INFANT MORTALITY IN THE CZECH REPUBLIC*

Renata Klufová¹, Marek Šulista¹

Abstract
The aim of the paper is to briefly examine the development and the geographical distribution of newborns and infants up to one year of age in the Czech Republic. Numbers of newborns obtained from the Czech Statistical Office were used in this analysis. Tract counts are assessed by means of the GIS software and software for spatial statistics. We focused on perinatal, neonatal and infant mortality rates according to districts of residence. Descriptive techniques of spatial statistics were used to assess the geographical distribution of mortality rates.

Key words: disease mapping, newborns, infants till 1 year of age, GIS, spatial variation, spatial statistics, spatial epidemiology, mortality rates

Introduction
The evaluation of the newborns and infants up to one year and their mortality rates can be done by means of the traditional statistic and demographic tools. When incorporating spatial characteristics of tracts, it is necessary to use special tools of spatial statistics in potential combination with geographic information systems (GIS). Using these tools, it is possible to produce maps of mortality rates.

The representation and analysis of maps of disease incidence and mortality data are now established as a basic tool in the analysis of regional public health. The development of methods for mapping disease incidence has considerably progressed in recent years (8). One of the first examples of disease mapping is the Snow map of cholera victims in 1854. Disease maps are used to appraise the need for spatial variation in health resource allocation or to examine a relation of incidence to explanatory variables. In the first case, the purpose of mapping is to produce a map “clean” of any random noise and any artifact of population variation. In the case of the relation of incidence to explanatory variables, specific hypotheses concerning incidence are assessed and additional information included in the analysis.

The study of spatial variation in disease incidence is a vital component of descriptive epidemiology. The production of attractive and informative disease maps complements any formal statistical analyses of spatial variation and often, because of this attractiveness and the resulting visual impact of the maps, they will influence the reader much more than the accompanying statistics (9).

Material and Methods
Data from the Czech Statistical Office (11) were used in this paper. The publication contains detailed data on births (live, still), dead children and infant mortality (till one year of age), classified by vitality, birth weight, sex, region and district of residence, age and cause of death. The data are spatially related to districts and regions of residence.

To assess the status of the area with respect to disease incidence, it is appropriate to evaluate the expected disease incidence in the particular area and then to compare the observed incidence with the expected values.
This approach has been traditionally used for the analysis of counts within tracts (in the Czech case regions or districts of residence). The Czech data is not adequate, as only data for regions or districts of residence are available (not case events), therefore it is necessary to focus on the methods of disease mapping for tract counts.

Like in the analysis of case events, it is usual to assess maps of count data by comparison of the observed with the expected counts regarding the structure at risk population. Traditionally, the ratio of observed to expected counts within tracts (districts of residence) is called the standardized mortality/morbidity ratio (SMR) and this ratio is an estimate of relative risk within each tract. The ratio describes the risk of being in the disease group rather than the background group (8). The alternative measure of relation observed and expected counts are their differences. However, using the raw observed rates may be misleading, since the variability of such rates will be a function of the values of the background population. Several alternatives have therefore been developed to attempt to highlight better anomalous areas whilst accounting for the variability of the underlying population in each of the areas (2).

One approach how to map diseases is based on the assumption of constant overall rate of occurrence for all areas and their dependency, where counts $E_i$ are observations on independent Poisson random variables with expected value $\theta_i$ and $P_i$ is corresponding population. This approach is particularly common in epidemiology or medical geography where the rates relate to mortality or morbidity from some disease. A measure of relative risk is obtained by comparing the rate at each location to the overall mean:

$$\hat{\theta} = \frac{\sum_{i=1}^{N} E_i}{\sum_{i=1}^{N} P_i} \quad (1)$$

where $N$ is the number of areal units in the study region. An estimate of the expected number of events can be computed as

$$\hat{E}_i = \hat{\theta} \cdot P_i \quad (2)$$

The map of the relative risk is created instead of mapping the raw rates. That is, we divide the observed count, $E_i$, by its estimated expected value, $\hat{E}_i$, and multiply by 100. This method can lead to absurdly variable results when rare events are being considered or populations are highly variable. Very varying relative risk can be obtained in areas with small populations.

The suggestion in such cases is to employ various rate smoothing - a procedure used to statistically adjust the estimate for the underlying risk in a given unit, by borrowing strength from the information provided by the other spatial units (1). This approach stems from Bayesian statistics. An Empirical Bayes (EB) smoother uses Bayesian principles to guide the adjustment of the raw rate estimate by taking into account information in the rest of the sample. The motivation for considering a smoothing technique is to assess the degree of stability of the results. The overall mean is conceptualized as a random variable with its own “prior” distribution. Assume this distribution is characterized by mean $\theta$ and a variance $\phi$. The Bayesian estimate for the underlying risk at I becomes a weighted average of the raw rate $p_i$ and the prior, with weights inversely related to their variance:

$$\hat{\theta}_i = w_i p_i + (1 - w_i) \theta \quad (3)$$
with
\[ w_i = \frac{\phi}{\phi + \theta P_i}. \] (4)

A spatial rate smoother (Anselin 2004) is based on the notion of a spatial moving average or window average. Instead of computing an estimate as the crude rate for each individual district of residence, it is computed for that unit together with a set of chosen neighbors. This contrasts with EB approach, where the smoothed rate is calculated as an average of the crude rate and some separately computed reference estimate. Neighbors are usually defined in similar fashion to the specification of spatial weights in spatial autocorrelation analysis. In the case of simple contiguity (common borders) the smoothed rate is:
\[ \hat{\pi}_i = \frac{E_i + \sum_{j=1}^{j_i} E_j}{P_i + \sum_{j=1}^{j_i} P_j}, \] (5)

where \( j \in S_i \) are the neighbors for \( i \).

One of the components of spatial data exploration should be an analysis of spatial autocorrelation which can be for example visualized by means of a Moran Scatterplot \( (I) \). This is a specialized scatterplot with the spatially lagged transformation of a variable on the y-axis and the original variable on the x-axis, after standardizing the variable. The spatial lag then becomes:
\[ [Wz] = \sum_j w_{ij} z_j, \] (6)

where \( w_{ij} \) are elements of a row-standardized spatial weight matrix. Moran’s \( I \) coefficient of spatial autocorrelation is then:
\[ I = \frac{\sum_i \sum_j z_i w_{ij} z_j}{\sum_i z_i^2}, \] (7)

The techniques discussed above are available for example in the free software GeoDa. 0.9.5-i Special epidemiological scripts written for ArcGIS were also used.

**Results and Discussion**

The numbers of newborns have been increasing in the Czech Republic since 2000, with the exception of 2001, when a small decrease was recorded. 97 929 children were born in 2004, including 97 664 live births. This is the largest number of newborns since 1994. The Czech Republic achieves one of the lowest values of total fertility rate (TFR) in Europe. In the year of 2004, the value of TFR was 1.23. Some of the post-soviet countries (Byelorussian, Armenia, Ukraine) and also Italy, Spain, and Germany (from countries of the former EU) have similar values of this indicator. The highest values of TFR can be found in Ireland, France, Denmark, Norway and Finland but not one of these countries reaches the limit of net reproductive rate (2.1 children per 1 woman) when the population size remains stable.

There were 265 still births in the Czech Republic in 2004, which is 2.71‰. This value of still birth rate is the lowest in the whole study period (since 1998). It can be attributed to more rigorous monitoring of pregnants and to growing quality of health science. Only few countries in Europe can be proud of a similarly low still birth rate (Finland, Denmark and Norway). Since the values of still birth rate are too small, perennial averages were used instead.

Infant mortality rate up to one year of age is one of the lowest in the world and is still decreasing. The lowest value
was reported just in 2004 – 3.75‰. Since 2000, its value has been fluctuating around 4‰. As in the case of still birth rate, perennial averages are more suitable for the assessment of infant mortality rate and their subgroups. 366 children die till 1 year of age (210 boys and 156 girls) in 2004.

Geographically, the lowest values of infant mortality rate were found in Jihomoravský and Jihoèeský districts (under 2.5‰) and the highest in Karlovarský, Královéhradecký and Ústecký regions (under 5‰). Infant mortality rate is sometimes divided into neonatal mortality rate (from birth to 28 days) and postneonatal mortality rate (from 28 days to 1 year of age). Neonatal mortality rate is further divided into early neonatal mortality rate (under 7 days) and late neonatal mortality rate (from 8 to 28 days). Notable quality of infant mortality rate in the Czech Republic is based on low neonatal mortality rate (2.29‰), mainly on very low neonatal mortality rate (1.33‰). There were 224 deaths to 28 days of life recorded in 2004, including 130 deaths to 7 days. In Europe, only Island and Finland have lower neonatal mortality rate than the Czech Republic. The lowest value of early neonatal mortality is just in the Czech Republic (comparison of recent available sources from the database Health for All WHO).

The diseases in chapter XVI of ICD-10 “Certain conditions originating in the perinatal period” are the most common causes of infants’ deaths. Over one half of all the infants’ deaths were caused by these diseases. Respiratory disorders specific for perinatal period, slow grow and malnutrition of fetus, bleeding states and haematologic disorders of a fetus or a newborn were classified as the most frequent diagnosis from this chapter. Other 15% of infants’ deaths were caused by the diseases included in chapter XVII of ICD-10 “Congenital malformations, deformation and chromosomal abnormalities”. As the most common diagnosis from this chapter were found out congenital malformations of the circulatory system. Also “Injury, poisoning and certain other consequences of external cause” (chapter XIX of ICD-10) represents an important group of causes of deaths of children to 1 year of age.

Crude measures of relative risk of infant mortality in the period of 2000 – 2004 were calculated with respect to the geographical variation producing extremely variable results. As Bailley and Gatrell (1995) suggest, the problem of areas with small populations or rare

**Fig. 1: Standardized mortality ratios for infant deaths 2000 - 2004**
events could be countered by probability map.

According to Figure 1 we can suggest the highest values of infant SMR mainly in the north-west of the Czech Republic, in districts of large towns (such as Ostrava – OV, Brno – BM), which can be probably related to the quality of the environment, the social situation and other conditions (for example health care accessibility and its quality). High values of infant SMR can be also seen in some districts of residence in Jihočeský district (Český Krumlov – CK, Jindřichův Hradec – JH, Písek – PI) and individual districts, such as Svitavy (SY), Bruntál (BR), Jičín (JC) and Jablonec nad Nisou (JN).

Smoothing procedures were chosen in the following steps to adjust the estimate for the underlying risk in a given units. Empirical Bayes smoothing and spatial smoothing methods were used with software GeoDa and ArcGIS.

As Anselin (2004) stated, in the Empirical Bayes smoothing method, the central role is played by the regional average to which the raw rates are shrunk. When the region is highly heterogenous (as in this case), the choice of overall regional average as the reference rate may be not be appropriate. In the case of the Czech infant mortality, almost all the high outliers discussed above disappeared.

Local’ forms of analysis focus on discerning local instabilities in overall spatial association. They are utilized to look for exceptions in overall spatial pattern and attempt to identify local spatial clusters or hot spots.

The spatial smoothing, shown in Fig. 3, tends to emphasize broad subregional trends. This map highlights an East-West division of the Czech Republic, suggesting spatial heterogeneity between Bohemia and Moravia. A visualization of spatial autocorrelation seems as a suitable tool for the exploratory analysis. Suitable tools of GeoDa software were used to evaluate spatial autocorrelation in the study data. When working with rates, the underlying assumption of stationarity may be violated by their intrinsic variance instability (I), when the population at risk varies considerably across observations. Thus, Moran’s I with EB Rate was used and also EB adjusted LISA (Local Spatial Autocorrelation) maps created confirming the above mentioned trend (fig. 4).
Permutation test of spatial autocorrelation gives a soft tendency to clustering (positive spatial autocorrelation) of raw rates of infant mortality. The value of Moran’s I 0.2204 is significant at the significance level of 0.05. Results in the case of EB smoothed rates are not statistically significant. The results of LISA analysis for raw rates and also EB smoothed rates are shown in Table 1.

All the three kinds of mortality show a tendency to cluster in the space. In the case of the EB smoothed rates, it seems that the early deaths tend more to cluster than death in the greater time period (infant mortality).

**Table 1: LISA analysis of infant, perinatal and neonatal mortality**

<table>
<thead>
<tr>
<th></th>
<th>Moran’s I</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant mortality (raw rates)</td>
<td>0.2204</td>
<td>0.003</td>
</tr>
<tr>
<td>Infant mortality (EB smoothed)</td>
<td>0.1605</td>
<td>0.18</td>
</tr>
<tr>
<td>Perinatal mortality (raw rates)</td>
<td>0.1914</td>
<td>0.007</td>
</tr>
<tr>
<td>Perinatal mortality (EB smoothed)</td>
<td>0.2853</td>
<td>0.01</td>
</tr>
<tr>
<td>Neonatal mortality (raw rates)</td>
<td>0.2565</td>
<td>0.002</td>
</tr>
<tr>
<td>Neonatal mortality (EB smoothed)</td>
<td>0.2853</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Conclusion

In this paper, modern tools of spatial statistics were used to explore the geographical distribution of infant mortality in the Czech Republic. The evaluation is based on tract counts data (in case of the Czech Republic these are districts of residence), as the events data were not available. The first insight to data implies that there is a need for a further and deeper study. It is necessary to pay attention to suitable explanatory variables, such the environment conditions, social status, local human behaviour or other local health determinants (e. g. accessibility of medical services). In searching for clusters, the ‘modifiable areal unit’ problem should also be taken into consideration. It is also absolutely necessary to assess data quality in further investigations.

References