5.3 Definition of Data Quality Indicators

Introduction

Complex factors influence the quality of data collected by a congenital anomaly registry. How, where and when are diagnoses of congenital anomaly made to residents of the region? How does the registry obtain and verify this information and how does it code it? Registry data are never a "perfect" description of which babies have congenital anomalies in the population and precisely what those anomalies are. Moreover, diagnostic definitions and methods change over time.

Registries may need to make difficult decisions about which types of information to prioritize in order to allocate limited resources, particularly whether to devote resources to collection of exposure information as well as diagnostic information. Some types of information are easier to obtain in some regions than others, depending on specialist referral systems for affected children, confidentiality restrictions, what is recorded in medical notes, availability of computer databases, and willingness to collaborate of key professionals.

EUROCAT's policy is to strive for high quality, accompanied by transparency as to strengths and weaknesses in data quality. Making Registry Descriptions and Data Quality Indicators (DQI) freely available will allow registries to evaluate their performance in relation to other registries, and will allow appropriate interpretations to be made of the results of data analyses. The DQI only have relevance in relation to the objectives of EUROCAT. Different data strengths are needed to participate in timely statistical monitoring in relation to environmental teratogenic exposures, or to evaluate the population sensitivity of prenatal screening for detecting affected pregnancies, or to establish the prevalence of rare genetic syndromes.

The first part of the evaluation of data quality is to read the detailed registry description (available on the EUROCAT website). Key questions to address in evaluating these registry descriptions include the following:

1. Does the registry describe its methods in enough detail? If not, the data emanating from the registry must be regarded as uninterpretable.
2. Is the registry fully population-based? If it covers residents of an area, how does it ensure coverage of residents delivering or seeking paediatric services elsewhere? If it is not entirely population-based, how does it avoid inclusion of selective referrals of high risk pregnancies into registry hospitals?
3. How does the registry ensure coverage of liveborn cases diagnosed after the early neonatal period?
4. How does the registry ensure coverage of late fetal deaths and stillbirths?
5. By what process are terminations of pregnancy following prenatal diagnosis identified in the population?
6. What specialist services are accessed for information, and are these services used for case identification or only for follow-up of known cases?

The second part of the evaluation of data quality is to look at the data generated by the registers.

The DQI of a particular registry can be compared to the EUROCAT average. Strong deviations on either side of the average should be examined. The set of data quality indicators has been produced under the headings:

Ascertainment
Accuracy of Diagnosis
Completeness of Information
Timeliness
Denominator Information

List of Data Quality Indicators (DQI)

Ascertainment

- **Total number of cases**
- **Total congenital anomaly prevalence (>200 per 10,000 births expected), with 95% confidence intervals**
  All cases in malformation chapter ICD10 (Q chapter) or ICD9 (range 740-759), including outside malformation codes (D215, D821, D1810, P350, P351, P371, 74, 75, 27910, 2281, 76076, 76280, 7710, 7711, 77121). EUROCAT minor anomalies are excluded.
- **Prevalence of anencephalus**
- **Prevalence of severe cardiac defects**
- **Prevalence of selected postnatal diagnosis**
  Includes codes for corpus callosum anomalies (Q040), cataract (Q120), coarctation of aorta (Q251), Hirschprung's disease (Q431) and craniosynostosis (Q750).
- **Prevalence of genetic syndromes and microdeletions**
- **Prevalence of malformed fetal deaths**
- **Down syndrome: Observed/Expected ratio by maternal age**
  This calculates the ratio of Observed to Expected Down Syndrome cases. Observed (O) is the number of livebirth (LB) + fetal death (FD) ≥20 weeks gestational age + the number of TOPFA corrected for probability of fetal survival to 20 weeks.

  The calculation is:
  O = LB + FD + (TOPFA corrected to 20 weeks gestational age)

  Expected (E) is based on EUROCAT average 5 year maternal age-specific estimates (LB +FD + TOPFA corrected to 20 weeks) for the time period of analysis applied to the maternal age profile of each registry birth population.

Accuracy of diagnosis

- **% potential multiples according to the flowchart variable**
- **% fetal deaths with post-mortem examination carried out**
- **% TOPFA (GA ≥ 15 weeks) with post-mortem examination carried out**
- **% chromosomal cases (except trisomy 13, 18 and 21) with karyotype text**
- **% Non-chromosomal potential multiple cases with known karyotype**
- **Prevalence of selected exact 4-digit Q-BPA codes**
  Selected Q-BPA codes = Q0000, Q0020, Q0400, Q0435, Q2110, Q2121, Q2510, Q2511, Q2620, Q3380, Q3911, Q4420, Q6141, Q6420, Q7131, Q8980.
- **Prevalence of selected unspecified Q codes**
  Selected unspecified codes = Q049, Q059, Q249, Q339, Q439, Q549, Q639, Q749, Q799, Q899, Q999
- **% livebirths with ASD, VSD, hydronephrosis, hypospadia or club foot with known data on surgery**

Completeness of information

Completeness of information describes the amount of **complete valid data** transmitted to Central Registry (eg not known values, invalid values, or missing/blank fields are counted as incomplete information)

- **Number of core variables 90% complete (out of chosen 11)**
  Variables are: sex, number of babies/fetuses delivered (nbrbaby), number of malformed in multiple set (nbrmalf), type of birth (type), birth weight (weight), length of gestation in completed weeks (gestlength), survival beyond one week of age (survival), when discovered (whendisc), if prenatally diagnosed, gestational age at discovery in completed weeks (agedisc), age of mother at delivery (agemo), civil registration status (civreg).
- **Number of non-core variables 80% complete(out of chosen 26)**
  date of death (death-date), condition at discovery (condisc), karyotype of infant/fetus (karyo), post mortem examination (pm), date of birth of mother (datemo), mother's residence code (residmo), total number of previous pregnancies (totpreg), mother's occupation at time of conception (occupmo), assisted conception (assconcept), illness before pregnancy (illbef), illness during pregnancy (illdur1), drugs1, consanguinity (consang), previous malformed siblings notified to EUROCAT (prevsib), sibling ID number notified to Central Registry (sib1), siblings with anomalies (sibanom), mother's family with anomalies (moanom), father's family with anomalies (faanom), first postitive prenatal test (firstpre), first surgical procedure for malformation (surgery), folic acid supplementation (folic), maternal education (matedu), socioeconomic status of mother (socm), socioeconomic status of father (socf), migrant status (migrant), aetiological classification of malformation (aetiology).
- **% TOPFA with civil registration known**
- **% live births with one week survival known**
- **Medication exposure recorded using 7 digit ATC codes**
  Yes or No
- **% of ATC codes with 7 digits and in correct format**
- **% genetic syndromes + microdeletions with syndrome text complete**
- **% malformation 1 text complete**
- **Number of unresolved data edits (excluding free text fields)**

Timeliness

- **Timeliness for February deadline**
  Yes or No

Denominator Information

- **Years with 80% of maternal age denominators (out of 5)**
- **Years with monthly denominators (out of 5)**