalies**

**Statistical analysis plan for the contributing BINOCAR registries (EMSYCAR, CAROBB, WANDA and NorCAS) and for the Wales registry**

**WP5 leads:** Judith Rankin (UNEW), Amanda Neville (UNIFE)

**Researchers:** Svetlana Glinianaia (UNEW)

**Wales registry team:** Anna Rawlings. Sue Jordan, Daniel Thayer (PHW, SU)

**Statistical advice**: Joan Morris, Joachim Tan (SGUL)

**The aim** of this analysis plan is to guide the registries in their local analyses and to detail the methods of analysis to be applied to all outcomes in WP5.

UNEW will summarise the analytic results from the three BINOCAR registries (analysed collectively), Wales registry and Finland registry (for 16-year-olds only) into coherent summary tables, write the WP5 deliverable report (D5.1) and a paper for publication on the educational outcomes (on Key stage (KS) 2 and KS4 results in 11-year-old and 16-year-old children respectively using BINOCAR and Wales data) and special education needs (SEN) in children with selected congenital anomalies (CAs) compared with the reference group.

In WP5, each contributing registry is responsible for writing their scripts and running their analysis.

The team agreed that the Wales registry team (AR and DT) will write the analysis scripts in R for Wales. These will include descriptive data, unadjusted and adjusted inferential analyses. The adjusted analyses will be binary logistic regression with pre-specified covariates.

**Inclusion / Exclusion Criteria**

For Wales, all live births (i.e. singletons and multiples) reported in the national congenital anomaly registry (CARIS) from 1998 to 2014, and still alive by their 12th & 17th birthdays, will be included in the study and will be linked to education data in the Secure Anonymised Information Linkage (SAIL) databank, where possible. The reference population identified before the linkage is all live born children without CAs (without EUROCAT codes for a major CA) born during the same time period as the cases and in the same geographical area covered by CARIS, linkable to SAIL education data, and still alive by their 12th & 17th birthdays.

For the BINOCAR registries, live born children with a major CA registered in the three registries (Thames Valley, Wessex and East Midlands) and known to be alive at age 11 (for KS2) and 16 years (for KS4) (cases) will be linked to the English National Pupil Database (NPD) education data. Referents will be a random sample from the background population who are not CA cases that are recorded in the NPD.

**Categorisation of Congenital Anomalies**

Children with CAs from 20 subgroups of structural CAs, and from four subgroups of chromosomal anomalies, will be included in the analyses.

For structural anomalies, only those defined as 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) (see chapter 3.4 of the EUROCAT Guide 1.41 and Garne et al, 20112) will be included. For chromosomal anomalies (coded as C in 'mult\_malf' variable), the following chromosomal syndromes will be included: Down syndrome (al89), Down syndrome with a congenital heart defect (CHD) (al89+al17), Down syndrome without CHD (al89 excluding al17) and Turner syndrome (al92) (see Table 1).

**Tables 0-1 – evaluating quality of linkage between the EUROCAT registry data and education data**

**Table 0. Quality of linkage between EUROCAT registries and education data by CA subgroup and year of birth –** using the number of referent children and children born with CAs and alive by age a) 12 years & b) 17 years (expected number) and those linked to education data, by CA subgroup and year of birth (starting from 1995 for English registries and from 1998 for Wales).

Linkage quality will be investigated locally by providing the expected number of children alive by school age (Table 0a) age 12 years & Table 0b) 17 years) and the number of children linked to education data by birth year (1998- 2007 for Wales and 1995-2007 for BINOCAR as the last birth year for 11-year-old children in 2017/18 school year is 2007, last birth year for 16-year olds - 2002) and CA subgroup for cases and for referent children and by calculating the percentage linked.

Data quality criteria for inclusion in the full analysis will be agreed for Wales and the BINOCAR registries as soon as the access to the BINOCAR data is obtained and the data examined.

**Table 1. Quality of linkage between EUROCAT registries and education data CA subgroup and educational stage –** using thenumber of referent children and children born with CAs, and assumed alive at the end of each educational stage (expected number) and those linked with education data, by CA subgroup.

Children known to have died (based on Office for National Statistics (ONS) mortality data from NHS Digital) at the end of each educational stage included (to an average of 11 years for Key Stage 2 and 16 years for Key Stage 4) will be excluded before the linkage for each Key stage.

The WP5 team have recently agreed that for the WP5 report due on 30th November 2021, the results of the analysis of educational outcomes at KS2 and KS4 only, as primary outcomes, will be reported. Therefore, the tables for other Key stages have been removed from this analysis plan. The analysis for younger children (Early Years Foundation Stage Profile (EYFSP) and KS1) and KS3 can be performed as a supplementary analysis at a later stage if time allows.

**Key Stage 2 (KS2) –11-year-old children**

**KS2 – Table A.** Sociodemographic status of linked children (typically aged 11 years) with Key Stage 2 (KS2) data (referents and cases) by specific congenital anomaly (CA) subgroup for children with CAs.

The descriptive table for child’s sex, gestational age (<32, 32-36, 37+ weeks and missing - for referents in Wales data only), major language, deprivation index (specified below), ‘free school meal eligibility’ and school absence by CA subgroup will be prepared for subgroups where numbers permit.

**KS2 – Table B**. KS2 test levels of attainment - number and percentage of children achieving level 4 and above (expected level) for children with congenital anomalies and referents by school subject (English/Welsh), Maths and Science).

**KS2 – Table C**. KS2 results - 1) Unadjusted and 2), 3) and 4) Adjusted ORs of achieving level 4 and above (expected level) by school subject (measured by test) - for children with congenital anomalies versus referents - logistic regression results.

1. Unadjusted
2. Adjusted for Free school meal eligibility (FSME – yes/no)
3. Adjusted for deprivation index (NPD Income Deprivation Affecting Children Indices (IDACI) tertiles or SAIL WIMD quintiles and Welsh lower super output area (LSOA) Townsend deprivation scores and ranks,3,4 birth year and gestational age (GA) group (birth year and GA group for Wales only)
4. Adjusted for school absence

The first analysis (1) will show if the unadjusted odds ratio of achieving level 4 and above (expected level) by school subject (English, Maths and Science) (measured by test) differ between children with a CA (by CA subgroup) and referents.

The second (2), third (3) and fourth (4) analyses will show if the results in Table 1 change when adjusted for deprivation indicators (FSME and IDACI for England or WIMD/Townsend score for Wales), school absence and other variables specified for Wales.

The analyses below should be performed on children with CAs only to investigate the effect of risk factors on the association between a specific CA and KS2 results for an individual school subject, i.e. 1) English/Welsh (KS2- Table D), 2) Mathematics (KS2 – Table E) and 3) Science (KS2 – Table F). Both the crude ORs (in the univariate analysis) and adjusted ORs (from multivariate models) will be presented for each factor. Either FSME or IMD (depending on the results of the univariate analysis) should be included in the multivariable model.

Due to expected small numbers for some specific CA subgroups, this analysis may be possible for the most common CA subgroups only (e.g. all CHD, severe CHD, orofacial clefts, Down syndrome (any)).

**KS2 – Table D.** Logistic regression results on the association between sociodemographic factors and KS2 results (achieving level 4 or above (expected level)) for *English or Welsh* for linked children with congenital anomalies

**KS2 – Table E**. Logistic regression results on the association between sociodemographic factors and KS2 results (achieving level 4 or above (expected level)) for *Maths* for linked children with congenital anomalies

**KS2 – Table F**. Logistic regression results on the association between sociodemographic factors and KS2 results (achieving level 4 or above (expected level)) for *Science* for linked children with congenital anomalies.

Last KS2 table (KS2 – Table G) compares the percentage of children receiving special education support at KS2 tests between children with a specific CA and referent children (as given in KS2\_TG\_1).

In NPD, for those with documented Special Education Needs (SEN), KS2\_TG\_2 provides the number and the percentage of children with specific SEN type for cases and referent children.

**KS2 – Table G**. Special Education Needs (SEN) at KS2 tests (KS2\_TG\_1) and by SEN type if SEN identified (KS2\_TG\_2) for linked children with congenital anomalies by CA subgroup compared with referent children

In NPD data, until 2014, SEN type included pupils at school action plus and those pupils with a statement of SEN, but not those pupils at school action. From 2015, this includes pupils on SEN support or with a statement of SEN or EHC plan.

We (UNEW) are going to then test the differences between cases and referents by calculating relative risks for each CA subgroup compared to the referent children.

**Key Stage 4 (KS4) –16-year-old children**

**KS4 - Table A.** Sociodemographic status of linked children (typically aged 16 years) with Key Stage 4 (KS4) data (referent children and cases) by specific congenital anomaly (CA) subgroup for children with CAs

The descriptive table for child’s sex, gestational age (<32, 32-36, 37+ weeks and missing - for referents in Wales data only), major language, Deprivation index and Free school meal eligibility by CA subgroup is included (for subgroups where numbers permit).

**KS4 – Table B.** KS4 - Table B. GCSE or equivalents - number and percentage of children achieving 5 or more GCSE/GNVQs at grades A\*-C (level 2) or A\*-G (level 1) - for children with congenital anomalies and referent children

This descriptive analysis will produce the percentage of children with CAs and referent children (from the number of eligible children for each group) who achieved level 2 (5 or more GCSE/equivalents at grades A\*-C), level 1 (5 or more GCSE/equivalents at grades A\*-G) or any pass in GCSE or equivalent.

For English NPD data, the variables describing the highest examination category achieved (KS4\_EXAMCAT (2004/05 - 2012/13), KS4\_EXAMCAT\_PTQ (2013/14) and KS4\_EXAMCAT\_PTQ\_EE (from 2014/15)) can also be used to describe children’s KS4 achievement.

**KS4 - Table C.** KS4 results - 1) Unadjusted and 2), 3) and 4) Adjusted ORs of achieving 5 or more GCSE/GNVQs at grades A\*-C or A\*-G - for children with congenital anomalies versus referents - logistic regression results

1. Unadjusted
2. Adjusted for Free school meal eligibility (FSME – yes/no)
3. Adjusted for deprivation index (NPD IDACI tertiles or SAIL WIMD quintiles and Welsh LSOA Townsend deprivation scores and ranks, birth year and gestational age group (the latter two for Wales only)).
4. Adjusted for school absence

The first analysis (1) will show if the odds of achieving 5 or more GCSE/GNVQs at grades A\*-C or A\*-G differ between children with a CA (by CA subgroup and any CA) and referents.

The second (2), third (3) and fourth (4) adjusted analyses will show if the results in Table 1 change when adjusted for deprivation indicators (FSME and IDACI for England or WIMD/Townsend score for Wales), school absence and other variables specified for Wales

The analyses below will be performed on children with CAs only to investigate the effect of risk factors on the association between a specific CA and KS4 results.

Due to expected small numbers for some specific CA subgroups, this analysis may be possible for the most common CA subgroups only (e.g. all CHD, severe CHD, orofacial clefts, Down syndrome (any)).

The primary outcomes are: 1) achieving 5 or more GCSE/GNVQs at grades A\*-C (Level 2) and 3) achieving 5 or more GCSE and equivalents at grades A\*-G (Level 1).

The secondary outcomes are: 2) achieving 5 or more GCSE and equivalents at grades A\*-C (Level 2) including English/Welsh and Maths as fundamental school subjects; 4) 5 or more GCSE and equivalents at grades A\*-G (Level 1) including GCSE English/Welsh and Maths.

**KS4 - Table D.** Logistic regression results on the association between sociodemographic factors and KS4 results (achieving 5 or more GCSE/GNVQs at grades A\*-C) - for children with congenital anomalies only

1. Achieved 5 or more GCSE/GNVQs at grades A\*-C (Level 2) (yes/no)

**KS4 - Table E.** Logistic regression results on the association between sociodemographic factors and KS4 results (achieving 5 or more GCSE and equivalents at grades A\*-C including English/Welsh and Maths) - for children with congenital anomalies only

1. Achieved 5 or more GCSE and equivalents at grades A\*-C (Level 2) including English/Welsh and Maths) (yes/no)

**KS4 - Table F.** Logistic regression results on the association between sociodemographic factors and KS4 results (achieving 5 or more GCSE and equivalents at grades A\*-G) - for children with congenital anomalies only

1. Achieved 5 or more GCSE and equivalents at grades A\*-G (Level 1) (yes/no)

**KS4 -Table G.** Logistic regression results on the association between sociodemographic factors and KS4 results (achieving 5 or more GCSE and equivalents at grades A\*-G including English/Welsh and Maths) - for children with congenital anomalies only

1. Achieved 5 or more GCSE and equivalents at grades A\*-G (Level 1) including GCSE English/Welsh and Maths (yes/no)

Last KS4 table (KS4 – Table H) compares the percentage of children receiving special education support at KS4 tests between children with a specific CA and referent children (as given in KS4\_TH\_1).

For those with documented Special Education Needs (SEN), KS4\_TH\_2 provides the number and the percentage of children with specific SEN type for cases and referent children

**KS4 – Table H.** Special Education Needs (SEN) at KS4 tests and by SEN type if SEN identified for linked children with congenital anomalies by anomaly subgroup compared to referent children.

In NPD data, until 2014, SEN type included pupils at school action plus and those pupils with a statement of SEN, but not those pupils at school action. From 2015, this includes pupils on SEN support or with a statement of SEN or EHC plan.

We (UNEW) are going then to test the differences between cases and referents by calculating relative risks by a CA subgroup.

**References**

**1**. EUROCAT. Chapter 3.4: Multiple Congenital Anomaly Algorithm (version 19.11.2014). In: *EUROCAT Guide 1.4: Instruction for the registration of congenital anomalies (Last update version 20/12/2016).* Newtownabbey, UK: EUROCAT Central Registry, University of Ulster; 2013.

**2**. Garne E, Dolk H, Loane M, et al. Paper 5: Surveillance of multiple congenital anomalies: implementation of a computer algorithm in European registers for classification of cases. *Birth defects research Part A, Clinical and molecular teratology.* 2011;91 Suppl 1:S44-50. doi: 10.1002/bdra.20777

**3**. Gartner A, Lester N. Briefing paper on LSOA Townsend deprivation scores calculated from unadjusted Census data. 2008. [http://www2.nphs.wales.nhs.uk:8080/hiatdocs.nsf/85c50756737f79ac80256f2700534ea3/17fdbca9920051368025772f003b5a35/$FILE/TownsendBriefing.pdf](http://www2.nphs.wales.nhs.uk:8080/hiatdocs.nsf/85c50756737f79ac80256f2700534ea3/17fdbca9920051368025772f003b5a35/%24FILE/TownsendBriefing.pdf). Accessed 09/03/2021.

**4**. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North.* London: Routledge; 1988.

**Table 1. Congenital anomaly subgroups to use in the WP5 Education analysis**

**(based on EUROCAT Subgroups of Congenital Anomalies (August 2016) with exclusions mentioned in chapters 3.2 and 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 |
| 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 |
|  Tetralogy of Fallot | Q213 | 7452 |  | al24 |
|  Coarctation of aorta | Q251 | 7471 |  | al32 |
| 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 |
| Ano-rectal atresia and stenosis | Q420-Q423 | 75121-75124 |  | al44 |
| Diaphragmatic hernia | Q790 | 75661 |  | al48 |
| Gastroschisis | Q793 | 75671 |  | al50 |
| Multicystic renal dysplasia | Q6140, Q6141 | 75316 |  | al54 |
| Hypospadias | Q54 | 75260 |  | al59 |
| Limb reduction defects | Q71-Q73 | 7552-7554 |  | al62 |
| Craniosynostosis | Q750 | 75600 |  | al75 |
|  |  |  |  |  |
| **Chromosomal anomalies** |  |  |  |  |
|  Down syndrome  | Q90  | 7580  | With or without al17 and al40 | Al89 |
|  Down syndrome with CHD | Q90  | 7580  |  | al89+al17 |
|  Down syndrome without CHD | Q90  | 7580  |  | al89 excluding al17 |
|  Turner syndrome  | Q96  | 75860, 75861, 75862, 75869  |  | Al92 |

**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: <https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/EUROCAT-ICD9-with-BPA-Extension.pdf>