Review Article

Paper 3: EUROCAT Data Quality Indicators for Population-Based Registries of Congenital Anomalies†

Maria Loane,[1]* Helen Dolk,[1] Ester Garne,[2] Ruth Greenlees,[1] and a EUROCAT Working Group‡

[1]EUROCAT Central Registry, Centre for Maternal, Fetal and Infant Research, Institute of Nursing Research, University of Ulster, Northern Ireland, United Kingdom
[2]Hospital Lillebaelt, Kolding, Denmark

Received 17 November 2010; Accepted 30 November 2010

The European Surveillance of Congenital Anomalies (EUROCAT) network of population-based congenital anomaly registries is an important source of epidemiologic information on congenital anomalies in Europe covering live births, fetal deaths from 20 weeks gestation, and terminations of pregnancy for fetal anomaly. EUROCAT's policy is to strive for high-quality data, while ensuring consistency and transparency across all member registries. A set of 30 data quality indicators (DQIs) was developed to assess five key elements of data quality: completeness of case ascertainment, accuracy of diagnosis, completeness of information on EUROCAT variables, timeliness of data transmission, and availability of population denominator information. This article describes each of the individual DQIs and presents the output for each registry as well as the EUROCAT (unweighted) average, for 29 full member registries for 2004–2008. This information is also available on the EUROCAT website for previous years. The EUROCAT DQIs allow registries to evaluate their performance in relation to other registries and allows appropriate interpretations to be made of the data collected. The DQIs provide direction for improving data collection and ascertainment, and they allow annual assessment for monitoring continuous improvement. The DQI are constantly reviewed and refined to best document registry procedures and processes regarding data collection, to ensure appropriateness of DQI, and to ensure transparency so that the data collected can make a substantial and useful contribution to epidemiologic research on congenital anomalies. *Birth Defects Research (Part A) 91:S23–S30, 2011.* © 2011 Wiley-Liss, Inc.

Key words: congenital anomalies; data quality; registries; ascertainment; completeness

INTRODUCTION

European Surveillance of Congenital Anomalies (EUROCAT) is a network of population-based registries established in 1979 to conduct epidemiologic surveillance of congenital anomalies in Europe. In 2010, membership extended to 41 registries in 21 countries, covering 31% of

Supported by the European Commission, under the framework of the European Union Health Programme, Grant Agreement 2006103 (Executive Agency for Health and Consumers).

†This is the 3rd of 6 papers of a supplement submitted to Birth Defects Research. Conflict of interest: none.

‡EUROCAT Working Group: Martin Haeusler (Styria, Austria); Vera Nelen (Antwerp, Belguim); Christine Verellen-Dumoulin (Hainaut-Namur, Belgium); Ingeborg Barisic (Zagreb, Croatia); Hanitra Randrianaivo (Ile de la Reunion, France); Babak Khoshnood (Paris, France); Berenice Doray (Strasbourg, France); Annette Queisser-Luft (Mainz, Germany); Anke Rissmann (Saxony Anhalt, Germany); Mary O'Mahony (Cork and Kerry, Ireland); Bob McDonnell (Dublin, Ireland); Carmel Mullaney (South East, Ireland); Elisa Calzolari, Emilia Romagna, Fabrizio Bianchi (Tuscany, Italy); Miriam Gatt (Malta); Marian Bakker (Groningen, Northern Netherlands); Kari Klungsoyr Melve (Bergen, Norway); Anna Latos-Bielenska (Wielkopolska, Poland); Carlos Matias Dias (Lisbon, South Portugal); Joaquin Salvador (Barcelona, Spain); Larraitz Arriola (San Sebastian, Basque Country, Spain); Marie-Claude Addor (Vaud, Switzerland); Wladimir Wertelecki (Khmelnytsky, Ukraine); Elizabeth Draper (East Midlands and South Yorkshire, United Kingdom); Judith Rankin (Newcastle, United Kingdom), Patricia Boyd, (Thames Valley, United Kingdom) David Tucker, (Swansea, Wales), Diana Wellesley (Wessex, United Kingdom).

*Correspondence to: Maria Loane, University of Ulster, Shore Road, Newtownabbey, Co Antrim, Northern Ireland, BT37 0QB.
E-mail: ma.loane@ulster.ac.uk

Published online 7 March 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/bdra.20779

all births in the European Union. The breadth of coverage encompassing regions with different health care, health care information systems, and access to resources (financial and expertise) present a challenge in terms of ensuring consistency, uniformity and quality of data collection throughout Europe. The EUROCAT data quality strategy aims at improvement, standardization, and transparency of data quality.

As Solomon et al. (1991) stated, ''A registry is a database of identifiable persons containing a clearly defined set of health and demographic data collected for a specific public health purpose . . . . There should be explicit standards and procedures for evaluating the system on a regular basis.'' As part of our data quality strategy, EUROCAT developed a set of data quality indicators (DQIs) in 2005. These DQIs are indicators calculated on the totality of data over one or several years for each registry, which serve to highlight the strengths and weaknesses of the data. The DQIs help to focus the attention of registries on areas needing improvement, and they are useful to data users for the interpretation of data. The DQIs cover the following areas: completeness of case ascertainment, accuracy of diagnosis, completeness of information on EUROCAT variables, timeliness of data transmission, and availability of population denominator information. DQIs have been published on the EUROCAT website each year since 2005 and are discussed at annual EUROCAT meetings.

The DQI set is only one of a number of components of the data quality strategy. Other components involving the EUROCAT Central Registry include:

1. Detailed registry descriptions (EUROCAT, 2010a) allowing the sources of information used by the registry, and the limitations of these data in terms of ascertainment and precision, to be understood
2. Investigation of the reasons for differences in prevalence of specific anomalies between registries for specific research projects
3. Annual statistical monitoring of trends and clusters revealing changes in data quality over time, which are investigated by each registry (EUROCAT, 2010b)
4. Case-by-case data validation through a standard validation program in the EUROCAT Data Management Program, which is used by all registries to transmit data to EUROCAT. This validation includes checks for duplicate cases and that logical values are entered (i.e., impossible or inconsistent data entries are queried).
5. Case-by-case data validation, particularly of malformation coding, through review of selected case sets for research projects and multiple malformation surveillance, with queries sent to registries

Individual registries also use other methods to improve accuracy and completeness. Sources of information for notification of each case are coded locally by registries and can be used by each registry for capture-recapture type analyses (EUROCAT, 2006) or analyses of trends in the proportion of notifications coming from each source.

In this article, we present the DQIs so that they can be used or adapted by congenital anomaly registries worldwide, and so that they can aid understanding and interpretation of EUROCAT data. We present the range of values of DQI in EUROCAT registries in the years 2004 to 2008 as an illustration.

MATERIALS AND METHODS

Summary of Methodology

Indicators are calculated for each registry and compared to the EUROCAT (unweighted) average. All registries are shown on a single output sheet (Table 1) to facilitate appreciation of the range of values and place of the registry within that range. Values that are either very high or very low compared with the EUROCAT average are deemed worthy of explanation.

The central registry calculates and publishes DQIs each year for the preceding 5 years of data, where *year* refers to year of birth, but DQI can also be applied to any time period.

DQIs are calculated for full-member EUROCAT registries that transmit individual anonymous data to the central database. In this article, we present the DQIs for 29 full-member registries for 2004 to 2008. DQIs are based on all cases of congenital anomaly coded to the Q chapter of World Health Organization International Classification of Diseases 10th revision ICD10, excluding cases with only minor anomalies according to the EUROCAT list of exclusions (EUROCAT, 2005a). EUROCAT cases include live birth (LB), fetal death from 20 weeks gestation (FD), and termination of pregnancy for fetal anomaly (TOPFA).

Definition of DQI

Table 2 shows the list of 30 individual DQIs classified by subset of DQI and the EUROCAT unweighted average DQI score.

Ascertainment

Total congenital anomaly prevalence per 10,000 births. The total prevalence is calculated as: Total cases (LB + FD + TOPFA) ÷ Total births in population.

This DQI assesses overall ascertainment of cases of major congenital anomaly within each registry. Registries can assess their performance in relation to the EUROCAT average, and prevalence below 2% is considered to represent underascertainment requiring an explanation in the registry description and improvement if possible.

Spina bifida–to-anencephaly ratio and prevalence of neural tube defects per 10,000 births. The DQI assessing the ratio of spina bifida cases to anencephalus cases is designed to indicate whether a registry may be underascertaining TOPFA. Registries with a ratio of high spina bifida to anencephalus and a lower than average prevalence of NTD may be missing TOPFA cases, because the rate of prenatal detection and termination for anencephaly is much higher than for spina bifida. The DQI is an indicator for further investigation only; it is possible that the true ratio may vary geographically.

Prevalence of selected cardiac anomalies per 10,000 births. The prevalence of five selected severe congenital heart defects (CHDs; i.e., hypoplastic left heart, transposition of great vessels, tetralogy of Fallot, coarctation of aorta, and common arterial truncus) represents CHD cases that should be consistently diagnosed across Europe in the prenatal or early neonatal period. A much lower than average prevalence suggests that registries have limited access to pediatric cardiology sources.

Prevalence of selected anomalies associated with postnatal diagnosis. Six congenital anomalies (corpus callosum anomalies, congenital cataract, coarctation of

Table 1
Results of EUROCAT Data Quality Indicators (DQI), 2004-2008, 29 registries

	EUROCAT average	Austria (Styria)	Belgium (Antwerp)	Belgium (Hainaut)	Croatia (Zagreb)	Denmark (Odense)	France (Ile de la Reunion)	France (Paris)	France (Strasbourg)	Germany (Mainz)	Germany (Saxony Anhalt)	Ireland (Cork & Kerry)	Ireland (Dublin)	Ireland (South East)	Italy (Emilia Romagna)	Italy (Tuscany)	Malta	Netherlands (Northern)	Norway	Poland (Wielkopolska)	Portugal (South)	Spain (Barcelona)	Spain (Basque Country)	Switzerland (Vaud)	Ukraine	East Midlands & South Yorkshire	UK (Northern England)	UK (Thames Valley)	UK (Wales)	UK (Wessex)
ASCERTAINMENT																														
Total no. of cases		1717	2173	1484	582	759	1860	4154	889	620	2779	478	2255	462	3436	2942	557	2183	6720	3333	1060	1051	1734	1360	2666	7026	3811	2264	5903	2891
Total congenital anomaly prevalence	239.3	331.5	221.6	236.52	183.6	282.7	251.6	311.0	348.8	488.2	319.9	276.6	183.7	128.8	174.4	195.4	281.5	238.8	383.6	231.3	111.2	177.9	216.0	368.6	233.2	200.3	234.8	183.5	348.3	206.0
Spina bifida / Anencephaly ratio	1.33	3.10	1.96	2.27	1.25	1.09	1.52	1.12	1.06	2.50	2.57	2.67	1.00	1.56	1.60	1.84	5.00	3.50	1.35	3.83	1.83	0.83	0.71	0.94	1.06	1.07	1.27	0.89	1.29	0.80
Neural Tube Defects prevalence	10.2	9.7	8.9	9.4	4.4	10.1	17.0	13.3	15.3	13.4	9.9	13.9	6.1	6.7	5.4	5.4	11.1	7.3	10.4	9.2	2.0	10.0	10.1	13.0	18.5	11.2	13.6	10.5	13.2	11.3
Selected cardiac prevalence	13.5	20.5	14.1	12.4	10.4	11.9	11.4	14.1	18.0	18.9	15.2	12.7	16.5	10.0	11.7	10.5	18.7	15.7	12.1	7.6	4.1	17.6	12.8	14.6	13.9	11.7	20.1	12.7	17.2	16.6
Selected postnatal diagnosis prevalence	18.0	33.2	24.4	14.7	9.5	19.4	19.6	26.8	33.0	28.3	29.0	20.3	18.7	7.8	15.8	15.0	18.7	18.5	11.8	11.0	4.7	21.2	28.0	39.6	16.5	11.8	19.0	15.6	28.0	15.3
Non-chromosomal syndrome prevalence	7.8	11.2	10.1	7.8	2.8	11.2	15.2	9.1	11.8	15.7	7.0	10.4	9.7	5.9	5.6	3.5	8.1	10.2	6.8	1.7	0.7	7.4	6.0	20.3	9.0	4.2	12.3	7.9	13.1	10.5
Down syndrome: Observed/Expected ratio by maternal age	0.87	1.01	0.62	-	0.63	1.17	-	1.44	1.53	1.42	0.94	0.96	0.90	0.70	0.66	0.56	1.15	0.72	0.92	0.72	0.33	0.90	1.04	1.42	0.99	0.89	1.14	0.96	1.12	1.11
Prevalence malformed SB 20-27 weeks	2.06	1.93	1.84	2.39	0.95	1.86	2.03	1.05	3.14	7.09	2.65	2.89	1.79	0.28	0.05	0.93	1.01	1.86	1.26	0.90	0.21	0.34	0.25	1.90	3.24	3.22	3.45	3.00	4.43	3.42
Prevalence malformed SB >= 28 weeks	3.02	1.54	3.16	1.91	0.63	3.72	3.52	2.99	2.75	6.30	2.53	9.84	11.08	1.67	0.81	0.93	5.56	3.17	2.85	1.25	0.21	1.35	1.12	2.44	6.03	2.82	3.02	4.13	2.89	4.63
% missing SB gestational age	1	0	6	4	17	0	0	0	0	0	0	0	0	0	11	7	0	0	0	3	0	17	0	0	1	0	0	0	1	0
ACCURACY OF DIAGNOSIS																														
% MM among non-syndromes/non-chromosomal anomalies	12.1	22.5	16.8	14.4	9.8	10.2	12.1	12.5	14.5	5.0	16.9	12.6	11.7	12.8	12.2	8.7	11.0	13.4	6.6	9.6	10.7	29.8	16.0	17.7	13.1	10.4	10.1	12.2	12.7	12.2
% SB with postmortem carried out	42	50	25	71	33	67	17	-	67	29	84	27	58	29	58	57	77	50	-	25	75	-	73	69	54	5	62	41	58	65
% SB with postmortem (results known)	35	39	19	71	33	67	15	-	67	29	67	23	58	29	58	47	77	50	-	22	50	-	73	69	51	5	62	40	53	0
% TOPFA with postmortem carried out	31	57	19	56	52	36	9	-	73	49	76	illegal	illegal	illegal	46	37	illegal	40	-	illegal	49	-	68	52	45	5	54	27	38	35
% TOPFA with postmortem (results known)	26	55	15	56	34	36	9	-	73	49	67	illegal	illegal	illegal	42	25	illegal	40	-	illegal	31	-	66	51	44	5	54	25	36	0
% Chromosomal anomalies with karyotype carried out	58	97	95	96	83	100	90	-	100	53	95	98	98	83	99	90	86	99	-	80	94	-	98	99	67	35	-	1	96	67
% Chromosomal anomalies with karyotype (results known)	56	95	31	96	67	98	90	-	100	51	95	87	96	79	98	88	86	99	-	76	91	-	98	99	66	34	-	0	93	67
% Chromosomal anomalies with karyotype text #	59	95	33	82	65	98	96	-	99	31	95	90	97	82	0	85	88	99	-	74	88	99	98	100	67	34	97	1	99	66
% Non-chromosomal MM with known karyotype	23	17	31	58	12	53	48	-	73	8	22	35	38	25	0	33	48	63	-	16	30	69	52	52	10	4	7	1	30	0
% Down syndrome with Congenital heart defects or duodenal atresia (Livebirths)	45	71	54	28	44	30	35	43	32	57	54	52	55	53	32	41	58	47	55	44	33	43	51	67	59	28	48	28	59	34
Prev selected ICD10 Q-chapter/BPA extension codes	12.8	35.5	12.6	NA	12.3	26.8	25.7	NA	21.2	6.3	44.7	17.4	6.1	3.1	11.2	8.8	47.0	11.7	0.6	0.2	2.2	20.0	20.6	48.0	15.3	9.7	15.0	5.6	24.8	15.4
Prev selected unspecified ICD10 Q-chapter codes	8.9	4.4	15.9	NA	2.5	1.5	2.8	NA	2.4	3.9	7.7	6.9	8.3	5.9	8.6	9.0	3.0	2.3	22.4	9.7	4.3	4.6	1.9	4.3	2.6	19.1	6.1	9.8	12.1	3.9
COMPLETENESS OF INFORMATION																														
No. core variables 90% complete (total=12)*	10	11	9	9	11	11	11	11	11	8	10	11	10	10	11	11	10	10	7	8	9	9	10	11	11	11	11	9	11	8
No. non-core variables 80% complete (total=26)*	9	7	4	6	15	18	14	2	18	6	7	16	7	15	10	12	15	9	2	4	6	0	14	17	19	2	5	3	10	8
% syndrome text complete	67.8	100	100	100	100	100	100	0.1	100	24	100	100	100	100	77	100	25	100	100	82	100	100	100	100	100	100	-	100	100	100
% malformation1 text complete	73	100	100	100	100	100	100	0	100	8	100	100	100	100	80	100	100	100	0	0	12	100	100	100	100	100	100	96	99	100
No other text information available	4							X							X				X			X								
TIMELINESS (Full members)																														
2008 data transmission by February 2010	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes			Yes		Yes	Yes	Yes	Yes	Yes	Yes			Yes				Yes	Yes	Yes	Yes	Yes	Yes
DENOMINATOR INFORMATION																														
80% of Maternal age denoms by 5yr groups 04-08	25	Yes	Yes		Yes	Yes		Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Availability monthly denoms 2004-2008	15		Yes	Yes	Yes			Yes			Yes	Yes	Yes		Yes	Yes	Yes		Yes			Yes	Yes	Yes	Yes	Yes				

Some registries code "Not known" for karyotype testing, yet have provided karyotype text information in the specify karyotype variable field.
"-" indicates no data available
X indicates no text information available for the remaining "specify text" variables
* years of data 2005-2008 only
Prev = prevalence per 10,000 births
SB= Stillbirths
TOPFA= Termination of pregnancy for fetal anomaly
MM = Multiple malformations
NA means Not available

Table 2
List of EUROCAT data quality indicators

Data quality indicator	EUROCAT average
Ascertainment	
Total no. of cases	
Total congenital anomaly prevalence per 10,000 births	239.3
Spina bifida:Anencephaly ratio	1.33
Neural tube defect prevalence per 10,000	10.2
Selected cardiac anomalies prevalence per 10,000	13.5
Selected postnatal diagnosis prevalence per 10,000	18.0
Nonchromosomal syndrome prevalence per 10,000	7.8
Down syndrome: Observed:Expected ratio by maternal age[a]	0.87
Prevalence malformed FD or SB 20–27 weeks' gestation per 10,000	2.06
Prevalence malformed SB $\geq$ 28 weeks' gestation per 10,000	3.02
Missing SB gestational age (%)	1
Accuracy of diagnosis	
MM among nonsyndromes or nonchromosomal (%)	12.1
SB with postmortem examination (%)	42
SB with postmortem examination (results known) (%)	35
TOPFA with postmortem examination (%)	31
TOPFA with postmortem examination (results known) (%)	26
Chromosomal anomalies with karyotype performed (%)	58
Chromosomal anomalies with karyotype perforemed (results known) (%)	56
Chromosomal anomalies with karyotype text (%)	59
Nonchromosomal MM with known karyotype (%)	23
Down syndrome with Congenital heart defects or duodenal atresia (live births) (%)	45
Prevalence selected ICD10 Q-chapter/BPA extension codes per 10,000	12.8
Prevalence selected unspecified ICD10 Q-chapter codes per 10,000	8.9
Completeness of information	
Average number of core variables 90% complete (of 12)[b]	10
Average number of noncore variables 80% complete (of 26)[b]	9
Syndrome text complete (%)	67.8
Malformation1 text complete (%)	73
Number of registries with no other text information available	4
Timeliness	
No. of registries that transmitted 2008 data by February 2010	21
Denominator information	
No. of registries with 80% of maternal age denominators by 5-year groups	25
No. of registries with monthly denominators	15

[a]New data quality indicator added in 2009–2010.
[b]Based on 2005–2008 data only.
FD, fetal death; SB, still birth; MM, multiple malformation; TOPFA, termination of pregnancy for fetal anomaly.

aorta, Hirschsprung's disease, unilateral renal agenesis, or craniosynostosis) covering a range of the major organ systems and more likely to be diagnosed after the neonatal period were selected to assess the ability of registries to ascertain cases after discharge from maternity units. A high prevalence indicates good access to pediatric sources.

Prevalence of selected nonchromosomal syndromes. This DQI assesses ascertainment of monogenic or teratogenic syndromes within each registry that usually are diagnosed within the first year after birth, as described in the EUROCAT Guide to Syndromes (EUROCAT, 2005b). A high prevalence indicates good access to medical genetic departments, high early referral to medical genetics services, or both, although syndrome diagnoses are made by pediatricians in some registries with DNA testing on request. A low prevalence indicates that cases within the database may have unrecorded syndromes or that these often postneonatally diagnosed cases are not reported to the registry. This DQI relates to accuracy of diagnosis and diagnostic coding and to case ascertainment.

Ratio observed to expected Down syndrome cases. In 2009 to 2010, a new DQI was added assessing Down syndrome (DS) ascertainment. The DS DQI calculates the ratio of observed to expected DS cases, assuming a uniform maternal age–specific DS prevalence across Europe. *Observed* (O) is the number of LBs plus the number of TOPFAs corrected for survival to LB (see Appendix for details). The calculation is: O = LB + (Corrected TOPFA $\times$ Registry-specific % LB DS of all DS births [LB + FD]). *Expected* (E) is based on 5-year-modeled, age-specific LB estimates (Savva et al., 2010) applied to the maternal age profile of each registry birth population (See Appendix). A low O:E ratio indicates that registries are missing cases, for example because they do not have direct notification of DS cases from cytogenetic laboratories serving their region's population.

Prevalence of malformed stillbirths by gestational age. There are two DQIs for assessing the ascertainment of fetal deaths associated with congenital anomalies, and they provide an essential part of perinatal mortality information based on registry data. The DQI for gestational weeks 20 to 27 assesses the extent of ascertainment of fetal deaths that may occur before the gestational age threshold for stillbirths in each country. The prevalence of fetal deaths at 28 weeks gestation and over represents those that should appear in stillbirth statistics in all countries. A DQI much lower than the EUROCAT average may indicate poor autopsy rates, poor cause of death recording on stillbirth certificates, or poor access by the registry to information about fetal deaths. A high TOPFA rate in the registry area may also lead to a low fetal death rate. Alternatively, misclassification of late TOPFA as fetal deaths or stillbirths can increase the fetal death and stillbirth rate. The number of fetal deaths is low compared with the total number of cases with congenital anomalies, so for smaller registries a few missing cases will have major influence on this DQI. To gauge the reliability of the information on fetal deaths by gestational age, we provide the proportion of fetal death cases with gestational age unknown.

Accuracy of Diagnosis

Proportion of multiple congenital anomaly cases among nonsyndromic and nonchromosomal cases.

EUROCAT has developed an algorithm to identify potential multiple congenital anomaly cases for surveillance (EUROCAT, 2005a; EUROCAT Report 9 paper). This DQI calculates the proportion of potential MM cases among cases that are not classified as syndromes or chromosomal anomalies by the algorithm. A high proportion of potential MM cases may indicate that codes for major anomalies are being used to classify minor features of a syndrome. A low percentage may suggest inadequate follow-up of cases diagnosed with an anomaly to record the complete diagnosis.

Proportion of stillbirths and terminations of pregnancy with postmortem examination performed. These DQIs assess the proportion of FD and TOPFA that have postmortem examinations performed and whether the postmortem results are known to the registry. Postmortem examination results will confirm a diagnosis of a congenital anomaly and may identify further (particularly internal) anomalies if not detected during prenatal screening or at birth. A low proportion suggests that accuracy of diagnostic information is low.

Proportion of chromosomal case with karyotype performed. These DQIs indicate the proportion of chromosomal cases with a karyotype performed and whether the karyotype results are known to the registry. The availability of karyotype results within a registry confirm accuracy of diagnosis and provide additional clinical information, such as presence of mosaicism or translocations.

Proportion of nonchromosomal multiple anomaly cases with known karyotype. This DQI indicates whether the presence of a chromosomal syndrome has been ruled out in potential multiple congenital anomaly cases using the algorithm described In EUROCAT Report 9, paper 5. A high proportion indicates that the cases are correctly classified as nonchromosomal. Lower proportions are more difficult to interpret, because accuracy depends on how well referral for karyotyping corresponds to the likelihood of a chromosomal syndrome given the malformations present.

Proportion of LB DS cases with congenital heart defects or duodenal atresia. This DQI assesses the extent to which the individual features of a syndrome are coded. EUROCAT recommends the coding of all major anomaly components of a syndrome.

Prevalence-selected ICD10 Q-chapter and BPA extension codes. The ICD10 codes for certain congenital anomalies have an extra BPA extension code added to allow more specificity of the anomaly. This code applies to the following selected anomalies: Q00.00 (anencephaly, acephaly, acrania, amyelencephaly), Q00.20 (iniencephaly, open), Q04.00 (agenesis of corpus callosum), Q04.35 (hydranencephaly), Q21.10 (ostium secundum atrial septal defect, type II), Q21.21 (common atrioventricular canal), Q25.10 (preductal coarctation of aorta), Q25.11 (postductal coarctation of aorta), Q26.20 (total anomalous pulmonary venous connection-subdiaphragmatic), Q33.80 (congenital [cysti]) adenomatoid malformation of the lung), Q39.11 (atresia of esophagus with fistula between trachea and lower esophageal pouch), Q44.20 (intrahepatic biliary atresia), Q61.41 (multicystic dysplastic kidney, bilateral), Q64.20 (congenital posterior urethral valves), Q71.31 (absence or hypoplasia of thumb), Q89.80 (caudal dysplasia sequence). A low prevalence suggests that the register is not systematically using the BPA extension code.

Prevalence-selected unspecified ICD10 Q-chapter codes. Precision of congenital anomaly coding is important, and the use of unspecified codes is discouraged. This DQI calculates the prevalence of use of the following unspecified codes: Q04.9 (brain malformation), Q05.9 (Spina bifida), Q24.9 (heart malformation), Q33.9 (lung malformation), Q43.9 (intestinal malformation), Q54.9 (hypospadias), Q63.9 (kidney malformation), Q74.9 (limb malformation), Q79.9 (musculo-skeletal malformation), Q89.9 (unspecified malformation), and Q99.9 (chromosomal abnormality). In the event that an unspecified code is entered in the EUROCAT Data Management Program, the validation routine will prompt users to enter a more specific code and to provide a text description of the case. This DQI identifies registries using unspecified codes and is the only DQI for which a small prevalence is desirable.

Completeness of Information

Core data. For a case to be accepted in the database, it must have a unique local identification number, date of birth, and at least one valid ICD10 congenital anomaly code (up to nine ICD10 codes can be given). EUROCAT has 12 additional core variables that are compulsory and must also be transmitted; these are: sex, number of babies (singleton, twin, or higher order), number of babies malformed in a multiple set, type of birth (e.g., LB, FD, TOPFA), civil registration status, birthweight, gestational age, survival to one week, age when anomaly was detected, gestational age at prenatal detection, age of the mother, and the McKusick code (Online Mendelian Inheritance in Man) if a Mendelian syndrome is present. The DQI specifies how many of the 12 core variables reach 90% completeness. Because the set of core variables was changed in 2005, the DQI is limited to the period 2005 onward.

Noncore data. There are 26 ''noncore'' variables that mostly provide information on process of diagnosis, maternal exposure, family history, and sociodemographic detail (EUROCAT, 2005a). Not all registries have the resources to collect complete and accurate information on all of these variables; therefore, they are not part of the compulsory dataset. The DQI specifies the number of noncore variables that reach at least 80% completeness, because this is generally the minimal requirement for inclusion in a EUROCAT study using the variable in question. Detailed tables of the percentage of missing, invalid, or unknown values for each variable accompany the DQI and are published on the membership-only section of the website (http://www.eurocat-network.eu)—the DQI seeks to summarize this information in a single value.

Proportion of cases with text describing syndromes or congenital anomalies. In addition to entering the ICD codes, EUROCAT encourages registries to provide written text describing the syndromes and congenital anomalies registered. This process is essential for some syndrome codes that represent more than one syndrome, and it validates congenital anomaly coding and may give more detail about the anomaly. These DQIs specify the proportion of cases with text completed for the syndrome and malformation variables, depending on whether a syndrome or malformation code was entered.

Availability of other text variables. In addition to the syndrome and malformation variables, eight other variables allow for additional text information (specify type of

twin, first prenatal test, karyotype, drug, consanguinity, sibling with an anomaly, mother or mother's family with an anomaly, and father or father's family with an anomaly). This DQI indicates the registries that do not provide any information on these text variables.

Timeliness. We expected to have the information on cases born in 2008 by February 15, 2010, allowing time for late diagnosis of cardiac defects and other anomalies and time for registration processes to be completed locally. Only registries that meet this deadline are included in statistical monitoring by the Central Registry (EUROCAT Report 9, Paper 4). This DQI specifies whether the registry met the February 15 deadline for data transmission that year (e.g., that met the 2008 data transmission deadline in 2010). An earlier deadline is currently under consideration, because considerable improvement has been made in the number of registries meeting the current deadline.

Denominator information. These DQIs identify the registries that provide the number of births by 5-year maternal age group for at least 80% of the population covered by the registry (maternal age denominators) and the number of registries that provide the number of births by month (monthly denominators), in the 5-year period covered by the DQI. Total birth denominators are an absolute requirement and therefore are not included in DQI.

RESULTS

Table 1 shows the full DQI results for 2004 to 2008 for each full member registry, which are also on the EUROCAT WEBSITE (EUROCAT, 2010c). Of the 29 registries, 21 transmitted their 2008 data by the February 15 2010 deadline; 27 provided maternal age denominators, and 15 provided monthly denominators; 28 registries had more than 90% complete information on at least eight core variables. Twelve registries provided more than 80% complete information on half of the noncore variables.

DISCUSSION

Complex factors influence the quality of data collected by a congenital anomaly registry. No registry has infinite resources; therefore, decisions are made at a local level about data collection to maximize the use of available resources and expertise within the registry (i.e. to do fewer things well, rather than more things poorly). Diagnostic issues are outside the control of the registry—for example, the autopsy or karyotyping rate, rate of referral to medical geneticists, or use of routine echocardiography—although some registry personnel also have a clinical role or feed-back information to improve diagnosis locally. The DQIs thus relate to both diagnostic and registry processes. Whereas cancer registries have developed and published some data quality indicators (Curado et al., 2009; de Angelis et al., 2009), none to our knowledge have been published for congenital anomaly registries.

EUROCAT's policy is to strive for high-quality data, accompanied by transparency as to strengths and weaknesses in data quality. The provision of DQIs allows registries to evaluate their performance in relation to other registries and to focus their efforts to improve data, and it allows appropriate interpretations to be made of the data collected. It is possible for a registry to score low on one or a few DQIs, but to be cost-effective in focusing on some areas of excellent data as a basis for selected areas of surveillance. As a network, EUROCAT has a broad set of surveillance objectives (EUROCAT, 2005a), but not all registries have an official capacity or remit to meet all these objectives. Some registries are best suited to participate in timely statistical monitoring in relation to environmental teratogenic exposures, whereas others are better suited to evaluate the population sensitivity of prenatal screening for detecting affected pregnancies, to establish the prevalence of rare genetic syndromes, or to participate in pharmacovigilance studies (EUROCAT, 2005a).

A distinction needs to be made between the underlying prevalence of congenital anomalies, and the prevalence of diagnosed congenital anomalies. The latter is affected by the quality of diagnostic services available and their accessibility to all those who may need them. A low prevalence of an underdiagnosed syndrome may on the one hand be considered poor quality data, but on the other hand be considered good quality data regarding what is diagnosed in the population. Ultimately, quality of data can only be properly assessed in relation to the specific purposes for which the data will be used. The EUROCAT DQI are related to the EUROCAT surveillance objectives, but will have different importance in relation to different types of surveillance activity.

Many of the DQI should be interpreted not as quantitative assessments of data quality, but as indicators that show where further investigation is required. It is particularly informative to look at the pattern of DQI, rather than any individual value – the pattern of DQI within a registry, or the pattern of a specific DQI across registries. A registry with values across many DQI which are much lower than the EUROCAT average needs a general approach to improving its registration sources and methods, which may require increased resources. Some of the variation in certain DQIs may reflect true geographic variation rather than data quality. For example, while a low NTD rate accompanied by a high spina bifida: anencephaly ratio may be an indication of missing terminations, there may also be real variation in ratio related to NTD prevalence (Dolk et al., 1991) and registries should examine their position within the pattern of these DQI. Similarly there may be some real variation in the prevalence of selected severe CHD, or selected syndromes.

The DS-related DQIs are an area of active development, and they have only been introduced in 2010. Where there is a high prenatal diagnosis and TOPFA rate, this may be selectively leading to termination of DS with severe associated structural anomalies (e.g., congenital heart defects), which is an area of ongoing research in EUROCAT. Alternatively, natural late fetal death rates even in countries without TOPFA are lower than reported in previous decades. The survival correction factors to LBs published in the literature (Morris and Savva, 2008) may therefore not be appropriate, thus leading to O:E ratios significantly above 1 for some registries. Nevertheless, a low O:E ratio is difficult to explain other than as a case ascertainment problem; this also has implications for the DQI assessing the proportion of CHD and duodenal atresia among live born infants with DS. This percentage may be lower when there is a high proportion of TOPFA in the population for the reasons explained previously. Again, the DQI serves as an indicator that

further investigation is required, and it is not an absolute assessment of data quality.

The DQI should be interpreted in the context of the standard registry descriptions outlining registry methodology (EUROCAT, 2010a). For example, if registries only ascertain cases diagnosed up to 1 week of age, then the prevalence of conditions associated with later postnatal diagnosis are likely to be low. If access to records from certain specialist services such as pediatric cardiology has not been granted, then this will be reflected in CHD-related DQI. Registry descriptions may also explain poor ascertainment. For example, transitions to different types of electronic reporting, or transitions to different data protection requirements, can create problems.

Some DQIs are limited in interpretation because of a lack of standardization in the health services. For example, the proportion of fetal deaths with autopsy performed would ideally be framed as the proportion performed by a specialist fetal pathologist, but this information is not routinely available or easy to standardize across countries. Similarly, the karyotyping rate as an indicator of quality of information on multiply malformed cases ignores the potential influence of indication for karyotyping in determining data quality.

The DQI were developed and first implemented in 2005 and first presented at the EUROCAT Registry Leaders Meeting in Poznan, Poland, June 2005. Since then, they have been presented and discussed each year at annual meetings and published on the EUROCAT Web site (EUROCAT, 2010c). It is now also possible to track data quality over time, both in terms of the average across the network, and for specific registries. Recent comparisons of 2000 to 2004 with 2004 to 2008 have shown favorable data quality trends across the network. For example, threefold as many registries now transmit recent data by the February 15 deadline. Fewer registries have an overall prevalence of congenital anomalies below the expected rate of 2% in 2004 to 2008, and the prevalence of selected, postnatally diagnosed anomalies has increased from 8.9 per 10,000 births to 18.0 per 10,000.

The development and use of DQI is not a static process because the clinical world may change dramatically over time; therefore, the DQIs need to be evaluated regularly and new DQIs should be developed as required. We believe that the DQIs presented in this article are a useful part of the entire data quality strategy, and we will be continuing to revise and refine the DQIs. Future versions will be available on the website (EUROCAT, 2010c).

ACKNOWLEDGMENTS

We thank the many people throughout Europe involved in providing and processing information, including affected families, clinicians, health professionals, medical records clerks, and registry staff.

REFERENCES

Curado MP, Voti L, Sortino-Rachou AM, et al. 2009. Cancer registration data and quality indicators in low and middle income countries: their interpretation and potential use for the improvement of cancer care. Cancer Causes Control 20:751–756.

De Angelis R, Francisci S, Baili P, et al. 2009. The EUROCARE-4 database on cancer survival in Europe: data standardisation, quality control and methods of statistical analysis. Eur J Cancer 45:909–930.

Dolk H, De Wals P, Gillerot Y, et al. 1991. Heterogeneity of neural tube defects in Europe: the significance of site of defect and presence of other major anomalies in relation to geographic differences in prevalence. Teratology 44:547–559.

EUROCAT. 2005a. EUROCAT Guide 1.3. Instructions for the registration and surveillance of congenital anomalies. Available at: http://www.eurocat-network.eu/ABOUTUS/DataCollection/GuidelinesforRegistration/Guide1_3InstructionManual. Accessed 27th January 2011

EUROCAT. 2005b. EUROCAT guide 6: definition and coding of syndromes. EUROCAT Central Registry, University of Ulster. Belfast, UK

EUROCAT. 2006. Using capture-recapture methods to ascertain completeness of a register, EUROCAT Central Registry, University of Ulster. Belfast, UK

EUROCAT. 2010a. Member registry descriptions. Available at: http://www.eurocat-network.eu/ABOUTUS/MemberRegistries/MembersAndRegistryDescriptions/AllMembers. Accessed 27th January 2011

EUROCAT. 2010b. EUROCAT statistical monitoring report 2007. Available at: http://www.eurocat-network.eu/content/Stat-Mon-Report-2007.pdf. Accessed 27th January 2011.

EUROCAT DQI. 2010c. Available at: http://www.eurocat-network.eu/ABOUTUS/DataCollection/DataQuality/DataQualityIndicators. Accessed 27th January 2011

Morris JK, Savva GM. 2008. The risk of fetal loss following a prenatal diagnosis of trisomy 13 or trisomy 18. Am J Med Genet A 146A:827–832.

Savva GM, Walker K, Morris JK, et al. 2010. The maternal age-specific live birth prevalence of trisomies 13 and 18 compared to trisomy 21 (Down syndrome). Prenat Diagn 30:57–64.

Solomon DJ, Henry RC, Hogan JG, et al. 1991. Evaluation and implementation of public health registries. Public Health Rep 106:142–150.

APPENDIX

Details of calculations for observed and expected DS cases

Indirect standardization (ISR) was used to compare the observed (O) number of DS cases corrected for survival to LB in each EUROCAT registry with the expected (E) number of DS cases based on age specific LB plus corrected TOPFA estimates applied to the registry population (ISR = O ÷ E).

Calculation of corrected observed DS cases. DS TOPFA cases include cases that are detected prenatally before 20 weeks gestation and that would not have survived to 20 weeks gestation, because of the significant

Table A1
Survival correction weights: probability of surviving up to 20 weeks gestation applied to each TOPFA case[a]

	Gestation (weeks)										
	10	11	12	13	14	15	16	17	18	19	20
Trisomy 21	0.89	0.91	0.92	0.93	0.97	0.97	0.98	0.98	0.99	1.00	1.00

The expected number of Down syndrome DS cases is based on 5-year age-specific live birth plus corrected TOPFA estimates based on data from England and (1989–2007).

[a]J.K. Morris, personal communication.

TOPFA, termination of pregnancy for fetal anomaly.

Table A2
Trisomy 21 live births plus corrected TOPFA estimates for 5-year maternal age groups based on data from England and Wales (1989–2007)

Maternal age (years)	Down syndrome live birth rate per 10,000 births
<20	6.678
20–24	6.992
25–29	8.395
30–34	14.764
≥35	64.980

Data from Savva GM, Walker K, Morris JK, et al. 2010. The maternal age-specific live birth prevalence of trisomies 13 and 18 compared to trisomy 21 (Down syndrome). Prenat Diagn 30:57–64.
TOPFA, termination of pregnancy for fetal anomaly.

spontaneous abortion rate of DS fetuses in the first and second trimesters. We therefore corrected the observed number of TOPFA cases at less than 20 weeks gestation for the likelihood that the fetus would otherwise have survived to 20 weeks gestation. We assume DS diagnosed after 20 weeks gestation will always be identified whether it is a TOPFA or an FD.

1. Adjust DS TOPFA pre-20 weeks to the proportion that would survive to 20 weeks gestation using the probability of surviving weights listed in Table A1. These cases are the ''corrected early TOPFA cases''.
2. Sum the number of corrected early TOPFA cases and the number of late TOPFA DS cases (>20 weeks gestation), multiply by the EUROCAT registry-specific proportion of LB DS cases of all DS births (LB + FD). This result is the proportion of TOPFA cases that would have resulted in an LB.
3. Add the number of LB DS cases to Step 2 to obtain the corrected observed number of DS cases.

Calculation of corrected observed DS cases. Calculation of Expected DS cases The expected number of DS cases is based on 5-year age-specific LB + corrected TOPFA estimates based on England & Wales data 1989-2007, see Table A2 (Savva, 2010).