Extended Abstract Submitted
to the Annual Meeting of the Population Association of America - 2009

The Road Towards Good Vital Registration Records: the case of Brazil

Everton E. C. de Lima
(everton@cedeplar.ufmg.br)

Bernardo Lanza Queiroz
(lanza@cedeplar.ufmg.br)

This work aims to study the spatial pattern of ill-defined causes of death (chapter XVIII of the International Classification of Diseases – ICD) across the geographical areas in Brazil. The strategy propose by this study is to search for spatial correlations measures, which can easily be applied to identify the spatial centers of badly-defined mortality causes.

In the developing countries, mortality estimates and the knowledge of the levels and trends of adult mortality is limited by the quality of available data,are limited by data quality. The most common problems faced by these countries are incomplete coverage of the vital registration system and errors in age declaration for both population and deaths counts (United Nations, 1997; 2002). It is argued that, in Brazil, ill-defined causes of deaths have a negative correlation with the vital registration coverage. That is, regions that presented high levels of under-registration in the past, are now collecting more complete number of events but with worse quality, thus increasing the number of ill-defined causes.

We analyze the spatial development in four different periods: 1986, 1996, 2000 and 2006. We make extensive use of the Ministry of Health database, called DATASUS. The data is collected at municipal level, but we aggregated it into comparable small areas (which are a sum of a few cities). These regions also do have not changed their spatial configuration during the period of analysis; becoming easy to compare the results through time. We analyze the recent evolution for 558 comparable small areas.

We analyzed the data using Kulldorff and Nagarwalla's (1995) method, which considers the rates as well as the geographical coordinate information linked to these rates. This ed method is the implemented in R Dcluster library (Gómez-Rúbio, et. al., 2005), which use different models and bootstraps (Davison and Hinkley, 1997) to compute the significance of the observed values of an event in space. The method proposed by Kulldorff and Nagarwalla (1995) is applied to detect regions centers of clusters. This method of clusters detection is based on a window of variable size that considers only the most likely cluster around a given region. Kulldorff's statistic works with the regions within a given circular window. The overall relative risk in the regions inside the window is compared to that of the regions outside the window. The hypothesis is that the two relative risks, inside and outside the window, are not equal and the relative risks inside the window are somewhat higher (Bivand et. al., 2008).

It is necessary to find out whether the heterogeneity of the relative risks is assessed or not. For this analysis, an initial Chi-square test is applied (see table 1 below). This test computes the observed and expected value of the event in each area, comparing and testing the significant differences between these two quantities (Bivand et. al., 2008). The initial Chi-square test shows that, for all periods, the relative risks are not homogeneous in space, being it more heterogeneous in 1996 than in recent years.

Then, it is applied a parametric test, in order to examine whether the spatial heterogeneity of the relative risks can be fitted or not in parametric way (see table 1 below). For all periods, we have evidence that this is case. Potthoff-Whittinghill test is also applied to study the homogeneity of the means of different variables using a Poisson distribution. This test analyzes the spatial homogeneity of the relative risks considering as alternative hypothesis that these relative risks are spatial heterogeneous according to a Negative Binomial Distribution.

Finally, Tango's test of general clustering defines the kind of parametric distribution used to compute the Kulldorff's statistic. The Tango's statistics of global clustering compares the observed and expected number of cases in each region. In order to estimate the cluster centers, the different types of interaction between neighboring regions are considered (Bivand et. al., 2008). The Tango's results show that in 1996 as well as in 2000 the relative risks of ill-defined causes of death follow a Negative Binomial distribution. In 2006, however, the risks are closer to a spatial Poisson distribution.

Table 1: Group of tests

Year	Chi-square test of Homogeneity	Chi-square parametric Bootstrap	Potthoff-Whittinghill test	Tango's test
1996	89180.5***	89180.5*** (a,b)	30007274129*** (a,b)	0.024*** (b,c)
2000	87372.25***	87372.25*** (a,b)	29778958230*** (a,b)	0.014*** (b,c)
2006	48399.5***	48399.5*** (a,b)	11216562996*** (a,b)	0.007*** b

Sig: * p < 0.05, ** p < 0.01, *** p < 0.001.

a: bootstrap simulation considering a Multinomial distribution

b: bootstrap simulation considering a Poisson distribution

c: bootstrap simulation considering a Negative Binomial distribution

Since we have defined a parametric distribution of the relative risks for each period, we apply the Kulldorff and Nagarwalla's method of cluster detection. In the next maps, we show the results of the Kulldorff and Nagarwalla's cluster detection for 1996, 2000 and 2006.

Figure 1: Cluster centers across Brazilian micro-regions

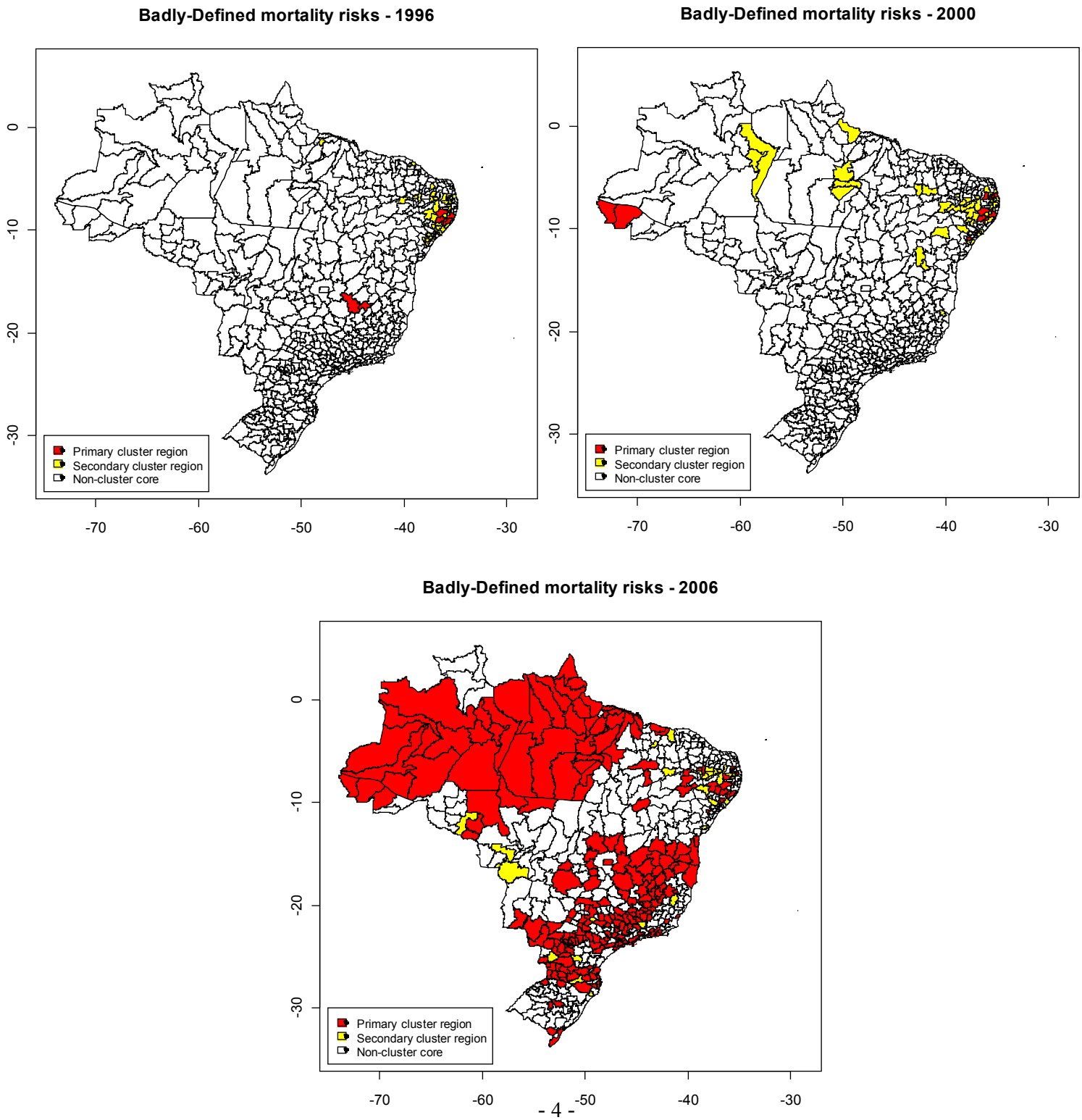

Figure 1 shows that the number of ill-defined causes of deaths has significantly increased through time. In 1996, we can barely detect cluster centers across the country. Only in the poorest locations of the country, the Northeast and North part of Minas Gerais state, there were several focuses of ill-defined risks. In 2000, the number of cluster centers increased slightly, reaching the Northern Amazon. However, in 2006, it looks like that the risks of ill-defined causes of death are spread homogeneously across the country.

We speculate that it has to do with vital registration under-registration levels, which may be reduced in the poorest regions of the country. Previous studies, in Brazil, showed that Northeast and Northern parts of the country show high levels of under-registration, reaching almost 50% in some areas in the 1980s and 30-40% in the 1990s. In recent years, 2006, the death coverage has improved significantly (for the country it is estimated to be about 90 to 95%). However, a better coverage does not come together with better registration. Our hypothesis is that deaths are now recorded but quality of information is bad, thus results in high proportion of ill-defined causes of deaths.

References

Bivand, Roger S., Pebesma, Edzer J. and Gómez-Rubio, Virgilio (2008) **Applied Spatial Data Analysis with R**. New York – NY, Springer.

Davison, A.C. and Hinkley, D.V. (1997) **Bootstrap Methods and Their Application**. Cambridge University Press, Cambridge.

Gómez-Rúbio, V., Ferrandiz-Ferragud, J. and López-Quílez A. (2005). Detecting clusters of diseases with R. **Journal of Geographical Systems**, 7:189-206.

Kulldorff, M. and Nargarwalla, N. (1995). Spatial disease clusters: Detection and inference. **Statistics in Medicine**, 14: 799-810.