

Spatiotemporal analysis of the relationship between socioeconomic factors and stroke in the Portuguese mainland population under 65 years old

André Oliveira,¹ António J.R. Cabral,¹ Jorge M. Mendes,² Maria R.O. Martins,¹ Pedro Cabral²

¹*Department of International Public Health and Biostatistics, Tropical Hygiene and Medicine Institute, Universidade Nova de Lisboa, Lisbon;* ²*Nova Information Management School, Universidade Nova de Lisboa, Lisbon, Portugal*

Abstract

Stroke risk has been shown to display varying patterns of geographic distribution amongst countries but also between regions of the same country. Traditionally a disease of older persons, a global 25% increase in incidence instead was noticed between 1990 and 2010 in persons aged 20–64 years, particularly in low- and medium-income countries. Understanding spatial disparities in the association between socioeconomic factors and stroke is critical to target public health initiatives aiming to mitigate or prevent this disease, including in younger persons. We aimed to identify socioeconomic determinants of geographic disparities of stroke risk in people 65 years old, in municipalities of mainland Portugal, and the spatiotemporal variation of the association between these determinants and stroke risk during two study periods (1992-1996 and 2002-2006). Poisson and negative binomial global regression models were used to explore determinants of disease risk. Geographically weighted regression (GWR) represents a distinctive approach, allowing estimation of local regression coefficients. Models for both study periods were identified. Significant variables included education attainment, work hours per week and unemployment. Local Poisson GWR models achieved the best fit and evi-

denced spatially varying regression coefficients. Spatiotemporal inequalities were observed in significant variables, with dissimilarities between men and women. This study contributes to a better understanding of the relationship between stroke and socioeconomic factors in the population <65 years of age, one age group seldom analysed separately. It can thus help to improve the targeting of public health initiatives, even more in a context of economic crisis.

Introduction

Global impact of stroke

Stroke is, globally, a public health problem of great magnitude, both in morbidity as well as in mortality. Comprising a range of *Cerebrovascular Diseases* from I60-I69 in the International Classification of Diseases and Related Health Problems, 10th revision (ICD-10, 2015), it ranked in 2010 as the second most common cause of death, as well as the third most frequent cause of disability-adjusted life-years (DALYs) worldwide (Krishnamurthi *et al.*, 2013), occurring more frequently in middle-age adults or older (WHO, 2014). In 2010, 16.9 million people suffered a first-time episode and 5.9 million stroke-related deaths were registered for this disease that was also responsible for 102 million DALYs lost and 33 million survivors. Moreover, a statistically significant increase was noticed in these indicators between 1990 and 2010 (Krishnamurthi *et al.*, 2013).

Stroke risk, defined as the probability of suffering a stroke (in terms of morbidity) or dying from one (mortality) has been shown to display varying patterns of geographic distribution, not only amongst countries but also between regions of the same country (Cox *et al.*, 2006; Addo *et al.*, 2012; Krishnamurthi *et al.*, 2013).

Stroke in the Portuguese context

Within the Portuguese context, cardiovascular diseases are currently the main cause of death, and within these, stroke was responsible for almost twice the number of deaths in 2012 (13,020) than ischemic heart disease (6605 deaths). In the same year, stroke also presented a higher standardized rate, with 61.4 deaths per 100,000 persons against 33.9 deaths per 100,000 persons due to ischemic heart disease. Furthermore, although there has been tendency for decrease in mortality caused by both diseases, stroke has experienced a lower decrease than ischemic heart disease, having decreased 19.10% between 2008 and 2012 (from 75.9 deaths per 100,000 persons in 2008), when compared with a 19.67% decrease in ischemic heart disease in the same period (from 42.2 deaths per 100,000 persons in 2008). Additionally, the proportion between stroke and ischemic heart disease is the inverse of what happens in most European countries,

Correspondence: André Oliveira, Department of International Public Health and Biostatistics, Tropical Hygiene and Medicine Institute, Universidade Nova de Lisboa, Lisbon, Rua da Junqueira 100, 1349-008 Lisbon, Portugal. Tel: +351213652600 - Fax: +351213632105. E-mail: miguel.oliveira@ihmt.unl.pt

Key words: Stroke; Working-age population; Geographically weighted regression; Multi-temporal analysis; Portugal.

Acknowledgements: we thank the Foundation for Science and Technology, under responsibility of the Portuguese Ministry for Education and Science, for funding this research (research grant n° SFRH/BD/63597/2009).

Received for publication: 28 March 2015
Revision received: 16 July 2015
Accepted for publication: 14 August 2015.

©Copyright A. Oliveira *et al.*, 2015
Licensee PAGEPress, Italy
Geospatial Health 2015; 10:365
doi:10.4081/gh.2015.365

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (by-nc 3.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



even the Mediterranean countries, and the reasons for this phenomenon are still not clear. Also, and more relevant for this investigation, there is clear evidence of inequalities in the spatial distribution of stroke amongst Portuguese regions (DGS, 2014b). In fact, in 2012, despite the overall national value presented above, standardized rate values amongst Portuguese regions of the Nomenclature of Territorial Units for Statistics (NUTS), level 2 (Eurostat, 2014) ranged between 53 deaths per 100,000 in the Algarve and 66,7 in the North (INE, 2012), and the reasons underlying these variations are not currently known (DGS, 2014b). Consequently, there is interest in identifying the determinants of these geographic disparities both in death and disease risk, with the aim of guiding prevention and mitigation efforts, focused not only on the elderly segments of the population, but also in younger persons. Regarding this issue, an incidence study, focused on patients of Pedro Hispano Hospital (located in Matosinhos, northern Portugal), concluded that stroke was an important pathology amongst young adults, accounting for 6% of all patients admitted with this diagnosis (Cardoso *et al.*, 2003).

Regarding Portuguese public policy, the importance of these issues and the need to address them are, in our view, comprised in the objectives of the Portuguese National Program for Cerebrovascular Diseases. Therefore, amongst its main objectives, this programme aims to improve the epidemiological and statistical knowledge on the determinants of cardiovascular pathologies, to improve the organization and rational supply of diagnostic and therapeutic care (with a special focus on cerebrovascular disease and ischemic heart disease) and to promote the evaluation of technologies in this domain (DGS, 2014a).

Determinants of stroke

Stroke is a multifactorial disease and its determinants can be discussed at the individual and ecological level. At present, much is already known about the individual determinants of cardiovascular disease in general, as well as those of stroke in particular, both modifiable or not (Sacco *et al.*, 1997). Amongst the non-modifiable, age, gender and genetic-related factors rank as relevant contributors. Hypertension is regarded as the most critical amongst modifiable determinants, closely followed by smoking, atrial fibrillation and physical inactivity (Brainin and Heiss, 2014). However, in a broad public health perspective, these individual determinants are generally the final link of a chain of events (Labarthe, 2011). Consequently, there is a need to focus on the *cause of the causes* (Marmot and Wilkinson, 2005), in other words to investigate and act on the beginning of a chain that can be long and complex. Regarding cardiovascular disease in general (and applicable also to stroke), this chain is commonly initiated by a mixture of unfavourable environmental and social situations that act as a trigger for patterns of harmful behaviour in the population. These behaviour, particularly unbalanced diets and physical inactivity will lead, in turn, to the development of individual risk factors like hypertension, obesity and others (Labarthe, 2011). The mixture of unfavourable environmental and social situations can be referred collectively as the person's socioeconomic status (SES) and comprises a different *breed* of determinants (Cox *et al.*, 2006), which includes education (Avendaño *et al.*, 2004; Lambert *et al.*, 2013; Wu *et al.*, 2013), professional occupation (Mendes, 1988; Franks *et al.*, 1991; Hayashi *et al.*, 1996; Kunst *et al.*, 1998; Cesana *et al.*, 2001) and also income and material ownership (Engström *et al.*, 2001; Li *et al.*, 2008). In a more strict sense, environmental determinants include exposure to air pollution and distance to hospital units (Busingye *et al.*, 2011; Labarthe, 2011). Links between SES and stroke have thus been identified, but the precise measure of the pathways through which SES affects stroke is still controversial (Cox *et al.*, 2006; Addo *et al.*, 2012). Consequently, understanding the

causal associations between SES and stroke can lead to better targeting and assessment of interventions by public health personnel and policy makers. The relationship between SES and stroke risk is indeed complex, and one of the aspects of this complexity is that it varies geographically, within and between populations worldwide (Addo *et al.*, 2012). In 2010, the incidence (in cases per 100,000 person-years) per country varied broadly amongst countries and their income settings (*e.g.* 60 cases in Kuwait and 504 in Lithuania). When considering the absolute number of new cases, in the 16.9 million cases that occurred worldwide, the majority (69%) took place in low- and medium-income countries. Also, between 1990 and 2010, there was a global and significant 25% increase in the incidence occurring in persons aged 20 to 64 years, and low- and medium-income countries were responsible for a great part of this increase. However, mortality presents a more complex scenario, since standardized rates now show a significant global decrease of 25% in both high and low income countries. Inversely, the absolute number of deaths increased on the same period by 26% (Krishnamurthi *et al.*, 2013).

Spatial variability of associations between stroke and its determinants

Since geographic differences in associations between SES determinants and stroke can be perceived, regression models used to measure these associations need to account for the fact that regression coefficients might vary similarly in space. Local modelling approaches allow investigators to estimate the relationships between determinants and disease risk with more accuracy, since they estimate regression coefficients for each location in the study area. Amongst these local approaches, Geographically weighted regression (GWR) modelling techniques calculate local regression coefficients, thus allowing the estimates of the associations between disease and explanatory variables to vary spatially and providing a more flexible modelling strategy than global methods (Weisent *et al.*, 2012). In this sense, GWR is a local spatial statistical technique suited to the analysis of spatial nonstationarity, on its turn defined as when the measurement of relationships among variables differs from location to location (Fotheringham, 2002). Since it estimates regression coefficients for each location in the study area, maps generated from these data play an important role in exploring and interpreting spatial nonstationarity (Mennis, 2006). With this in mind, several studies have already implemented this method, aiming to improve our understanding of the determinants of geographic disparities of Health (St-Hilaire *et al.*, 2010; Cheng *et al.*, 2011; Helbich *et al.*, 2012; Weisent *et al.*, 2012). There is some controversy surrounding GWR, with some authors stating that this technique is more suitable for exploratory analysis. However, others defend the flexibility of this model type to investigate spatially varying relationships (Helbich *et al.*, 2012), which is one of the main objectives of this research.

The objectives pursued in this study were: i) to identify the socioeconomic determinants of geographic disparities of stroke risk at the municipality level, in mainland Portugal; ii) to investigate whether regression coefficients for the associations between socioeconomic factors and stroke risk demonstrate spatial variability; iii) to compare the performance of several regression methods in modelling these associations, namely negative binomial and global as well as local Poisson models; and iv) to assess the spatiotemporal variation of the associations between determinants and stroke risk.

Materials and Methods

Study area, socioeconomic variables and data sources

This study was implemented in the Portuguese continental territory, encompassing 275 municipalities in 1991 and 278 in 2001 bearing a total population of approximately 8.1 million in 1991 (INE, 1991a) and 9.9 million in 2001 (INE, 2001a). The municipalities are the Local Administrative Units, level 2 according to the European Commission, and can also be regarded as statistical geographic units (formerly level 5 of the NUTS system; Eurostat, 2014). The municipalities are relatively heterogeneous in terms of population, ranging in 2001 from 1924 inhabitants in Barrancos to 564,657 in Lisbon.

The choice of the municipality as statistical unit was made after counting the number of stroke deaths for the two time periods in the study population. In fact, given the two populations under study (persons of each gender <65 years old), the use of a higher-scale unit, such as the parish, would aggregate much fewer occurrences with a very high number of parishes registering no stroke deaths, even over the course of several years. An excessive number of units with zero occurrences can imply that the probability distribution would not be a Poisson one, but instead a mixture of distributions (Barceló *et al.*, 2008). Since most methods used in this study are based on the Poisson probability distribution or similar, such as Negative Binomial (although being able to accommodate overdispersion up to a certain limit) this could be a problem. On the other hand, the choice of a lower-scale unit, *e.g.* NUTS level 3, would allow higher aggregated values but lower the spatial resolution and therefore the number of statistical units substantially in the analysis, since mainland Portugal is divided in 28 of these units against 278 municipalities. Thus, this *lower granularity* would render some of the statistical models used here (such as GWR) less interesting and lower the statistical power, assuming a common rule of thumb stating that for the central limit theorem to apply, a sample size of 30 would be satisfactory in most practical situations (Daniel, 2009).

Variables representing socioeconomic determinants were included in four categories, the first three being *Occupation*, *Education* and *Income*, as suggested by Cox *et al.* (2006). Labarthe (2011) also stresses the importance of *Environmental* determinants, both in a broad sense, such as distance to nearest hospital but also in a more strict sense, such as exposure to atmospheric pollution due to area of residence. We therefore assumed them as factors that can contribute to increase the degree of social disadvantage of an area following the general idea that *the neighbourhood where you live is a risk factor for stroke* (Balamurugan *et al.*, 2013). Regarding the link between stroke mortality with air pollution, Hu *et al.* (2008) concluded, in a study focused in the State of Florida, USA, that a high risk of stroke mortality was found in areas with low income and high air pollution levels and low levels of exposure to green space.

The choice of variables for the period 1991 and 2001 to include in the four categories was based on the available literature, but also conditioned by the availability of the related data from official sources. Thus, the first category encompassed several occupation-related issues available in census data, including broader issues transversal to any occupation such as the percentage of men or women working more than 45 hours a week in the total working municipal population. This information was obtained from the Portuguese National Statistics Institute (INE) (1991a, 2001a). Working overtime can contribute to the onset of hypertension, thus increasing stroke risk (Hayashi *et al.*, 1996). The lack of occupation, namely the percentage of unemployed men or women in the municipal population, obtained from INE (1991a, 2001a),

was also included. The relationship between unemployment and stroke has been signalled in works such as Franks (1991), who reported high and significant correlations with stroke risk, particularly regarding male unemployment. Spasato *et al.* (2012), in a study focusing on early ischemic stroke, also reports an association between unemployment and a higher risk of adjusted in-hospital mortality.

Occupation in a more strict sense was included, namely the percentage of manual plus agricultural workers in the total working population (no gender separation), obtained from INE (1991a, 2001a). Kunst *et al.* (1998) reported that in all countries, including Portugal, manual workers had higher stroke mortality rates than non-manual ones. Additionally, farmers and farm labourers in Portugal had much higher stroke death rates than the rest of the population. Cesana *et al.* (2001) reported, with respect to myocardial infarction, an increasing rate of case-fatality for decreasing levels of socio-occupational class in a relatively young population (35-64 years old male residents). An education variable, obtained from INE (1991a, 2001a) was included, namely the percentage of men or women in the municipal population having attained basic school. Avendaño *et al.* (2004), report that educational differences in stroke mortality were observed across Europe during the 1990s. Cox *et al.* (2006) point out educational attainment as one of the most commonly used forms of measure SES in the context of stroke studies.

Regarding income, in some countries such as USA, variables such as the median household income or the percentage of population in poverty are currently available at the county (municipal) level through official sources such as the Area Resource File (HRSA, 2015), and have been used in studies focused on stroke such as that by Schieb *et al.* (2013). In Portugal, however, this is not available at the municipal level, not even as part of the census information. Thus, in our view, the closest available surrogates to income at the municipal level were variables related to purchasing power. Following this rationale, we included the index of municipal per capita purchasing power (INE, 1993a, 2000a) and the municipal share (percentage) of purchasing power of each municipality in the data obtained from INE (1993b, 2000b). These are more complex indicators than others used in this investigation, although other complex indicators have been used to measure SES at the area level, such as the deprivation indexes of Carstairs or Townsend (Cox *et al.*, 2006). The municipal average household rental value (available only for 2001), obtained from INE (2001a), was also included, based on the hypothesis that municipalities with higher average household rental values can be regarded as more affluent ones.

Regarding the environmental determinants, the distance in minutes from the municipal town hall to the nearest hospital was included. Travel distances by road were calculated based on the 2011 road network and were thus evaluated only within the 2002-2006 mortality models. In USA, nearly half of all stroke deaths in 1999 occurred before transportation or by the time of arrival at the emergency department (Labarthe, 2011), a fact that reinforces the need for rapid assistance to stroke occurrences. The number of municipal-based health infrastructures (available only for 2001), obtained from INE (2001a), was also included on an experimental basis, in the sense that the greater availability of health assets represents an increased protective factor. In fact, even if this variable includes all health infrastructures in the municipality, in most municipalities these structures were (at the time periods this study refers to) mainly National Health Service health centres and related facilities (extensions) aimed at covering the municipal territory. Thus, these represent a protective factor in the sense of continuously providing services that improve the residents' health, since these services are mainly focused on essential care, prevention or healing (Portuguese Ministry of Health, 2015).



The percentage of persons, living in predominantly rural areas (men or women) within the municipality, was included as an environmental variable, as we hypothesized that residents in these areas are less exposed to urban atmospheric pollution. Calculations were based on 2008 data, namely predominantly rural areas within municipalities, obtained from the Portuguese Institute for Support to Small and Medium-Sized Companies and Innovation (IAPMEI, 2008) and population estimates from INE (2008), and this variable was therefore tested within the 2002-2006 mortality models only. Variables were discriminated by gender whenever this information was available. This is a relevant issue, since it is known that the epidemiology of human stroke is sexually dimorphic (Lang and McCullough, 2008). A recent review, focused on discussing the various pathologic mechanisms of ischemic stroke that may differ according to gender, also concludes that gender does in fact have an important role in ischemic stroke, a relevant fact since 85% of these events are of the ischemic type (Gibson, 2013).

The mortality data covered two five-year periods, the first from January 1, 1992 to December 31, 1996 and the second between January 1, 2002 and December 31, 2006. Amongst persons less than 65 years old, a total of 9601 deaths were recorded in the first 5-year period and 5,852 deaths in the second. Mortality data was used both directly as death counts, and also for the calculation of mortality rates, as detailed under the following subheading. Stroke mortality data for the two periods was kindly provided by INE. Population denominator data at the municipality level for each period, and standard population data for standardized rates, were also obtained from INE (1994, 2001a, 2004). These time periods were chosen primarily in order to understand the evolution of the relationship between socioeconomic factors and stroke between two consecutive and relatively recent decades, and also bearing in mind that the first period registers much more mortality in persons <65 years old (almost the double) than the second. These periods were also devised in order to assess the relationship between socioeconomic data from the 1991 and 2001 census and the aggregated mortality in ensuing years, not spanning more than half a decade from each census year, as suggested in works by Thrift *et al.* (2006) and Schieb *et al.* (2013). A third period would have been useful, but the necessary data were not available, since mortality data were available only up to 2012 following the 2011 census and the situation at the time of data gathering. Finally, the cartographic boundary files of municipalities were obtained from INE (1991b, 2001b).

Calculation of stroke risk, mapping and initial spatial analysis

Stroke death risk was first used to assess the frequency and spatial distribution of this disease among municipalities. This stroke death risk was computed as the number of stroke deaths in men or women <65 years of age (all analysis were performed for men and women separately) reported per municipality during each 5-year study period, divided by the estimates of corresponding municipality population (men or women) <65 years of age in the mid-period of the years 1994 and 2004, respectively. Due to potential confounding of varying age structures between municipalities, the stroke death risk was age-standardized (Siegel, 2012) using the Portuguese mainland population in 2001 as standard population. This procedure ensured that differences in the geographic distribution of stroke risk were not affected by geographic differences in the distribution of age in the population. The calculated age-standardized stroke death risk for each 5-year study period was then presented as the number of stroke deaths per 100,000 population.

Choropleth maps (Pfeiffer *et al.*, 2008) were produced for socioeconomic variables and age-standardized stroke death risk to allow visual

assessment of the comparative spatial distribution of these variables. All maps were produced using ArcMap (ESRI, 2014) based on Jenk's optimization classification method (Jenks, 1967) to determine suitable intervals for the spatial display of variables. Stroke death risk and socioeconomic variables of interest were then assessed for spatial clustering at the municipality level using Global and Local Moran's I (Lloyd, 2010) in Geoda (GeoDa Center, 2014), allowing a first exploratory analysis of the presence of spatial autocorrelation. Pearson's linear correlation coefficients were computed to identify highly correlated pairs of explanatory variables. Several criteria were defined to guide this phase of model-building. The first was to retain in the model, as much as possible, only those variables with a correlation coefficient inferior to 0.6 among them (Weisent *et al.*, 2012). Beyond this threshold, only one of a pair of highly correlated variables (*i.e.*, with $r \geq 0.6$) was retained for further investigation, thus reducing the number of independent variables to include in the subsequent modelling process. Second, each model should retain, correlation permitting, at least one variable representing each category. Based on these two criteria, a subset of socioeconomic variables was filtered from the initial dataset containing all variables, for both male and female models, before regression analysis.

Univariate Poisson regression analysis

Socioeconomic status variables were investigated for potential association with stroke death risk. From the total set of these variables, only those remaining after multicollinearity assessment, as discussed in the previous step, were evaluated at this phase and beyond. The dependent variable specified in the univariate Poisson regression models was the number of stroke deaths reported in each municipality in men or women <65 years of age in each study period. An offset was specified, corresponding to the estimation of the municipality population of men or women <65 years of age in 1994 or 2004, for the 1992-1996 and 2002-2006 periods, respectively.

When dealing with disease count data, particularly for diseases with a reduced number of cases, the Poisson model can be a suitable regression model. Other disease analysis studies dealing with small areas have applied the Poisson model to describe a disease distribution (Cheng *et al.*, 2011). Thus, univariate (simple) ordinary Poisson regression models were then fitted to the data using the generalized linear model (Dobson, 2002) function using the R statistical software (R project, 2014d). The assessment of dispersion detected significant overdispersion (R project, 2014b) in the univariate Poisson regression models, implying that these were inappropriate for the data. Consequently, negative binomial regression models were used for all subsequent non-spatial multivariate regression modelling and final comparisons with the other model approaches tested in this study.

Multivariate negative binomial models

As in univariate regression analysis, the dependent variable specified in the multivariate negative binomial regression models was the count of stroke deaths reported for each municipality in men or women for each period and the offset variable was the 1994 or 2004 estimations of municipality populations of men or women less than 65 years old. Thus, the regression equation for the negative binomial model can be seen on eq. 1.

$$\ln(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad \text{eq. 1}$$

where λ is $E(Y)$ with Y as the dependent variable, β_s parameter estimates (regression coefficients), while the X_s pertain to the vector of K socioeconomic variables under investigation (Weisent *et al.*, 2012).

The appropriate number of variables to be included in negative binomial regression models was previously selected, by applying McHenry's All Possible Variables selection method (McHenry, 1978) in the NCSS software (NCSS, 2014) to the set of non-collinear independent variables available for each study period. Besides this method, empirical experiments were also performed with several different models, to obtain further information on model parsimony.

An appropriate R function (R project, 2014e) was then used to determine all possible models bearing parsimonious combinations of variables as signalled in the previous step, from the pool of previously defined non-collinear independent variables sets. Negative binomial multiple regression model building was then performed in R (R project, 2014c). All possible parsimonious combinations of variables were tested, and the best models were chosen primarily on the basis of the best (lowest) Akaike's Information Criterion (AIC) (R project, 2014a), and secondly on the statistical significance (at least $P < 0.05$) of the variable regression coefficients.

Global and local Poisson geographically weighted regression

Global models estimate one coefficient per explanatory variable, averaged over all locations, while local GWR models estimate a set of coefficients for each location, i.e. for each municipality in this study. Eq. 2 contains the equation for the local GWR model.

$$Y_i(\mu) = \beta_{0i}(\mu) + \beta_{1i}(\mu) X_{1i} + \beta_{2i}(\mu) X_{2i} + \dots + \beta_{ki}(\mu) X_{ki} \quad \text{eq. 2}$$

The $\beta_{ki}(\mu)$ denote regression coefficients for the relationship between an explanatory variable and the dependent variable around a location μ and are therefore unique to that location, while the X_s are the independent variables included in the model (Fotheringham, 2002). The local GWR model allows the investigator to calculate a specific regression coefficient for each location, when the relationship between the dependent and independent variable is evaluated. Therefore, by assuming a causal relationship, the investigator is able

to evaluate how the impact of a specific risk factor on the outcome changes by location. Using the Poisson distribution within the GWR structure is presently the most suitable strategy for analysing areal disease counts available, particularly when low numbers are involved (Weisent *et al.*, 2012). Just as in the negative binomial models, the dependent variable specified in the model was the number of stroke deaths reported in each municipality in men or women <65 years of age in each study period, and the offset was the estimation of the population of men or women <65 years of age resident in the municipality in 1994 or 2004. Both models were fitted using the spatial statistical software GWR, developed solely for the purposes of this type of analysis (GeoDa Center, 2014). For the local Poisson GWR, an adaptive kernel method was selected, to account for differences in the density and diversity of municipalities across mainland Portugal. Dealing with irregularly shaped municipalities is particularly important, as shapes, sizes and density vary widely between metropolitan and rural areas. The kernel varies the size of the analysis window, allowing the same number of municipalities to be incorporated in each local estimate, with a zero weight value applied to all municipalities outside the analysis window in each local regression analysis. For each model, a manual iterative approach based on the AIC was used to identify the number of nearest neighbours (municipalities) to be used as the optimal model bandwidth. To analyse the spatial variability in the association between stroke risk and independent variables, the estimated regression coefficients from the local GWR were displayed as choropleth maps.

Several criteria were defined for producing the GWR coefficient maps, essentially based on the works of Weisent *et al.* (2012) and Muniz (2009). Thus, the initial classification for each map was obtained using the Jenk's classification scheme, in which classes are defined according to apparently natural groupings of data values. These breaks can also be enforced using break points that are known to be relevant to a particular application (Longley *et al.*, 2005). Thus, an important requirement in this study was that each coefficient map should be able to clearly illustrate the spatial distribution of negative *versus* positive coefficients. To achieve this effect, the basic Jenks scheme was modified to include a zero breakpoint in the scale. Also, blue was used

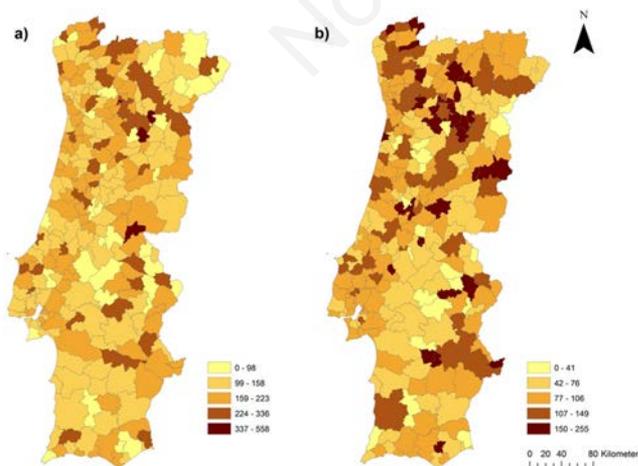


Figure 1. Municipal age standardized stroke death risk per 100,000 population in men (a) and women (b) under 65 years old, 1992-1996.

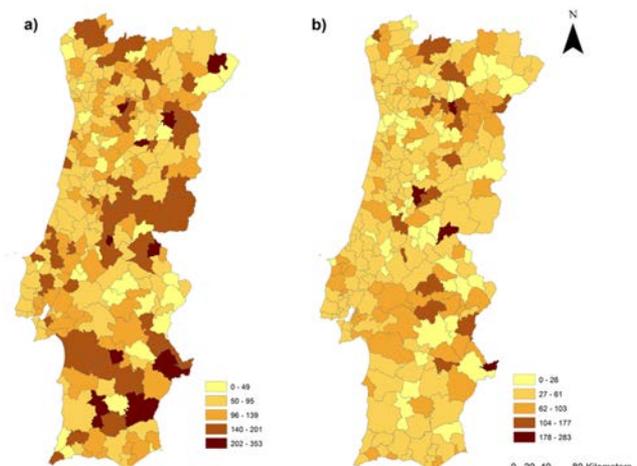


Figure 2. Municipal age standardized stroke death risk per 100,000 population in men (a) and women (b) under 65 years old, 2002-2006.



as base colour for coefficients with a negative signal, in order to illustrate a positive relationship (in the broad sense that the socioeconomic factor is protective in a municipality) and, inversely, positive signal coefficients are painted using red as base colour. Finally, assessment of goodness-of-fit of the negative binomial, global and local Poisson GWR models was done using the AIC.

Results

Spatial distribution of stroke risk and socioeconomic factors

Age adjusted municipal stroke risk estimates presented a great degree of variation, in both men and women, with a general tendency for decrease between the periods of 1992-1996 and 2002-2006, and also for consistently lower values in women. Hence, in 1992-1996, risk values in men ranged from 0 (2 municipalities) to 558 deaths per 100,000 population, with a median of 160 and a standard deviation of 70. In 2002-2006, risk values in men ranged already from 0 (7 municipalities) to 353, with a median of 93 and a standard deviation of 53. Stroke risk in women presented proportionally lower values, ranging in 1992-1996 from 0 (7 municipalities) to 255 deaths per 100,000 population, with a median of 91 and a standard deviation of 43. However, in 2002-2006, the values ranged from 0 (29 municipalities) to 283 deaths per 100,000 population, with a median of 52 and a standard deviation of 41, thus presenting a different picture, with the upper value of risk increasing and less change in variation between the two periods.

Regarding spatial distribution, stroke risk mapping shows evidence of geographic inequalities in risk across the study area, between gender and time period. In 1992-1996, the map of stroke risk in men indicates that municipalities with higher stroke risk seemed to shift away from the coast, with more aggregation of high values in municipalities of north-eastern Portugal, whereas the remaining mainland had mostly lower values, again with some increase towards the East and away from the sea (Figure 1A). As for the spatial distribution of stroke risk in women, the pattern of geographic disparities is slightly distinct, with higher values occurring mostly in municipalities across the northern half of the territory with some high values in the South, again occurring mostly in inland municipalities (Figure 1B).

However, ten years later, the pattern seemed almost reversed between men and women. Thus, the 2002-2006 map of stroke risk in men shows that municipalities with higher stroke risk seemed to shift away from the coast towards the North and Centre, whereas in the South (mainly in municipalities within the upper Alentejo area), they presented a different pattern that resembled a belt from the sea to the south-eastern border of the country (Figure 2A). As for the spatial distribution of stroke risk in women, the pattern of geographic disparities was much smoother, with the highest values in municipalities located almost in a line from centre-north to mid-southeast continental Portugal (Figure 2B). Summary statistics of socioeconomic factors also evidenced important disparities among municipalities, periods and gender (Table 1). For instance, in more than 50% of the municipalities, a great part of the population (70-88%) had manual labour professions in 1991, with the situation changing to 60-76% in 2001.

Based on the statistically significant Moran's I statistic results ($P=0.001$), the spatial distribution of socioeconomic factors showed spatial clustering (Figures 3-6). A concentration of higher values for the percentage of both male and female population working more than 45 hours a week in 1991 can be observed in the inner northernmost municipalities, and also along the north-eastern border (Figure 3A and

B). In 2001, this pattern seemed to soften, but a concentration of higher values remained in the inner northernmost area (Figure 4A and B). Regarding the spatial distribution of the percentage of population having attained basic school in 1991, concentrations of higher values were consistently observed (both for men and women) along the coast and towards the South, with some high values inland, mostly corresponding to non-coastal major cities, such as Évora and Beja (Figure 5A and B). In 2001 this pattern remained, but was even more concentrated towards coastal areas (Figure 6A and B).

When examining socioeconomic variables for multicollinearity, significant ($P<0.001$) and strong positive correlations ($r\geq 0.6$) were indeed found between variables in both periods, and therefore some variables were excluded from the analysis due to this bias. The remaining variables were used in the next modelling step.

Socioeconomic determinants of geographic distribution of stroke risk

After the elimination of multicollinear variables in the previous step, the remaining socioeconomic factors were investigated for univariate associations with stroke risk, and most had highly significant ($P<0.001$) risk associations (Table 2). However, significant overdispersion was detected, making ordinary Poisson models inappropriate for dealing with associations between these datasets. Since negative binomial models fit the data better than the ordinary Poisson models (by accommodating overdispersion better), the analysis beyond this point proceeded with negative binomial-based models.

Based on McHenry's All Possible Variables selection method (McHenry, 1978), and also on empirical experiments with several different models, a number of variables for the most parsimonious models were established. Specifically, the optimal final models for the 2002-2006 period should contain 5 explanatory variables each. As for the 1992-1996 period, the models should contain 2 or 3 variables. At this point, all possible combinations with the established number of variables were determined, departing from the pool of non-collinear variables previously determined for each model. All possible combinations for each model were then run within a negative binomial regression framework and the best male and female models chosen for each time-frame, based on the lowest AICs. Finally, variables whose coefficients did not attained the $P<0.05$ significance criterion were sequentially excluded and the models run again iteratively until all their remaining variables attained at least $P<0.05$ significance. From this point on, the four final models were used to compare the performance of the 3 modelling approaches (negative binomial, global Poisson GWR and local Poisson GWR) tested in this study.

Negative binomial models for the 2002-2006 period evidenced a tendency for stroke risk to be significantly lower in municipalities with higher proportions of men or women having attained basic school, a tendency already present for women in the 1992-1996 period (Table 3). In contrast, the stroke risk appeared to be significantly higher in municipalities with a higher unemployment rate for women in the 2002-2006 period. As for the 1992-1996 period, the stroke risk seemed to be significantly higher in municipalities with a higher percentage of men working more than 45 hours a week (Table 3).

An evaluation of Pearson's standardized residuals from negative binomial models evidenced the presence of positive spatial autocorrelation in the residuals of all models (Moran's I: 0.13, $P=0.002$ in the male 2002-2006 model; 0.07, $P=0.03$ in the female 2002-2006 model; 0.14, $P=0.001$ in the female 1992-1996 model; and 0.09, $P=0.01$ in the male 1992-1996 model). This implies that, even if the negative binomial model performed better than the ordinary Poisson model in this situation, it still contained residual spatial autocorrelation and thus did

Table 1. Summary statistics of socioeconomic factors investigated for potential associations with municipal stroke death risk.

Category	Variable	Mean		SD		Median		Minimum		Maximum	
		1991	2001	1991	2001	1991	2001	1991	2001	1991	2001
Social status	Men working more than 45 hours a week (%)	34.8	24.7	9.6	5.5	34.0	24.5	14.7	11.6	67.9	46.4
	Women working more than 45 hours a week (%)	21.9	11.5	8.1	3.6	20.7	10.9	5.8	5	56.0	29.1
	Manual and agricultural workers (%)	67.9	57.9	10.7	9.7	70.3	60.2	27.6	23.1	87.5	75.5
Employment	Unemployment rate, men (%)	3.8	4.6	1.8	1.6	3.4	4.5	0.9	1.6	14.7	13.6
	Unemployment rate, women (%)	11.2	10.9	7.1	4.7	9.4	9.6	2.4	3.4	45	33.1
Education	Men having attained basic school (%)	14	13	3.6	1.8	13.2	13	6.2	7.8	24.3	17
	Women attained finished basic school (%)	11.7	10	3.2	1.7	11.4	10	5.2	5.3	21.1	14.4
Income	Index of municipal purchasing power per capita (%)	61.5	66.6	33.9	28.8	52.0	58.6	18.9	33.7	291	305.2
	Municipal purchasing power (%)	0.4	0.4	1.4	1.2	0.1	0.1	0.01	0.01	20	16.4
	Average municipal household rental value (EUR)	-	117.5	-	31.8	-	116	-	48.0	-	256
Environment	Municipal health infrastructures (Count)	-	8.6	-	8.4	-	6	-	1	-	93
	Distance to nearest hospital (min)	-	26.3	-	18.4	-	25	-	1	-	89
	Men living in predominantly rural areas (%)	-	42.7	-	60	-	43	-	0	-	85.3
	Women living in predominantly rural areas (%)	-	31.6	-	25	-	29	-	0	-	84.8

SE, standard deviation.

Table 2. Results of univariate (simple) associations between stroke risk and selected variables.

Category	Variable	Estimate ^o			AIC	Estimate			AIC
		Men under 65 years old (1992-1996)				Women under 65 years old (1992-1996)			
Social status	Working more than 45 hours a week (%)	0.0061 (0.0032, 0.0090)	0.0015	0.0001	1792	0.0122 (0.0076, 0.0167)	0.0023	0.0001	1421
Employment	Unemployment rate (%)	0.0277 (0.0121, 0.0432)	0.0079	0.0005	1797	0.0031 (-0.0178, 0.0239)	0.0106	0.7670	1447
Education	Having attained basic school (%)	-0.0071 (-0.0133, -0.0009)	0.0032	0.0245	1804	-0.0269 (-0.0355, -0.0184)	0.0044	0.0001	1409
Income	Municipal purchasing power (%)	0.0164 (0.0116, 0.0212)	0.0025	0.0000	1767	0.0082 (0.0016, 0.0146)	0.0033	0.0134	1441
		Men under 65 years old (2002-2006)			Women under 65 years old (2002-2006)				
Social status	Working more than 45 hours a week (%)	0.0012 (-0.0060, 0.0083)	0.0037	0.3260	1597	0.0373 (0.0205, 0.0539)	0.0085	0.0001	1242
	Manual and agricultural workers (%)	-0.0006 (-0.0030, 0.0019)	0.0012	0.6460	1597	-0.0014 (-0.0047, 0.0019)	0.0017	0.4080	1260
Employment	Unemployment rate (%)	0.0320 (0.0135, 0.0504)	0.0094	0.0007	1586	0.0352 (0.0225, 0.0475)	0.0064	0.0001	1233
Education	Having attained basic school (%)	-0.0783 (-0.0984, -0.0582)	0.0103	0.0001	1540	-0.0768 (-0.1050, -0.0484)	0.0144	0.0001	1233
Income	Municipal purchasing power (%)	0.0051 (-0.0039, 0.0138)	0.0045	0.2570	1596	0.0136 (0.0021, 0.0246)	0.0057	0.0178	1256
	Average municipal household rental value (EUR)	-0.0030 (-0.0041, -0.0019)	0.0006	0.0001	1568	-0.0022 (-0.0037, -0.0007)	0.0008	0.0041	1253
Environment	Distance to nearest hospital (min)	0.0059 (0.0036, 0.0082)	0.0012	0.0001	1573	0.0071 (0.0039, 0.0103)	0.0016	0.0001	1242

SE, standard error; AIC, Akaike's information criterion. ^o95% confidence interval.**Table 3. Comparison of negative binomial, global and local geographically weighted Poisson models.**

Model	Variable	Negative binomial model		Global Poisson GWR model		Local Poisson GWR model		
		Coefficient (P)	AIC	Coefficient (P)	AIC	Min	Max	AIC
2002-2006 model (men)	Intercept	-5.7782 (0.0001)	1470	-5.9075 (0.0001)	520	-7.4228	-3.7536	431
	Men having attained basic school (%)	-0.0853 (0.0001)		-0.0783 (0.0001)		-0.2616	0.0376	
2002-2006 model (women)	Intercept	-7.2131 (0.0001)	1215	-7.2182 (0.0001)	467	-7.2760	-6.5239	354
1992-1996 model (men)	Intercept	-6.7290 (0.0001)	1657	-6.6724 (0.0001)	616	-7.4739	-5.6543	534
	Men working more than 45 hours a week (%)	0.0089 (0.0002)		0.0061 (0.0001)		-0.0327	0.0354	
1992-1996 model (women)	Intercept	-6.6254 (0.0001)	1378	-6.6884 (0.0001)	397	-7.4714	-5.3142	322
	Women having attained basic school (%)	-0.0307 (0.0001)		-0.0269 (0.0001)		-0.1324	0.0290	

GWR, geographically weighted regression; AIC, Akaike's information criterion.

not eliminate the spatial autocorrelation totally in any model. Therefore, the presence of significant spatial autocorrelation in all model residuals requires the use of a more suitable spatial model.

Comparative evaluation of modelling approaches and spatial analysis of local coefficients

Using the AIC goodness-of-fit statistic for comparing and evaluating model performance, the model with the lowest AIC value had the best model fit. This criterion was used to compare the 3 modelling approaches tested upon the 4 final models. In all models, the local Poisson GWR model showed the best fit, followed by the global Poisson GWR model and lastly by the negative binomial model (Table 3). It is relevant to point out that all 4 local Poisson GWR models presented evidence of non-stationarity in the regression coefficients, with the sole exception of the unemployment rate in women for the 2002-2006 period.

This is shown by the comparison between the inter-quartile ranges of the local regression coefficients, which were all larger than twice the standard errors of the regression coefficients of the global Poisson GWR model, except for the unemployment rate in women for the 2002-2006 period (Table 4). This indicates that regression coefficients for the majority of the variables included in the local GWR models were not constant but varied across the municipalities of the Portuguese mainland. This implies that the strength of the associations between stroke risk and each of the final explanatory variables varies depending on the spatial location, except for the unemployment rate in women for the 2002-2006 period. Assuming a causal relationship, this implies that the effects of determinants are not static throughout the study area but are, on the contrary, greatly dependent on geographical location. Figures 7-10 display the spatial patterns of local GWR municipal regression coefficients for the explanatory variables that showed evidence of

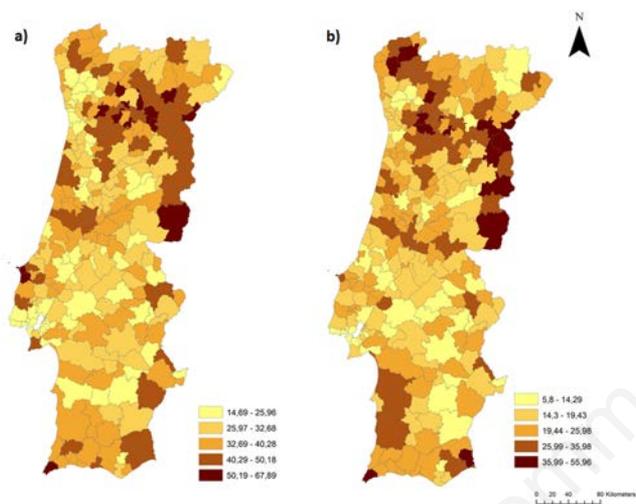


Figure 3. Municipal percentage of men (a) and women (b) working more than 45 hours a week, 1991.

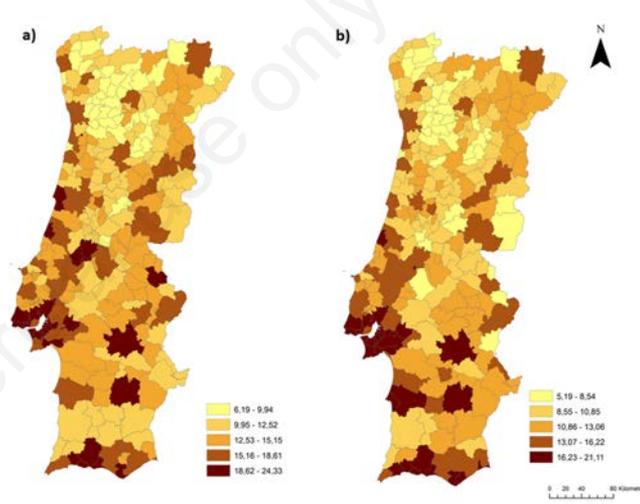


Figure 5. Municipal percentage of men (a) and women (b) having attained basic school, 1991.

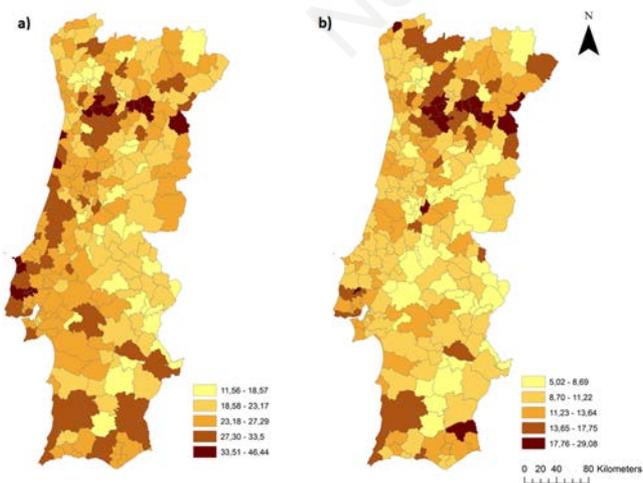


Figure 4. Municipal percentage of men (a) and women (b) working more than 45 hours a week, 2001.

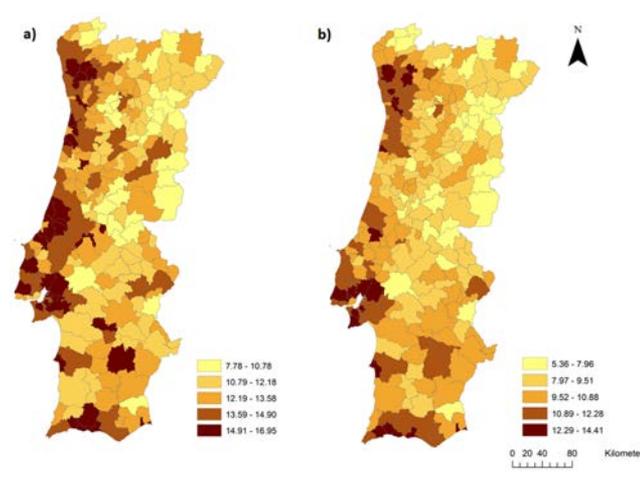


Figure 6. Municipal percentage of men (a) and women (b) having attained basic school, 2001.

spatial non-stationarity in the final models. The statistical evidence for their non-stationarity can be observed in Table 4. In order to analyse the spatial drift of local GWR regression coefficients, and since each model only had one final highly significant and non-stationary variable, some of the variables that were discarded in the final refinement of negative binomial regression models (but had at least a 90% confidence level) were used as ancillary variables to provide spatial and temporal context. Age-standardized stroke mortality was also used for this purpose.

Thus, according to the local Poisson GWR model for men in the 1992-1996 period, stroke risk tended to be higher in municipalities that had relatively higher percentages of persons working more than 45 hours a week (33.4%). Those areas were mainly located in the more remote inland areas of the Northeast along a strip extending inland from the middle of the south-eastern border towards south and in a much smaller area by the sea (Figure 7). These areas also tended to have higher values for age-standardized mortality (153.73 deaths per 100,000 individuals) and lower unemployment rates (4.46%). On the other hand, areas with comparatively lower percentages of persons working more than 45 hours a week (28.54%) had the strongest negative association

with stroke risk. These areas had, on average, lower age-standardized mortalities (150.06 deaths per 100,000 individuals), slightly higher unemployment rates (4.76%). The largest such area stretched along the coast between the north of Lisbon and the mid-Alentejo coast. Assuming a causal relationship, the risk seemed to decrease between some of the most remote areas in inner Portugal towards the sea where the largest cities are located, although some intermediate low-risk areas remain a challenge.

In the male model for the 2002-2006 period, the stroke risk had evolved to significant associations with basic school attendance, those associations being negative in most of the territory with the exception of a strip of land extending along a northwest-southeast axis between central Portugal near the sea and the south-eastern border, and also in an isolated municipality near the mid-eastern border. Interestingly, the areas with the most negative associations were in municipalities located along the middle and northern border with Spain. There was also an inland cluster in the Góis municipality, situated almost in the centre of the country (Figure 8). These municipalities are, on average, more distant from hospitals (44 minutes *vs* 27 in municipalities with positive associations). They also had a higher age-standardized mortality than

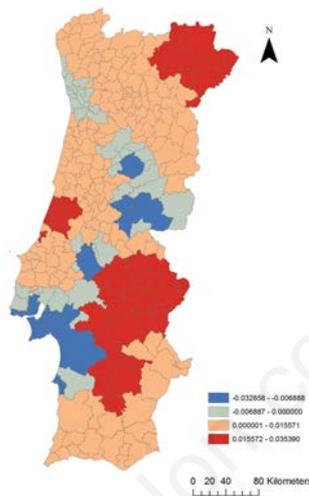


Figure 7. Municipal geographically weighted regression coefficient estimates for the percentage of men working more than 45 hours a week, 1992-1996.

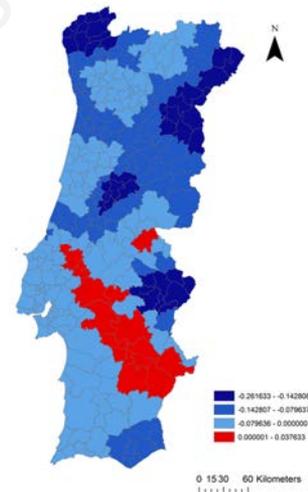


Figure 8. Municipal geographically weighted regression coefficient estimates for the percentage of men having attained basic school, 2002-2006 model.

Table 4. Evaluation of local Poisson geographically weighted regression model coefficients stationarity.

Model	Variable	Global Poisson GWR		Local Poisson GWR	Non-stationarity of regression coefficient IQR>2SE
		SE	2×SE	IQR	
Women under 65 years old (2002-2006)	Women having attained basic school (%)	0.015	0.031	0.048	Yes
	Unemployment rate, women (%)	0.007	0.014	0.003	No
Women under 65 years old (1992-1996)	Women having attained basic school (%)	0.004	0.009	0.058	Yes
Men under 65 years old (2002-2006)	Men having attained basic school (%)	0.01	0.021	0.073	Yes
Men under 65 years old (1992-1996)	Men working more than 45 hours a week (%)	0.002	0.003	0.013	Yes

GWR, geographically weighted regression; SE, standard error; 2SE, 2× standard error; IQR, interquartile range.

those with positive associations (111.93 deaths per 100,000 individuals versus 105.63, respectively).

Regarding the female model for 1992-1996, the associations between women having attained basic education and stroke risk were negative in most of the territory, but switched to positive in an area centred on the metropolitan area of Porto, the second most largest city in Portugal, and in a second much smaller nearby area in the municipality of Santa Marta de Penaguião. The Lisbon municipality and metropolitan area, on the other hand, were among the areas with the strongest negative associations, a clear pattern of divergence between the two main cities (Figure 9). Positive association municipalities had, on average, less women having attained basic schooling (10.54%) than those areas with the strongest negative associations (12.53%) and higher age-standardized mortality. However, areas with stronger negative associations presented a relatively higher average of women working more than 45 hours a week (21.45 *vs* 20.88%) and a higher unemployment rate (9.17 *vs* 7.71) when compared with the positive association areas.

Finally, in the female model for 2002-2006, associations between women having attained basic education and stroke risk were negative in all of Portugal's mainland municipalities, while the most extreme negative values occurred in the Northeast away from the sea with values increasing towards the sea and the South (Figure 10). It is also interesting to signal that ancillary indicators seemed to worsen in the reverse order, since on average municipalities with highest negative associations are more distant to hospitals (33 *vs* 28 min), have a higher percentage of women working more than 45 hours a week (13.47 *vs* 10.96%), a higher unemployment rate (12.85 *vs* 12.48%) and also higher age-standardized mortality with 62.61 *vs* 59.05 deaths per 100,000 individuals.

Discussion

Health geography and spatial analysis in epidemiology are emerging areas of study in the Portuguese context, as noticed in studies focusing in the spatiotemporal clustering of the tuberculosis incidence (Nunes,

2007) and risk factors for this disease (Couceiro *et al.*, 2011) in Portugal. More closely related works have focused on the effects of exposure to cold (due to poorly heated dwellings in winter) on persons with acute coronary syndromes, such as Vasconcelos *et al.* (2011). To our knowledge, and after bibliographic searches in scientific indexation engines, no studies in Portugal have used a local Poisson GWR approach to investigate the geographic variations of the association between stroke risk in the population <65 years old and socioeconomic factors, combined with a multitemporal analysis approach based on two distinct time periods, such as the one tested here. Following this lead, this study also aims to contribute in advancing knowledge on this subject, in the Portuguese context and internationally.

Local GWR models have existed for some time now, and can be a valuable tool to the explore complex relationships between health outcomes and socioeconomic risk factors, with some authors clearly stating the need to investigate the spatial variations in regression coefficients in order to ensure that suitable disease control programmes are used, regardless of the disease and geographical area under scrutiny (Weisent *et al.*, 2012). This is especially true for a multifactorial disease as stroke, whose associations to socioeconomic factors have been mentioned in the literature for some time (Cox *et al.*, 2006; Addo *et al.*, 2012), albeit the precise measure of these associations remains fuzzy to this day.

A marked difference was observed between coefficients of global multivariate models and those of local models in this study. Indeed, global models offer a *one-size-fits-all* approach to the associations between socioeconomic factors and risk. For instance, global approaches signalled that municipalities with high percentages of persons having attained basic school tended to have a consistent lower stroke risk in models for men and women in the 2002-2006 period, and also for women in the 1992-1996 period. In addition, municipalities with high percentages of men working more than 45 hours a week in the 1992-1996 period tended consistently to have a higher stroke risk. Conversely, local Poisson GWR models unfolded a diverse range of regression coefficients for the associations between stroke risk and their potential socioeconomic determinants throughout the municipalities of mainland Portugal. This suggests that global models are less

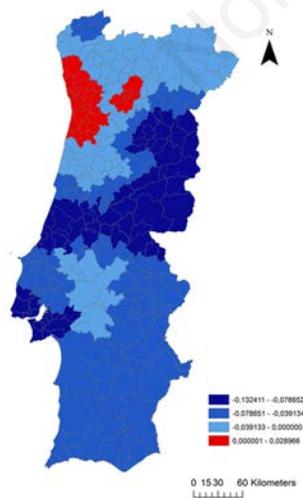


Figure 9. Municipal geographically weighted regression coefficient estimates for the percentage of women having attained basic school, 1992-1996.

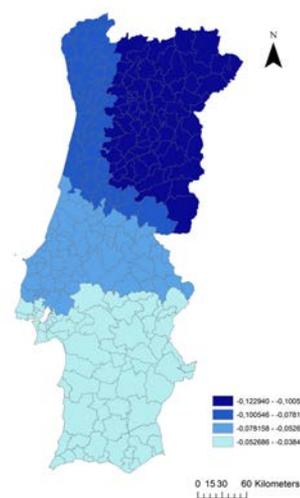


Figure 10. Municipal geographically weighted regression coefficient estimates for the percentage of women having attained basic school, 2002-2006 model.

accurate and reliable, since regression coefficients of determinants ranged from negative to positive throughout the study area. Thus, local models seem to present a better approach to assess the nature of the relationships between determinants and stroke risk, at least within the study area. These results convey the idea that the strength of association between stroke and its determinants changes by location and that this is something that needs to be taken into account when preventive measures are planned for this disease. Indeed, a socioeconomic factor may be more relevant as determinant of stroke in some municipalities (or group of municipalities) and less in others. Local Poisson GWR results successfully revealed the existence of spatial patterns in the spatially varying coefficients in this study, as already observed in other studies following this methodology for other diseases (St-Hilaire *et al.*, 2010; Cheng *et al.*, 2011; Helbich *et al.*, 2012; Weisent *et al.*, 2012). This pattern was also apparent in a very recent communication focusing on geographic disparities with respect to neighbourhood for heart attack and stroke in Tennessee, USA (Odoi and Busingye, 2014). For instance, in models for men and women in the 2002-2006 period, and also for women attaining basic school in the 1992-1996 period, there seemed to be a stronger lowering of the stroke risk in more remote municipalities, and this was especially noticed in the model for women in the 2002-2006 period. We hypothesize that the influence of having attained this level of education in lowering stroke risk is more relevant in these areas, which typically contain less and also older people, while the effect becomes diluted towards coastal areas, where the population is more numerous and where more persons have achieved higher educational levels. Assuming that the ancillary socioeconomic factors used for context contribute to worsen stroke risk, their spatial distribution also seems to strengthen this hypothesis, since the more remote areas are on average more distant from hospitals and have a relatively higher percentage of women working more than 45 hours a week, in addition to higher unemployment rates and age-standardized mortality.

Paradoxically, the use of a methodology including local Poisson GWR analysis also helps to confirm the general tendency of the final variables in lowering or increasing stroke mortality risk, conveyed by global models. Thus, even in the 1992-1996 model for men (the one with more coefficient heterogeneity), coefficients for the percentage of men working more than 45 hours a week were positive in the majority of Portuguese mainland municipalities (77%). This reinforces results of other studies focusing on the influence of this factor on stroke risk in working-age men (Hayashi *et al.*, 1996). Also, in the remaining 3 models, the percentage of persons having attained basic school has a negative coefficient in most municipalities, more precisely 94% (in the 2002-2006 model for men), 87% (in the 1992-1996 model for women) and even 100% in the 2002-2006 model for women. This also reinforces the general idea that higher degrees of education lead to better health outcomes (Avendaño *et al.*, 2004; Henriques *et al.*, 2009; Wu *et al.*, 2013).

Regarding the spatiotemporal variation of models, it is relevant to note that, although there is a shift in male models, not only in the significant factors at play, but also in their patterns of spatial distribution, some factors remain consistently significant. This is the case for both female models. Thus, assuming a causal relationship, the percentage of women having attained basic school remained a protective factor from one decade to another in most Portuguese mainland municipalities. Furthermore, some southern municipalities bear strong positive coefficients in the male models in both study periods, suggesting that those municipalities should be further investigated.

Some patterns remain, however, more difficult to explain, such as the isolated cluster of municipalities centred in the area of Porto in the 1992-1996 female model and the northwest-southeast group of munic-

ipalities in the male 2002-2006 model, all with positive coefficients for the association between stroke risk and persons having attained basic school. We hypothesise that there may be other factors at play in these areas, beyond those analysed in this study, that can eventually be strong enough to confound or obfuscate the general tendency for stroke risk to become lower as the percentage of population having attained basic school increases.

Several courses of action remain open to exploration in this study's subject, some derived from limitations within this study, some more closely related to the Portuguese context. One issue pertains to the very nature of the GWR technique, which brings advantages but also risks. On one hand, local regression coefficients contain extensive information on the non-stationary processes under study, on the other its calibration is a complex process, in which, among others, the choice of bandwidth is an important and sensible issue. Effectively, smaller bandwidths fit the data more effectively, but can be less reliable due to the lack of degree of freedom in the local model fitting, while large bandwidths may lead to meaningful spatial variation in the coefficients being neglected when the true distribution of the coefficient is spatially varying, making bandwidth selection a trade-off process (Kemp, 2008). Thus, despite being, as Fotheringham (2002) pointed out, a tool for the analysis of spatially varying relationships, GWR is mostly an exploratory tool (Bivand, 2014); very much fit for hypothesis generation. In this sense, we believe that GWR was quite successful in unveiling the non-stationarity present in the spatial relationships investigated here, and, as expected, the spatial drift of its coefficients raised several hypotheses; then again, further testing (Matthews and Yang, 2012), for instance with Bayesian hierarchical models as used by Odoi and Busingye (2014), could be applied to better confirm and/or explain the patterns revealed by the analysis of the spatial coefficient drift.

Another issue is the lack of a systematic and direct measuring of poverty and income in Portugal (including their spatial distribution), which leads investigators to use surrogate variables, as was done here. We hypothesize that the use of more explicit variables for poverty and income would further strengthen these models, and that those variables could have a greater potential for significance. Some variables could gain with better temporal alignment with the reference periods for mortality. For instance, distance to nearest hospital in the 2002-2006 models was calculated over the 2011 road network, since it was not possible to obtain a more up-to-date road dataset, this was the only option available. Another limitation relates to the evolution of stroke mortality itself. Suffice to say that, since the absolute number of deaths due to stroke have generally diminished between these two periods, so have the numbers per municipality (in general), which brings a need for future model calibration in order to deal with the increasing effect of small numbers or even zero cases per territorial unit, particularly in the population <65 years old, the focus of this study.

Conclusions

The relationship between stroke and socioeconomic factors is very complex, and definitely not an easy one to assess. However, this study provides several contributions to a better understanding of the nature of this relationship arising from its focus in the critical segment of population <65 years of age, which is seldom analysed separately. By splitting the analysis into different dimensions, namely the geographical and temporal dimensions as well as segmentation by gender, this study provides a more diversified picture. Given the importance of the latter, the option of making separate analyses by gender has also provided



interesting insights. Although various official epidemiological and statistical texts on stroke do provide a number of separate statistics for men and women, we feel that more attention could be given to this matter in terms of official policy.

By going beyond the typical frame used in studies based on global regression models, we not only aim to contribute to a better understanding of the true nature of the relationship between stroke and socioeconomic determinants, but also to improve targeted public health initiatives, particularly those with a preventive focus. Our study is in line with promoting the evaluation of technologies in this domain, and since this work helps to improve the epidemiological and statistical knowledge on the determinants of cardiovascular pathologies, we feel that this study has contributed in an area relevant to the objectives of the Portuguese National Program for Cerebrovascular Diseases. Concerning the need to improve the organization and rational supply of diagnostic and therapeutic care, one issue to be considered is the territorial distribution of health assets and initiatives, in both prevention and mitigation of stroke, and our spatial analysis also provides a contribution to this point.

It will be most interesting not only to further improve this analysis by implementing some of the improvements previously suggested in the discussion, but also to extend this analysis to the coming timeframe of 2012-2016. Since this study analyses two timeframes that precede the present socioeconomic crisis, initiated in 2008, we believe that a spatiotemporal portrait including a period most affected by this crisis has the potential to further enhance the benefits of this analysis.

References

- Addo J, Ayerbe L, Mohan KM, Crichton S, Sheldenkar A, Chen R, Wolfe CDA, McKeivitt C, 2012. Socioeconomic status and stroke: an updated review. *Stroke* 43:1186-91.
- Avendaño M, Kunst AE, Huisman M, van Lenthe F, Bopp M, Borrell C, Valkonen T, Regidor E, Costa G, Donkin A, Borgan J-K, Deboosere P, Gadeyne S, Spadea T, Andersen O, Mackenbach JP, 2004. Educational level and stroke mortality: a comparison of 10 European populations during the 1990s. *Stroke* 35:432-7.
- Balamurugan A, Delongchamp R, Bates JH, Mehta JL, 2013. The neighbourhood where you live is a risk factor for stroke. *Circ Cardiovasc Qual Outcomes* 6:668-73.
- Barceló AM, Saez M, Cano-Serral G, Martínez-Beneito MÁ, Martínez J, Borrell C, Ocaña-Riola R, Montoya I, Calvo M, López-Abente G, Rodríguez-Sanz M, Toro S, Alcalá JT, Saurina C, Sánchez-Villegas P, Figueiras A, 2008. Métodos para la suavización de indicadores de mortalidad: aplicación al análisis de desigualdades en mortalidad en ciudades del Estado español (Proyecto MEDEA). *Gac Sanit* 22:596-608.
- Bivand R, 2014. Geographically weighted regression. Available from: <http://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf>
- Brainin M, Heiss W-D, 2014. *Textbook of stroke medicine*. Cambridge University Press, Cambridge, UK.
- Busingye D, Pedigo A, Odoi A, 2011. Temporal changes in geographic disparities in access to emergency heart attack and stroke care: are we any better today? *Spat Spatiotemporal Epidemiol* 2:247-63.
- Cardoso T, Fonseca T, Costa M, 2003. Acidente vascular cerebral no adulto jovem. *Acta Medica Port* 16:239-44.
- Cesana G, Ferrario M, Gigante S, Sega R, Toso C, Achilli F, 2001. Socio-occupational differences in acute myocardial infarction case-fatality and coronary care in a northern Italian population. *Int J Epidemiol* 30:53-8.
- Cheng EM, Atkinson PM, Shahani AK, 2011. Elucidating the spatially varying relation between cervical cancer and socio-economic conditions in England. *Int J Health Geogr* 2011:10-51.
- Couceiro L, Santana P, Nunes C, 2011. Pulmonary tuberculosis and risk factors in Portugal: a spatial analysis. *Int J Tuberc Lung D* 15:1445-54.
- Cox AM, McKeivitt C, Rudd AG, Wolfe CDA, 2006. Socioeconomic status and stroke. *Lancet Neurol* 5:181-8.
- Daniel WW, 2009. *Biostatistics: a foundation for analysis in the health sciences*. 9th ed. Wiley & Sons, Hoboken, NJ, USA.
- DGS, 2014a. Direção-Geral da Saúde. Programa Nacional para as Doenças Cérebro-Cardiovasculares. Available from: <http://www.dgs.pt/paginas-de-sistema/saude-de-a-a-z/programa-nacional-para-as-doencas-cerebro-cardiovasculares/paginas-de-sistema/saude-de-a-a-z/programa-nacional-para-as-doencas-cerebro-cardiovasculares/objetivos.aspx>
- DGS, 2014b. Portugal. Doenças Cérebro-Cardiovasculares em números. 2014. Available from: <http://www.dgs.pt/estatisticas-de-saude/estatisticas-de-saude/publicacoes/portugal-doencas-cerebro-cardiovasculares-em-numeros-2014.aspx>
- Dobson AJ, 2002. *An introduction to generalized linear models*. 2nd ed. Chapman & Hall, Boca Raton, FL, USA.
- Engström G, Jerntorp I, Pessah-Rasmussen H, Hedblad B, Berglund G, Janzon L, 2001. Geographic distribution of stroke incidence within an urban population relation to socioeconomic circumstances and prevalence of cardiovascular risk factors. *Stroke* 32:1098-103.
- ESRI, 2014. ArcGIS for desktop. Available from: <http://www.esri.com/>
- Eurostat, 2014. Local administrative units. Available from: http://ec.europa.eu/eurostat/c/porta/layout?P_l_id=345247&P_v_l_s_g_id=0
- Fotheringham AS, 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, Chichester and Hoboken, NJ, UK and USA.
- Franks PJ, Adamson C, Bulpitt PF, Bulpitt CJ, 1991. Stroke death and unemployment in London. *J Epidemiol Commun H* 45:16-8.
- GeoDa Center, 2014. GeoDa Center for geospatial analysis and computation. Available from: <https://geodacenter.asu.edu/>
- Gibson CL, 2013. Cerebral ischemic stroke: is gender important? *J Cerebr Blood F Met* 33:1355-61.
- Hayashi T, Kobayashi Y, Yamaoka K, Yano E, 1996. Effect of overtime work on 24-hour ambulatory blood pressure. *J Occup Environ Med* 38:1007-11.
- Helbich M, Leitner M, Kapusta ND, 2012. Geospatial examination of lithium in drinking water and suicide mortality. *Int J Health Geogr* 2012:11-19.
- Henriques FC, Rodrigues TF, Martins MFO, 2009. Ageing, education and health in Portugal: prospective from the 19th to the 21st century. *Hygiea Internationalis* 8:81-96.
- HRSA, 2015. AHRF: area health resources files. Available from: <http://ahrh.hrsa.gov/>
- Hu Z, Liebens J, Rao KR, 2008. Linking stroke mortality with air pollution, income, and greenness in northwest Florida: an ecological geographical study. *Int J Health Geogr* 2008:7-20.
- IAPMEI, 2008. Listagem de Freguesias Rurais. Available from: http://www.iapmei.pt/resources/download/Freguesias_Rurais.xls
- ICD-10, 2015. International classification of diseases, version 2015. Available from: <http://apps.who.int/classifications/icd10/browse/2015/en>
- INE, 1991a. Censos. Resultados definitivos Portugal, 1991. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=65588377&PUBLICACOESstema=00&PUBLICACOESmodo=2

- INE, 1991b. Censos 1991. Importação dos principais dados alfanuméricos e geográficos (BGRI). Available from: <http://mapas.ine.pt/download/index1991.phtml>
- INE, 1993a. Poder de compra per capita por localização geográfica. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0001720&contexto=bd&selTab=tab2
- INE, 1993b. Proporção de poder de compra (% - no total do País) por localização geográfica. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0001355&contexto=bd&selTab=tab2
- INE, 1994. População residente (N.º) por local de residência. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003182&contexto=bd&selTab=tab2
- INE, 2000a. Poder de compra per capita por localização geográfica. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0001720&contexto=bd&selTab=tab2
- INE, 2000b. Proporção de poder de compra (% - no total do País) por localização geográfica. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0001355&contexto=bd&selTab=tab2
- INE, 2001a. Censos. Resultados definitivos Portugal, 2001. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=133411&PUBLICACOESstema=00&PUBLICACOESmodo=2
- INE, 2001b. Censos 2001. Importação dos principais dados alfanuméricos e geográficos (BGRI). Available from: <http://mapas.ine.pt/download/index2001.phtml>
- INE, 2004. População residente (N.º) por local de residência. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003182&contexto=bd&selTab=tab2
- INE, 2008. População residente (N.º) por local de residência. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003182&contexto=bd&selTab=tab2
- INE, 2012. Taxa de mortalidade padronizada por doenças cerebrovasculares por 100 000 habitantes (N.º) por local de residência. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003785&contexto=bd&selTab=tab2
- Jenks GF, 1967. The data model concept in statistical mapping. *International Yearbook of Cartography* 7:186-90.
- Kemp KK, 2008. *Encyclopedia of geography information science*. SAGE Publications, Los Angeles, CA, USA.
- Krishnamurthi RV, Feigin VL, Forouzanfar MH, Mensah GA, Connor M, Bennett DA, Moran AE, Sacco RL, Anderson LM, Truelsen T, O'Donnell M, Venketasubramanian N, Barker-Collo S, Lawes CMM, Wang W, Shinohara Y, Witt E, Ezzati M, Naghavi M, Murray C, 2013. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet Glob Health* 1:e259-81.
- Kunst AE, Rios M del, Groenhouf F, Mackenbach JP, 1998. Socioeconomic inequalities in stroke mortality among middle-aged men an international overview. *Stroke* 29:2285-91.
- Labarthe DR, 2011. *Epidemiology and prevention of cardiovascular diseases: a global challenge*, 2nd ed. Jones & Bartlett Learning, Burlington, MA, USA.
- Lambert C, Vinson S, Shofer F, Brice J, 2013. The relationship between knowledge and risk for heart attack and stroke. *J Stroke Cerebrovasc* 22:996-1001.
- Lang JT, McCullough LD, 2008. Pathways to ischemic neuronal cell death: are sex differences relevant? *J Transl Med* 6:33.
- Li C, Hedblad B, Rosvall M, Buchwald F, Khan FA, Engström G, 2008. Stroke incidence, recurrence, and case-fatality in relation to socioeconomic position: a population-based study of middle-aged Swedish men and women. *Stroke* 39:2191-6.
- Lloyd C, 2010. *Local models for spatial analysis*. CRC Press, Oxford, UK.
- Longley PA, Goodchild, MF, Maguire, DJ, Rhind, DW, 2005. *Geographic information systems and science*. 2nd ed. John Wiley & Sons, Chichester, UK.
- Marmot M, Wilkinson R, 2005. *Social determinants of health*. 2nd ed. Oxford University Press, Oxford, UK.
- Matthews SA, Yang T-C, 2012. Mapping the results of local statistics: using geographically weighted regression. *Demogr Res* 26:151-66.
- McHenry CE, 1978. Computation of a best subset in multivariate analysis. *Appl Stat* 27:291-6.
- Mendes R, 1988. The impact of occupation on workers' health: I. Morbidity. *Rev Saude Publ* 22:311-26.
- Mennis J, 2006. Mapping the results of geographically weighted regression. *Catogr J* 43:171-9.
- Muniz JO, 2009. Spatial dependence and heterogeneity in ten years of fertility decline in Brazil. Available from: paa2006.princeton.edu/papers/60162
- NCSS, 2014. *Statistical software*. Available from: <http://www.ncss.com/>
- Nunes C, 2007. Tuberculosis incidence in Portugal: spatiotemporal clustering. *Int J Health Geogr* 6:30.
- Odoi A, Busingye D, 2014. Neighborhood geographic disparities in heart attack and stroke mortality: comparison of global and local modeling approaches. *Spat Spatiotemporal Epidemiol* 11:109-23.
- Pfeiffer DU, Robinson TP, Stevenson M, Stevens KB, Rogers DJ, Clements ACA, 2008. *Spatial analysis in epidemiology*. Oxford University Press, Oxford, UK.
- Portuguese Ministry of Health, 2015. Centro de saúde. Available from: http://www.hospitalfeira.min-saude.pt/index.php?option=com_content&view=article&id=106:2-centro-de-saude&catid=39:guia-online-do-utente&Itemid=201
- R project, 2014a. Akaike's an information criterion. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/AIC.html>
- R project, 2014b. Dispersion test. Available from: <http://artax.karlin.mff.cuni.cz/r-help/library/AER/html/dispersiontest.html>
- R project, 2014c. Fit a negative binomial generalized linear model. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/glm.nb.html>
- R project, 2014d. Fitting generalized linear models. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/glm.html>
- R project, 2014e. Generate all combinations of n elements, taken m at a time. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/utils/html/combn.html>
- Sacco RL, Benjamin EJ, Broderick JP, Dyken M, Easton JD, Feinberg WM, Goldstein LB, Gorelick PB, Howard G, Kittner SJ, Manolio TA, Whisnant JP, Wolf PA, 1997. Risk factors. *Stroke* 28:1507-17.
- Schieb LJ, Mobley LR, George M, Casper M, 2013. Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and healthcare characteristics. *Stroke* 44:146-52.
- Siegel JS, 2012. *The demography and epidemiology of human health and aging*. Springer, New York, NY, USA.
- Sposato LA, Ioli P, Povedano G, Esnaola y Rojas MM, Saposnik G, 2012. Unemployment: a social risk factor associated with early ischemic stroke mortality? Results from the Argentinean National Stroke Registry (ReNACer). *J Stroke Cerebrovasc* 21:679-83.
- St-Hilaire S, Mannel S, Commendador A, Mandal R, Derryberry D, 2010. Correlations between meteorological parameters and prostate can-



- cer. *Int J Health Geogr* 2010:9-19.
- Thrift AG, Dewey HM, Sturm JW, Paul SL, Gilligan AK, Srikanth VK, Macdonell RAL, McNeil JJ, Macleod MR, Donnan GA, 2006. Greater incidence of both fatal and nonfatal strokes in disadvantaged areas: the Northeast Melbourne stroke incidence study. *Stroke* 37:877-82.
- Vasconcelos J, Freire E, Morais J, Machado JR, Santana P, 2011. The health impacts of poor housing conditions and thermal discomfort. *Procedia Environ Sci* 2010:158-64.
- Weisent J, Rohrbach B, Dunn JR, Odoi A, 2012. Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modeling approaches. *Int J Health Geogr* 2012:11-45.
- WHO, 2014. STEPwise approach to stroke surveillance. Available from: <http://www.who.int/chp/steps/stroke/en/>
- Wu SH, Woo J, Zhang X-H, 2013. Worldwide socioeconomic status and stroke mortality: an ecological study. *Int J Equity Health* 2013:12-42.

Non commercial use only