Small-scale geographical variation in multiple sclerosis: A case-control study using Danish register data 1971–2013

Kristine Bährmann⁎, Nete Munk Nielsen, Melinda Magyar, Nils Koch-Henriksen, Rikke Baastrup Nordsborg, Annette Kjær Ersbøll

⁎ Corresponding author.
E-mail address: akri@niph.dk (K. Bährmann).

1. Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system. Its etiology is largely unknown, but commonly assumed to be a complex interaction between genes and environment. Although some controversy does exist, migration studies have suggested that environmental exposures during the first two decades of life could be of particular relevance for MS (Gale and Martyn, 1995; Dean and Elian, 1997). Furthermore, there seems to be a month-of-birth effect and a maternal effect in MS, which could point towards an effect of in utero or perinatal environmental exposures (Torkildsen et al., 2012; Ebers, 2008). By evaluating the spatial variation in the occurrence of MS according to residence at birth, age 15, and onset of disease, potential environmental exposures and a possible timing of exposure might be identified.

The occurrence of MS seems to vary between continents (Koch-Henriksen and Sorensen, 2010), and a number of studies have reported spatial differences in both prevalence and incidence of MS within countries and regions (Fromont et al., 2010; Pugliatti et al., 2002; Torabi et al., 2014). In most previous studies, information on residence of MS patients was limited to residence within administrative region, and incidence studies were primarily based on residence at onset of disease or diagnosis. However, spatial differences according to residential municipality at age 5 to 15 years have been reported (Pugliatti et al., 2002).

In Denmark, a previous study found spatial differences in MS incidence rates between counties in 1948–64 (Koch-Henriksen and Hyllested, 1988). A spatial cluster of 8 MS patients within a small Danish community has been reported (Haahr et al., 1997), but otherwise, to our knowledge, a spatial analysis of more recent Danish data has not been carried out. However, nationwide Danish registers now offer a unique possibility for utilizing accurate information on complete residential history on all Danish citizens. That enables spatial analyses based on the exact residence of MS patients at any time in life, including childhood.

The objective of this study was to evaluate the spatial variation in the risk of MS in Denmark based on residence at birth, age 15, and onset of disease, respectively. The objective was addressed in a population-
based case-control study.

2. Materials and methods

Since April 1968, all Danish citizens have been assigned a unique personal identification number which is recorded in the Danish Civil Registration System (DCRS) (Pedersen, 2011). This identification number enables individual-level linkage of data from different registers. DCRS includes information on all current and historical residential addresses of persons living in Denmark.

The Danish Multiple Sclerosis Registry (DMSR) (Koch-Henriksen et al., 2001; Brønnum-Hansen et al., 2011) was established in 1956 and collects information on Danish MS cases. The register has shown a completeness of 91% and a validity of 94% (Koch-Henriksen et al., 2001). Information in DMSR includes year of diagnosis and year of clinical onset of disease.

The study area consisted of the Scandinavian country Denmark (43 000 km²). The study population included all 13 968 MS cases (hereby referred to as cases) with clinical onset of disease 1971 (from 1971 the registration of residential addresses in DCRS is complete). DCRS includes information on all current and historical residential addresses of persons living in Denmark.

2. Materials and methods

Since April 1968, all Danish citizens have been assigned a unique personal identification number which is recorded in the Danish Civil Registration System (DCRS) (Pedersen, 2011). This identification number enables individual-level linkage of data from different registers. DCRS includes information on all current and historical residential addresses of persons living in Denmark.

The Danish Multiple Sclerosis Registry (DMSR) (Koch-Henriksen et al., 2001; Brønnum-Hansen et al., 2011) was established in 1956 and collects information on Danish MS cases. The register has shown a completeness of 91% and a validity of 94% (Koch-Henriksen et al., 2001). Information in DMSR includes year of diagnosis and year of clinical onset of disease.

The study area consisted of the Scandinavian country Denmark (43 000 km²). The study population included all 13 968 MS cases (hereby referred to as cases) with clinical onset of disease 1971–2013 (from 1971 the registration of residential addresses in DCRS is complete). Cases should fulfil the contemporary diagnostic criteria for MS defined by Allison and Millar (~1993), Poser (1994–2004), and McDonald (2005-) (Brønnum-Hansen et al., 2011; Przybek et al., 2015). Immigrants and descendants (N = 704), cases born outside Denmark (N = 124), and cases with no Danish address at onset of disease (N = 147) were excluded. Thereby, the study population was reduced to 12 993 cases.

Initially, 100 age- and sex-matched controls per case were randomly selected by incidence density sampling (Rothman, 2002) from DCRS. Subsequently, these controls were subject to the same exclusion criteria as the cases and should in addition be alive and MS free (no onset of disease) at the onset date of the corresponding case.

Information on the residential history of cases and controls was obtained from DCRS, and addresses were geocoded, i.e. translated into UTM (Universal Transverse Mercator, EUREF89, zone 32N) coordinates, using a nationwide address database established and maintained by the Danish Agency for Data Supply and Efficiency in cooperation with the Danish municipalities. In total, 96% of the addresses were geocoded.

For the main analyses, three data sets including information on residence at birth, age 15 (15th birthday), and onset of MS (defined as July 1st in the year of clinical onset), respectively, were generated. All cases and controls with a geocoded address at the relevant time point were included in each data set. Hence, three data sets with different number of cases and different sets of controls were generated. This was done in order to include as many cases as possible at each time point, and to ensure that all controls had a geocoded address at the relevant time. For the analyses, the number of controls per case was reduced to 5 (by random sampling) in order to reduce computing time. Cases with onset of disease before age 15 were not included in the age 15 data set. Furthermore, cases with residence at three small islands (Læsø, Anholt, Samsoe) at birth (N = 20), age 15, (N = 17) or onset of disease (N = 13) were excluded from the relevant data sets.

The main analyses were stratified according to gender and birth cohort (in decades). The latter in order to account for potential changes over time in environmental exposure. These analyses were only reported for age 15.

Sub-analyses of cases with a substantial change in residence were run in attempt to distinguish the potential effect of environmental exposure at birth, age 15, and onset of disease. A substantial change was defined as ≥25 km between addresses, with sensitivity analyses using 20 km as limit. This criterion was met by only 770 cases when comparing addresses at birth and age 15. Therefore, only age 15 and onset of disease were considered in these sub-analyses. New sets of controls fulfilling the same distance criterion were sampled for these analyses.

The spatial variation in risk of developing MS was estimated by kernel regression using a Generalised Additive Model (GAM) approach to take the matched design into account (Jarner et al., 2002). At location x, the risk p(x) of developing MS was modelled using a logit link

\[ \text{logit}(p(x)) = g(x) \]

where g is an unknown smooth function. Hence, mapping g will provide an illustration of the spatial variation in the (log-) odds of developing MS. Only relative differences should be inferred from g, since the actual size of the risk of developing MS cannot be estimated from the available case-control data.

A Gaussian kernel with correction for edge effects was used for kernel regression. The bandwidth of the kernel, the main determinant of the smoothness of the estimated map (i.e. the g function), was chosen to be 10 km. This was based on visual inspection of the estimate to achieve a reasonable trade-off between details and smoothness, since various methods for determining the bandwidth did not provide a finite value. Several bandwidths were tried to make sure the overall results were not affected by the choice of bandwidth.

The hypothesis of no spatial variation (constant g) was tested using a Monte Carlo approach, and p-values were based on 500 random permutations of cases and controls. Permutations were done within case/control sets to account for matching. High-/low-risk areas were identified as areas with estimated odds above/below the point wise 97.5-2.5-percentile of the distribution based on the random permutations of cases and controls.

Analyses were run in R (Core Team, 2016) using the spatstat package version 1.46–1 (Baddeley et al., 2015).

This study was based entirely on register data. According to Danish law, register-based studies can be carried out without consent from the data subjects. The study was approved by the Danish Data Protection Agency (approval number: 2015-57-0008).

3. Results

In total, 12 993 cases with onset of MS in 1971–2013 were included in the study. The median age at onset was 35 years (25% quantile Q1:27, 75% quantile Q3:42) for women and 36 years (Q1:28, Q3:43) for men. The number of cases with a geocoded residential address at birth, age 15, and onset of disease, respectively, is given in Table 1. At birth, 78% had a geocoded address. At age 15, the proportion was 85%, and at onset of disease, 99%. In total, 9912 (76%) cases had a geocoded address at all three time points.

In general, the analyses included twice as many women as men. Most cases were born in the 1950’s and 1960’s. Among cases with a geocoded address at both birth and age 15, 60% did not change their residence between birth and age 15. At age 15, the average time at the current residence was 11.2 years (median = 15, 25%-quantile of the distribution based on the random permutations of cases and controls.

Analyses were run in R (Core Team, 2016) using the spatstat package version 1.46–1 (Baddeley et al., 2015).

This study was based entirely on register data. According to Danish law, register-based studies can be carried out without consent from the data subjects. The study was approved by the Danish Data Protection Agency (approval number: 2015-57-0008).

3. Results

In total, 12 993 cases with onset of MS in 1971–2013 were included in the study. The median age at onset was 35 years (25% quantile Q1:27, 75% quantile Q3:42) for women and 36 years (Q1:28, Q3:43) for men. The number of cases with a geocoded residential address at birth, age 15, and onset of disease, respectively, is given in Table 1. At birth, 78% had a geocoded address. At age 15, the proportion was 85%, and at onset of disease, 99%. In total, 9912 (76%) cases had a geocoded address at all three time points. In general, the analyses included twice as many women as men. Most cases were born in the 1950’s and 1960’s.

Among cases with a geocoded address at both birth and age 15, 60% did not change their residence between birth and age 15. At age 15, the average time at the current residence was 11.2 years (median = 15 years), and only 10% resided >15.8 km away from their residence at birth. Among cases with a geocoded address at both age 15 and onset of disease, 50% resided <6.4 km away from their residence at age 15, whereas 25% resided >28.9 km away. At onset of disease, the average time at the current residence was 9.5 years (Q1 = 1.7, median = 5, median = 4 years, and 25%-quantile of the distribution based on the random permutations of cases and controls.

Table 1

<table>
<thead>
<tr>
<th>Year of birth</th>
<th>Birth</th>
<th>Age 15</th>
<th>Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1940</td>
<td>927 (9.2)</td>
<td>927 (8.4)</td>
<td>1202 (9.4)</td>
</tr>
<tr>
<td>1940–49</td>
<td>1872 (18.5)</td>
<td>1872 (16.9)</td>
<td>2349 (18.3)</td>
</tr>
<tr>
<td>1950–59</td>
<td>2225 (22.0)</td>
<td>2130 (21.1)</td>
<td>3135 (24.4)</td>
</tr>
<tr>
<td>1960–69</td>
<td>2347 (23.2)</td>
<td>3036 (27.5)</td>
<td>3134 (24.4)</td>
</tr>
<tr>
<td>1970–79</td>
<td>1879 (18.5)</td>
<td>2065 (18.7)</td>
<td>2119 (16.5)</td>
</tr>
<tr>
<td>≥1980</td>
<td>880 (8.7)</td>
<td>826 (7.5)</td>
<td>887 (6.9)</td>
</tr>
</tbody>
</table>
below the 97.5-/2.5-percentile. Spatial variation was found in all maps, but especially based on residence at age 15 and onset of disease (residence at birth: p = 0.04, age 15: p = 0.002, onset: p = 0.002). The map scale had a range of about 1, i.e. areas of the highest risk had about two times higher odds of MS compared to areas of the lowest risk. The maps based on residence at birth and age 15 appeared quite similar. Low-risk areas were found on Lolland, in central and eastern Jutland, Northern Jutland and the capital Copenhagen. High-risk areas were found in Eastern Jutland. Some differences were also observed: high-risk areas on Funen and around Mors were only found in the birth map, whereas high-risk areas on Zealand and around Esbjerg were only found in the age 15 map. Based on residence at onset of disease, a high-risk area was again found in Eastern Jutland, but also in south-eastern Jutland, and furthermore around Mors (as in the birth map) and Esbjerg (as in the age 15 map). Low-risk areas were found in Northern Jutland, around Copenhagen, and on Lolland extending to southern Zealand.

Gender-stratified estimates based on residence at age 15 indicated spatial variation in the odds of MS among both women (p = 0.006) and men (p = 0.008) (Fig. 2, top row). Some differences between the two estimated maps were seen, e.g. in the western part of Southern Jutland where opposite results between women (low-risk) and men (high-risk) were found. High-risk areas around Aarhus and Esbjerg, and low-risk areas on Lolland and around Copenhagen were found in both maps. Stratifying according to birth cohort (restricted to 1940–1979 as very few cases were born before/after, Table 1), the spatial variation in the odds of MS based on residence at age 15 was only found in the 1950’s (1940’s: p = 0.67, 1950’s: p = 0.04, 1960’s: p = 0.17, 1970’s: p = 0.64) (Fig. 2, middle/bottom rows). Therefore, indication of high-/low-risk areas was only given in the 1950’s map. The map indicated low-risk areas in Western Jutland and high-risk areas in Northern Zealand and Western Funen which were not found in the overall age 15 map (Fig. 1, middle). Otherwise, the patterns in the two maps were similar.

Analysing the 2926 cases with ≥ 25 km between residences at age 15 and onset of disease (Fig. 3), spatial variation in the odds of MS was found based on residence at onset (p = 0.002), but not at age 15 (p = 0.49). At age 15, however, the estimate did indicate unexpectedly high odds of MS in northern Eastern Jutland and low odds in central Jutland and Copenhagen. Similar overall results were obtained using 20 km as limit.

4. Discussion

This was a nationwide, register-based case-control study including 12,993 cases of MS in Denmark. We identified spatial variation in the risk of MS according to residence at birth, age 15, and onset of disease. Several high- and low-risk areas were identified across the country with some variation between birth, age 15 and onset of disease.

The first geographical description of the occurrence of MS in Denmark was reported in 1934 (Gram, 1934) based on disability insurance claims. In 1956, MS prevalence rates were studied (Hyllested, 1956). Both studies may be misleading in terms of assessing the risk of MS, as previously noted regarding prevalence rates (Koch-Henriksen and Hyllested, 1988). However, except for Funen, the overall pattern in these early results corresponded to the results of the present study in terms of identifying areas with high/low occurrence of MS. Another Danish study found spatial differences in MS incidence rates 1948−64 based on county of residence at onset of disease (Koch-Henriksen and Hyllested, 1988). Except for low incidence in Northern Jutland, their results did not correspond to the results of the present study. This may reflect a change over time, but also the discrepancy between the methods used in the two studies. The previous study aggregated data within counties which may influence the results, known as the Modifiable Areal Unit Problem (Openshaw and Taylor, 1979).

Several studies have reported spatial differences in MS prevalence within countries and regions (Fromont et al., 2010; Pugliatti et al.,...
Fewer studies considering MS incidence also found spatial differences according to residential area at onset of disease (Fromont et al., 2010; Torabi et al., 2014). Spatial differences according to residential municipality of later MS patients at age 5 to 15 years have been found in Sassari, Italy (Pugliatti et al., 2002).

Spatial variation in MS may reflect both genetic and environmental differences. The Danish population, however, has been very homogeneous regarding race and ethnicity, and immigrants and their descendants were all excluded from the analyses. Clustering of MS within families has been reported (Nielsen et al., 2005), but still the observed spatial differences within Denmark are more likely to reflect environmental rather than genetic differences. Environmental differences may also include differences in e.g. life style.

Sunlight exposure has been mentioned as an explanation of North-South gradients in MS prevalence. A meta-analysis, however, found no correlation between MS incidence and latitude in western Europe (Koch-Henriksen and Sorensen, 2010). In contrast, a recent study reported a persistent North-South gradient in MS incidence in Ireland (O’Connell et al., 2017), but in the present study, no indication of systematic latitudinal variation in the risk of MS was seen.
Although not conclusive, other studies have suggested exposure to chemicals in soil (Monti et al., 2016), water (Irvine et al., 1989), and air (Bolviken et al., 2003) as potential explanations of spatial differences in MS. Other environmental factors which have been associated with the risk of developing MS in susceptible individuals include Epstein-Barr virus, vitamin D status, smoking, obesity in early life (Ascherio and Munger, 2016), urban residence (Green et al., 2013), socioeconomic position during childhood (Nielsen et al., 2013), education (Riise et al., 2011), and fewer childbirths (Magyari et al., 2013). None of these exposures present an obvious explanation of the observed spatial variation in MS within Denmark. The latter exposures are life style factors which may be associated with urbanization. The spatial variation in MS, however, did not immediately correspond to variation in urbanization in Denmark. For example, the capital Copenhagen and the predominantly rural area Lolland were both low-risk areas. Further studies are needed to assess the association between specific exposures and the observed spatial variation in MS.

The observed spatial patterns in MS according to residence at birth and age 15 were similar. However, change of residence during childhood was limited, which may explain the similarity. Moreover, it makes it difficult to distinguish between the effect, if any, of environmental exposure at birth and age 15. Most cases tended to reside relatively close to their childhood address at onset of disease. Therefore, spatial differences at onset might reflect childhood exposure, and vice versa. To eliminate that, we looked at cases who had a substantial change in residence from age 15 to onset of disease. In this sub-population, we found spatial variation in MS according to residence at onset of disease, but not at age 15. This does not support the hypothesis that childhood is the most vulnerable time for environmental exposures relevant for MS. These results were, however, based on a relatively small, highly selected subgroup and should therefore be interpreted with caution. Also, the results did indicate a few areas with unexpectedly high/low risk of MS among those who resided there at age 15 and later moved away. This may suggest some effect of childhood exposure.

In DMSR, clinical onset of MS is determined retrospectively, based on review of the clinical records by MS expert neurologists. Thus, it does not rely on the patient's recollection, which is a strength of the study.

Our study may have some limitations. In the main analyses, any temporal component in exposure was ignored by aggregating data from more than 40 years. We stratified analyses according to birth cohort in attempt to reveal changes in the spatial pattern over time, but the results were based on small samples and may be subject to low power. Furthermore, the last cohort was still relatively young and further MS cases may develop. We did not take into account the duration of stay at a certain residence or other residences than those at birth, age 15, and onset. Further studies aiming at determining the age of susceptibility to environmental exposure and possible changes in environmental factors over time should include a space-time analysis utilizing the complete residential history available for all Danish citizens. In other populations, previous studies have investigated space-time clustering according to the age of later MS patients with contradicting results (Pugliatti et al., 2006; Riise et al., 1991; van Buuren et al., 1998; Ashley and Mackenzie, 1970).

In Denmark, the health-care system is public and available free of charge to all citizens. However, utility of the health-care system is suspected of varying across the country. That could potentially lead to geographical differences in the diagnosis of MS, which would affect results on the spatial variation in disease risk. However, since DMSR is estimated to be 91% complete (Koch-Henriksen et al., 2001), lack of diagnosis is not expected to explain the observed spatial differences.

In this study, a case-control design was used in order to reduce computing time. The Danish registers in principle allow for conducting a cohort study, since residential information is available for all citizens, but this was deemed infeasible for computational reasons. The main results were very consistent despite using different sets of controls at each considered time point. This adds to the strength of the results since it is unlikely that a random finding would be reproduced when the data change. Initial analyses indicated that increasing the number of controls did not affect the results.

In general, the uncertainty of the estimated maps depends on population density and varies across the country. The uncertainty is not apparent from the maps, but it is accounted for in the determination of high-/low-risk areas.

In conclusion, we found small-scale geographical variation in the risk of MS in Denmark, which suggests that local environmental risk factors could be at play. These risk factors may be related to life style, but further studies are needed to confirm that.

Funding

The work was supported by the Danish Multiple Sclerosis Society (Grant no. A29759). The funding source was not in any way involved in the study or the submission of the article.

Conflicts of interest

Melinda Magyari M.M: has served on scientific advisory board for Biogen, Sanofi, Teva, Roche, Novartis, Merck, has received honoraria for lecturing from Biogen, Merck, Novartis, Sanofi, has received support from Biogen, Sanofi, Teva, Roche, Novartis, Merck, has received grant support from Biogen, Sanofi, Teva, Roche, Novartis, Merck, has received honoraria from Biogen, Sanofi, Teva, Roche, Novartis, Merck, has received support from Biogen, Sanofi, Teva, Roche, Novartis, Merck, has received grant support from Biogen, Sanofi, Teva, Roche, Novartis, Merck.
for congress participation from Biogen, Genzyme, Teva, Roche.

References