Eurocat

European Surveillance of Congenital Anomalies

Supported by the EU-Commission Public Health Directorate, Public Health Programme
WHO Collaborating Centre for the Epidemiologic Surveillance of Congenital Anomalies

EUROCAT Statistical Monitoring Protocol

(Uploaded to EUROCAT website February 2007)

EUROCAT Central Registry
University of Ulster
Newtownabbey, Co Antrim
Northern Ireland, BT37 0QB

Tel: +44 28 9036 6639
Fax: +44 28 9036 8341
Email: eurocat@ulster.ac.uk
Web: www.eurocat.ulster.ac.uk

EUROCAT PROTOCOL FOR STATISTICAL MONITORING AND INVESTIGATION OF TRENDS AND CLUSTERS

CONTENTS

Acknowledgements:
Alan Kelly and Conor Teljeur (Trinity College Dublin) adapted and implemented the scan technique for statistical monitoring, in discussion with Helen Dolk (EUROCAT Project Leader, University of Ulster) with first results presented to the Registry Leader's Meeting (RLM) in Heidelberg 2003. Subsequently the statistical code was incorporated into the EUROCAT Data Management Programme (EDMP) by James Densem (Biomedical Computing Ltd). Statistical monitoring was co-ordinated by Araceli Busby (LSHTM) until 2005, and by Maria Loane (Central Registry, UU) since then. In 2006, data analysis using the standard software was transferred from Dublin to Central Registry. This protocol has been written by all those involved.

1. Background: monitoring strategy and timetable

Statistical Monitoring relates to two of EUROCAT's objectives:

- to provide essential epidemiologic information on congenital anomalies in Europe
- to co-ordinate the detection of, and response, to clusters and early warning of teratogenic exposures

EUROCAT was established in 1979 in the wake of the thalidomide epidemic, and statistical monitoring still aims at early detection of any new teratogenic drug. Since then, interest has also widened to other environmental chemicals as potential teratogens (see EUROCAT Special Report: a Review of Environmental Risk Factors for Congenital Anomaly 2004 www.eurocat.ulster.ac.uk/pubdata/envrisk.html). Statistical monitoring is only one part of surveillance of teratogenic exposures, to identify potential cause for concern where there is no specific prior hypothesis about the exposure. It is essentially a screening method to scrutinise data regularly and systematically, to detect any previously unrecognised increases in frequency. Where there are specified hypotheses about new teratogens (e.g. which drug, when, where), other direct approaches for analysis should be undertaken which are not the subject of this protocol.

The EUROCAT Coding and Classification Committee select EUROCAT congenital anomaly subgroups for statistical monitoring. All subgroups are defined in EUROCAT Guide 1.3.

Currently, statistical monitoring is conducted to detect changes in time within each registry. Analysing data across registries, and incorporating the spatial dimension are envisaged in future developments.

Statistical methods have been chosen to be relatively straightforward for public health authorities to understand and communicate, and which can be supplied in the EUROCAT software (EDMP) to member registries for use by non-statisticians.

The elements of the current monitoring strategy are:

- Common user-friendly statistical software for use centrally and locally (EDMP)
- Annual statistical monitoring for trends and clusters at central level, 15 months after last date of birth (e.g. year 2004 births included in monitoring in March 2006).
- More frequent and/or earlier statistical monitoring locally (by member registries)
- Use of EUROCAT communications and special data analyses to respond to news about clusters identified locally or outside the monitoring system
- A clear and prompt system of investigation and reporting of results
- Use of statistical monitoring additionally as a data quality control system

All full member registries able to meet the February annual data transmission deadline participate in cluster detection monitoring. All full member registries are included in trend analysis. In future, associate members will be included in trend analysis. Associate

member registries are invited to use EDMP locally for cluster detection and report their results to Central Registry.

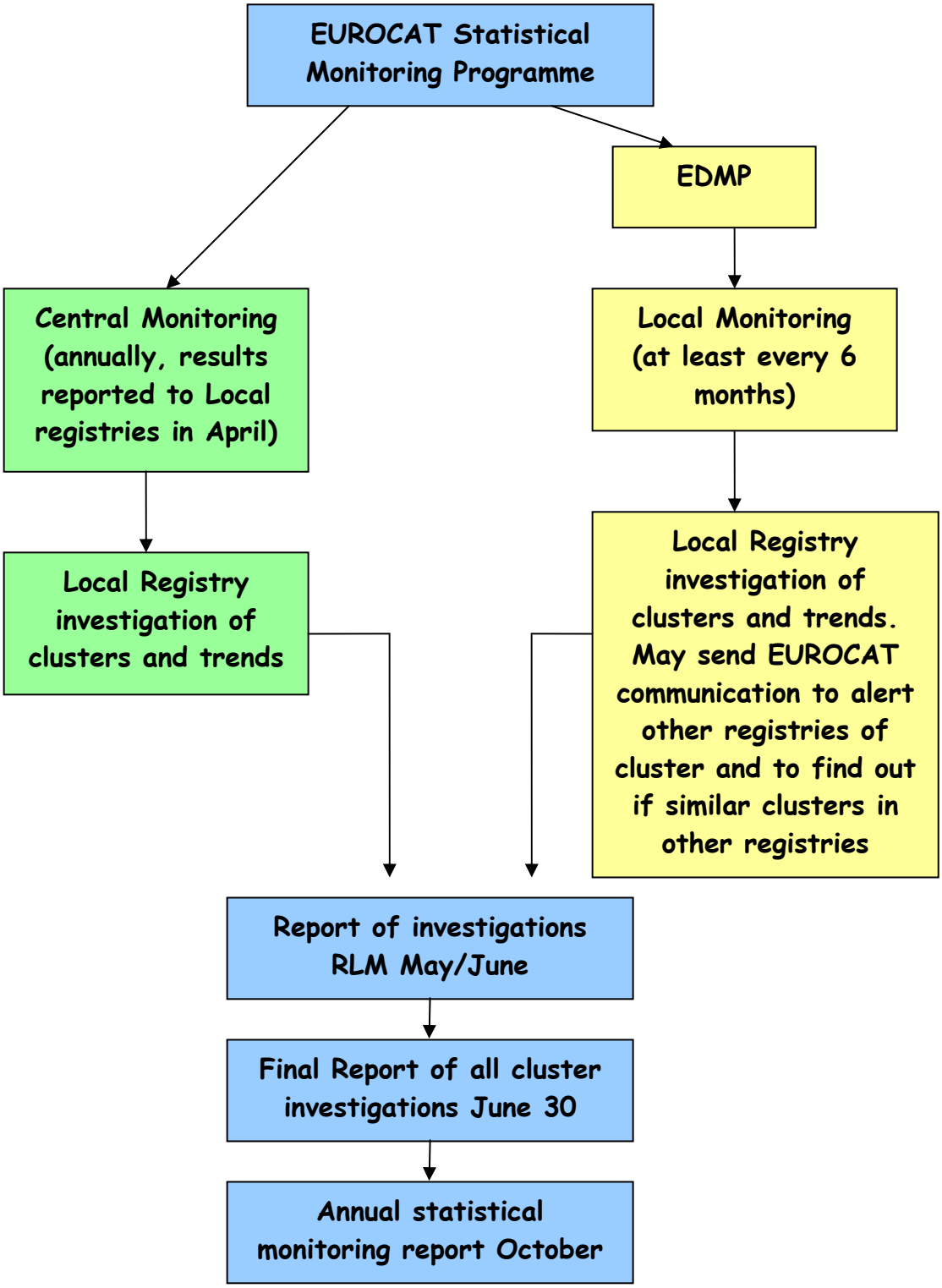

Figure 1: Monitoring and reporting centrally and locally.

2. Statistical Methods

2.1 Statistical methods for the detection of trends

A chi squared test for trend using the number of cases per year of birth and the number of births per year. Both overall year-to-year heterogeneity and the trend component are calculated. The output shows the direction and slope of any linear trend and the statistical significance (p value). Since the Chi squared test is based on conventional probabilistic statistics, 5% of the output will be statistically significant by chance.

2.2. Statistical methods for the detection of clusters

Cluster detection is based on a moving window test described by Naus and Nagarwalla[1] (see Appendix 1 for statistical details). The method uses a moving window of a given number of cases (window size), measuring the length of time between the first and last case. The method detects whether the given number of cases has occurred in a shorter time than would be expected by chance. The method is not robust with a window size of less than 5 cases. A minimum of 7 cases over the study period of interest is needed to run the analysis. All window sizes from a minimum of 5 to a maximum of the total number of cases minus 2 are tested.

Each registry and anomaly subgroup is tested independently and the analysis takes into account the number of tests with different window sizes being performed within registry/anomaly to allocate a likelihood to each cluster. The statistical significance for each identified cluster is derived from a fixed number of Monte Carlo simulations (999 iterations are used). The output will give the "p-value range" showing the credible interval within which the p-value is expected to lie (by virtue of simulation this is not a single point estimate). As a first step, a 'quick' check is performed to test whether the most significant cluster is significant at $p<=0.1$. If no significant cluster then move on to next cluster. If significant cluster is found then a full check is performed for all clusters with a first iteration p-value less than 0.06. Further iterations then give the p-value range, and any cluster where the p-value range is not either entirely or partly below 0.05 are excluded.

The method gives each cluster identified a "lambda" value denoting its likelihood or significance. Many clusters may overlap in time, the inclusion or exclusion of individual cases changing their significance. The "most significant" cluster is chosen (lowest p-value). All other significant clusters ($p<0.05$) for which 75% of cases overlap with the "most significant" cluster are considered to belong to the same cluster group. The method then looks for the second most significant cluster not already allocated to a cluster group and proceeds similarly identifying the second cluster group. The output identifies the

---

[1] Naus, J.I. (1965). The distribution of the size of the maximum cluster of points on a line. Journal of the American Statistical Association 60, 532-538

Nagarwalla, N. (1996). A scan statistic with a variable window. Statistics in Medicine 15, 845-850

most significant cluster in each cluster group, and the first and last case in time belonging to any cluster within the cluster group.

Following identification of all cluster groups over the time period of analysis, cluster groups overlapping with the last two years of data, and in which the most significant cluster is less than 18 months in length, are chosen for output and investigation. An option is also provided to scan the most significant cluster in all years of data scanned, rather than just the last two years. This is possible in Central Registry for research purposes. Central Registry also has the options of selecting the percentage missing gestational age and the cluster length.

Currently Central Registry performs annual statistical monitoring using the most recent 5 years of data. More than 5 years may tend to identify trends rather than clusters, and will be computationally slower. Less than 5 years will fail to detect if the most recent years are unusual compared to preceding years.

Since it is exposure during early pregnancy (organogenesis) that is relevant, estimated date of conception is used where possible rather than date of birth. Thus cases of different gestational ages (and terminations of pregnancy of low gestational age) are related to a common time when they passed through organogenesis. Whether analysis is by date of birth or by date of conception, the time period of analysis begins on 1 January of the first year of the five year period of analysis.

The validity of the scan method depends on having complete or virtually complete data for all months and years scanned. Where date of conception is used as a basis for cluster detection, the conception period for statistical monitoring must end 9 months before the last birth month where data collection is complete. Where full years of data are used for surveillance, this means ending the period of the scan at 31 March (date of conception) of the last year of data collected. Cluster detection based on date of conception requires the earlier cut-off point, as we do not have all of the birth outcomes for cases conceived after 31 March in any calendar year.

The Scan method identifies both significant excesses and deficits of cases.

The data requirements are:

- Full individual dataset including date of birth, outcome of pregnancy (live/still/TOP), gestational age, malformation codes and their derived anomaly subgroups (see EUROCAT Guide 1.3).

- Population change (no. births) must be less than 10% per year, as the method is based on case counts. Future modification will allow for changing population on a year on year basis.

- Where gestational age is missing, it is estimated on the basis of the average gestational age in the registry, by year, anomaly subgroup, and outcome of pregnancy. Gestational age is not estimated if it is missing for more than 10% of cases for the registry and anomaly subgroup, in which case cluster detection is based on date of birth.

- Cases without date of birth or only partial date of birth are excluded from cluster detection.

- Each dataset used for monitoring should be archived, both by Central Registry and by local registries, as the database is dynamic. Registries need to save a copy of the database file used for the statistical monitoring exercise and to rename the file eg edmpdata_statmon06

3. Use of the EDMP by local registries

Registries may wish to use the EDMP to expand the trend tests run centrally e.g. use a different period of years (e.g. since beginning of registry), congenital anomaly subgroups outside the selected subgroups for monitoring (these "user defined subgroups" can be defined within the EDMP)

Local monitoring for clusters can be run more frequently using the most recently ascertained data. As a guide, local registries should run the EDMP statistical monitoring program at least every 6 months and as early as possible after data collection is complete (apart from late diagnosed cases).

Registries must choose the time period for monitoring. Central Registry uses the last five years (reasons given above). Registries may choose to lengthen the period. Registries may also choose to end the period before the end of the year, in order to scan the most recent data. Remember dates of conception included in the monitoring must end 9 months before the last complete month of birth. Which clusters are detected and their statistical significance will depend on the number of years scanned, which is an a priori protocol decision to be made before running the software (i.e. do not "make the cluster disappear" by trying many different time periods).

Clusters and trends identified locally should be investigated in the same way as those identified by central monitoring (see section 4), and reported at the following Registry Leaders Meeting and in the Annual Statistical Monitoring Report.

Following preliminary investigations, plausible clusters and trends not resulting from data quality errors can be reported to central registry for communication to all local registries for further investigation. In this way the situation across Europe may be monitored for early detection of possible new teratogens.

The scan statistic is based on a fixed number of iterations, and if the same scan is run multiple times on the same data, slightly varying results regarding the most significant clusters, or the p value, may be obtained.

4. How to investigate trends and clusters identified by statistical monitoring

The results of all cluster investigations in local registries will be part of our report to the EC and will be available on the open website (http://www.eurocat.ulster.ac.uk/clusteradservice.html).

Investigations vary according to the anomaly of interest and local conditions. Elisabeth Robert has prepared notes to guide investigations of clusters arising from statistical monitoring, based on long experience in the Central East France register and elsewhere (Appendix 2).

Note that clusters arising from statistical monitoring are a special situation. Guidelines for investigating clusters more generally (e.g. those reported to you from the community or health professionals) are part of the EUROCAT Cluster Advisory Service (http://www.eurocat.ulster.ac.uk/clusteradservice.html).

4.1 Guidelines for the preliminary investigation of trends

Consider the following, consulting with clinicians or others where appropriate:

- Case verification
  - Confirmed and accurate diagnoses?
  - Duplicates?
  - Resident within region? (truly population-based?)
- Print out a graph of the yearly prevalence – is the trend gradual or a step change? When does the trend appear to begin?
- Has there been a change in definition or diagnosis, or diagnostic methods e.g. increasing use of prenatal or postnatal ultrasound.
- Have there been changes in how cases are reported to the register?
- Is the trend found across all reporting hospitals?
- Is the trend found in isolated or multiply malformed cases?
- Have there been changes in register population?
- Are there any known changes in the risk factors for this anomaly?
- Have any other registers experienced a similar trend for this anomaly?

4.2 Guidelines for the preliminary investigation of clusters

- Case verification
  - Confirmed and accurate diagnoses?
  - Duplicates?
  - Resident within region? (truly population-based?)
- Diagnostic dimension:
  - How heterogeneous are the diagnoses?
  - Do any other anomalies have clusters at the same time?
- Space dimension:
  - Are they clustered near each other within region?
  - Do other regions have a cluster at a similar time? (use EUROCAT communication to query other registries).
- Time dimension:
  - Is the cluster part of a longer term trend (investigate as trend rather than as cluster)
  - When does the increased risk appear to start and end – a longer period than the dates of the cluster itself? Look at other clusters in the cluster group to get an idea of the extent of the cluster.
  - If cluster is based on date of birth rather than date of conception, is it likely making assumptions about gestational age of cases within and outside the cluster that the cluster would also appear if analysed by date of conception?
- Aetiological factors:
  - Which factors have been investigated?– choice of factors to investigate depends on type of anomaly
  - Which factors have been looked at within the registry database (list variables) and outside the registry database and do any of these appear to explain the cluster?
- Local context:
  - Was there local awareness of the cluster before it was found by central statistical monitoring, either by local EDMP monitoring or other means?
  - Are there any local concerns about environmental exposures which may need investigation?

If investigation of clusters identifies data errors (e.g. incorrect diagnoses, incorrect dates of birth) these errors should be corrected and updated data included in the next data transmission to Central Registry. A record of these errors should be included in the local cluster report sent to Central Registry (ma.loane@ulster.ac.uk).

4.3 Reporting

Your report on the preliminary investigation of a cluster should include:

1) What you did to verify and describe the cases in the cluster – e.g. confirmation of diagnosis, confirmation that there are no duplicate cases, confirmation that mothers are resident in registry area (or conform to population definition), describe cases as isolated or multiply malformed or syndromes, describe available data on family history, describe whether excess cases come from entire registry area or one geographical part, describe whether excess cases born in one selected hospital or all notifying hospitals;

2) the methods and results of any investigations as to whether changes in diagnostic or reporting practice might have contributed to the cluster;

3) an account of local concerns about exposures and how they came to your attention, and methods and results of any investigations of exposures carried out;

4) whether anyone in your region (e.g. local community or health professional) had previously been aware of the cluster;

5) the basis for your decisions to conduct the investigation in the way you did.

6) What is the next step? Do you consider the trend/cluster "explained" by your preliminary investigation, or does it require a further period of surveillance to verify cause for concern, or do you intend to proceed to an aetiologic or other investigation? Note that guidance on further cluster investigation can be found on www.eurocat.ulster.ac.uk/clusteradservice.html

APPENDIX 1.

Scan statistic – procedure for calculation

Scan statistic formula based on the formula defined by Nagarwalla (1996).

Data
The date for each case is provided. A start and end date are specified and each case date is represented as a fraction of the period of interest.

$$fraction_{case} = \frac{date_{case} - date_{start}}{date_{end} - date_{start}}$$

Methods

Calculation of lambda
For a given subset of cases, it is possible to calculate the test statistic lambda as follows:

$$\lambda = \left(\frac{n}{r}\right)^{n}\left(\frac{r-n}{n}\right)^{r-n}\left(\frac{1}{d}\right)^{n}\left(\frac{1}{1-d}\right)^{r-n}$$

Where r is the total number of cases, n is the number of cases in the subset, also called the scanning window, and d is the date fraction spanned by the subset of cases. Typically a minimum scanning window of 5 is used. Lambda is calculated for every possible subset of n consecutive cases for $n = 5,\ldots, r$

The first step is to generate simulated datasets with random numbers. For each simulation, the lambda for each subset is calculated and the largest lambda is recorded. After, for example, 999 iterations the lambdas are ordered by size from smallest to largest and the 95th percentile is noted as λ_{sig}.

For the real dataset, the lambda for each subset of cases is calculated. If the lambda is greater than or equal to λ_{sig} then the cases are designated as a significant cluster and the details of the cluster, such as start date and duration, are recorded. The p-value is recorded as the position of lambda within the 999 simulated lambdas.

Bootstrapping
Bootstrapping is then used to determine and upper and lower bound for the p-value associated with each cluster. For each of 1000 iterations, 999 lambdas are sampled, with replacement, from the simulated lambdas. The position of the cluster lambda, λ_{obs}, is determined and a p-value calculated and stored. The standard deviation of the 1000 p-values is used to calculate the upper and lower confidence intervals for the original p-value. The p-value is given in the output table is the mean p-value generated in the bootstrapping exercise along with the confidence intervals.

Group allocation
The final step is to allocate clusters to groups based on their overlap. The cluster with the largest lambda is identified and is allocated to the first group. Each ungrouped cluster is then tested for overlap with the first cluster. Overlap is calculated as the number of cases appearing in both clusters divided by the number of cases appearing in either cluster. If the overlap is greater than or

equal to 0.75 then both are given the same group label. If any clusters remain ungrouped then the cluster with the largest lambda is identified and given a new group label and the procedure is repeated until all clusters are grouped.

Nagarwalla, N. (1996). A scan statistic with a variable window. Statistics in Medicine 15, 845-850

APPENDIX 2

Investigating a cluster of birth defects
Elisabeth Robert-Gnansia, Lyon FRANCE

A cluster can be defined as a space-time aggregation of occurrences, in this context of birth defects. An example of such a cluster is the occurrence of 5 infants with a given malformation born in a geographic area during 4 months and with 3,000 annual births. If the overall rate for this malformation in the region is 1/6,000, it means that the expected number of cases during the 4 months is 0.17, and 5 were found. Such aggregations can have many causes. The reason for being interested in them is that the cases may have been caused by some local source of exposure. One often thinks of industrial pollution as a cause, but other factors may explain the cluster, including many artefactual events.

Clusters may be identified as a local observation by people living in a specific area, by health care workers, or by the routine surveillance of congenital malformations or other pregnancy outcomes. In rare cases, the detection of a cluster is forwarded from the news media, or a pressure group who may perceive that a specific source of pollution may cause birth defects and therefore a cluster must have occurred around that source. It then becomes our responsibility to investigate the identified cluster. Some review articles and papers on statistical methodology have questioned the usefulness of investigating clusters, and of monitoring programs in general for the detection of a newly introduced teratogen.

There are many arguments against cluster investigations. First, there is a great statistical probability that many clusters occur by chance alone, because there are many serious congenital malformations monitored routinely and repeated evaluations. Second, in evaluating situations in which an environmental agent has been evidenced as harmful, it appears that most clusters have occurred in areas where no registry was operative or where the hypothesis came from observations made by astute clinicians. Third, clusters are often alleged in relationship to a perceived environmental hazard, such as a toxic waste site, and the investigating team often must extend the investigation far beyond what is indicated scientifically in order to respond to community concerns.

Among the possible explanations for a cluster, one thinks first of a local harmful factor, but the actual cause may be different: local temporal variations in diagnosis, in ascertainment, local aggregation of genes, e.g. due to population migrations, demographic or socioeconomic variations, and finally a random phenomenon. Clusters are different to long time trends: a steep increase of urinary flow anomalies has occurred from the early 1980s, obviously due to better and earlier diagnosis through fetal ultrasonography. After

having excluded all artefactual explanations, sometimes one nevertheless may be unable to answer the question "Is there a cluster?".

Examples can be given of surveillance events that may or may not be clusters: a marked increase in the prevalence at birth of hypospadias was observed in several registries in the world from the end of the 1980s, but no clearcut explanation can be given: improved ascertainment of minor forms or real increase? An sharp increase of the incidence of gastroschisis has been observed as well in many western registries since the beginning of the 90s, the reality of which is much less questionable than the one of hypospadias, but no explanation has been found so far.

When a cluster is considered as likely, other questions are to be answered:

1. To follow up or not to follow up? The decision to follow-up means spending time and money, and run the risk of spreading rumours.

2. How to follow up?

This should be performed in 4 stages, which might be called (a) nosodetermination, (b) chronodetermination, (c) geodetermination, (d) etiodetermination.

The first stage (*nosodetermination*) consists in confirming the case status, with exact diagnoses, and definition of inclusion and exclusion criteria: it should be checked that the cluster is not due to inclusion of minor forms of the studied malformations, which were not included in the baseline data. This is especially true for limb reduction defects (inclusions of missing phalanges?) or hypospadias (distal forms). Also diagnostic methods or medial personnel might have changed (e.g. more trained ultrasonographist or neonatologist), and explain an increased number of cases.

The second stage (*chronodetermination*) consists in checking the clustering in time: dates of birth or of termination of pregnancy are to be clearly assessed. Gestational ages are to be taken into consideration, especially when dealing with a cluster arising over a short period of time.

The third stage (*geodetermination*) should check the clustering in space. Some malformed babies might be transfered in utero for a better management of a surgical malformation at birth. One should then eliminate the possibility of a concentration of cases in university hospitals that attract patients with rare malformations, by assessing correctly the place where the child's family actually lived during pregnancy.

The fourth stage (*etiodetermination*) is the most difficult one, and consists in collecting detailed exposure information on cluster cases. A more detailed form than the routine one is to be used in order to collect all this information. If an environmental exposure is suspected, one should try to study similarly exposed areas, and to set up a case-control study within the area (3). In any case, it is essential to continue the surveillance.

The experience of registries that have investigated numerous clusters of congenital malformations shows that:

- Clusters will occur and will be found if looked for
- Most clusters have technical explanations or are artefacts
- Most clusters will go away when you look them "in the eyes"
- Nevertheless, clusters may be important for the identification of hazards in the environment and we have to investigate them

To decide whether one should follow up a cluster or not is difficult and depends mainly on available resources. If one investigates a cluster, most likely no explanation will be found. A unique example of a routine cluster investigation that led to an environmental factor as a possible explanation is given by a study led in the county of Östergötland, in Sweden. A higher prevalence at birth of congenital malformations was noticed (1), and looking at data revealed that cardiovascular malformations (CVM) explained the differences in rates. A case-referent study was performed (2), which retrieved several known risk factors for CVM (maternal diabetes, body mass index over 29, involuntary childlessness, previous spontaneous abortion, first trimester exposure to thyroid drugs and NSAIDs). All the significantly high ORs were higher in the studied county than in the rest of Sweden. Geographic comparisons of CVM were made within the county (3), and the hypothesis arose that the clustering in time and space showed that drinking water chlorination might be a mild risk factor for CVM (4).

Acknowledgement
This summary was adapted from a lecture by Pr Bengt Källén (Lund), with permission.

References
1. Blomberg M, Selbing A, Kallen B.Congenital malformations in the southeast of Sweden--a registry study with validation. Acta Paediatr. 2000 Oct;89(10):1238-43 2. Cedergren MI, Selbing AJ, Kallen BA. Risk factors for cardiovascular malformation--a study based on prospectively collected data. Scand J Work Environ Health. 2002;28:12-7.
3. Cedergren M, Selbing A, Kallen B. Geographic variations in possible risk factors for severe cardiac malformations. Acta Paediatr. 2002;91:222-8.
4. Cedergren MI, Selbing AJ, Lofman O, Kallen BA. Chlorination byproducts and nitrate in drinking water and risk for congenital cardiac defects. Environ Res. 2002;89:124-30.

APPENDIX 3.

List of congenital anomaly subgroups for monitoring.

The EUROCAT congenital anomaly subgroups are defined in EUROCAT Guide 1.3. The following list shows which subgroups are to be analysed in the following ways:

1. Prevalence by outcome of pregnancy, registries and/or years combined.
2. Prevalence by outcome of pregnancy, by registry and year
3. Analysis of trends, all cases combined
4. Detection of clusters, all cases combined.

Chromosomal and non-chromosomal cases will be analysed separately for trends and clusters.

In preparation. Meanwhile, see Annual Statistical Monitoring Report 2004.

APPENDIX 4.

EUROCAT Data Management Programme (EDMP) output – an explanation

In preparation. To include pictures of the tables and graphs.