Deliverable D8.3: Report to EU institutions hosting health care databases with guidelines for improving the quality of the congenital anomaly coding

Project Acronym:	EUROlinkCAT
Project Title:	Establishing a linked European cohort of children with congenital anomalies
Funding Scheme:	European Union's Horizon 2020 Research and Innovation programme
Grant Agreement Number:	733001
Participants in WP8 Deliverable D8.3	Ester Garne, Paediatric Department, Hospital Lillebaelt, Kolding, Denmark Joan Morris, Population Health Research Institute, St George's, University of London, UK Marian Bakker, UMCG University Medical Center Groningen, Netherlands
Lead Author's Contact Details	Dr Ester Garne Phone: +45 7636 2238 Email: ester.garne@rsyd.dk
Project Website Address:	https://www.eurolinkcat.eu/

Partners in WP8 D8.3

Acronym, (partner's number)	Partner/Registry	Names of contributors
SGUL; QMUL (1;23)	St George's, University of London; Queen Mary University of London	Joan Morris
UU (2)	Ulster University	Maria Loane
RSD (3)	Denmark: Funen	Ester Garne
UNIFE (5)	Italy: Emilia Romagna	Amanda Neville
KDB (6)	Croatia: Zagreb	Ingeborg Barišić, Ljubica Boban
CNR-IFC (7)	Italy: Tuscany	Anna Pierini
UMCG (8)	Northern Netherlands	Marian Bakker, Hermien de Walle
PHW NHS; SU (9;22)	UK: Wales	David Tucker, Sue Jordan
FISABIO (11)	Spain: Valencian Region	Clara Cavero-Carbonell
PUMS (12)	Poland	Anna Latos-Bieleńska, Anna Jamry Dziurla
THL (13)	Finland	Anna Heino, Mika Gissler
OVGU (15)	Germany: Saxony-Anhalt	Anke Rissmann
INSA (16)	South Portugal	Paula Braz
PIH (18)	Belgium: Antwerp	Elly Den Hond

Contents

Executive Summary and Recommendations

Electronic health care databases are potentially valuable data sources for the surveillance of congenital anomalies and for studies evaluating the risk of congenital anomalies following exposures to risk factors such as medications, viruses/infections and environmental factors in the first trimester of pregnancy. For example, regional and national health care databases could be used to monitor primary prevention measures such as the reduction in the risk of a neural tube defect due to the fortification of flour with folic acid. The EUROlinkCAT study evaluated the quality of the coding in hospital databases in eleven European regions in eight countries.

The EUROlinkCAT study concluded that although the hospital databases do contain important data, they currently cannot be used as the only data source for the surveillance of congenital anomalies. The fundamental weakness is that the type of anomaly and organ system involved are not reported for the majority of terminations of pregnancy for fetal anomaly (TOPFAs). This is essential as over 21.0% of major congenital anomalies reported in EUROCAT from 2015-2019 resulted in a TOPFA. In addition, for livebirths some anomalies are overreported in hospital databases whilst others are underreported.

The study also concluded that data routinely collected in electronic health care databases should be improved to enable the data to be used in the surveillance and research of congenital anomalies. Codes for classifying and reporting the congenital anomalies resulting in TOPFAs in electronic health care databases need to be developed. In addition, the accuracy of the coding of congenital anomalies in all births should be improved and algorithms to accurately discriminate between major congenital anomalies and suspected or minor anomalies should continue to be refined. Some countries, mainly in southern and eastern Europe, who do not currently have electronic health care databases, should be supported in establishing these databases in order to enable effective surveillance to occur across Europe.

The following recommendations are made to enable the potential of the data in electronic health care databases to be fully exploited in the primary prevention of congenital anomalies.

For national registration bodies

- Develop registration systems to assign a permanent unique identification (ID) number to each baby as soon as possible after the birth to ensure that ICD codes and procedure codes for the first days after birth can be linked to the baby.

- In the mother's record, report ICD codes for terminations of pregnancy for fetal anomalies. Create an additional record, possibly in a separate database, that reports each specific anomaly diagnosed in the fetus and which can be linked to the mother's ID number.

For health care database designers

- Include outpatient contacts in health care databases as less severe congenital anomalies may not be visible in hospital discharge databases if surgery is not required.
- Allow the use of more than one diagnosis code for both hospital discharges and outpatient contacts.
- Include records, possibly in a separate database, that report each specific anomaly diagnosed in the fetus and which can be linked to the mother's ID number.
- Allow codes to be revised within a certain time frame after the initial coding, as:
    - the coding may be amended by more experienced doctors and coders
    - the coding may be refined by results of diagnostic examinations/tests arriving after the TOPFA or after the child has left the hospital.

For health care database coders

- Always use the most specific code available for the congenital anomaly and avoid using codes for "other" or "unspecified" congenital anomaly.
- Use extended versions of ICD for the coding of rare congenital anomalies or use other coding systems to make the rare diagnoses visible in health care databases.
- Continually undergo training to ensure optimal coding quality and consistency of the data on congenital anomalies.

For researchers using health care databases

- Link to data from congenital anomaly registries wherever possible.
- Use, if possible, validated algorithms for identifying congenital anomalies in health care databases.
- Collaborate with the people working with the health care databases as they know their data.
- Discuss all results with the people working with the health care databases as an aid to interpretation and quality improvement.

Introduction

This report will describe the coding of congenital anomalies in hospital databases and discuss the problems if the surveillance or research related to congenital anomalies is based solely on hospital discharge diagnosis. Recommendations for improving the coding will be given.

Background

Congenital anomalies are a major burden for morbidity and mortality in infancy and childhood. Major congenital anomalies affect 2-3% of all fetuses (Boyd et al, 2011). Surveillance of congenital anomalies has become an important public health activity since the thalidomide disaster, aiming to prevent similar or smaller incidents (Khoury et al, 1994). Such surveillance will prevent congenital anomalies occurring by detecting new environmental, medication, or lifestyle teratogenic exposures as early as possible (Dolk et al, 2015).

EUROCAT congenital anomaly registries originally depended on individual case identification, either sent to the registry by clinicians or actively searched for in clinical records. However, the last decade has seen an increase in the use of electronic health care records as an additional data source for the registries. There is also an increasing public concern about access to complete medical records due to privacy, and therefore the focus has turned to the possibility of using routinely collected data in electronic health care databases for the surveillance of congenital anomalies. However, these databases vary in their aim, function and coding systems used, and these characteristics will influence the validity of the data for use in the surveillance of congenital anomalies.

Hospital discharge databases include ICD9 or ICD10 codes for contacts with the hospital by each patient, primarily for financial purposes, which may result in over-reporting of some anomalies. In addition, a diagnosis may be suspected and coded at discharge, but may not be confirmed after referral to another hospital. Examples are hip dislocation or hypospadias suspected by the midwife at birth, but not confirmed until after referral to the surgical departments for evaluation.

The quality of the coding of diseases in health care databases is dependent on a number of factors: the quality of the coding system used and how detailed it is, the clinical knowledge of the coder, the time available for coding for the persons that carry out the coding for the database, and the diagnostic details available about the patient. In some countries, the medical doctors code all diagnoses for the discharge

letter. In other countries, trained coders are employed to read the medical record after discharge and they add the relevant codes and all procedures performed for each hospital stay.

There is very limited published information about the clinical coding systems used in health care databases in European countries. A survey published back in 2001 showed that most countries in Europe use ICD9 or ICD10 for clinical coding, but in some countries these are used only for secondary care (Lusignan et al, 2001).

Registries and researchers using databases to study congenital anomalies may use algorithms to discriminate between true cases and suspected or minor cases (Astolfi et al, 2016). As there are many grey areas in the definition of major congenital anomalies, access to medical records including results of specific examinations (MRI scan, echocardiography, genetic tests, post-mortem examinations) will still be necessary for the correct interpretation of all cases (Tairou et al, 2006). An example of this is atrial septal defect (ASD) where an echocardiography performed in the neonatal period in most cases will show a flow over the atrial septum, as the foramen ovale from fetal life has not yet closed. Many clinicians will code this as an ASD in the discharge letter despite the benign nature of this finding (Garne et al, 2012).

An important issue in the surveillance of congenital anomalies and in aetiologic studies is that all fetuses with congenital anomalies should be included (Charlton et al, 2010). As terminations of pregnancy for fetal anomaly (TOPFAs) are more frequent among the more severe congenital anomalies, a significant proportion of these cases would be excluded from the analysis if only liveborn infants are included (Charlton et al, 2014). In the beginning of EUROCAT (1980-89), 6.3% of all reported cases with major congenital anomalies were TOPFAs. In a more recent 5-year period (2015-19), 21.0 % of all cases reported to EUROCAT from full member registries were TOPFAs (www.eurocat-network.eu data on 01.04.2022). A significant proportion of the most severe cases will be missed if surveillance and research are performed on livebirths only.

In hospital databases, TOPFA cases do not have their own personal identity. The TOPFA procedure is coded under the maternal ID. Known diagnosis before TOPFA may have been coded in relation to maternal contacts with the health system, but ICD9 and ICD10 have limited codes for this purpose. In ICD10 there are codes in subchapter O35: O350 “Maternal care for (suspected) central nervous system malformation in fetus” and O351 “Maternal care for (suspected) chromosomal abnormality in fetus”. The WHO version of

ICD10 does not specify codes for medical abortions, but there are two subchapters in chapter O that can be used for countries to define their own codes (O04 and O05).

An additional problem for the registration and surveillance of TOPFAs is that the results of examinations performed after a TOPFA for confirmation of the final diagnosis (genetic tests, post-mortem examinations) will not be visible in the hospital databases as there is no identity to record them for. A few very late TOPFAs (gestational age after viability, usually 22 weeks of gestation) will be recorded as livebirths and for these there may be more specified coding. With a very early diagnosis of an anomaly such as anencephaly or Down syndrome, the procedure may have been performed before week 12 or week 16, which is the legal limit for a TOPFA without special permission in most countries. Here the TOPFA may just be coded as O04: "legal termination of pregnancy" without mentioning any anomalies. In twin pregnancies, a prenatal diagnosis of a severe anomaly in one fetus may be followed by a fetal reduction procedure. This procedure will usually be coded as an outpatient contact during the pregnancy. There is no ICD9 or ICD10 code for this situation and the subsequent birth of the co-twin may be recorded as a birth of a singleton. Some countries have defined new codes within their coding systems to overcome some of the problems in the coding of TOPFAs.

Published Validation studies

The coding of congenital anomalies in hospital discharge databases in the US has been validated and found to be incomplete (Boulet *et al.*, 2006). Hexter *et al.* (1990) found that diagnoses in the hospital discharge diagnosis index in California were imprecise, which resulted in many anomalies being categorised as 'unspecified' or 'other'. Metcalfe *et al.* (2014) showed that in-hospital data were adequate in ascertaining most, but not all congenital anomalies, while other sources of administrative data, particularly data from out-patient physician visits, were not able to do this. Salemi *et al.* (2018a) showed that expansion from 10 to 31 diagnosis code fields improved ascertainment by preventing the loss of 2.5% of congenital anomaly cases with defect-related diagnoses appearing only in code positions 11 to 31, but with major differences by type of anomaly. In their study, for example, the ICD9 code '742.1' for microcephaly was listed outside the first 10 codes 20% to 25% of the time; therefore, any changes in the number of diagnosis code fields available for use for each patient would confound the trends assessment for microcephaly. Recent studies in the US estimated that 93% of babies with any congenital anomaly would be identified (Salemi *et al.*, 2016, Wang *et al.*, 2010), but that the proportion identified with specific anomalies is much lower with 54% of limb reduction defects reported as an example (Salemi *et al.*, 2016). Andrade *et al.* (2013) found only

37% of pregnancies affected with anencephaly were recorded. A Canadian study reported slightly higher accuracy, but this was based on a very restricted set of congenital anomalies (Blais *et al.*, 2013).
A study from Australia compared the diagnosis in the congenital anomaly registry to the hospital discharge data for liveborn infants (Schneuer et al, 2021). They found 84% overall agreement between the congenital anomaly registry and the hospital discharge diagnosis and more than 93% agreement for cardiac, abdominal wall and gastrointestinal anomalies. Among the children only visible in the discharge diagnosis a high proportion had skin anomalies or unspecified codes.

To our knowledge there are no validation studies from Europe comparing the coding in a congenital anomaly registry to the hospital discharge diagnosis. Results from the EUROlinkCAT study on accuracy of the coding of congenital anomalies in hospital databases will be presented in this report.

Results from EUROlinkCAT

Four EUROlinkCAT studies were performed to investigate the availability of hospital databases and the accuracy of the coding of congenital anomalies in these databases. Firstly, in Work Package 6, the accuracy of coding of congenital anomalies of livebirths in hospital databases was evaluated. Secondly, also in Work Package 6, the visibility of TOPFAs and the coding of their anomalies in the hospital databases was investigated. Thirdly, in Work Package 4, livebirths in congenital anomaly registries were linked with their data in hospital discharge databases to investigate hospitalisations and surgeries, and information about the validity of congenital anomaly codes that arose from this work is reported here. Finally, in Work Package 8, a short questionnaire with questions on coding both at discharge from hospital and for outpatients was sent to clinicians.

Availability of national electronic health care databases

The aim of the EUROlinkCAT study was to link data on liveborn children with congenital anomalies born from 1995 to 2014 to health care databases to obtain information about morbidity during their first 10 years of life. The databases in Scandinavia had accurate linkage for all birth years of the EUROlinkCAT study. The database in Wales had accurate linkage from 1998 and those in England from 2003. In the Netherlands, accurate linkage occurred from 1995, but the database changed in 2012 with some errors in the records for 2013. The databases in Italy and Spain were only able to provide accurate data for the children born in the last 10 birth years of the study (2005-2014). Linkage to hospital databases in Croatia and Poland was not possible when the EUROlinkCAT study started in 2017. However, such linkage did occur in Poland in 2021. The EUROCAT registries in Portugal and Ukraine were not able to link to the relevant health care data. It

was a general finding that the quality of the data in the health care databases improved over time, particularly during the first years after they were established.

Accuracy of coding of congenital anomalies for liveborn children

Eleven EUROCAT registries linked their congenital anomaly data to regional or national hospital databases. We focused on 17 specific anomaly groups, including anomalies detectable at birth, anomalies with a high prenatal detection rate and anomalies diagnosed after discharge from the maternity unit. For these specific anomaly groups, we compared the diagnosis codes from the hospital database to the codes registered in the EUROlinkCAT cases. The sensitivity (the proportion of EUROlinkCAT children correctly identified in the hospital data with the same congenital anomaly code as they have in EUROlinkCAT) and the positive predictive value (PPV) (the proportion of children in hospital data correctly identified in EUROlinkCAT with exactly the same congenital anomaly code as they have in the hospital data) were calculated for the anomaly groups within each registry. A low sensitivity indicates that the registration of these congenital anomalies is incomplete in hospital databases. A low PPV indicates that the registration of these congenital anomalies is not sufficiently precise in hospital databases.

Registries linked between 58% and 99% of their liveborn cases to hospital data, of which eight registries linked more than 90% of their cases. For the linked EUROlinkCAT cases, the proportion with a congenital anomaly code recorded in the hospital data in the first year of life varied from 49% in Zagreb (manual linkage) to 96% in the Valencian Region.

In most registries, sensitivity was high (>80%) for Hirschsprung's disease, abdominal wall defects, cleft lip with or without cleft palate and Down syndrome. Low sensitivity (<50%) was frequently observed for clubfoot and congenital hydronephrosis. The PPV was high (>80%) for gastroschisis and Down syndrome and low (<50%) for ASD. The comparison between congenital anomaly coding in hospital databases and the EUROlinkCAT cases, highlighted differences between the hospital databases possibly due to differences in the health care systems. Also, the set up and purpose of the hospital database, including coding practices, affected the sensitivity and PPV.

Congenital anomalies that were not recorded accurately in the hospital database were, for instance, congenital anomalies with a high termination rate, mild anomalies that do not require hospitalisation or severe anomalies that are treated in specialist centers outside the region of coverage. If available,

outpatient data and other data sources such as pathology reports or cytogenetic databases should be used to improve completeness.

It is therefore important to have a firm knowledge of coding practices in hospital databases and to be informed on the specific codes that are used for specific anomalies. These may not always be the correct code according to the EUROCAT guidelines. Coding and identification of children with complex and multiple anomalies may be challenging, which is important to acknowledge when evaluating the aetiology and outcomes in children with congenital anomalies. Information on related factors such as gestational age at birth, which can differentiate between anomalies at term vs. normal aspects of development in preterm births, is often missing. Children with a congenital anomaly with a low PPV need to have the congenital anomaly code validated or confirmed using other data sources, before the congenital anomaly can be considered as correct.

In conclusion, congenital anomaly registries where experts validate and code the congenital anomaly based on all available information, are still the most appropriate data source to monitor the prevalence of congenital anomalies, evaluate health care policies and study possible risk factors. However, electronic health care databases are potentially valuable data sources that could enhance the data in congenital anomaly registries. Algorithms to automatically discriminate between major congenital anomalies and suspected or minor congenital anomalies in health care databases are being developed to optimise the use of these databases (Astolfi et al, 2016).

Terminations of pregnancy for fetal anomalies

Data from three EUROCAT registries were included in the study on the evaluation of the coding of TOPFAs. The EUROlinkCAT study planned to include many more registries, but there was very restricted access to the information on TOPFAs in the hospital databases in many countries and therefore the coding could not be evaluated.

Almost all pregnancies ending with a TOPFA were visible in the hospital databases (100% in two countries and 78% in one country). The percentage of cases for whom there was a code for a congenital anomaly was 90%, 67% and 44% in the three countries. A more specific code for the anomaly was mainly given for neural tube defects and chromosomal anomalies. For other anomalies there was no code given to state in which organ system the anomaly was.

The overall conclusion from the study is that hospital databases cannot be used for studies on risk factors/pregnancy exposures for congenital anomalies or for surveillance as a high proportion of the most severe anomalies do result in a TOPFA and hence will not be included in these studies.

Experiences from the EUROlinkCAT study on hospitalisations and surgeries

Children with some specific congenital anomalies can be identified in hospital discharge databases from the codes for the neonatal surgery that they require for survival. However, the EUROlinkCAT study on hospitalisations and surgeries for children with congenital anomalies found that there were problems with missing data in the neonatal period. In most countries it takes some time before the newborns have their permanent name and ID number/health care number. This may explain why the procedure codes for surgery were missing for some children with anomalies requiring neonatal surgery for survival.

For example, for esophageal atresia, in data from seven congenital anomaly registries there were 399 liveborn children without other anomalies identified in hospital discharge databases that were alive 28 days after the birth. For 91 (23%) of these children, the code for the surgery was not visible in the databases. For several registries further investigation of these individual cases was undertaken and reasons for these results included:

i. Child transferred to a specialist hospital outside the region
ii. Child treated at a private hospital
iii. Age at time of surgery was incorrectly coded
iv. Surgery incorrectly coded
v. Child died before surgery, but the death not coded in the hospital discharge database

Results from questionnaires to clinicians

A short questionnaire with questions on coding at discharge from hospital and for outpatients was sent by email to all EUROlinkCAT partners having a EUROCAT registry on 13th November 2019. The covering letter requested the registry to ask several doctors in some of the hospitals the registry collected information from to complete the questionnaire. Proposed departments were obstetrics, neonatology, paediatrics, cardiac surgery, gastro-intestinal surgery or others that the registry had strong connections with.

A total of 73 questionnaires were received from 11 registry areas (Table 1). One registry filled in the questionnaire themselves (Finland) and one registry filled in the questionnaire based on replies from five

clinicians (South Portugal). Additional detailed information on coding status in 2020 was provided by two clinicians (see Appendix).

Table 1 Registries that completed questionnaires and number per registry

EUROCAT Registry	**Number of completed questionnaires received**
Belgium – Antwerp	2
Croatia – Zagreb	1
Denmark – Funen	4
Finland	1 (registry leader)
Germany – Saxony	7
Italy – Tuscany	4
Northern Netherlands	1
Poland	45
South Portugal	1 (based on 5 replies)
Spain – Valencian Region	2
UK – SGUL	1

The questionnaires were mainly filled in by medical doctors working in paediatrics and neonatology.
All countries except Italy, used ICD10 for coding of diagnoses in hospitals. Two questionnaires from Poland said that both ICD9 and ICD10 were used. The Valencian Region reported use of a Spanish version of ICD10. South Portugal and Northern Netherlands reported use of ICD9 in hospitals up to 2014 and hospitals in the Valencian Region used ICD9 up to 2015. In Poland, a restricted list of ICD10 codes in hospital was implemented in 2008.

All replies, except two from Tuscany, stated that several ICD codes could be given at discharge. For one department in Italy not all patients were given an ICD9 diagnosis at discharge and another department said that only one code was given.

For outpatients the replies differed. Seven replies said that they had no outpatients in their department. Replies from five departments in Italy, Spain, Germany and Netherlands stated that no ICD codes were given for an outpatient visit. Five doctors said that only one code could be given and the remaining stated that several codes could be given.

The six questionnaires from Antwerp, Northern Netherlands, the Valencian Region and UK-SGUL said that trained coders were responsible for the coding of discharge diagnosis (Table 2). A further three departments in Saxony and one in Poland also used trained coders for discharge diagnosis together with the medical doctors. In the remaining four departments in Saxony and in South Portugal the coding was done by both medical doctors and trained coders. In Croatia, Denmark, Finland, Tuscany and most departments in Poland the medical doctors were responsible for the coding of diagnosis at discharge, sometimes with help from nurses or secretaries and in four departments in Poland with help from automatically generated diagnosis. For outpatient contacts, the picture was the same but with fewer responses (Table 2).

Table 2. Results of coding questionnaire

EUROCAT Registry	**Number coded by trained coder**	**Number coded by doctors**	**Number without diagnosis from outpatient visits**
Belgium – Antwerp	2/2	0/2	1/1
Croatia – Zagreb	0/1	1/1	0 /1
Denmark – Funen	0/4	4/4	0/4
Finland	0/1	1/1	0/1
Germany -Saxony	7/7	4/7	1/3
Italy – Tuscany	0/4	4/4	1/3
Northern Netherlands	1/1	0/1	0/1
Poland	2/45	43/45	0/44
South Portugal	5/5	5/5	0/5
Spain – Valencian Region	2/2	0/2	2/2
UK – SGUL	1/1	0/1	0/1

Nine departments replied that a given code could not be changed. 56 departments said that a given code could be changed with very variable time interval where this was possible (ranging from days to years).

One department In Tuscany and all departments in Poland reported that the data on diagnosis were kept locally. All others said that the information was shared with regional or national authorities. The aim of recording the diagnosis was mainly for financial purposes in Germany and Poland. For other areas, the data

were also used for administrative and statistical purposes and less frequently for research (*note that clinicians may not know how data are distributed to other databases and used).*

Conclusions from the questionnaire.

In most countries included in the survey the hospital databases used ICD10 for coding of diagnoses. The coding at discharge was done by either the doctors or trained coders. In most databases more than one ICD code could be given for the hospital stay and it was possible to change the code at a later stage. In some hospitals not all patients were given an ICD code for an outpatient visit.

Conclusions

Data routinely collected in electronic health care databases should be improved to enable the data to be used in the surveillance and research of congenital anomalies. Codes for classifying and reporting the congenital anomalies resulting in TOPFAs in electronic health care databases need to be developed. In addition, the accuracy of the coding of congenital anomalies in all births should be improved and algorithms to accurately discriminate between major congenital anomalies and suspected or minor anomalies should continue to be refined. Some countries, mainly in southern and eastern Europe, who do not currently have electronic health care databases, should be supported in establishing them, to enable surveillance to occur across Europe. A set of recommendations have been made to enable the potential of the data in electronic health care databases to be fully exploited in the primary prevention of congenital anomalies: please see pages 4 and 5.

References

- Astolfi G, Ricci P, Calzolari E, Neville A, Pironi V, Santoro M, Bianchi F. [Validation of an algorithm for identifying cases with congenital malformations by using hospital discharge records]. Epidemiol Prev. 2016 Mar-Apr;40(2):124-30. Italian.
- Blais L, Berard A, Kettani FZ, Forget A. Validity of congenital malformation diagnostic codes recorded in Quebec's administrative databases. *Pharmacoepidemiol Drug Saf* 2013;**22**:881-889.
- Boulet SL, Correa-Villasenor A, Hsia J, Atrash H. Feasibility of using the national hospital discharge survey to estimate the prevalence of selected birth defects. *Birth Defects Res A Clin Mol Teratol* 2006;**76**:757-761.
- Charlton RA, Neville AJ, Jordan S, Pierini A, Damase-Michel C, Klungsøyr K, Andersen AM, Hansen AV, Gini R, Bos JH, Puccini A, Hurault-Delarue C, Brooks CJ, de Jong-van den Berg LT, de Vries CS. Healthcare databases in Europe for studying medicine use and safety during pregnancy. Pharmacoepidemiol Drug Saf. 2014 Jun;23(6):586-94
- Boyd PA, Haeusler M, Barisic I, Loane M, Garne E and Dolk H. Paper 1:The EUROCAT Network – organisation and processes. Birth Defects Research (Part A) 2011; 91: S2-S15
- Charlton RA, Weil JG, Cunnington MC, de Vries CS. Identifying Major Congenital Malformations in the UK General Practice Research Database (GPRD). Drug Saf 2010; 33 (9): 741-750
- De Lusignan S, Minmagh C, Kennedy J, Zeimet M, Bommezijn H, Bryant J. A Survey to Identify the Clinical Coding and Classification Systems Currently in Use Across Europe. Medinfo 2001; 86-89
- Dolk H, Loane M, Teljeur C, Densem J, Greenlees R, McCullough N, Morris J, Nelen V, Bianchi F, Kelly A. Detection and investigation of temporal clusters of congenital anomaly in Europe: seven years of experience of the **EUROCAT** surveillance system. Eur J Epidemiol. 2015 Nov;30(11):1153-64.
- Garne E, Olsen MS, Johnsen SP, Hjortdal V, Andersen HØ, Nissen H, Søndergaard L, Videbaek J; Danish Register of Congenital Heart Disease. How do we define congenital heart defects for scientific studies? Congenit Heart Dis. 2012 Jan-Feb;7(1):46-9.
- Hexter AC, Harris JA, Roeper P, Croen LA, Krueger P, Gant D. Evaluation of the hospital discharge diagnoses index and the birth certificate as sources of information on birth defects. *Public Health Rep* 1990;**105**:296-307.
- Khoury MJ, Botto L, Mastroiacovo P, Skjaerven R, Castilla E, Erickson JD. *Monitoring for multiple congenital anomalies: an international perspective*. Epidemiologic Review (1994) 16:335-350. https://doi.org/10.1093/oxfordjournals.epirev.a036157

- Metcalfe A, Sibbald B, Lowry RB, Tough S, Bernier FP. Validation of congenital anomaly coding in Canada's administrative databases compared with a congenital anomaly registry. *Birth Defects Res A Clin Mol Teratol* 2014;**100**:59-66.
- Salemi JL, Rutkowski RE, Tanner JP, Matas J, Kirby RS. Evaluating the impact of expanding the number of diagnosis codes reported in inpatient discharge databases on the counts and rates of birth defects. *J Am Med Inform Assoc* 2018a;**25**:1524-1533.
- Salemi JL, Tanner JP, Sampat D, Anjohrin SB, Correia JA, Watkins SM, Kirby RS. The Accuracy of Hospital Discharge Diagnosis Codes for Major Birth Defects: Evaluation of a Statewide Registry With Passive Case Ascertainment. *J Public Health Manag Pract* 2016;**22**:E9-E19.
- Schneuer FJ, Lain SJ, Bell JC, Goldsmith S, Mc Intyre S, Nassar N. The accuracy of hospital discharge data in recording major congenital anomalies in Australia. Births Defects Research 2021: 1-11
- Tairou F, De Wals P, Bastide A. Validity of death and stillbirth certificates and hospital discharge summaries for the identification of neural tube defects in Quebec City. Chronic Dis Can. 2006;27(3):120-4.
- Wang Y, Cross PK, Druschel CM. Hospital discharge data: can it serve as the sole source of case ascertainment for population-based birth defects surveillance programs? *J Public Health Manag Pract* 2010;**16**:245-251.

Appendix

National Coding Systems in 2020 in Denmark and England

Denmark

Coding system: ICD10 national version since 1994, continuously updated with new extended codes

When: ICD10 codes given at any discharge and for all referrals to other department/hospital and for each out patient visit.

How many: there is no upper limit of number of codes. One main code (A- code) is given for the hospital stay or outpatient contact is given and additional codes (B-codes) are added when appropriate (chronic diseases, other health problems)

Who codes: codes mainly given by medical doctors assisted by secretaries and nurses

Aim: main aim is administrative and financial, but data is also available in Statistics Denmark for research.

Who decides: the national coding system is administrated by National Board of Health and includes ICD10 codes for disease registration, procedure and surgery codes, classification of injuries, administrative codes and codes for hospitals and departments. Health persons and institutions can apply for new codes to be implemented.

Written by Ester Garne

Coding of congenital anomalies in national health care and clinical databases and by the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) – England

Coding systems: The WHO ICD-10 classification codes are used by the health care databases in the English National Health Service (NHS), such as the Hospital Episodes Statistics (HES) database which contains details of all admissions, Accident and Emergency attendances and outpatient appointments at NHS hospitals in England. In addition, congenital anomalies are coded using ICD-10 in some specialist clinical databases, such as the neonatal intensive care clinical system, which is used by all hospitals providing maternity care services in England. The European Paediatric Cardiac Coding (EPCC) system is used for the coding of congenital cardiac conditions and is used in cardiology clinical databases. ICD-10 codes are also used for the cause of death registration purposes by the Office of National Statistics (ONS).

The WHO ICD-10 classification codes and the British Paediatric Association Classification of Diseases (BPA) code extensions are used for the registration of congenital anomalies in England by NCARDRS. NCARDRS also uses UM codes, which were initially created by BINOCAR and further developed by the NCARDRS Coding Group. The UM codes are used for the coding of antenatal scan findings. Only the ICD-10 BPA codes are used for reporting purposes. Other coding systems such as Orphanet, OPCS, the European Paediatric

Cardiac Coding (EPCC), ICD-11 are referred and mapped to in order to provide more precise coding guidance wherever relevant and as necessary for the development of electronic data feeds systems.

When: The data that feeds into HES is initially collected during a patient's time at hospital from admission to discharge and for each stay / visit to a hospital either as an inpatient or outpatient.

Data flows to NCARDRS from multiples sources from individual notification forms completed by clinical staff (antenatal and postnatal) to bulk data feeds received electronically from HES, the ONS death registration data as well as from other local and national clinical systems. The coding of anomalies by the NCARDRS registration team happens at any stage along the patient pathway i.e. from the initial suspicion of the diagnosis through to the diagnosis being confirmed and further refined i.e. from postnatal clinical findings and diagnostic tests wherever relevant. Anomalies are coded as either suspected, probable or confirmed. Only probable and confirmed cases are submitted to EUROCAT for reporting and analysis purposes.

How many: There is no upper limit of the number of codes per case used either by HES or by NCARDRS. In NCARDRS, codes are added to the registered congenital anomaly cases as and when new clinical evidence is supplied to the Register from the notifying professionals or hospital trusts / health care services (as per the notification sources briefly outlined above)

Who codes: Trained clinical coders at each hospital code all clinical information that feeds into HES. Coding is also done by medical doctors in the specialist clinical systems. ONS uses the IRIS software, version 2013, to code cause of death, which had incorporated all updates to ICD-10 approved by the WHO.

In NCARDRS coding is performed by registration officers who are trained and specialise in the coding and classification of congenital anomalies. NCARDRS has developed a standardised, national coding guidance – the NCARDRS Coding Tool, which provides detailed guidance on congenital anomaly conditions, based on the EUROCAT coding guidance and adapted to the granular level necessary as informed from specific scenarios of diagnostics pathways, confirmation criteria and kept up to date as the clinical evidence base is evolving in the English and UK-wide health care settings.

Aim: The HES data is used for non-clinical purposes, such as research and planning health services. NCARDRS aims to provide a comprehensive, accurate, quality and timely congenital anomalies dataset to be used to meet the overall NCARDRS objectives in supporting clinical practice, patient experience and research.

Who decides: The Terminology and Classifications Delivery Service at NHS Digital decides the national standards for recording and categorising information to support care delivery, statistical analysis, research and the reimbursement of health and care providers.

The NCARDRS coding guidance is developed by the NCARDRS Coding Group, the membership of which includes disease coding and clinical expertise from across the UK. Mechanisms are in place for liaising with

external coding guidance bodies (i.e. EUROCAT, Orphanet, WHO, NHS Digital) and communicating national coding guidance developments to contribute to consistency and standardisation in the congenital anomalies and rare disease coding and case classification on international level.

Written by Sylvia Stoianova for Public Health England