medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis cohort population structure and disease etiology

Short title: CCPGSMS cohort population structure and disease etiology

Alison M. Pagalilauan[1,2], Elif Everest[1,2], Suzanna Rachimi[1,3], Daniel Reich[4], Alex D. Waldman[1,5,6], A. Dessa Sadovnick[7], Carles Vilariño-Guell[7], Michael J. Lenardo[1,2*]

[1] Molecular Development of the Immune System Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America

[2] Clinical Genomics Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America

[3] Immunology Graduate Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America

[4] Translational Neuroradiology Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, United States of America

[5] Academic Unit of Neuropathology, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, United Kingdom

[6] Emory University MD/PhD Program, Emory University School of Medicine, Atlanta, Georgia, United States of America

[7] Department of Medical Genetics, University of British Columbia, Vancouver, Canada

* Corresponding author

E-mail: lenardo@nih.gov (MJL)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Background: Previous genetic and epidemiological studies have examined subpopulations from the Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis (CCPGSMS) patient cohort, but an encompassing analysis of the study population has not yet been carried out.

Objective: This study examines patterns of multiple sclerosis (MS) prevalence in 13,663 cohort members, including 4,821 patients with MS or suspected MS and 8,842 family members.

Methods: We grouped participants into epidemiologic subgroups based on age of MS onset, clinical stage at diagnosis, symptom type at disease onset, sex, proband status, disability as measured by the EDSS, and ancestry based on reported ethnicity.

Results: We observed a 2.7:1 MS prevalence ratio of women to men, though disease severity was greater for male patients. Variation in the age of disease onset between patients was only slightly associated with sex and strongly associated with disease type. Specific types of clinical symptoms at disease onset were associated with the prognosis. Regional residence did not correlate with disease onset, type, or severity.

Conclusion: Population trends, as presented here, are not explained by environmental factors alone, highlighting the need for a comprehensive genetic analysis to understand disease variance across families.

Keywords: epidemiology, demographics, multiple sclerosis, population trends, prognosis, risk factors

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Multiple sclerosis (MS) is the leading cause of non-traumatic brain injury in young adults and is characterized by immune-mediated damage to the myelin nerve sheath and central nervous system (CNS) degeneration [1]. MS has varying forms and stages of the disease, including relapsing-remitting MS (RRMS, ~85% of cases), secondary-progressive MS (SPMS, ~60–70% of RRMS cases develop into this form), and primary-progressive MS (PPMS, ~10–15% of cases) [2,3]. Many patients first present with an acute “clinically isolated syndrome” (CIS). If the patient’s initial magnetic resonance imaging (MRI) demonstrates CNS lesions, CIS will likely progress to relapsing disease [4]. MS pathogenesis is believed to involve environmental and genetic factors [5,6]. While the increasing prevalence of MS may be due to changes in diagnostic criteria and improvements in disease awareness in health systems around the world [7], alterations in environmental factors may also play a role. Distance from the equator, vitamin D deficiency, and infection with Epstein-Barr virus (EBV) are significant MS susceptibility factors [8,9]. In the United States, the highest MS prevalence is in northeastern and midwestern states, generally at higher latitudes [2]. Prevalence in Manitoba, Canada, is among the highest worldwide [10,11]. The correlation between higher latitudes and increased risk of MS may involve decreased exposure to solar ultraviolet (UV) radiation [12]. Reduced exposure to UV radiation was first proposed as an environmental susceptibility factor for MS to explain how vitamin D protects from MS [12]. Vitamin D controls the promoter region of the human leukocyte antigen (HLA) allele *HLA-DRB1*, the main susceptibility locus for MS [13]. The specific haplotype is *HLA-DRB1*1501*, dominant in Northern European populations [13]. Like MS, EBV prevalence is also increased at higher latitudes and a risk factor for MS [14,15]. However, the specific pathogenic mechanisms of environmental factors are unknown, leading to interest in defining genetic factors.

The most strongly associated genetic risk factors identified with MS are certain HLA II genes, such as DQA1, DQB1, and DRB1 [16,17]. Genome-wide association studies (GWASs) have identified a host of genes, mainly immunological, that may contribute to MS [18]. How these genes contribute to MS pathogenesis remains unknown. Given the

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

persisting questions regarding MS etiology and pathogenesis, it is valuable to examine historical cohorts using modern epidemiological techniques and genomic analyses. One classic study of MS is the Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis (CCPGSMS), for which information is still publicly available [5,21]. A total of 13,663 patients and family members were included before enrollment ended. Exome sequencing analyses have been performed for a subset of CCPGSMS families, revealing candidate risk variants associated with MS [19,20]. Here, we provide information on population structure and disease etiology of the complete CCPGSMS cohort.

Methods

This study is a retrospective analysis of data assembled by the Canadian Collaborative Study Group. The investigation uncovered MS patients with pediatric onset of disease, families with multiplex MS disease, and MS co-occurring with other autoimmune diseases. Participants provided written informed consent under an IRB protocol approved by the Research Ethics Review Board (REB # H08-01669) at the University of British Columbia. Data were collected over 18 years, providing a large patient and family member cohort with thorough background information [21]. Enrollment ended approximately 10 years ago. Genetic screens, geography questionnaires, and co-morbidity questionnaires were administered to all patients.

When the study first began in 1993, patients were diagnosed using the Poser criteria. These now obsolete criteria required evidence of two or more relapses lasting over 24 hours, occurring a minimum of one month apart, in addition to evidence of CNS lesions [22]. Where possible, patients and family members were later re-evaluated using the McDonald criteria, which requires evidence of CNS damage over time [23,24]. Patients and their families were categorized as “affected,” “unaffected, blood-related,” and “unaffected, married-in” based on disease status and relationship to the proband.

The investigators of the present study did not collect the original specimens or clinical information, which were de-identified for use in the present study. NIH reviewed the

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

study and determined an IRB-approved protocol was not necessary because it did not meet the basic requirements to be considered human subject research. Data collected during the original CCPGSMS study were further divided by age of MS onset (i.e., age at date of diagnosis for clinical symptoms), clinical stage at diagnosis, symptom type at disease onset, sex, proband status, disability as measured by the EDSS, and ancestry based on reported ethnicity.

The mean ages of onset in different groups of patients were compared using the Kruskal-Wallis test. The differences in the types of symptoms at disease onset in different patient groups that were defined based on disability scores and clinical stages were tested using the chi-squared test for trend. All statistical analyses were performed with R (v4.2.1) in RStudio (v2023.06.1). The package ms.sev (v1.0.4) was used to perform global age-related MS severity (ARMSS) score calculations [25]. We used ggplot2 (v3.3.6) and plotrix (v3.8.2) to generate plots and display the results of statistical tests. The distribution of age at MS onset was analyzed in 10-year increments (Table 1; Fig 1). Using cancensus [26] (v0.5.5) and tidycensus (v1.4.4), we mapped the probands across Canadian provinces. The package cancensus allowed us to access Canadian census data available via Statistics Canada (2022).

Table 1. Summary of population data included in the study, subdivided based on disease status, blood-relation or marriage status, proband status, and sex (n=13,663).

MS Affected		**Familial Samples**	
4,821		8,842	
Proband	**Not Proband**	**Unaffected, blood-related**	**Unaffected, married-in**

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2,473		2,348		7,770		1,072	
M	**F**	**M**	**F**	**M**	**F**	**M**	**F**
658	1,815	647	1,701	3,435	4,335	551	521

M, male; F, female.

EDSS scores were converted to ARMSS scores [25]. This disability measure normalizes EDSS scores relative to disease duration to detect disability differences between different age groups. We used an ARMSS-based disability metric instead of the MS Severity Score (based on disease duration), as well as the EDSS (snapshot disability indicator), to represent disease severity status better since disease duration information can be incorrect due to the possibility of subclinical MS stages with varying periods. ARMSS scores were only calculated for patients with known age at EDSS collection available (n=1,746). ARMSS scores were used to generate global ARMSS (gARMSS) scores for each patient using the age of data collection and EDSS data. Because sample sizes were not equal between the male and female proband groups, a Monte Carlo simulation was used to randomly sample 475 patients from the female proband group over a series of 10,000 iterations with the male population and calculate an average gARMSS within a sample size equal to that of the male proband group (n=475).

Results

The population data gathered by the CCPGSMS represented a total of 13,663 individuals: 4,821 patients with MS (approximately 35%), subdivided into probands (51%) and affected relatives (49%), and 8,842 unaffected family members (Table 1). Patients included male=27% and female=73%. Unaffected family members were grouped by blood relation (88%) or marital status (12%) and again divided by sex (male=45%, female=55%) (Table 1). We found that although the sex ratio of unaffected

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

individuals was comparable (females/males=1.2), the affected individuals were skewed towards women (females/males=2.7). We next examined patterns of MS incidence (Table 2; Fig 1), which was highest among individuals 21 to 30 years old, independent of sex (Fig 1), followed by those aged 31 to 40 years. MS was extremely rare under the age of 10 (n=7) and over 60 (n=19), probably due to lower awareness of early- and late-onset MS at the time of the sample collection period.

Table 2. Proportion of the affected population grouped by age at the time of MS diagnosis (n=3,787).

Age Stratum	*n*	Proportion (%)	Cumulative Proportion (%)
4–10	11	0.29	0.29
11–20	449	11.86	12.15
21–30	1490	39.35	51.50
31–40	1147	30.29	81.78
41–50	532	14.05	95.83
51–60	139	3.67	99.50
61–70	18	0.50	99.80
71–75	1	0.03	100

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

RRMS was the most common clinical form (34.8%), followed by chronic-progressive MS (CPMS) (10.4%) and relapsing-progressive MS (RPMS) (7.5%). Benign diagnoses describing a less severe version of RRMS with few mild relapses were grouped with the larger RRMS group. Though “chronic progressive” is no longer commonly used, we believe it is important to reflect the cohort demographics of the original study. We also did not have enough clinical information to associate CPMS data with more specific categories such as SPMS or PPMS. Patients with benign phenotypes were defined as having an EDSS score equal to or lower than 3.0 after 15 or more years from disease onset and having never received MS therapies [19,27]. SPMS (4.1%) and PPMS (1.7%) were the least frequently reported clinical forms. Affected patient diagnoses that were unreported at the time of data collection were considered “unknown” (41.5%) (Fig 2A). Disease type contributed strongly to the variation in age of onset between patients. Mean age of onset of patients with a progressive disease was significantly higher than that of patients with benign and relapsing forms of MS (both $p<0.0001$).

The ancestry breakdown was examined for proband patients for whom data were available (n=905). Most participants reported ancestry from Europe (92.5%), far exceeding those reporting ancestry from Asia (5.7%), Africa (1%), and Indigenous North America (0.05%) (Fig 3). Reported ancestry did not significantly contributed to variation in age of onset (p=0.06).

Residency at the time of data collection was also examined and plotted by province to show the distribution of proband patients across Canada as a dot density map (Fig 4A). The total population per province, as estimated by the 1996 Census, was indicated on

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the same dot density map to highlight the influence of geographic distribution on overall cohort demographics. Patients were mainly from Ontario (33.2%) and Quebec (17.7%), the most populous provinces, followed by British Columbia (15.2%), Nova Scotia (10.8%), and Saskatchewan (10.6%) (Fig 4A). There appeared to be an ascertainment bias towards patients recruited in British Columbia and Ontario, as the study was spearheaded in parallel by MS clinics in Vancouver and London, though we observed enrichments in probands (individuals with familial MS) in Nova Scotia and Saskatchewan (Fig 4B). Provincial residency did not appear to contribute significantly to variation in onset age.

To interpret differences between subsets of patients with MS, disability status was examined using ARMSS scores. Affected males had significantly higher gARMSS scores compared to affected females (p=0.00095) (Fig 5A). To account for the discrepancy in sample size between male and female proband subsets, this sex-biased difference in disease severity was further corroborated by a probabilistic model with a Monte Carlo simulation generating 10,000 simulated data sets built from female proband EDSS data equaling the sample size of the male proband EDSS data (n=475). The mean ARMSS of each simulated data set was compared to the mean gARMSS score of the male proband group ($p<2.2\times10^{-16}$). Across all iterations of the simulation, the mean difference in gARMSS scores between groups remained small but constant, with male proband scores maintaining an average 0.53 increase over female proband scores (Fig 5B).

To assess associations between symptoms at disease onset and prognosis, patients with available gARMSS scores were stratified into two groups: those with a gARMSS score ≥5 (indicative of an unfavorable disease course, n=1,264) and those with a gARMSS score <5 (indicative of a favorable disease course, n=495). We used this cutoff based on a previous study indicating that the cutoff well-distinguished the patients based on their disability statuses [28]. Indeed, the mean gARMSS score in the current patient cohort was significantly higher in the unfavorable disease course group (p<0.0001), showing that the disability scores of each group were not closely aggregated below and above the cut-off value of 5. It is important to note that the higher

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

number of patients with a gARMSS score ≥5 than those with a gARMSS score <5 is likely due to the long follow-up period (>18 years) of the cohort, given that the development of a progressive course is likely an age-related phenomenon [29]. The symptoms at disease onset were categorized as follows: fatigue, diplopia, retrobulbar neuritis, balance problems, gait disturbances, upper limb ataxia, sensory-face, sensory-spinal cord, pain, motor-acute, motor-slow, transverse myelitis, bladder dysfunction, Lhermitte's sign, and vertigo. Among individuals with available data on symptoms at disease onset, the correlation between the type of symptom and disease course was examined using the chi-squared test for trend, comparing the unfavorable and favorable course groups. A significantly higher proportion of patients in the unfavorable course group presented with gait problems and slow motor dysfunction at disease onset compared with those in the favorable course group (p=0.0027 and p=0.005, respectively), while sensory spinal cord symptoms were more prevalent at disease onset in the favorable course group (p=0.017) (Fig 6A). Consistently, within the subgroup of patients with a well-defined disease course history (benign/relapsing-remitting with complete remission vs. definitive progressive course), a greater percentage of individuals with progressive MS exhibited balance problems (p=0.022), gait disturbances (p=0.012), and slow motor dysfunction (p=0.026) at disease onset. Conversely, patients with a benign course more frequently manifested sensory spinal cord symptoms as their initial disease presentation (p=0.04) (Fig 6B). Given the higher disease severity observed in male patients, we tested whether there was a difference between prognosis-associated symptoms at disease onset in males and females. This revealed that sensory spinal cord symptoms were significantly more frequent in females at disease onset (p<0.0001). Other comparisons did not show a difference in severity that could shed light on MS pathogenesis. For example, disease severity in different provinces appeared to relate mainly to the population density of different provinces (S2 Fig).

Discussion

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Our study population of 13,663 participants comprises one of the largest familial specimen sets assembled for an autoimmune disease. Patterns of MS incidence found in this cohort corroborate those examined in previous studies. Among CCPGSMS patients, peak incidence was observed between ages 21 and 40. In both males and females, patients with CPMS and PPMS were diagnosed at slightly older ages than those with other types of MS (Fig 3B), aligning with the notion that older age predisposes patients to more progressive phenotypes or that progression reflects the natural history of the disease as patients age. We also observed that RRMS was the most common form of MS; 82% of patients in this study were younger than 41 years at the time of diagnosis, which is consistent with existing reports [15,30,31].

Over a third of proband patients reported residency in Ontario (32.7%). The remaining proband population was largely from Quebec (17.4%), British Columbia (16.7%), Nova Scotia (9.4%), and Saskatchewan (9.3%). Though the 1996 Census reported that less than 6% of the total population of Canada resided in Nova Scotia and Saskatchewan at the time, more proband patients of the CCPGSMS cohort were found to be residents of these provinces than residents of Alberta (7.9%) or Manitoba (5.2%). Studies of MS prevalence in Canada have reported the Canadian prairies (SK, AB, MB) as having the highest incidence rates of all regions [10,11]. Proband incidences in this study were proportionally comparable to the relative populations of most Canadian provinces, apart from Alberta and Nova Scotia. Notably, the relative proportion of probands in Nova Scotia was three times greater than the relative proportion of Canadians residing in the province in 1996. It is also important to note that many rural communities have reduced access to medical care, which sometimes forces affected patients to move to urban centers for better access. Conversely, some patients with increased severity may be limited in their ability to move to major centers.

Increased disease severity in males was observed in the present cohort, validating previous observations that although women with RRMS relapse more frequently than men with RRMS, the latter group exhibits more rapid disability accumulation following each relapse [32] and thus a worse overall outcome. The ratio of affected females to affected males in our study is comparable to the 3:1 ratio measured globally [33]. This

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

demographic imbalance in sex is relatively constant across age strata in this study and has been observed in numerous other cohorts [31,32,33]. However, different hypotheses have been proposed to explain this distribution [32,34]. Regarding the differences in clinical presentation at onset, there are a limited number of studies investigating differences in attack locations between female and male patients with MS, yielding inconsistent results [35,36]. A comprehensive study involving 14,969 patients with MS examined the relationship between 49,279 phenotypically characterized MS attacks and associated demographic and clinical parameters [36], showing that sensory attacks are more commonly observed in female patients and associated with a non-progressive course. This is in line with our current observations that sensory spinal cord symptoms were more frequent in female patients and in those with a lower disability score and benign course, and male patients had higher disability scores. The reasons underlying the differences in clinical presentation and progression between sexes are currently unclear; however, cellular and animal studies have indicated possible molecular differences between sexes, involving differences in oligodendrocyte differentiation, remyelination capacity, or axonal vulnerability [37,38].

In addition to sex, ancestry is a risk factor for MS. Here, we report a skewed number of affected patients with ambiguous European ancestry, confounding risk attributable to proximity to the equator. This limitation is due to the biased geographical ascertainment in this study and the small sample size of proband patients for which complete demographic information was provided (n=905). In our analysis, the sample size was strictly limited to proband patients for which all criteria (i.e., sex, age, clinical stage, ancestry, and residence) were available. Though most enrolled participants reported European ancestry, it is unclear whether this distribution is indicative of trends unique to the etiology of MS, as ~70% of Canadians are ethnically European [39]. This ascertainment bias is a weakness of this dataset. Regional distance from the equator was previously shown to be a risk factor in MS, but alone, it does not determine disease status. Reduced vitamin D due to more limited sunlight exposure, rather than proximity to the equator, has been discussed as the relevant correlated risk factor. In rural areas where more populations engage in outdoor work such as farming, vitamin D exposure may not always accurately correlate with latitude [40]. Confounding studies of MS

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

prevalence in Canada have reported a higher incidence among patients from the Canadian Prairies, which are north of other Canadian regions but also experience the most average annual sunlight hours [11]. Further, many countries across the world have seen an increase in the prevalence of MS, regardless of geographical location [41]. Interestingly, more significant increases in female-to-male ratios are associated with a latitudinal gradient, with the ratio being more drastically skewed towards female patients with RRMS in northern latitudes [32,42]. Findings such as these have led many researchers to reassess the idea of a latitudinal gradient as a causative factor, further suggesting that this gradient effect has been attenuated [33].

A major limitation of the current study is that MS diagnostic criteria have changed since the data were initially collected, resulting in non-identical, multilateral patient diagnoses and severity assessments. Another limitation is that the effect of disease-modifying treatments was not accounted for in the analysis of MS severity. Lastly, there is a possible selection bias in different sub-comparisons since we could not include the whole cohort in some analyses due to missing data. The next steps for this CCPGSMS cohort will be combining the population data with genomic DNA sequencing. As multiplex MS families are common in this cohort, we plan to focus on identifying high-risk genetic variants in MS.

Acknowledgments

We thank Justin Gabrielski for critical reading of the manuscript.

References

1. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. N Engl J Med. 2018;378(2):169-80.
2. Dilokthornsakul P, Valuck RJ, Nair KV, Corboy JR, Allen RR, Campbell JD. Multiple sclerosis prevalence in the United States commercially insured population. Neurology. 2016;86(11):1014-21.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3. Dutta R and Trapp BD. Relapsing and progressive forms of multiple sclerosis: insights from pathology. Curr Opin Neurol. 2014;27(3):271-8.
4. Doshi A and Chataway J. Multiple sclerosis, a treatable disease. Clin Med. 2016;16(6):s53-9.
5. Sadovnick AD. Genetic background of multiple sclerosis. Autoimmun Rev. 2012;11(3):163-6.
6. Sadovnick AD, Yee IM, Criscuoli M, DeLuca GC. Genes and environment in multiple sclerosis: Impact of temporal changes in the sex ratio on recurrence risks. Mult Scler J. 2022;28(3):359-68.
7. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler. 2020;26(14):1816-21.
8. Hedström AK. Risk factors for multiple sclerosis in the context of Epstein-Barr virus infection. Frontiers in Immunology. 2023 Jul 24;14:1212676.
9. Sabel CE, Pearson JF, Mason DF, Willoughby E, Abernethy DA, Taylor BV. The latitude gradient for multiple sclerosis prevalence is established in the early life course. Brain. 2021 Jul 1;144(7):2038-46.
10. Marrie RA, Yu N, Blanchard J, Leung S, Elliott L. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. Neurology. 2010;74(6):465-71.
11. Beck CA, Metz LM, Svenson LW, Patten SB. Regional variation of multiple sclerosis in Canada. Mult Scler. 2005;11:516-9.
12. Hayes CE, Cantorna MT, DeLuca HF. Vitamin D and Multiple Sclerosis. Exp Biol Med. 1997;216(1):21-7.
13. Handunnetthi L, Ramagopalan SV, Ebers GV. Multiple sclerosis, vitamin D, and HLA-DRB1*15. Neurology. 2010;74(23):1905-10.
14. Disanto G, Pakpoor J, Morahan JM, Hall C, Meier UC, Giovannoni G, et al. Epstein-Barr virus, latitude and multiple sclerosis. Mult Scler. 2013;19(3):362-5.
15. Gbaguidi B, Guillemin F, Soudant M, Debouverie M, Mathey G, Epstein J. Age-period-cohort analysis of the incidence of multiple sclerosis over twenty years in Lorraine, France. Sci Rep. 2022;12(1001).

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

16. Fogdell A, Hillert J, Sachs C, Olerup O. The multiple sclerosis- and narcolepsy-associated HLA class II haplotype includes DRB5*0101 allele. Tissue Antigens. 1995;46:333-6.
17. Alcina A, Abad-Grau MD, Fedetz M, Izquierdo G, Lucas M, Fernandez O, et al. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. PLoS One. 2012;7(1):e29819.
18. International Multiple Sclerosis Genetics Consortium*†, ANZgene, IIBDGC, WTCCC2. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019 Sep 27;365(6460):eaav7188.
19. Wang Z, Sadovnick AD, Traboulsee AL, Ross JP, Bernales CQ, Encarnacion M, et al. Nuclear Receptor NR1H3 in Familial Multiple Sclerosis. Neuron. 2016;90(5):948-54.
20. Gil-Varea E, Urcelay E, Vilariño-Güell C, Costa C, Midaglia L, Matesanz F, et al. Exome sequencing study in patients with multiple sclerosis reveals variants associated with disease course. J Neuroinflammation. 2018;15(1):265.
21. Sadovnick AD, Risch NJ, Ebers GC. Canadian collaborative project on genetic susceptibility to MS, phase 2: rationale and method. Can J Neurol Sci. 1998;25(3):216-21.
22. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: Guidelines for research protocols. Ann Neurol. 1983;13:227-31.
23. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for MS: Guidelines from the international panel on the diagnosis of MS. Ann Neurol. 2001;50:121-7.
24. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2017;17(2):162-73.
25. Manouchehrinia A, Westerlind H, Kingwell E, Zhu F, Carruthers R, Ramanujam R, et al. Age related multiple sclerosis severity score: Disability ranked by age. Mult Scler. 2017;23(14):1938-46.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

26. Von Bergmann J, Jacobs A, Shkolnik D. cancensus: R package to access, retrieve, and work with Canadian Census data and geography. 2022. v0.5.5.
27. Amato, MP, Zipoli, V, and Goretti, B. et al. Benign multiple sclerosis. J Neurol. 2006;253:1054-9.
28. Everest E, Uygunoglu U, Tutuncu M, Bulbul A, Onat UI, Unal M, et al. Prospective outcome analysis of multiple sclerosis cases reveals candidate prognostic cerebrospinal fluid markers. PLoS One. 2023;18(6):e0287463.
29. Tutuncu M, Tang J, Zeid NA, Kale N, Crusan DJ, Atkinson EJ, et al. Onset of progressive phase is an age-dependent clinical milestone in multiple sclerosis. Mult Scler J. 2013;19(2):188-98.
30. Westerlind H, Ramanujam R, Uvehag D, Kuja-Halkola R, Boman M, Bottai M, et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. Brain. 2014;137(3):770-8.
31. Rollot F, Fauvernier M, Uhry Z. Effects of age and disease duration on excess mortality in patients with multiple sclerosis from a French nationwide cohort. Neurology. 2021;97(4):e403-13.
32. Ysrraeliy MC, Correale J. The impact of sex hormones on immune function and multiple sclerosis development. Immunology. 2018;156:9-22.
33. Koch-Henriksen N, Sørensen PS. The changing demographic pattern of multiple sclerosis epidemiology. The Lancet Neur. 2010;9(5):520-32.
34. Voskuhl RR, Sawalha AH, Itoh Y. Sex chromosome contributions to sex differences in multiple sclerosis susceptibility and progression. Mult Scler. 2018;24(1):22-31.
35. Tsantes E, Leone MA, Curti E, Cantello R, Vecchio D, Granella F. Location of first attack predicts the site of subsequent relapses in multiple sclerosis. J Clin Neurosci. 2020;74:175-9.
36. Kalincik T, Buzzard K, Jokubaitis V, Trojano M, Duquette P, Izquierdo G, et al. Risk of relapse phenotype recurrence in multiple sclerosis. Mult Scler J. 2014;20(11):1511-22.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

37. Marin-Husstege M, Muggironi M, Raban D, Skoff RP, Casaccia-Bonnefil P. Oligodendrocyte progenitor proliferation and maturation is differentially regulated by male and female sex steroid hormones. Dev Neurosci. 2004;26(2-4):245-54.
38. Li WW, Penderis J, Zhao C, Schumacher M, Franklin RJ. Females remyelinate more efficiently than males following demyelination in the aged but not young adult CNS. Exp Neurol. 2006;202(1):250-4.
39. Statistics Canada. Table 98-10-0033-01 Broad age groups, gender and census year: Canada, provinces and territories [Data table]. 2022.
40. Kearns PK, Paton M, O'Neill M, Waters C, Colville S, McDonald J, et al. Regional variation in the incidence rate and sex ratio of multiple sclerosis in Scotland 2010-2017: findings from the Scottish Multiple Sclerosis Register. J Neurol. 2019;266(10):2376-86.
41. Urru SA, Antonelli A, Sechi GM, MS Working Group. Prevalence of multiple sclerosis in Sardinia: A systematic cross sectional multi-source survey. Mult Scler J. 2020;26(3):372–80.
42. Trojano M, Lucchese G, Graziano G, Taylor BV, Simpson Jr S, Lepore V, et al. Geographical variations in sex ratio trends over time in multiple sclerosis. PLoS One. 2012;7(10):e48078.

Figure legends

Fig 1. Age of MS onset distribution by sex (n=3,787). Data represent all patients for whom age at initial diagnosis was provided.

Fig 2. (A) Distribution of MS forms for all patients diagnosed as affected from the patient cohort (n=4,821). (B) Faceted boxplots characterizing age of onset by sex and MS clinical stage at the time of data collection (n=3,787). Data represent all patients for whom age at initial diagnosis was provided. CP, chronic-progressive; PP, primary-progressive; RP, relapsing-progressive; RR, relapsing-remitting; SP, secondary-progressive; UNK, Unknown.

Fig 3. Distribution of proband patients grouped by reported ancestry (n=905).

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig 4. (A) Relative percentage distribution of proband patients by Canadian province (n=2,372). Datapoints are proportionally sized to the relative percentage of probands and overlayed with total population per province as estimated by the 1996 Census (accessed via Statistics Canada). Probands from unknown provinces (n=101) were excluded from this distribution. (B) Relative percentage of MS proband patients per province compared to the relative percentage of the total Canadian population per province. BC, British Columbia; AB, Alberta; SK, Saskatchewan; MB, Manitoba; ON, Ontario; QC, Quebec; NL, Newfoundland; NB, New Brunswick; NS, Nova Scotia.

Fig 5. (A) Average gARMSS of proband patients by sex with error bars indicating standard of the mean (n=1,746; p=0.00095). Data represent all proband patients for whom EDSS at initial diagnosis was provided. Two-tailed paired Student t-test p-value indicate statistical significance (***, p<0.001). (B) Histogram of gARMSS means of proband patients by sex, randomly sampled by Monte Carlo simulation with 10,000 iterations to equalize sample size. EDSS, Expanded Disability Status Scale; gARMSS, global age-related MS severity.

Fig 6. Differences in the symptoms at disease onset (A) between patients with a gARMSS score lower than 5 (favorable disease course) and those with a gARMSS score equal to higher than 5 (unfavorable disease course) and (B) between patients with a benign or progressive MS type. Chi-squared test for trend (*, p<0.05; **, p<0.01). gARMSS, global age-related MS severity.

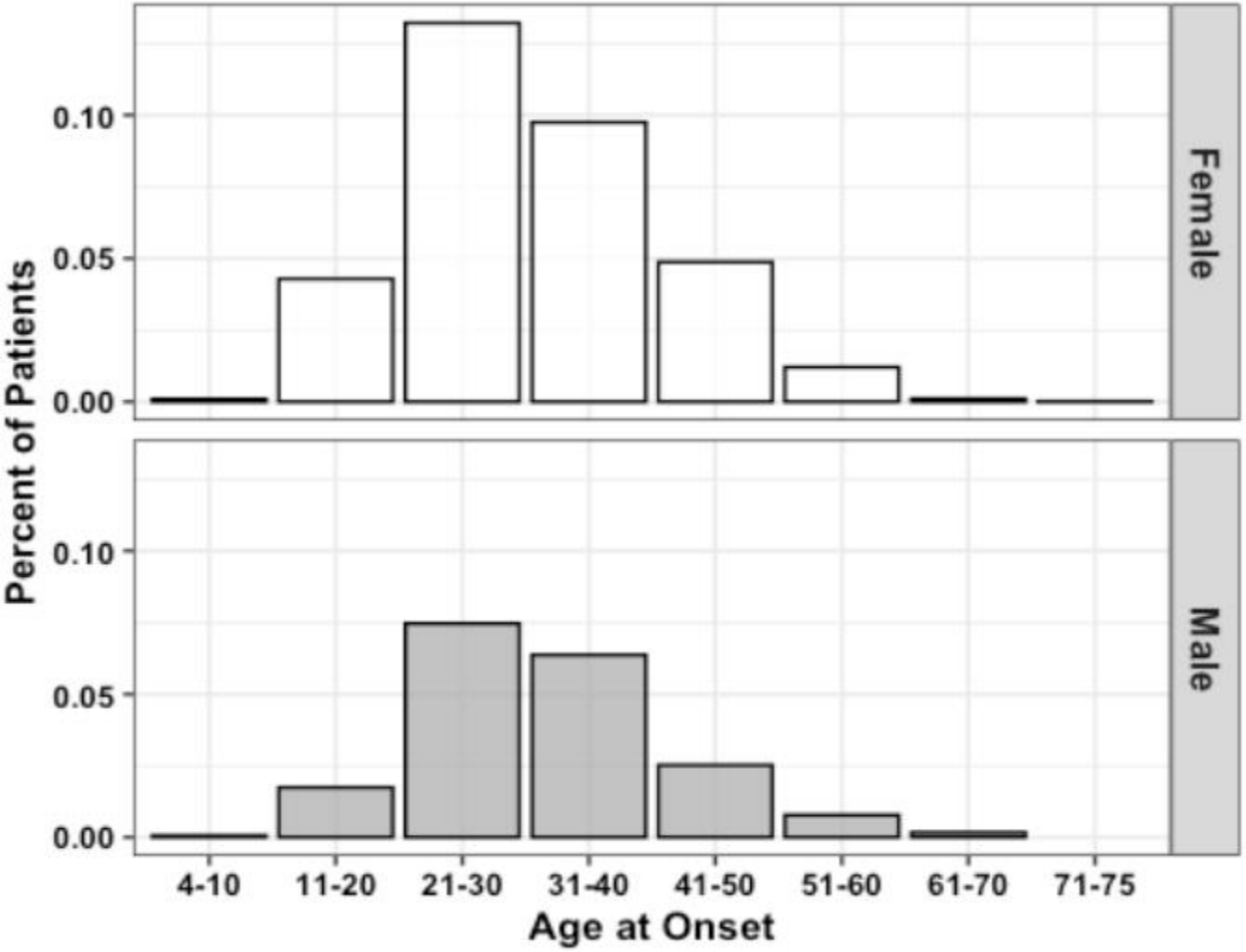
Supporting information

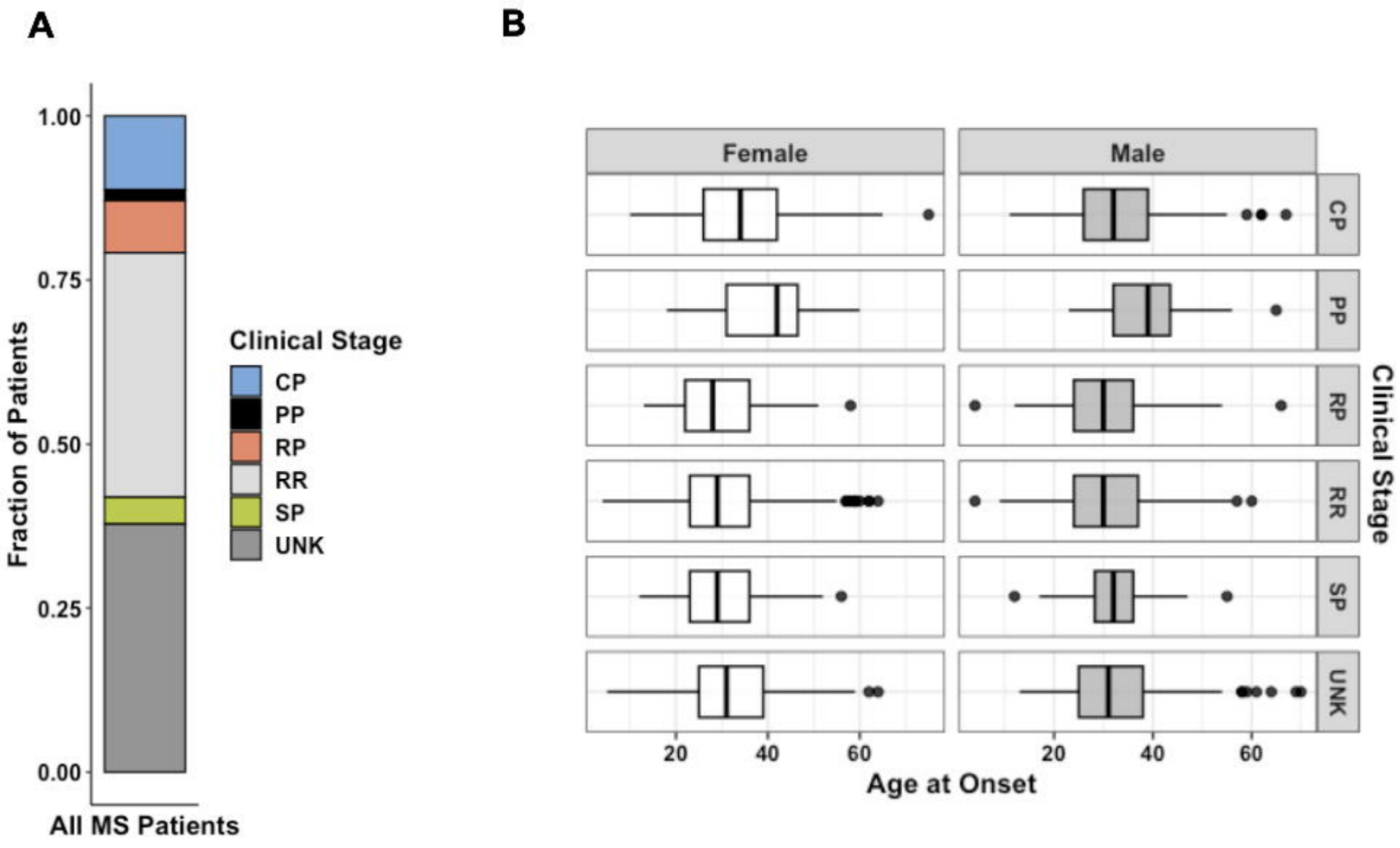
S1 Fig. Age of MS onset distribution by sex (n=3,787). Data represent all patients for whom age at initial diagnosis was provided.

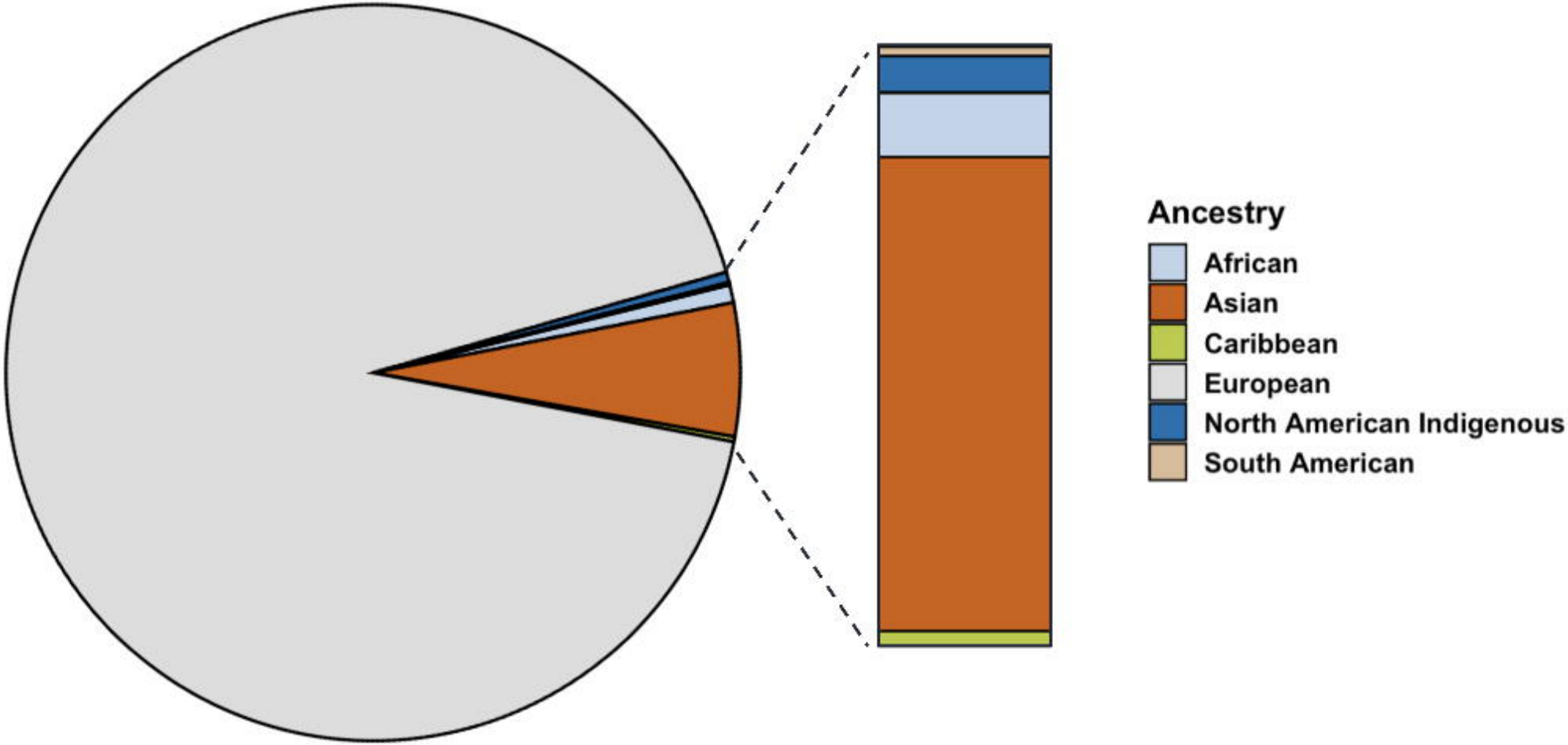
S2 Fig. Heatmap distribution of ARMSS scores for proband patients by province (n=2,473). Darker shading indicates higher incidence of a particular gARMSS score stratum in a particular province. AB, Alberta; BC, British Columbia; NL, Newfoundland;

medRxiv preprint doi: https://doi.org/10.1101/2024.04.18.24305992; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

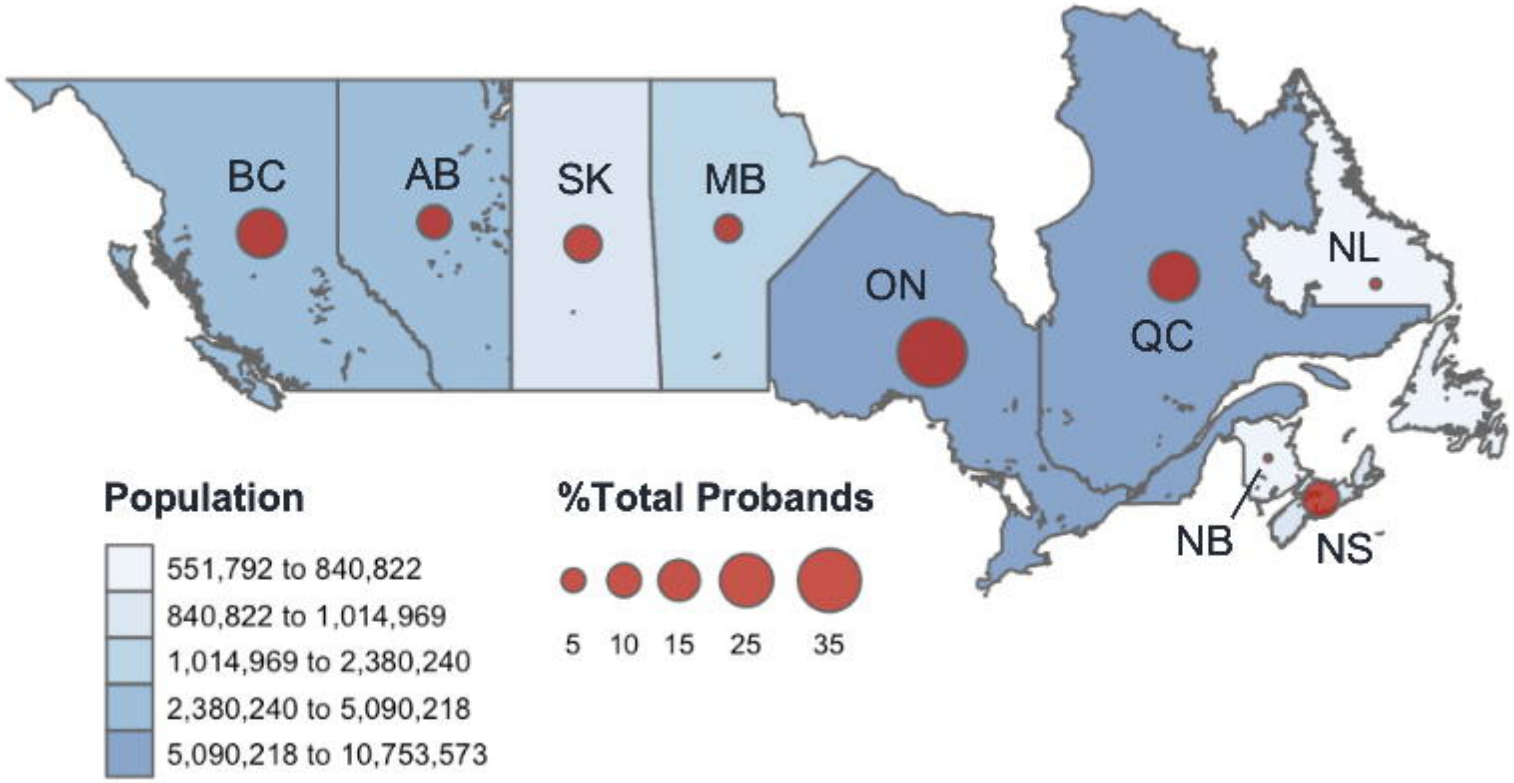
NS, Nova Scotia; ON, Ontario; SK, Saskatchewan; QC, Quebec; MB, Manitoba; UNK, Unknown; NB, New Brunswick. gARMSS, global age-related MS severity.


Percent of Patients
0.10
0.05
0.00
Female
Male
4-10
11-20
21-30
31-40
41-50
51-60
61-70
71-75
Age at Onset

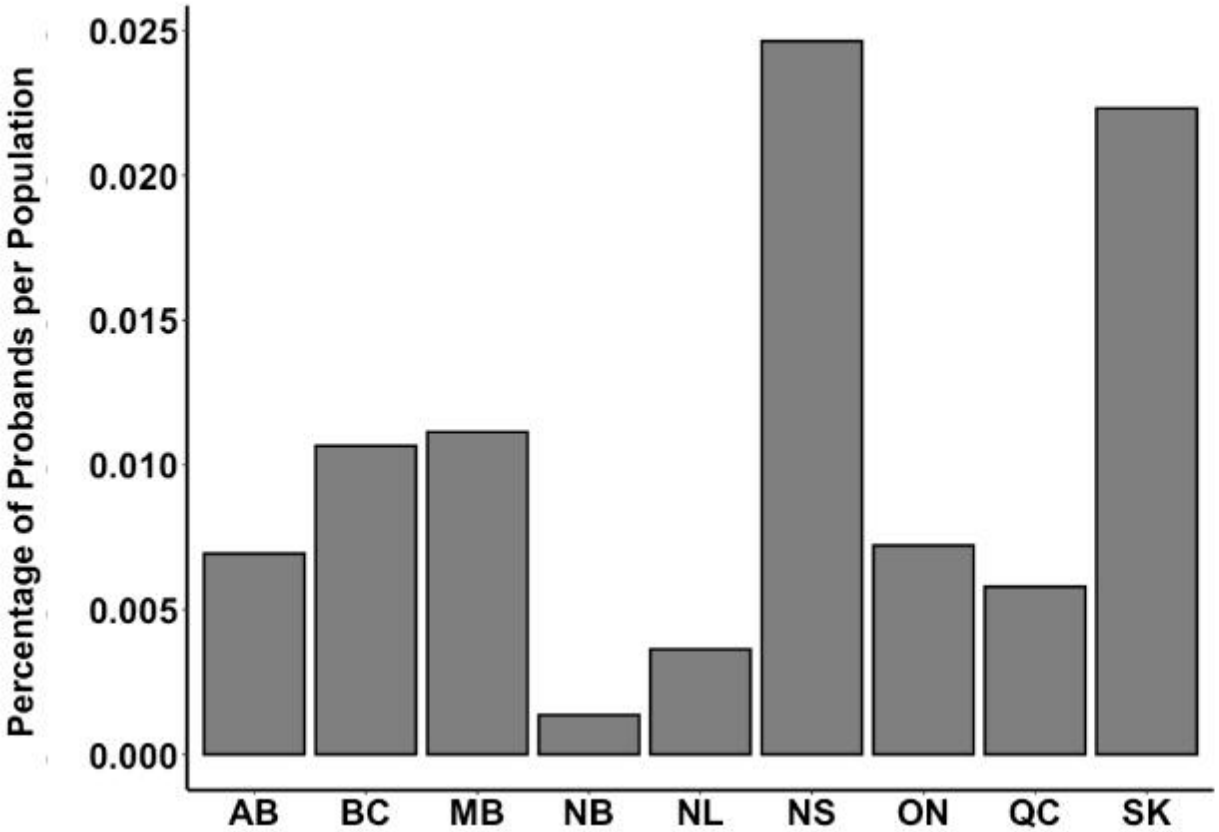

A
B
1.00
0.75
0.50
0.25
0.00
Fraction of Patients
All MS Patients
Clinical Stage
CP
PP
RP
RR
SP
UNK
Female
Male
20
40
60
Age at Onset
Clinical Stage

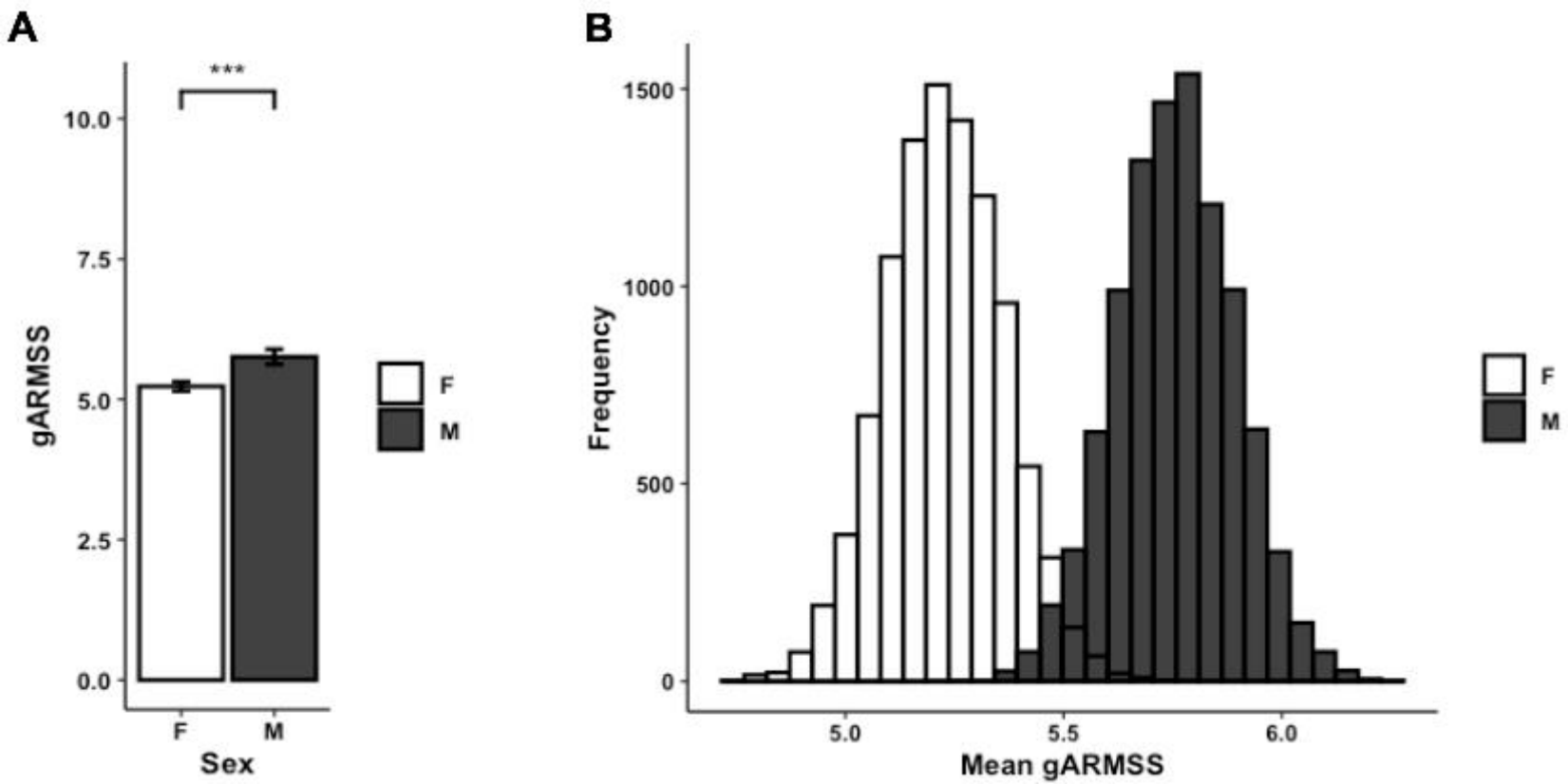

Ancestry
African
Asian
Caribbean
European
North American Indigenous
South American

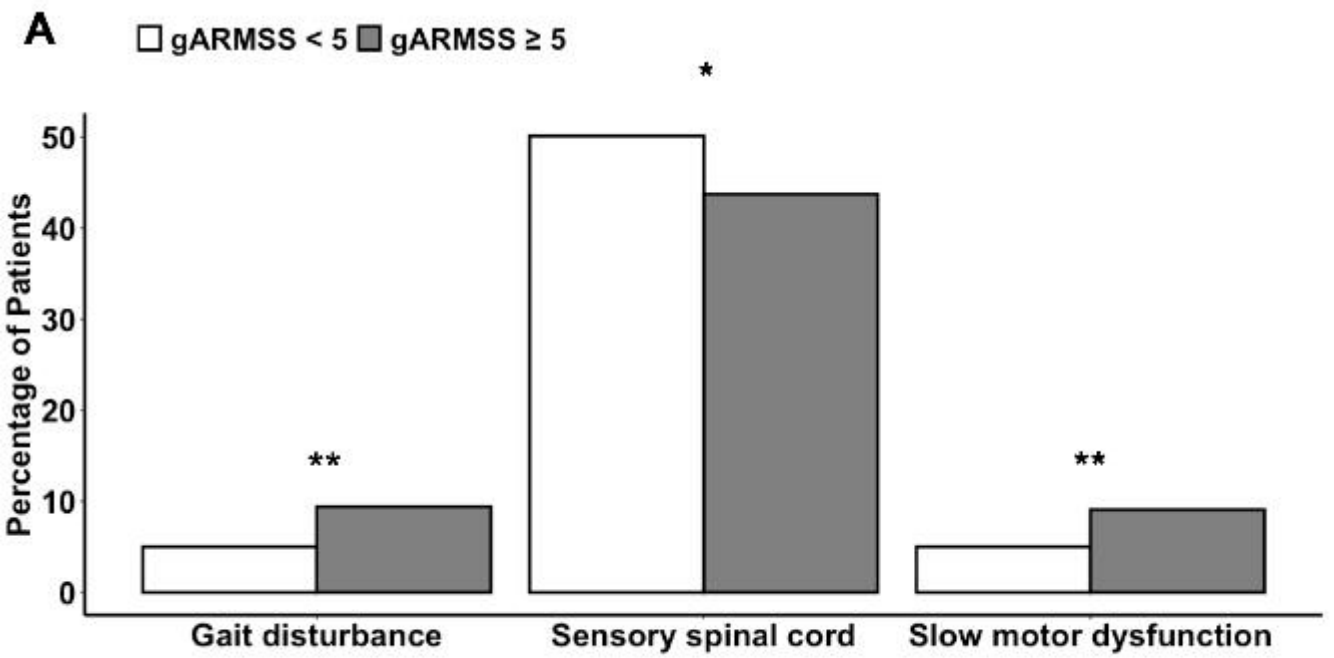
A

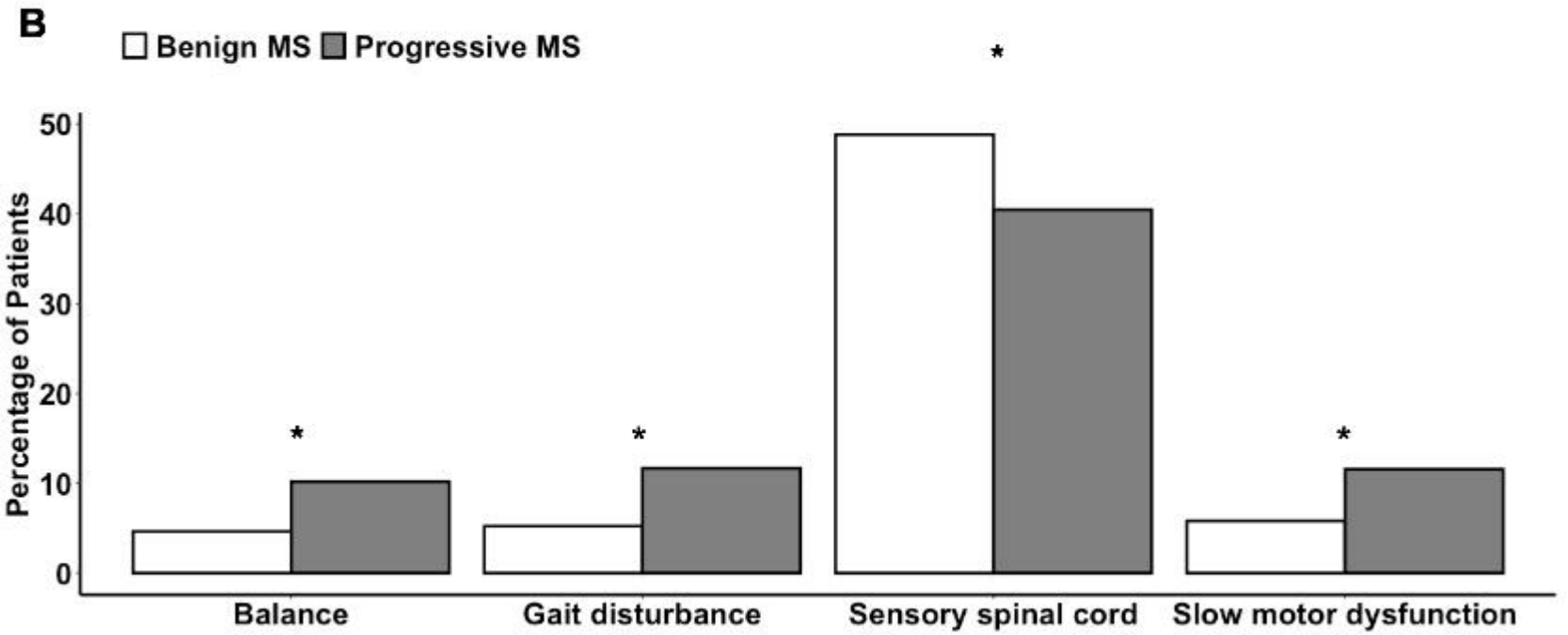

BC
AB
SK
MB
ON
QC
NL
NB
NS
Population
551,792 to 840,822
840,822 to 1,014,969
1,014,969 to 2,380,240
2,380,240 to 5,090,218
5,090,218 to 10,753,573
%Total Probands
5
10
15
25
35

B


Percentage of Probands per Population
0.025
0.020
0.015
0.010
0.005
0.000
AB
BC
MB
NB
NL
NS
ON
QC
SK


A
***
10.0
7.5
5.0
2.5
0.0
gARMSS
F
M
Sex
F
M
B
1500
1000
500
0
Frequency
5.0
5.5
6.0
Mean gARMSS
F
M


A
gARMSS < 5
gARMSS ≥ 5
Percentage of Patients
50
40
30
20
10
0
*
**
**
Gait disturbance
Sensory spinal cord
Slow motor dysfunction


B
Benign MS
Progressive MS
Percentage of Patients
50
40
30
20
10
0
*
*
*
*
Balance
Gait disturbance
Sensory spinal cord
Slow motor dysfunction