Simple visualization techniques and statistical data analysis in prenatal diagnosis

Paula Barroso[1]; Teresa Oliveira[2]; Amílcar Oliveira[2] Avelino Rolha[3]
[1]Hospital Distrital de Santarém, Santarém, Portugal
[2]Universidade Aberta and CEAUL, Lisbon, Portugal
[3]Hospital de Santarém, Santarém, Portugal

Corresponding author: Paula Barroso, email: pbecografia@gmail.com

Abstract

Pregnant women have always believed that an unpleasant event during her pregnancy - an accident, a shock, a great sadness - could lead to the birth of an imperfect child, or worse, a "monster" (Daston, Park, 1998; Duden, 1991). Since the 70s, the improvement of prenatal diagnosis (PND) changed this situation, since it is not more than vague apprehensions, but often the prediction of an event very precise. To determine the health and condition in a fetus or embryo before it is born, the rule of PND is crucial. By employing a variety of techniques, the aim of PND is to detect birth defects, abnormalities and other conditions, since without this knowledge there could be an inauspicious outcome for both, the fetus and the mother. In our work we consider the PND data base of administrative area of Santarém, Portugal, to study the differences between the various municipalities in the region and to try to identify the key factors that may contribute to the birth disorders. Visualization is crucial to understand very large data sets and, together with statistical analysis, enables to quickly screen compounds and to obtain greater insight on factors affecting properties. In this work simple visualization techniques are applied in order to explore data and to identify trends. Software R is used on the statistical data analysis, namely on time series representation and on tests for proportion comparisons, which are performed to identify the differences/similarities among municipalities and among years, aspiring to provide an improvement of the rules for decision support, prevention measures and medical action.

Keywords - Congenital Anomalies; Prenatal diagnosis; Statistical models; Seasonality; Series; Trend.

1.Introduction

Prenatal Diagnosis (PND) is the set of Medical Procedures to evaluate the proper formation and development of the fetus before birth - it has assistencial, psychological and economical consequences - and for the population it is relevant to evaluate this cost effectiveness. Congenital malformations, congenital abnormalities and congenital defects are expressions used to describe fetal anomalies at time of birth. Congenital comes from Latin word *congenitus* that means existing from time of birth. They are morphological defects of part or full organ or even a major zone of the human body as result from intrinsic abnormal development.

During the past 50 years, studies concerning the human congenital malformations have been performed in many countries. We can see for example the paper by Queißer-Luft and Spranger(2006), where prevalence, early warning and risk factor data along with selected results are presented, based on the Mainz data (Geburtenregister Mainzer Model), from Germany. Another study, referring to data from Netherlands and comparing it with data from other countries in Europe, is presented on the thesis by Cohen-Overbeek (2008), aiming to determine the impact of prenatal detection of fetal spina bifida, to study the outcome of prenatally versus postnatal detected anomalies that are amenable to postnatal treatment and to establish prenatal cut-off levels for postnatal referral of mild renal. Rahmani et al.(2010), presents a study based on data from Iran, analysing one of the main sources of genetic disorders, the consanguineous marriages. However, in Portugal, not much attention has been given to this issue, therefore this work is justified.

Malformations are due to imperfect genetic programming (damaged by mutation) or environmental teratogenic agents or even both. Based on that, malformations are classified in three main groups: genetic, environmental and multifactor. The first group of malformations is hereditary and may occur in different generations of one family; the second occur occasionally, for more information we refer to Brent(2004) and the third group is a mix of the other two.

Besides the currently research efforts, the conjecture that about half of the congenital malformations are still unknown is a reality. From the other half, approximately 25% have chromosome defects with genetic origin and the proportion of environmental origin, like Environmental Teratogenic Agents, Physical or Chemical factors, may be about 15% of total congenital abnormalities, see Czeizel (2005).

Some defects are more frequent in specific geography, as we can see at
http://www.nhlbi.nih.gov/health/dci/Diseases/Sca/SCA_WhoIsAtRisk.html

Sickle cell disease (SCD) affects millions of people throughout the world and is particularly common among those whose ancestors came from sub-Saharan Africa; Spanish-speaking regions in the Western Hemisphere (South America, the Caribbean, and Central America); Saudi Arabia; India; and Mediterranean countries such as Turkey, Greece, and Italy. Also it is known that cerebral abnormalities are more frequent in Ireland and West of United Kingdom. Other defects are geographically independent but are dependent of other factors, e.g. mother's age in mongolism.

It is crucial to understand the currently known fetal malformations, taking out the most of the prevention multidisciplinary modeling, within its limitations in the context of prenatal diagnosis. Improving the handling of prevention factors, is expected that this will reflect a good influence on the results of malformation occurrence risk.

2.Objectives

The main aims of our work are:

(i) to test the hypothesis that fetal malformations have an homogeneous proportion distribution among municipalities.

(ii) to test the following hypothesis that fetal malformations have an homogeneous proportion distribution among years.

(iii) to study if there is some seasonal trend (month) on the fetal malformations.

(iv) to analyze the trends on the number of pregnancies and on the number of congenital effects, based on the graphical visualization of the time series.

In this context, the evaluation of the fetal malformation distribution in Santarém District, Portugal, between 2000 and 2011 will be performed. Data were available at RENAC (*Registo Nacional de Anomalias Congenitas* – Congenital Anomalies National Registry) and includes those collected in each Prenatal Diagnosis consulting at Hospital Distrital de Santarém. Number and kind of malformations will be compared among the municipalities.

3.Methods and analysis

After extracting data from the database relative to births and malformations within the studied period, visualization, time series analysis and statistical tests for proportions comparison were performed. The aim is to draw conclusions and outline future actions for prevention and genetic advise for future mothers.

Sample data overview

Within [2000-2010] period, there were 18988 pregnancies and 556 cases of malformations, with the following geographical distribution:

Table 1. Pregnancies and Malformations per Municipality

	Pregnancies	Malformations
Almeirim	2498	85
Alpiarça	788	17
Azambuja	701	23
Benavente	335	7
Cartaxo	2281	76
Chamusca	635	17
Coruche	1322	28
Rio Maior	1340	34
S. de Magos	1829	38
Santarém	6274	196
Others	985	35
Total	18988	556

We used R to the test the hypothesis that fetal malformations have an homogeneous proportion distribution among municipalities and as result at the significance level of 5% we obtain the p-value = 0.08715. So there is no statistical evidence of significant differences between municipalities.

We also compared the proportions among years on the proportions of malformations and the resulting p-value was 1.172e-13. So, at the significance level of 5% there are significant differences among the years. The year 2010 was the one with higher malformation proportion.

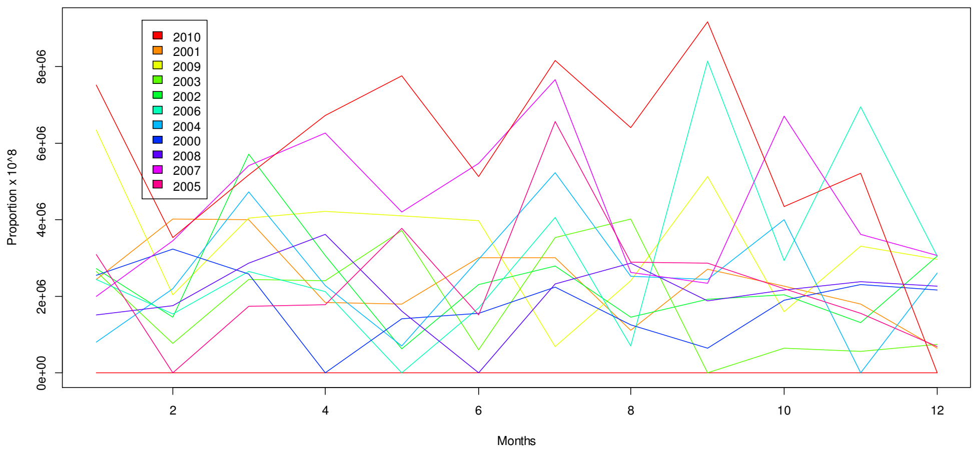

Figure 1. Malformation proportions along years 2000-2010

In figure 1 we can observe that in almost all the eleven years of observations the months of february and december are those with smallest proportion of malformation births.

In next figure we represent all the time series with the proportion of malformations for every month along years 2000-2010. Along the first 8 years we can observe some stability on the trend, except in 2004, and from 2008 on we can observe a small trend to an increase of the proportion of malformations. We are then motivated for further analysis to better understand the factors associated with this trend.

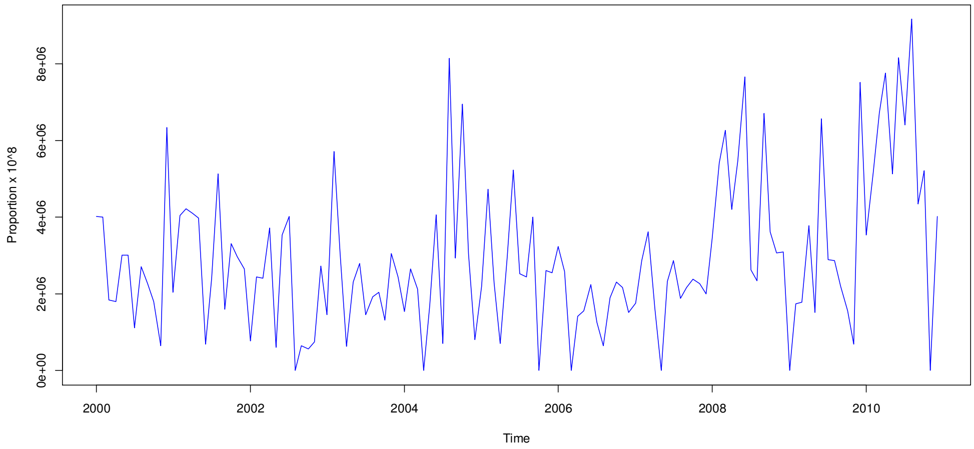

Figure 2. Monthly representation of Malformation proportions from year 2000-2010

Separating pregnancies from malformations and representing both in the graphic of figure 3, we visualize a decreasing in the number of pregnancies, so we should expect a decreasing also in the number of malformation cases.

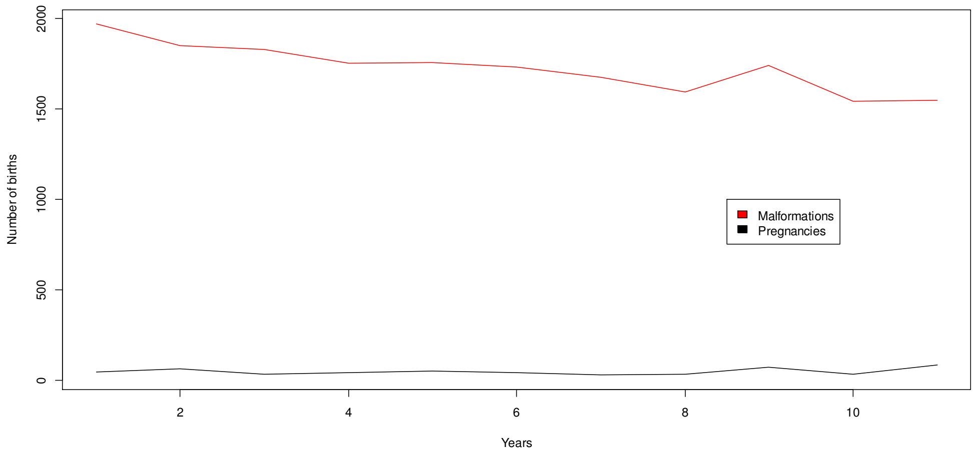

Figure 3. Pregnancies and malformation totals along years 2000-2010

However, we can observe a small increase in the number of malformations, while the number of pregnancies decreases and this is a matter of concern, given the character of the issue. It is therefore crucial to continue to explore the data in order to look for associations and key factors that may modify this troubling trend. Namely in a future work we will explore data using multivariate techniques.

5. Considerations and future work

Visualization is a must to understand very large data sets. Using visualization tools together with statistical tools enables us to quickly screen compounds and obtain greater insight on factors affecting properties. The malformation proportion is homogeneous in the observed area of 3262km^2 with a current population of 183000 habitants. However it is also important to analyse the differences/similarities among municipalities in various other aspects, aspiring to provide an improvement of the rules for decision support, prevention measures and medical action. For example in a future work we intend to investigate if the fetal malformations have an homogeneous morphologic distribution among the municipalities and, if not, it is mandatory to understand the significance of the variations. Also we conclude that birth rate is decreasing but malformations have a slightly trend to increase. What are we doing wrong? Are we in some way negatively influencing our own evolution? We need to look for answers, since an ever increasing number of genetic disorders appear as a big challenge in the coming future.

References

1. Brent R. L.(2004) "Environmental Causes of Human Congenital Malformations: The Pediatrician's Role in Dealing With These Complex Clinical Problems Caused by a Multiplicity of Environmental and Genetic Factors". PEDIATRICS. Volume113. No4. 957-968, April 2004.

2. Cohen-Overbeek, T.E.(2008) "Detection of congenital anomalies before or after birth; does it make a difference?". Rotterdam University.

2. Czeizel AE. (2005) "Birth Defects Are Preventable". *Int J Med Sci* 2005; 2(3):91-92. doi:10.7150/ijms.2.91. Available fromhttp://www.medsci.org/v02p0091.htm

3. National Heart, Lung, and Blood Institute. Disease and conditions index. Sickle cell anemia: who is at risk? Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute; 2009. Available from:
http://www.nhlbi.nih.gov/health/dci/Diseases/Sca/SCA_WhoIsAtRisk.html

4. Queißer-Luft A., and Spranger, J.(2006) "Congenital Malformations". *Dtsch Arztebl* 2006; 103(38): A 2464–71.

5. Rahmani SA, Aboualsoltani F, Pourbarghi M, Dolatkhah H, Mirza Aghazade A(2010) "The Frequency of Consanguineous Marriages and Their Effects on Offsprings in Tabriz City". *Shiraz Shiraz E-Medical Journal*, Vol. 11, No. 1, January 2010

6. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.