RESEARCH ARTICLE

Genetic and Environmental Susceptibility to Multiple Sclerosis

Author

Douglas S. Goodin, MD

MS Center at the University of California, San Francisco, 675 Nelson Rising Lane, Suite 221D. San Francisco, CA 94158

Phone: (415) 514-2464

Fax : (415) 514-2470

E-mail: douglas.goodin@ucsf.edu

Abstract

OBJECTIVE: To explore the nature and basis of environmental and genetic susceptibility to multiple sclerosis (MS).

BACKGROUND Susceptibility to multiple sclerosis (MS) is complex but clearly involves both environmental events and genetic factors. Certain epidemiological observations regarding MS (e.g., proportion of women among MS patients, population-prevalence of MS, impact of birth-month and migration patterns on the likelihood of MS, recurrence-risks for MS in siblings and twins, and time-dependent changes in MS-prevalence and the female to male sex-ratio) are well-established.

DESIGN/METHODS: We define the "genetically-susceptible" subset (*G*) to include everyone with any non-zero life-time chance of developing MS. Individuals who have no chance of developing MS, regardless of their environmental experiences, belong to the mutually exclusive "non-susceptible" subset (*G–*). We consider the implications that these well-established epidemiological observations have regarding the genetic and environmental basis of susceptibility to MS. In addition, we use the change in the female to male sex ratio, observed over a 35-year interval in Canada, to construct the response curves relating an increasing likelihood of MS to an increasing probability of a susceptible individual experiencing an environmental exposure sufficient to cause MS.

RESULTS: Environmental susceptibility to MS requires at least three different events – one occurring during the intrauterine or early post-natal period, another during childhood or adolescence, and a third (or more) many years later. Vitamin D deficiency and Epstein-Barr viral infections are likely involved. Moreover, we demonstrate that only a very small fraction of the general populations throughout Europe and North America is susceptible to MS. The vast majority of individuals in these populations has no chance whatsoever of developing MS, regardless of their environmental experiences. Even among carriers of the *HLA-DRB1*15:01~HLA-DQB1*06:02~a1* haplotype, only a small minority can possibly be members the (*G*) subset. Also, despite the preponderance of women among MS patients, compared to men, women are less likely to be susceptible and have a higher environmental threshold for developing MS. Nevertheless, the penetrance of MS in susceptible women is substantially greater than it is in men. Moreover, MS-probability in susceptible individuals increases with an increasing likelihood of a sufficient environmental exposure, especially among women. However, these response-curves plateau at under 50% for women and at a significantly lower level for men.

CONCLUSIONS: The pathogenesis of MS requires both a genetic predisposition and a suitable environmental exposure. Nevertheless, genetic-susceptibility is rare in the population and requires specific combinations of non-additive genetic risk-factors. By contrast, a sufficient environmental exposure (however many events are involved, whenever these events need to act, and whatever these events might be) is common, currently occurring in, at least, 76% of susceptible individuals. In addition, the environmental response-curves (especially in men) plateau well below 50%, which indicates that disease pathogenesis is partially stochastic.

Introduction

Multiple sclerosis (MS) is a recurrent inflammatory disease of the central nervous system and it is one of the most disabling diseases of young adults.[1] Episodic bouts of inflammation, which typically last days to weeks and occur unpredictably, cause injury to the myelin sheaths, to the oligodendrocytes, and in some cases, to the nerve cells and axons. In northern parts of Europe and the Americas the prevalence is between 0.1 and 0.25% of the population and the disease occurs predominantly in women. Thus, in most contemporary samples, women account for 65-75% of individuals with MS. In the large majority of cases the clinical-onset of disease occurs between the ages of 15 and 45 years. Nevertheless, now that magnetic resonance imaging (MRI) has become widely available, it is clear that the actual disease-onset can precede its first clinical manifestations by a decade or more and, in some cases, the clinical onset may never occur. For example, several pre-MRI autopsy studies reported that ~0.1% of individuals (without known symptoms during life) are found, incidentally, to have pathological evidence of MS at the time of death.[2-5]

Susceptibility to multiple sclerosis (MS) is complex but clearly involves both environmental events and genetic factors.[6-9] Considerable recent progress has been made in understanding both aspects of this susceptibility. On the genetic side, several genome-wide association screens (GWAS), which incorporate large arrays of single nucleotide polymorphisms (*SNPs*), have now identified many common MS-risk variants, located in scattered genomic regions, as being associated with MS.[10-15] For example, a recent GWAS from the International MS Genetics Consortium,[15] found 233 independent MS-associated *SNPs*, of which 32 were located within the major histocompatibility complex (*MHC*), and one was located on the *X*-chromosome. Most of these MS-associated *SNPs* are close to (or within) genes involved in the adaptive and innate arms of the immune system. Nevertheless, despite this recent increase in the number of genetic associations, the relationship of MS-susceptibility to the *HLA-DRB1*15:01~HLA-DQB1*06:02* haplotype of the human leukocyte antigens (*HLA*), inside the *MHC*, has been known for decades.[12,16-21]

Moreover, several well-established epidemiological parameters (e.g., the concordance rates in twins and siblings, the proportion of women among MS patients, the population prevalence of MS, the month-of birth for individuals who develop MS, and the time-dependent changes in MS-prevalence and in the female to male sex-ratio) have important implications with regard to the nature of both environmental and genetic susceptibility to MS. Importantly, each of these parameter values, at least theoretically, is directly observable for any population and, in actuality, have been observed in several population-based studies out of Canada.[22-26] It is the purpose of this manuscript, therefore, to review how these population-based epidemiological observations can be used to infer the values of non-observable parameters such as the population probability of being genetically susceptible, the likelihood that a susceptible person will actually develop MS, the proportion of susceptible individuals who are women, the timing, number, and nature of the environmental events necessary for MS pathogenesis, the likelihood that a susceptible individual will experience an environmental exposure sufficient to cause MS, and the probability that a susceptible individual who receives a sufficient environmental exposure will actually develop the disease.

Environmental Events in MS Pathogenesis

When considering environmental events involved in MS pathogenesis, it is convenient to divide an individual's environmental experiences into three time brackets – the intrauterine (*IU*) and early post-natal environments shared exclusively by twins, the familial micro-environment shared by siblings (including twins), and the remaining environments shared by the population generally. Notably, the impact of the familial micro-environment on MS-risk seems to be minimal. Thus, studies in conjugal couples, brothers and sisters of different birth order, adopted individuals, and in siblings and half-siblings raised together or apart, have generally indicated that MS-risk is unaffected by these micro-environmental influences.[22,23,25-29] If so, then the relevant environmental events for most MS cases act either at the shared IU environment of twins or at the population level.

Environmental Events near Birth The importance of the *IU* and early post-natal environments for MS pathogenesis is suggested by the so-called "maternal effect" in MS.[28] Support for such a "maternal effect" is provided by three independent lines of evidence. First, half-siblings, who are concordant for MS, are twice as likely to share the mother as they are to share the father.[22,28] This suggests that MS susceptibility is being transmitted from mother to child through something other than nuclear genes. An environmental event, occurring either in the *IU* period or soon thereafter, is one possibility. Once the child becomes independent of their mother, however, such a maternal effect would be unexpected for any environmental event.

Alternatively, such a maternal effect might result from mitochondrial inheritance, genetic-imprinting favoring expression of certain maternal genes, or other epigenetic factors. With regard to these other possibilities, however, there has been some speculation about the possibility of a so-called "Carter effect" in MS.[30,31] This hypothetical effect might occur if men were to be less susceptible to MS than women and if, as a result, men were to have more "potent" susceptibility genes when they actually develop the disease. In such a circumstance, paternal transmission of MS should be more common when the father's side is "genetically loaded" compared to maternal transmission when the mother's side is similarly "loaded". One report found weak evidence ($p = 0.032$) for such a "Carter effect" [30] whereas a larger study did not.[31] Neither study, however, provided evidence for the excessive maternal transmission expected if mitochondrial genes, genetic imprinting, or epigenetic factors were the basis of the "maternal effect" in MS.[28] By contrast, if an environmental event were responsible for this "maternal" effect, these studies would not demonstrate it because the *IU* and early post-natal environments are the same regardless of which parent transmits the MS-risk.

Second, the MS concordance rate for fraternal-twins consistently exceeds that for full siblings. For example, in a large population-based study from Canada,[24] the concordance rate for MS in full-siblings was 2.9% compared to a concordance-rate in dizygotic (*DZ*)-twins of 5.4%. Studies in other populations generally support the same conclusion.[32-34] Such a disparity cannot be attributed to mitochondrial inheritance, genetic

imprinting, or epigenetic factors because these factors should be similar for both siblings and *DZ*-twins sharing the same biological parents. Rather, this discrepancy must be due to environmental events occurring during the shared *IU* or in the early post-natal period.

The third line of evidence relates to the month-of-birth effect for MS, which has been reported in studies from Canada, northern Europe, and Australia.[35-39] Thus, combining patients from the northern hemisphere (Canada, Denmark, and Sweden), the peak MS-risk was for babies born in May and the nadir was for those born in November, compared to other months of the year.[35] Several other studies have also reported a similar birth-of-month effect in the northern hemisphere. [36-39] In the southern hemisphere, by contrast, this effect is reversed such that MS-risk is maximum in November/December and has its minimum in May/June.[38]

Some authors have suggested that this month-of-birth effect might be artifactual due to a failure to adjust properly for the place and year of birth.[40,41] However, in the Canadian study,[24] one of the control groups used consisted of unaffected siblings of the MS proband (which should correct for both of these confounders) and, in a study from Norway that specifically took these confounders into account, there still was a significantly increased MS-risk for babies born in April.[38] Moreover, a recent systematic review and meta-analysis concluded that the month and season of birth were significantly associated with MS.[39] Regardless, if this month-of-birth effect is genuine, this provides unequivocal evidence for an early environmental event, involved in MS pathogenesis that is time-locked both to the birth and to the solar cycle. This *circa annum* periodicity to MS susceptibility could be due to seasonal variations in maternal sun exposure (and therefore vitamin D_3 levels) while the child is *in utero.*[35] Alternatively, seasonal infections have a *circa annum* periodicity and might be account for such a month-of-birth effect. Nevertheless, because *intra-uterine* infections of the child by these organisms are uncommon, any association with seasonal infections would probably have to be a secondary phenomenon.

Environmental Events during Adolescence A second environmental event during adolescence is suggested by observations in people who relocate from one geographical region to another and who experience a different MS-risk compared to that in their home country.[42-47] For example, if an individual makes a relocation prior to their adolescent years from an area of high MS prevalence to an area of low prevalence (or *vice versa*), their MS-risk becomes similar to that of the region to which they relocated. By contrast, when they make the same relocation after this adolescent period, their MS-risk remains similar to that of the region from which they relocated. Moreover, the children of immigrants from low-prevalence areas who are born in "high-prevalence" regions have an MS-risk similar to their birth country rather than their country of ethnic origin.[44] These observations suggest an environmental event, involved in MS pathogenesis, that occurs sometime between childhood and the adolescent years.

Environmental Events during Adult Life Third, the initial clinical symptoms in MS are generally delayed considerably (often by decades) following the period when the maternal factor and the migratory factor take place. It is possible that these early environmental events, by themselves, are sufficient to cause MS although, in that case, the long delay between these events and the typical onset of clinical MS seems somewhat

difficult to rationalize. Consequently, it seems likely that subsequent environmental events are responsible for the timing of symptom onset.

Specific Environmental Events

Many potential environmental triggers for MS have been suggested over the years.[47] These suggestions have included trauma, stress, vaccinations, obesity, tobacco, vitamin deficiencies, low-sunlight, cosmic-rays, occupational hazards, living with domesticated animals, dietary habits, and toxic exposures. They have also included a variety of specific infections such as Epstein-Barr virus (EBV), human herpes virus 6, typhoid, smallpox, chicken pox, Chlamydia, and others. Of these potential environmental events, EBV infection, vitamin D_3 deficiency, tobacco, and obesity have attracted the greatest current interest as having a potential role in MS pathogenesis.[1] Nonetheless, several of these other factors continue to have strong proponents and no single factor has yet been proven conclusively to be either related or unrelated. Nevertheless, many of the proposed associations lack credible scientific evidence, biological plausibility, or both. Here we will focus our attention on the possible role that EBV and vitamin D_3 deficiency might play.

Epstein - Barr virus EBV is a DNA virus of the herpes family. It is a very common infection of humans, with over 90% of the population becoming infected over their lifetime.[48-59] In many parts of the world the initial EBV infection occurs during early childhood and is either asymptomatic or it produces non-specific symptoms indistinguishable from many other childhood illnesses. However, if the initial infection is delayed until adolescence or young adulthood, the syndrome of infectious mononucleosis (glandular fever) develops in 35 to 50% of cases. The viral infection seems to specifically target the epithelial cells of the oropharynx and the B-cells.

Following the initial lytic phase of the infection, a latent infection of B-cells by EBV ultimately predominates and, in these cells, the virus persists indefinitely. Periodically, EBV can become reactivated, resulting in further cell lysis and producing fresh viral particles. During the incubation period or early in the acute illness, antibodies to antigens associated with the process of viral replication, such as the viral capsid antigen (VCA) and the diffuse and restricted early antigens (EA), are found in the serum.[60] The antibodies to VCA persist for the lifetime of the individual. Antibodies to EA are generally taken as a sign of active infection although, in approximately 20-30% of patients, theses antibody titers persist for years. The EBV nuclear antigens (EBNA 1 to 5) are expressed in latently infected B-cells, and antibodies to these antigens also persist for the lifetime of the individual.

Table 1. Antibodies to EBV in the sera of MS cases and controls.

Study	EBV+ MS Cases (%)	EBV+ Controls (%)	*p-value*
Sumaya, [48] ‡	155/157 (98.7%)	76/81 (93.8%)	0.05
Bray, [92] ‡	309/313 (98.7%)	363/406 (89.4%)	0.0001
Larson, [92] ‡	93/93 (100%)	78/93 (83.9%)	0.0001
Sumaya, [49] *	104/104 (100%)	23/26 (88.5%)	0.007
Shirodaria, [92] ‡‡	26/26 (100%)	24/26 (92.3%)	-
Munch, [92] †	137/138 (99.3%)	124/138 (89.9%)	0.0004
Myhr, [92] *	144/144 (100%)	162/170 (95.3%)	0.008
Wagner, [92] †	107/107 (100%)	153/163 (93.9%)	0.01
Ascherio, [50] ††	143/144 (99.3%)	269/287 (93.7%)	0.008
Sundström, [52]	234/234 (100%)	693/702 (98.7%)	ns
Haahr, [53] †	153/153 (100%)	50/53 (94.3%)	0.05
Ponsonby, [54] ‡‡	136/136 (100%)	252/261 (96.6%)	0.05
Abrahamyan, [59] ‡‡	610/610 (100%)	4134/4343 (95.2%)	0.0001
Total	**2351/2359 (99.7%)**	**6401/6749 (94.8%)**	**$p < 10^{-25}$**

* Study measured antibodies against the Epstein-Barr nuclear antigens (EBNA), the viral capsid antigen (VCA), and the early antigens (EA).
‡ Study measured antibodies only against VCA
† Study measured antibodies only against EBNA and EA
†† Study measured antibodies only against EBNA and VCA. One person was antibody negative to each antigen but it is unclear from the text whether they were the same person. The review by Haahr [53] suggests they were not.
‡‡ Study measured antibodies only against EBNA and VCA

EBV infection has been consistently linked to MS, especially when it causes symptomatic mononucleosis.[48-59] Indeed, the evidence a prior EBV infection in adult-onset MS is present in essentially 100% of cases and the odds ratio (*OR*) for cases compared to controls is highly significant (*Table 1*). Even in those rare MS patients who test negatively for prior exposure to EBV, this finding could easily be a false negative result because, in every such case, the antibody response wasn't measured against the entire set of EBV antigens (*Table 1*). Also, the prior nature of the EBV infection is supported both by the presence of IgG (not IgM) antibodies to EBV antigens and by the unequivocal evidence (when it has been assessed) of infection years prior the onset of clinical symptoms.[48-59] In this context, the word "prior" is being used to mean before the clinical-onset of MS, which, as discussed earlier, may follow the actual disease-onset by many years.

Moreover, this ~100% prevalence of EBV antibodies in adult onset MS cannot be ascribed to either false negative tests in the general population or false positive tests in MS patients. Also, it cannot be ascribed a general "hyper-immune" state in MS patients because

the antibody responses in MS patients to other common pathogens (e.g., measles, mumps, chicken pox, herpes simplex, cytomegalovirus, etc.) are not similarly increased.[1,48,50,51] Therefore, the ~100% association of MS with a "prior" EBV infection (if correct) seems to indicate that EBV is a necessary (but not a sufficient) condition for adult MS to develop and, if so, EBV must be a part of the causal pathway leading to MS.

Likely, however, EBV is not the factor responsible for the "maternal effect" discussed earlier because EBV infection does not occur either *in utero* or during the early post-natal period. Moreover, because of the association of MS with late EBV infection and with mononucleosis, it seems likely that EBV acts during late childhood or adolescence and, thus, would be a better candidate for the second environmental event. Regardless, however, it seems clear that EBV infection plays some role in MS pathogenesis.

Vitamin D Deficiency The production of active vitamin D requires the two-step conversion of 7-dehydro-cholesterol into active vitamin D.[61-64] The first step is conversion into vitamin D_3, which is catalyzed by the exposure of 7-dehydro-cholesterol in the skin to ultraviolet B (UVB) radiation. Subsequently, vitamin D_3 is hydroxylated to form active vitamin D in the tissues. The dietary intake of vitamin D_3 can circumvent the UVB-dependent part of this pathway and, thus, maintain normal vitamin D_3 serum levels in the absence of UVB radiation. Nevertheless, vitamin D_3 is found in only a few natural dietary sources such as oily fishes and reindeer. Interestingly, two human populations with a notably low MS-risk [65-68] are the Inuit or Eskimos (who consume large quantities of oily fish) and the Sami or Lapps (who eat reindeer meat regularly). Other human populations, by contrast, require sufficient exposure of the skin to UVB radiation in order to maintain adequate vitamin D_3 serum levels throughout the year. Biologically, vitamin D acts (together with its receptor and the retinoid X receptor) as a transcription factor that controls the expression of thousands of nuclear genes throughout the genome. Notably, one of these vitamin D regulated genes is the MS-associated *HLA-DRB1*15:01* allele discussed earlier.[69] Moreover, the critical importance of vitamin D to human health is suggested by the fact that, in temperate regions of the earth, the prevalence of lighter skin tones (in diverse ethnic groups) is thought to reflect a convergent evolutionary adaptation to needing adequate vitamin D_3 in these areas.[70,71]

As latitude increases (both north and south of the equator), the amount of UVB radiation reaching the Earth's surface is reduced and adequate the UVB exposure necessary for vitamin D_3 synthesis may not be unavailable for some (or many) months of the year. For example, it has been estimated that the level of UVB radiation at the US-Canadian border during most months of the year is insufficient to produce an adequate amount of vitamin D_3.[70-73] Moreover, maps of UVB availability around the world are strikingly similar to comparable maps of MS prevalence.[1,70,71]

Vitamin D_3 is important for the maturation of the immune system and for a variety of immune functions including cell proliferation, differentiation, and immunomodulation and, moreover, the vitamin D_3 receptor (VDR) is expressed on cells throughout the body, including activated T and B cells and on macrophages.[63,74-78] In addition, vitamin D_3 deficiency seems to play a role in the pathogenesis of several autoimmune diseases.[74,75,77] With this background, there have been several studies, which have explored, more directly,

the possible relationship of vitamin D_3 deficiency to MS and these studies have provide some support for the relationship of vitamin D_3 deficiency in either childhood or adolescence and MS.[79-82]

Vitamin D_3 deficiency would, consequently, seem to be a good candidate for the "maternal" factor in MS pathogenesis discussed earlier. Vitamin D_3 levels are coupled to the solar cycle, it is involved in maturation of the immune system, its deficiency has been associated with other autoimmune disorders, its the world-wide distribution of MS mirrors that for reduced UVB radiation around the globe, and extreme northern populations with high dietary intake of vitamin D_3 have a low MS-prevalence. However, regardless of any connection with the "maternal" factor in MS pathogenesis, vitamin D_3 deficiency could also act during childhood, during adolescence, later in life, or even at multiple different times. Indeed, the direct data supporting a role for vitamin D_3 in MS actually suggests that there may be an impact during childhood or adolescence.[79-82]

Changing Environmental Exposures

MS epidemiology has changed over the past several decades. Thus, the prevalence of MS seems to be increasing, especially among women.[26,83-90] As a result of this change, the female to male (*F:M*) sex ratio for MS in Canada has increased during every 5-year increment except one between 1941-1980.[26] Over the entire interval, the ratio has increased from 2.2 in (1941-1945) to 3.2 in (1976-1980). These changes seem far too rapid to be genetically based. It is conceivable, however, that this observed *F:M* sex-ratio change might be artifactual. For example, if women were more likely than men to have minimally symptomatic MS, then, now that these patients are being diagnosed by our improved imaging and laboratory methods, women might represent a disproportionate number of these newly diagnosed cases. Alternatively, in previous times, vague symptoms of MS in women may have been written off as "non-organic" more often than they were in men. Nevertheless, four lines of evidence argue strongly against this change being artifactual. First, this increase in the sex ratio began before, and continued up to, the advent of modern imaging and laboratory methods.[26] Second, among asymptomatic individuals, incidentally, found to have MS by MRI, the *F:M* sex ratio is approximately the same as current estimates for symptomatic MS and 80% of the those with spinal cord lesions are women − i.e., those lesions having the greatest odds for progression to "clinical" MS.[61] Third, the increasing prevalence among women has been observed world-wide.[26,83-90] And finally, the greater penetrance of MS in women is confirmed independently by the *MZ*-twin data (*see below*). Therefore, the observed change in the *F:M* sex ratio seems, almost certainly, to reflect a change in the environmental conditions related to MS pathogenesis.

Although many wide-spread environmental changes are known to be taking place (e.g., increasing atmospheric concentrations of CO_2, CH_4, and other pollutants; increasing global temperatures; a depletion of stratospheric ozone; a greater dietary consumption of trans-fats and processed foods, etc.), one recent change (relevant to a possible role for vitamin D_3 deficiency) is that people are increasingly encouraged to use either sun-avoidance or sun-block as a means of preventing skin cancers.[91] Notably, sun-block with sun-protective-factor (SPF)-15 blocks ~94% of the incoming UVB radiation and higher SPF levels block even more.[91] As a result, any wide-spread use of sun-block and/or sun-avoidance will exacerbate any population deficiency of vitamin D_3 synthesis and will likely increase the occurrence of diseases related to it.

In summary, the current epidemiological evidence seems to support the existence of three (or more) environmental events that contribute to MS pathogenesis. The first event occurs near birth, the second occurs during childhood or adolescence, and the third (or more) occurs long after the first two have already taken place. At present, the two best candidate factors identified are vitamin D_3 deficiency and EBV infection. Indeed, as discussed above, these two factors seem particularly well-suited to the first two environmental-events in MS pathogenesis. Nevertheless, even if EBV infection and vitamin D_3 deficiency are part of *some* pathway leading to adult MS, they need not be on the *same* or the *only* pathway. Indeed, these two environmental-events might interact in several possible ways to cause MS.[1,92] No pathway can be excluded entirely although, if a prior EBV infection is a necessary (but not a sufficient) condition for MS to develop (*see above*), this suggests that these two events must act sequentially to form part of the environmental cascade, which leads to adult MS.[1,92]

Genetic Factors in MS Pathogenesis

The risk of developing MS for individuals who have an affected family member increases in rough proportion to the amount of shared genetic-information between themselves and the proband.[22,23,31-35,47,93,94] Thus, for example, siblings of an MS proband (50% genetic similarity) have a 20-30 fold increased risk compared to the general population whereas *MZ*-twins (100% genetic similarity) have a risk ~10 times greater and cousins (25% genetic similarity) have a risk ~5 times less than the MS-risk in siblings.[24,32 93,95-98] These observations, by themselves, unequivocally, implicate genetic factors as playing an important role in the pathogenesis of MS.

Indeed, as noted earlier, there have now been 233 independent genomic locations (many within or near immune-related genes) that are associated with MS.[15] Of particular interest for many years has been the association of MS with certain alleles within the MHC. Typically, these studies have focused on establishing the relationship between genetic susceptibility and specific alleles at specific *HLA* loci. In individuals of European descent, it has long been known that there is an increased MS-risk associated with carrying either *HLA-DRB1*15:01* or *HLA-DRB1*03:01* alleles and that there is a "protective" effect of carrying the *HLA-A*02:01* allele.[12,16-21] For example, in the large Wellcome Trust Case Control Consortium (WTCCC) dataset,[14,99] the odds ratio (*OR*) of MS for individuals possessing one or more of these alleles is highly significant – for *HLA-DRB1*15:01* ($OR=3.24;\ p<<10^{-300}$); for *HLA-DRB1*03:01* ($OR=1.27;\ p<10^{-11}$); and for *HLA-A*02:01* ($OR=0.69;\ p<10^{-53}$).

Despite this focus on single alleles of specific genes, however, these *HLA* alleles don't really exist in isolation. Thus, within the *MHC* region, most *HLA* alleles are in tight linkage disequilibrium with each other and, overall, the *HLA* region consists of a relatively small collection of highly conserved extended haplotypes (*CEHs*), which stretch (at least) across the "classical" *HLA* genes (*HLA-A*, *HLA-C*, *HLA-B*, *HLA-DRB1*, and *HLA-DQB1*) – a distance spanning nearly 3 *mb* of DNA.[99-101] For example, in the predominantly European WTCCC, the most frequent 250 *CEHs* accounted for 57% of all *CEHs* present.[100-101] This haplotypic structure is found in all human populations.[101] Nevertheless, the actual *CEH* compositions, which account for this population structure, are markedly divergent from one region to the next.[99-101] Thus, it seems that these *CEHs* are under a

strong selection pressure and, presumably, such divergence is due to specific environmental and/or biological pressures that vary with time, with geographic location, or with both.[100-101]

In the *HLA* Class II region, this linkage disequilibrium is especially strong between (at least) the *HLA-DRB1* and *HLA-DQB1* loci. For example, in the predominantly European data from the WTCCC, 97.5% of the *HLA-DRB1*15:01* alleles (the most common *DRB1* allele in Europeans; control frequency=13.0%) are linked to the *HLA-DQB1*06:02* allele. Similarly, 98.4% of the *HLA-DRB1*03:01* alleles (control frequency=11.8%) are linked to the *HLA-DQB1*02:01* allele. Similar tight linkages are found for most other *DRB1~DQB1* combinations.[100] In addition, we have described a collection of *SNP*-haplotypes that are composed of unique combinations of the *SNPs* (*rs2395173*; *rs2395174*; *rs3129871*; *rs7192*; *rs3129890*; *rs9268832*; *rs532098*; *rs17533090*; *rs2187668*; *rs1063355*; and *rs9275141*), and which span 0.25 *mb* of DNA surrounding the *HLA-DRB1* locus.[99-101] Ten of these *SNPs* are within intergenic regions whereas *rs1063355* is within exon 5 of the *DQB1* gene. One such 11-*SNP* haplotype (*a1*) adds further specificity to the *HLA-DRB1*15:01~HLA-DQB1*06:02* haplotype.[99-101] Thus, 99% of (*a1*) *SNP*-haplotypes carry the *HLA-DRB1*15:01~HLA-DQB1*06:02* haplotype and, conversely, 99% of these *HLA*-haplotypes carry the (*a1*) *SNP*-haplotype.[100] This (*DRB1*15:01~DQB1*06:02~a1*) haplotype is referred to as the (*H+*) haplotype. Nevertheless, because, in the WTCCC, 93.4% of *HLA-DRB1*15* alleles are actually the *HLA-DRB1*15:01* allele, and because 99% of *HLA-DRB1*15:01* carriers also carry the full (*H+*) haplotype, each of these designations will be used interchangeably as (*H+*).[100]

Regardless of such strong linkage disequilibrium in the Class II region, however, there are nuances to susceptibility that accrues because of the *CEH* structure. For example, in persons of European descent, the Class II *HLA-DRB1*03:01~ HLA-DQB1*02:01* haplotype comes in two forms. The first (present in 84% of the WTCCC controls) is coupled to the (*a6*) *SNP*-haplotype and the second (present in 15% of the WTCCC controls) is coupled to the (*a2*) *SNP*-haplotype.[100] Each form has a distinct relationship to susceptibility. For (*a2*) carriers, among non-(*H+*)-carrying individuals, a single copy is consistently associated with an increased MS-risk.[27] By contrast, for (*a6*) carriers, the risk associated with carrying a single copy varies from being "risky" to being "protective" depending upon the Class I portion of the *CEH* being considered.[100] Similarly, all carriers of the (*H+*) haplotype have an increased MS-risk, although the degree of association varies depending upon the *CEH* involved.[100] By contrast, some *HLA-DRB1*15:01~ HLA-DQB1*06:02* haplotypes that don't also carry the (*a1*) *SNP*-haplotype, seem not to be associated with any MS-risk.[100] And, finally, although the *HLA-A*02:01* allele is "protective" when considered as a single allele, some of the *CEHs* on which this allele is present seem to have little impact on MS-risk whereas on other *CEHs* this allele seems to have a "protective" effect .[100,101]

Given both this strong linkage disequilibrium within the Class II region, in addition to the superimposed the *CEH* structure of the *MHC*, it is unclear what gene (or genes) within a "risk" haplotype is responsible for the associations with MS-susceptibility that are observed. Similar concerns apply to all of the 233 genetic associations that have been reported [15] and it is, thus, unclear what constitutes the basis of susceptibility to MS. This is the topic considered in the following section, the detailed mathematical development of which is available in an earlier publication.[102]

Genetic and Environmental Susceptibility to MS

Despite this undoubted importance of genetic factors and environmental events in MS-pathogenesis, susceptibility to MS might be envisioned in number of different ways. In order to highlight some issues that might be involved in MS pathogenesis, we can consider, as examples, disease states for which we understand (or think we understand) the underlying pathophysiology.

The first is sickle cell disease (*SCD*), which occurs in ~3% of individuals in certain sub-Saharan regions of Africa.[103] All individuals with *SCD* are homozygous for the *HbS* mutation of the hemoglobin gene. Even though certain environmental events (e.g., high-altitude, infection, strenuous exercise, and dehydration) can impact the clinical expression of *SCD*, fundamentally, *SCD i*s thought of as a genetic disorder.

The second is the flu, which affects 5−20% of the population in North America each year.[103] Although one person may be more or less susceptible than another to a particular year's variant given their genetic make-up, presumably, everyone could become sick if they had a sufficient exposure to the influenza virus. Thus, despite the possible genetic differences in susceptibility, fundamentally, the flu is an environmental (infectious) disease.

The third is breast cancer, for which the life-time probability in the US is ~12.5% in women and ~0.1% in men. Individuals who have the *BRCA1* or *BRCA2* mutations (<1% of the population) have a risk of breast cancer 4-7 times that in the general population.[103] Nevertheless, there is likely a baseline risk of breast cancer such that no one is completely risk-free. Although the genetic make-up (including gender) influences the baseline risk and the environment likely affects the penetrance of the *BRCA* mutations, fundamentally, some breast cancer cases are genetic and others are fundamentally environmental (possibly due to exposures such as by radiation, toxins, pregnancy, or other occurrences).

The fourth is infection by the human immunodeficiency virus (*HIV*). Anyone in the population can acquire this virus although individuals who engage in high-risk behaviors (e.g., unprotected anal-receptive sex or intravenous drug use and needle-sharing) are particularly vulnerable. Among persons of northern European decent, ~1% are homozygous for the *Δ-32 mutation* of the *CCR5* gene and these individuals are almost completely resistant to *HIV* infection.[103] Consequently, fundamentally, *HIV* is an environmental disease (infectious) with an interaction between environmental factors (i.e., the virus and specific high-risk behaviors). However, certain genetic traits (*e.g., the Δ-32 mutation*) can be decisive in determining the degree of susceptibility.

Whether MS-susceptibility resembles any of these disease-states (or some other) is unknown although several basic epidemiological observations in MS bear directly on the different possibilities. In this section, we utilize directly observable, and well-established, "population parameters" (e.g., the concordance rates in twins and siblings, the proportion of women among MS patients, the population prevalence of MS, the time-dependent changes in the sex-ratio, etc.) to logically infer the values of other non-observable parameters of interest (e.g., the population probability of being genetically susceptible, the likelihood that a susceptible person actually develops MS, the proportion of susceptible individuals who are women, the likelihood that a susceptible individual experiences a sufficient environmental exposure, etc.).

Methods

For the purposes of this section, we will define five parameters. The first, *P*(*MS*), is the expected life-time probability that an individual from the general population, selected at random, will develop MS. This parameter is the expected penetrance of MS.

The second, *P*(*G*), is the expected probability that an individual from the general population is also a member of the (*G*) subset. We define the (*G*) subset, in turn, to include everyone who has any non-zero chance of developing MS (i.e., regardless of how small that risk might be). Everyone who is not a member of the (G) subset is, by definition, a member of the mutually exclusive $(G\text{-})$ subset, consisting of non-susceptible individuals, who have no chance, whatsoever, of getting MS, regardless of the environmental exposures that they experience during their life-times. The subset (G) can also be partitioned into two mutually exclusive sub-subsets, (*G*1) and (*G*2), suitably defined, such that the sub-subset (*G*1) has an expected penetrance greater than that for (*G*2). If the expected penetrance is statistically different between these two sub-subsets, our analysis will be restricted to those circumstances, in which both sub-subsets, (*G1*) and (*G2*), considered separately, each has a distribution of penetrance values that conforms to the Upper Solution (*see #4 below*).

The third, *P*(*E*), is the probability that a member of the (*G*) subset will experience an environmental exposure, sufficient to cause MS, given the environmental conditions of the time (whatever these conditions might be). By this definition, everyone who ultimately develops MS must have had a sufficient environmental exposure, even for those individuals who have a "purely genetic" form of MS (i.e., those for whom any environmental exposure is sufficient).

The fourth is a set of terms, $P(MS \mid MZ_{MS})$, $P(MS \mid DZ_{MS})$, and $P(MS \mid S_{MS})$. The first two, $P(MS \mid MZ_{MS})$ and $P(MS \mid DZ_{MS})$, are the expected life-time probability of developing MS for a person who is part of either a monozygotic or a dizygotic twin-ship, given that their co-twin either has or will develop MS. These probabilities are estimated by the observed proband-wise concordance rate for either *MZ*-twins or *DZ*-twins.[104] The last, $P(MS \mid S_{MS})$, is the expected life-time probability of developing MS for a sibling (*S*), given the fact that their co-sibling either has or will develop MS.

The final term, $P(MS \mid IG_{MS})$, is the adjusted proband-wise concordance rate for *MZ*-twins. Such an adjustment may be necessary because concordant *MZ*-twins, in addition to sharing identical genotypes (*IG*), also share their intrauterine (*IU*) and certain other (particularly early) post-natal environments. Thus, perhaps, these environments, shared by *MZ*-twins, might similarly impact the likelihood of developing MS in the future for both individuals. One method to adjust for this possibility is to consider the difference in concordance rates between non-twin siblings and fraternal twins (i.e., siblings who have the same genetic relationship with each other but who are divergent in their *IU* and early environmental experiences).[102]

From these parameters, using the epidemiological data from Canada circa 2000–2010 (*Table 2*), we can logically estimate the value of the another, non-observable, parameter, $P(MS \mid G)$, which is the conditional life-time probability of developing MS for a member of the (*G*) subset. This parameter is the expected penetrance for the (*G*) subset. Clearly, by the definition of the (*G*) subset (*above*), everyone who actually develops MS

during their life-time must be a member of this subset. From this observation, and from the definition of conditional probability:

$$P(MS \mid G) = P(MS, G) / P(G) = P(MS) / P(G)$$

This equation can be re-arranged to yield: $P(G) = P(MS) / P(MS \mid G)$

Once the value of $P(G)$ is established, this can then be used to assess the nature of MS pathogenesis. For example, if: $P(G) = 1$, then everyone can develop MS under the right environmental circumstances and, from this, we would conclude that MS must be caused, at least in some cases, by "purely environmental" factors (*e.g., flu, HIV, breast cancer*). Naturally, any such a conclusion does not preclude the possibility that genetic factors also have a significant impact upon the likelihood of disease (*e.g., HIV, breast cancer*).

By contrast, if $\{P(G) < 1\}$, then the development of MS is possible only for certain individuals (*e.g., SCD*) and, therefore, we would conclude that MS must be a genetic disorder (i.e., unless someone has the proper genetic constitution, they have no chance of getting the disease, regardless of their environmental exposures). Naturally, again, any such a conclusion does not preclude the possibility that disease pathogenesis also requires the co-occurrence of specific environmental events. In addition, how we characterize genetic susceptibility, will depend upon the degree to which $P(G)$ is less the unity and upon the magnitude of any differences between the "high" and "low" penetrance subgroups. For example, in *HIV*, if homozygous *Δ-32* mutations protected an individual completely from disease, then: $P(G) = 0.99$. In this circumstance, however, we would probably characterize *HIV* as fundamentally environmental and homozygous *Δ-32* mutations as "protective" rather than characterizing every non-homozygous individual as "susceptible". By contrast, in *SCD*, where: $P(G) = 0.03$, we would consider homozygous *HbS* mutations as the defining trait for (G) subset membership. Even if it were possible, in extremely rare circumstances, for a non-homozygous individual to develop *SCD*, we would probably still characterize *SCD* as a fundamentally genetic disorder.

1. MS Penetrance – *P*(*MS*)

There are three methods available for estimating $P(MS)$. The first is to use the observed population prevalence. Taking into account the fact the clinical-onset of MS almost always occurs between the ages of 15 and 45 years, leads to the conclusion $P(MS)$ is approximately twice the population prevalence.[102] In the northern Europe and the Americas, most prevalence estimates are between 100 and 250 cases per 100,000 population or 0.1–0.25% so that, by this method, we would estimate:

$$P(MS) \approx 0.002 - 0.005$$

A second method is to measure MS prevalence within the age-band of 45-55 years. In this age-band, most MS patients will have already experienced their clinical onset and few will have experienced their expected excessive mortality. Therefore, the MS prevalence in this age-band should estimate the penetrance of MS.[102] Using published estimates of MS prevalence within this age-band from Sweden [88] and the US [105] leads to the estimate that:

$$P(MS) \approx 0.003 - 0.0034$$

A third method is to use population-based death data. Because, by the time of death, every case of clinically-evident MS must have already declared itself, we can equate MS penetrance with the percentage of death certificates that mention the diagnosis of MS.[102] Using this data from a population-based study out of Canada,[106] leads to the estimate that:

$$P(MS) \approx 0.0028$$

Thus, all three of these methods of estimation are quite consistent with each other and each lends support to the conclusion that, in the northern parts of Europe and the Americas:

$$P(MS) \approx 0.003$$

2. MS Penetrance among Women and Men – $P(MS \mid F)$ & $P(MS \mid M)$

The proportion of women among MS patients in the Canadian twin dataset (*Table 2*) is 66%.[24] In the WTCCC dataset this proportion is 72%.[102] In the study of Orton and colleagues [26] out of Canada, in the most recent epoch, the proportion of women among MS patients is 76%. In a recent estimate from the United States, the proportion of women among MS patients is 74%.[105] To determine these penetrance values, we can use the relationship that:

$$P(MS \mid F) = P(F \mid MS) * P(MS) / P(F)$$

and the Canadian data from *Table 2*: $P(F) = P(M) = 0.5$. In this case, it follows directly from *#1* (*above*) that:

$$P(MS \mid F) = P(F \mid MS) * (0.003 / 0.5) = 0.006 * P(F \mid MS)$$

Similarly: $P(MS \mid M) = P(M \mid MS) * (0.003 / 0.5) = 0.006 * P(M \mid MS)$

Consequently: $P(MS|F) \geq (0.66/0.34) * P(MS|M) = 1.94 * P(MS|M)$

Table 2. Epidemiological Data for Multiple Sclerosis in Canada circa 2000 –2010 *

Population Data

$P(H+) = 0.24$ $P(F) = P(M) = 0.5$

Family Data

$P(MS \mid MZ_{MS}) = 0.253$ $P(MS \mid DZ_{MS}) = 0.054$

$P(MS \mid S_{MS}) = 20 / 692 = 0.029$

Gender Data

$P(F \mid MS) = P(F \mid MZ_{MS}) = 88 / 133 = 0.66$

$P(F \mid MZ_{MS}) / P(F \mid MS) = 0.92 / 0.66 = 1.39$

$P(F \mid MS) / P(F) = 0.66 / 0.5 = 1.32$

$P(F \mid MS, MZ_{MS}) = 22 / 24 = 0.92$

$P(MS \mid F, MZ_{MS}) = 0.34$

$P(MS \mid M, MZ_{MS}) = 0.067$

HLA-DRB1*15 (H+) Data

$P(H+ \mid MS) = P(H+ \mid MZ_{MS}) = 40 / 93 = 0.43$

$P(H+ \mid MS) / P(H+) = 0.43 / 0.24 = 1.79$

$P(H+ \mid MS, MZ_{MS}) / P(H+ \mid MS) = 0.45 / 0.43 = 1.05$

$P(H+ \mid MS, MZ_{MS}) = 9 / 20 = 0.45$

$P(MS \mid H+, MZ_{MS}) = 0.31$

$P(MS \mid H-, MZ_{MS}) = 0.29$

Sex Ratio Data

Time Period (*#1*) -- 1941–1945: $P(F \mid MS)_1 / P(M \mid MS)_1 = 2.2$

Time Period (*#2*) -- 1976–1980: $P(F \mid MS)_2 / P(M \mid MS)_2 = 3.2$

*The value for $P(H+)$ – *see Text* for the definition of the $(H+)$ haplotype – was provided by Dessa Sadovnick, was based on 400 Canadian controls, and the rate was confirmed in a large transplant database (*personal communication*). The *F:M* sex-ratio in the general population of Canada was taken from the 2010 Canadian census. Recurrence risks for monozygotic (*MZ*) twins, dizygotic (*DZ*) twins, siblings (*S*) and the other summary data were taken from the study of Willer et al.[24] The *F:M* sex-ratio among Canadian MS patients at each of the 5-year time-periods (1941–1945 & 1976–1980) was taken from the study of Orton et al.[26] {*NB: By the definition of subset (G), in all circumstances:*

$$P(MS \mid MZ_{MS}) = P(MS \mid G, MZ_{MS}) = P(MS, G \mid MZ_{MS})\}$$

3. Adjusting for the Shared *IU* Environment of *MZ*-twins – $P(MS \mid IG_{MS})$

Using the Canadian population-based data (*Table 2*) on the recurrence risks in non-twin siblings, *DZ*-twins, and *MZ*-twins (concordance rates for siblings=2.9%; concordance rates for *DZ*-twins=5.4%; concordance rate for *MZ*-twins=25%) to make this adjustment [102] leads to the estimate of:

$$P(MS \mid IG_{MS}) = (2.9/5.4) * 0.25 = 0.134$$

4. MS Penetrance in Susceptible Persons – $P(MS \mid G)$

We define the set $\{X\}$ to include the expected penetrance of every member of the subset (G). In this case, for notational simplicity, we can define the following terms:

$$x = P(MS \mid G); \quad x' = P(MS \mid IG_{MS}); \quad \text{and:} \quad \sigma_X^2 = Var(X)$$

Using these definitions, it can be shown [102] that:

$$x^2 - (x')x + \sigma_X^2 = 0$$

which is a quadratic equation solved by:

$$x = \frac{(x') \pm \sqrt{(x')^2 - 4\sigma_X^2}}{2}$$

This last equation has real solutions only when the variance (σ_X^2) range is restricted such that:

$$0 \le \sigma_X^2 \le (x'/2)^2$$

Moreover, this maximum variance, $(x'/2)^2$, occurs when the distribution of penetrance values in the set $\{X\}$ is bimodal,[107,108] such that half the (G) subset has a penetrance of (0) and the other half has a penetrance of (x'). From this point of maximum variance, the variance of the $\{X\}$ subset decreases both when:

$$x \to x' \quad \text{and:} \quad x > x'/2 \qquad (\textit{the Upper Solution})$$

and when: $x \to 0$ and: $x < x'/2$ (*the Lower Solution*)

By definition, every member of (G) has an expected penetrance greater than zero. Therefore, the Upper Solution limits become: $x'/2 < x \le x'$

And the Lower Solution limits become: $0 < x < x'/2$

Moreover because: $x' = x + \sigma_X^2/x$. Therefore, if: $\sigma_X^2 = 0$; then: $x' = x$

The Upper Solution, as: $(x \to x')$, reflects the gradual transition from the bimodal distribution (*described above*) to a unimodal distribution and, finally, to a distribution where every genotype in (G) has exactly the same penetrance (i.e., $x = x'$). By contrast, the Lower Solution as: $(x \to 0)$, reflects an increasingly assymetric, non-unimodal, distribution of penetrance values within (G).

5. MS Penetrance in Susceptible Women and Men – $P(MS \mid G,F)$ & $P(MS \mid G,M)$

The set $\{X\}$ of penetrance values for members of the (G) subset is, at least, bimodal. Thus, from the *MZ*-twin data (*Table 2*) out of Canada:

$$P(MS \mid F, MZ_{MS}) = 0.34 >> 0.067 = P(MS \mid M, MZ_{MS})$$
$$\chi^2 = 8.5; \quad p = 0.0035$$

Consequently these sub-subsets of women (F) and men (M) have significantly different expected pentrances. Therefore each, considered separately, are assumed to follow the Upper Solution (*see Methods & #4 above*). Adjusting for the similar *IU* environment of *MZ*-twins (*see #3, above*), it follows that:

$$0.093 < P(MS|F,G) \leq 0.187$$
and: $$0.017 < P(MS|M,G) \leq 0.034$$

These ranges for men and women don't overlap, which indicates that susceptible women must have a greater MS-penetrance than susceptible men.

6. Genetic Susceptibility in Women and Men – $P(G \mid F)$ & $P(G \mid M)$

From the relationship derived in the *Methods* (*above*), it follows that:

$$P(G \mid F) = P(MS \mid F) / P(MS \mid F,G)$$
and: $$P(G \mid M) = P(MS \mid M) / P(MS \mid M,G)$$

From *#2 & #5* (*above*) and using the *MZ*-twin data from Canada (*Table 2*), it follows that:

$$0.021 = (0.006 * 0.66)/0.187 \leq P(G|F) < (0.006 * 0.66)/0.093 = 0.043$$
and: $$0.06 = (0.006 * 34)/0.034 \leq P(G|M) < (0.006 * 0.34)/0.017 = 0.12$$

Again, these ranges don't overlap so that men are more likely to be susceptible than women. If our estimate for the proportion of women among MS patients were increased to 73%, these ranges would just barely overlap. Although this percentage is certainly possible (*see #2 above*), four lines of evidence support the conclusion that, even in such a circumstance, men are still more likely than women to be members of the susceptibile subset (G). First, it seems inappropriate to use the *MZ*-twin dataset (*Table 2*) to estimate the twin concordance rates but to use a different dataset to estimate the proportion of women among MS patients. Second, in making the above calculation, we are positing an extreme and tri-modal distribution for the set $\{X\}$. Thus, this calculation, envisions a penetrance distribution where half of the women have a uniform penetrance of slightly more than zero and half have a uniform penetrance of 0.34 – i.e., women have the maximum variance possible – and, in which every man has a uniform penetrance of 0.034, which is intermediate between these two extremes for women – i.e., men have a zero variance. Third, it is not possible that the variance of penetrance values for the (F,G) subset to be at its maximum value because this value exceeds the maximum total variance possible for the entire (G) subset.[102] And fourth, some of the maximum possible variance

in the $\{X\}$ set must be accounted for just by the separation of penetrance values between men and women (*see #5 above*). Each of these considerations will decrease our estimate for the upper limit for $P(G \mid F)$.

7. Genetic Susceptibility in the Population – *P*(*G*)

Based on the relationship in women that: $P(G,F) = P(G \mid F) * P(F) = 0.5 * P(G \mid F)$ and a similar relationship in men, we can use *#6* (*above*) to estimate that:

$$P(G) = P(G,M) + P(G,F) < (0.043 + 0.12)/2 = 0.082$$

If the Upper Solution (*see #4, above*) applies to the full set $\{X\}$, we can estimate that:

$$0.022 \leq 0.003/0.134 = P(MS)/P(MS|G) = P(G) < 0.003/0.067 = 0.044$$

Thus, under any circumstance, only a very small fraction of the population has any chance of developing MS regardless of their environmental experiences. In this sense, like *SCD*, MS is a genetic disease (*see Methods, above*).

In addition, it is of note that, in Canada, the liklihood of carrying the (H+) haplotype for the general population is 24% (*Table 2*). Even taking the largest of the above estimates for $P(G)$, fewer than (8.2/24)=34% of (H+)-carriers could possibly be members of the (G) subset.[102] Moreover, considering that only half of MS patients carry the (H+) haplotype, and considering that 8.2% is and upper-bound, likely far fewer than 34% of (H+)-carriers are in the subset (G). In this circumstance, genetic susceptibility to MS must arise from a combination of this haplotype together with "susceptible states" at other genetic loci.[102] By itself, the (H+)-haplotype poses no risk and, indeed, more generally, genetic susceptibility to MS seems to require specific combinations of non-additive risk-factors.[102]

8. Environmental Factors in MS

We can define (E_T) to be the prevailing environmental conditions (whatever these conditions are) experienced by a population during some time-period (T). We also define (E_i) to be the environmental exposure, which is sufficient for MS to develop in the i^{th} susceptible individual (whatever these events might be, whenever these events need to act, and however many events might be involved) – i.e., in order for MS to develop in the (i^{th}) individual requires that both events (E_i and G_i) occur jointly. If there are (m) members of the subset (G), the probability of a sufficient environmental exposure, $P(E)$, in the (G) subset at time-period (T) is:

$$P(E) = P(E|\,\mathrm{G}, E_R) \sum_{i=1}^{m} P\left(E_{i,}\ G_i, |\ G, E_T\right) = \sum_{i=1}^{m} P(G_i\ |\mathrm{G}, E_r) * P(E_i|G_I, G, E_r)$$

where: $P(G_i \mid G, E_T) = P(G_i \mid G) = 1/m$

Using the standard methods of survival analysis,[109] we define the cumulative survival $\{S(u)\}$ and failure $\{F(u)\}$ functions in addition to the hazard-rate functions $\{h(u)\}$ and $\{g(u)\}$ in susceptible men and women (respectively) for developing MS at different levels of environmental exposure. These hazard-rate functions are assumed to be proportional. The implications of non-proportionality are considered elsewhere.[102] However, assuming proportionality, then:

$$g(u) = R * h(u)$$

where: $u = P(E)$ and (R) is the proportionality constant.

For men, we transform exposure from (u) units into (a) units, by defining {$H(u)$}to be the definite integral of the hazard-function {$h(u)$} from a (u) level of exposure to a (0) level of exposure and, then, by defining the (a) units to be:

$$a = H(u) = \int_0^u h(u)du$$

where $da=h(u)\,du$

Because these (a) units are arbitrary, we can assign "1 unit" of environmental exposure to be the difference in exposure level between any two time points (e.g., a_1 and a_2) such that:

$$a_2 - a_1 = 1$$

Similarly, for women, we can transform exposure into so-called "apparent" exposure units (a^{app}) such that:

$$a^{app} = R * a$$

and where "1 unit" of environmental exposure (on this scale) is now defined such that:

$$a_2{}^{app} - a_1{}^{app} = 1$$

A standard derivation from the methods of survival analysis,[109] demonstrates that survival curves are exponentially related to the hazard function, such that, in this circumstance, it can be shown [102] that:

For men : $F(a) = 1 - e^{-a}$

and, for women: $F(a^{app}) = 1 - e^{-a^{app}}$

In considering the probability of developing MS (i.e., of failure), we will use subscripts (*1*) and (*2*) to denote the failure probabilities and the values of other parameters at the 1st and 2nd time-periods respectively (i.e., 1941–1945 & 1976–1980, *see above*). Importantly, unlike true survival where everyone fails given a sufficient amount of time, the probability of developing MS may not reach 100% as the probability of a sufficient environmental exposure increases to unity. Moreover, the limiting value for the cumulative probability of developing MS for men (***c***) may not be the same as it is for women (***d***).

Consequently, the failure probability for susceptible women and men at the 1st time period can be expressed as:

$$F(a^{app})_1 = P(MS, E \mid G, F)_1 = d * \{1 - e^{-a1^{app}}\} \qquad (\textit{for women})$$

and: $$F(a)_1 = P(MS, E \mid G, M)_1 = c * \{1 - e^{-a1}\} \qquad (\textit{for men})$$

From the definitions of "1 exposure unit" (*see #8 above*), at the 2nd time point, these equations become:

$$F(a^{app})_2 = P(MS, E \mid G, F)_2 = d * \{1 - e^{-a1^{app}+1}\} \qquad (\textit{for women})$$

and: $$F(a)_2 = P(MS, E \mid G, M)_2 = c * \{1 - e^{-a1+1}\} \qquad (\textit{for men})$$

The values for these failure functions at time-periods (*1*) and (*2*) represent two points on the exponential response curves for women and men. Because any two points on an exponential curve uniquely and completely defines that curve, the observations regarding the change in the (*F:M*) sex-ratio over time in Canada (*Table 2*), can be used to construct these two response curves (*see Figure 1*). From these curves, four conclusions

can be drawn.[102] First, as can be seen in the *Figure*, the environmental threshold at which MS begins to develop in susceptible individuals is greater for women than it is for men. The magnitude of this threshold difference depends upon some of the parameter values chosen. However, in all circumstances, this threshold is greater in women if the hazards are proportional.[102] Second, it can be shown that:

$$P(E \mid G,F)_2 = P(MS,E \mid G,F)_2 / d > 0.76$$

and: $$P(E \mid G,M)_2 = P(MS,E \mid G,M)_2 / c > 0.83$$

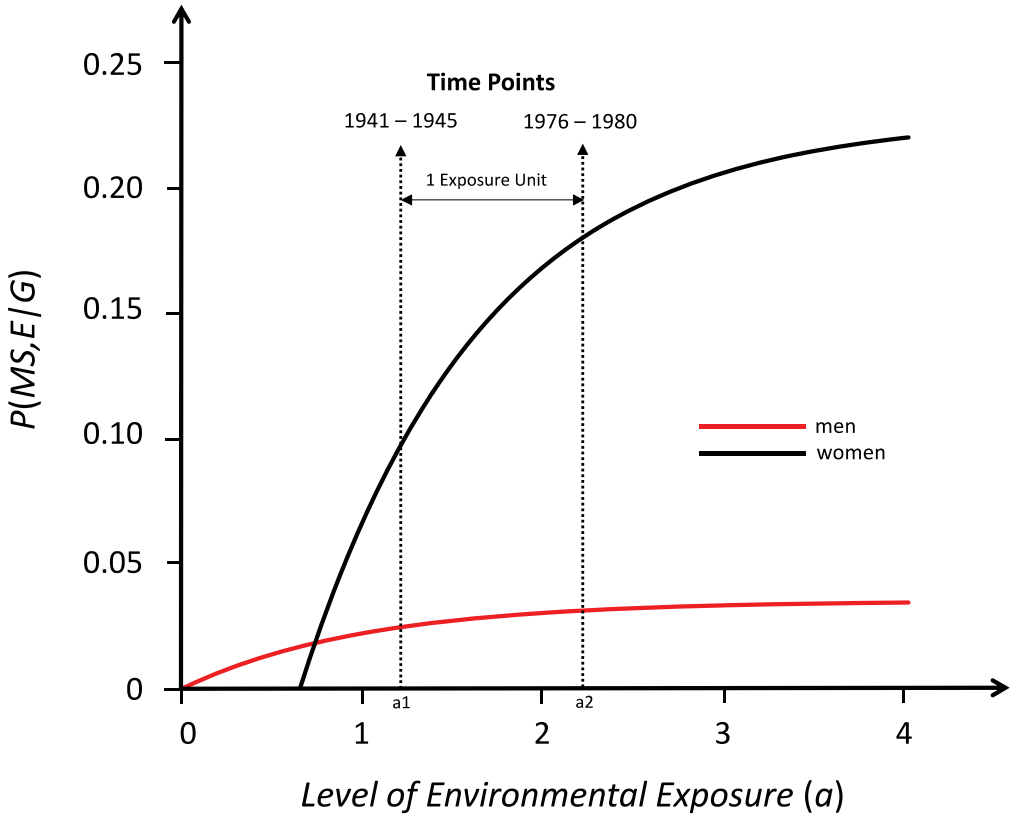

Figure 1. Response curves for the likelihood of developing MS in genetically susceptible men and women with an increasing probability of a sufficient environmental exposure $\{P(E)\}$, assuming proportional hazards ($R=1$). Response curves are derived from the change in the *F:M* sex-ratio over time in Canada.[26] The probability of getting MS in a genetically-susceptible individual – i.e., $P(MS,E \mid G)$ – is shown on the *y-axis*. The exposure level $\{P(E)\}$ for the population is shown on the *x-axis* using transformed "exposure units" (a) – *see Text*. One "exposure unit" is defined arbitrarily as: $(a_2 - a_1)$ for men and $(a_2^{app} - a_1^{app})$ for women (*see Text*). In the graph, because we chose ($R=1$), these two scales are the same. This need not be the case and the hazard may not be proportional.[102] Plots have been constructed using the values provided in the *MZ*-twin study from Canada.[24] together with the estimates: $\{P(MS)_1 / P(MS)_2 = 0.6\}$ & $\{P(G) = 0.044\}$.

Thus, the large majority of the susceptible population is currently experiencing an environmental exposure sufficient to cause MS. Moreover, the relevant environmental exposures, especially if these are multiple (*see above*), must currently be occurring at population-wide levels.[102] Third, because the (*F:M*) sex ratio has changed between the two time periods (*see above*) we can define a constant (*C*) and, thereby, estimate that:

$$C = P(MS)_1 / P(MS)_2 < P(M \mid MS)_2 / P(M \mid MS)_1 = 0.238 / 0.313 = 0.76$$

Therefore, the prevalence of MS in Canada must have increased by at least 32% between these two time periods. And fourth, it can be shown that the theoretical limits for (***c***) and (***d***) are: $c \approx P(MS|M, MZ_{MS})$ and $d \approx P(MS|F, MZ_{MS})$.[102] Therefore, the curves, as they are depicted in *Figure 1*, must be inaccurate because, for these particular curves:

$$c = 0.035 < 0.067 = P(MS \mid M, MZ_{MS})$$

and:

$$d = 0.228 < 0.34 = P(MS \mid F, MZ_{MS})$$

There are several variables that can be adjusted to match these constraints. To analyze this, we considered, iteratively, parameter combinations, which covered a wide range of plausible values: $(0.25 \leq C \leq 0.75)$, $(0.2 \leq R \leq 5.0)$, $(0.001 \leq P(G) \leq 1.0)$, $(0.18 \leq P(G|F) \leq 0.70)$, and $(0.002 \leq P(MS) \leq 0.006)$. Moreover, in this analysis, the estimates for (***c***) and (***d***) were required to be within (± 15%) of their observed proband-wise *MZ*-twin concordance rates (*Table 2*). In this analysis, there were many combinations that matched these constraints. The solution space covered by these matching combinations included the full range of possibilities for the parameters of *C*, *R*, and *P*(*MS*), By contrast, the ranges for both *P*(*G*) and $P(F \mid G)$ were restricted such that: $\{0.02 \leq P(G) \leq 0.055\}$ and $\{0.33 \leq P(F|G) \leq 0.5\}$.This restricted range for *P*(*G*) fits within the framework developed previously and confirms the conclusion that developing MS is not a possibility for a large majority of the population (*see #7 above*). Similarly, this analysis confirms that women are less likely than men to be in the (*G*) subset (*see #6 above*).

Discussion

The analysis provides considerable insight to the nature and basis of MS and to the role that genetic and environmental determinants play in MS pathogenesis. The fact that only a very small fraction of the general population are members of the genetically-susceptible subset (*G*) indicates that the vast majority of the population has no chance whatsoever of developing MS, irrespective of the environmental conditions that these individuals experience.[102] Having the proper genetic constitution is essential to disease pathogenesis. In this sense, MS is a genetic disorder. Nevertheless, this genetic susceptibility is complex. Single genes or single haplotypes do not seem to contribute much. For example, (*H+*) haplotype is the genetic trait with the largest (by far) MS-association of any in the genome (for the WTCCC: *OR*=3.28; $p<<10^{-300}$). Nevertheless, despite this strong association with MS, only a small minority of individuals who carry this haplotype have any MS-risk at all.[102] In such a circumstance, it must be that genetic susceptibility is related to carrying this haplotype together with other genetic traits. Notably, also, this haplotype is only a portion of much several much longer *CEHs*, which span the entire *MHC* region.[99-101] However, genetic susceptibility cannot be explained on the basis of the state of the *MHC*. Thus, despite the large number highly selected *CEH*, and despite a significant variability in MS-association observed for different *CEHs*, every (*H+)*

carrying *CEH* (regardless of its rarity) seems to be strongly MS-associated [99-101] and, consequently, most of individuals who carry these *CEHs* are not members of the subset (*G*).

In addition, it seems clear that, despite the fact that certain genetic combinations increase the likelihood being a member of the (*G*) subset, these combinations are heterogeneous. Thus, considering all the associated genetic regions identified so far, every person (including both patients and controls) has a unique genotype and, moreover, only a very small fraction of individuals (who actually develop MS) share even the same 4-locus genetic combination.[102] This suggests that, although genetically-based, susceptibility to MS is largely idiosyncratic.

Despite the conclusion that MS is a genetic disease, however, MS is equally an environmental disorder. Specific environmental exposures are also necessary for disease-patho*g*enesis. Indeed, the fact that the (*F:M*) sex-ratio has increased steadily from 1941 to 1980 in Canada, indicates that a sufficient environmental exposure is required for MS to develop (*Figure 1*). If a person is not exposed to a sufficient environment, they cannot develop MS, irrespective of their genetic constitution. However, neither environment nor genetics alone is sufficient for disease pathogenesis. Thus, the basis of this genetic susceptibility is complex and requires an interaction between genetic and environmental events in order for the disease to develop.

As discussed earlier, at least three environmental events, probably sequential, seem to be implicated as necessary for MS to develop is a genetically susceptible individual.[1,92,102] The first environmental event (or "maternal" factor) occurs during the *IU* or early post-natal period. Support for this factor comes from the discrepancy in recurrence-rates between twin and non-twin siblings, from the fact that concordant half-twins are twice as likely to share the mother than the father, and from the periodic, circa-annum, effect that month-of-birth has on the subsequent likelihood of developing MS. As noted earlier, in the northern hemisphere, this periodicity to MS-susceptibility peaks just before the summer months and dips to its nadir just before winter. This pattern is inverted southern hemisphere.[22,24,28,32-39] Each of these observations implicates an environmental event, involved in MS pathogenesis, that is occurring near birth. The circa-annum periodicity to susceptibility implies that this environmental event, whatever it may be, is coupled to the solar cycle.

A second environmental event is implied by the migration data whereby an individual who relocates (prior to adolescence) from an area of high-prevalence to an area of low prevalence (or *vice versa*), has an MS risk, which is similar to that of the area to which they moved. [42-47] By contrast, when they make the same relocation later, their MS risk is similar to that of the area from which they moved.[42-47] These observations implicate an environmental event, involved in MS-pathogenesis, which occurs at or around puberty. And third, because the onset of clinical MS generally occurs long after the first and second environmental events have already taken place, it seems that one or more additional environmental events are also necessary for clinical MS to develop.

Naturally, there is no guarantee that the environmental events, which are sufficient to cause MS in one person, are the same as those that are sufficient in another. Nevertheless, those factors or events, which have been implicated in MS-pathogenesis so far, appear to

affect a large proportion of susceptible individuals in a similar manner. Thus, the fact that we even have evidence for the first two factors (as described above) suggests this. In addition, a prior EBV infection has been strongly linked to MS, especially when this infection occurs during adolescence and results in symptomatic mononucleosis. [48-59] Indeed, such an infection prior to clinical onset occurs in ~100% of MS cases (*Table 1*) and, if this is the case, this would indicate that EBV exposure is a 'necessary factor' in the causal pathway leading to MS. Moreover, if this factor is necessary, it must be occurring sequentially with the "maternal" factor because the "maternal" factor acts long before adolescence. Finally, there is a considerable amount of circumstantial evidence, which suggests a role for vitamin D_3 deficiency in this causal pathway. Because late *EBV* infection typically occurs during or after adolescence, EBV seems a much better candidate for the second (rather than the first) environmental event. By contrast, Vitamin D_3 deficiency, which is coupled to the solar cycle, is involved immune system maturation, and associated with autoimmunity,[70-78] seems to be a much better candidate for the first environmental event.

Naturally, it is possible that those environmental events, which are sufficient to cause MS for one individual, are different than those that are sufficient for another. Despite this possibility, however, the same environmental events seem to affect large proportions of susceptible individuals in a similar manner. Indeed, the fact that we even have evidence for the "maternal" and "migratory" factors suggests this. Moreover, as noted above, a prior *EBV* infection seems to occur in ~100% of MS cases (*Table 1*) and, if so, this would indicate that *EBV* exposure was a 'necessary factor' in the causal pathway leading to MS.[1,92,102] Additionally, this would indicate that every MS patient has, at least, this environmental exposure in common and, thus, that no one has "purely genetic" MS (i.e., no one can develop MS under <u>any</u> environmental conditions).

Nevertheless, even when an individual with the proper genetic composition experiences an environmental exposure sufficient to cause MS in that person, still, over half of such individuals will not develop clinical disease (*Figure 1*). Very likely, some of these individuals will be found to have subclinical disease.[2-5,110] Nevertheless, although this might increase our estimate for $P(MS)$ by as much as 50-100%, this is still insufficient to account for the fact that the plateau of the response curves (especially for men) never even approach 100% (*Figure 1*). Importantly, this circumstance cannot be ascribed to any "unidentified" environmental occurrences because we have already defined a sufficient environmental exposure very broadly to include both those environmental events that are known or suspected in addition to those that are completely unknown. Consequently, this failure to reach 100%, even when: $P(E)=1$ in susceptible individuals, indicates that stochastic processes must also be involved in disease-pathogenesis.

References

1. Goodin DS. The epidemiology of multiple sclerosis: Insights to a causal cascade. *Handb Clin Neurol.* 2016;138:173-206.
2. Vost A, Wolochow D, Howell D. Incidence of infarcts of the brain in heart diseases. *J Path Bact.* 1964;88:463-470.
3. Georgi VW. Multiple Sklerose: Patholiogisch-Anamtomische Befunde multiple Sklerose bei klinisch nicht diagniostizierte Krankbeiten. *Schweiz Med Wochenschr*. 1966;20:605-607.
4. Gilbert J, Sadler M. Unsuspected multiple sclerosis. *Arch Neurol*. 1983;40:533-536.
5. Engell T. A clinical patho-anatomical study of clinically silent multiple sclerosis. *Acta Neurol Scand.* 1989;79:428-430.
6. Gourraud PA, Harbo HF, Hauser SL, Baranzini SE. The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev.* 2012;248:87–103.
7. Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. *Biochim Biophys Acta*. 2014;1842:1889-1895.
8. Goodin DS. The nature of genetic susceptibility to multiple sclerosis: Constraining the Possibilities. *BMC Neurology.* 2016;16:56.
9. Goodin DS. The Genetic and Environmental Bases of Complex Human-Disease: Extending the Utility of Twin-Studies. *PLoS One.* 2012;7(12): e47875.
10. Herrera BM, Cader MZ, Dyment DA, et al. Multiple sclerosis susceptibility and the X chromosome. *Mult Scler.* 2007;13:856–8.
11. Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet.* 2009;18:767-778.
12. Sanna, S. Pitzalis M, Zoledziewska M, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nature Genet.* 2010;42:495–497.
13. The International Multiple Sclerosis Genetics Consortium & the Wellcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011;476:214-219.
14. International Multiple Sclerosis Genetics Consortium (IMSGC). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis *Nat Genet.* 2014;45:1353-60.
15. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365 (6460). doi:10.1126/science.aav7188.
16. Dyment DA, Herrera BM, Cader Z, et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum Mol Genet* 2005;14:2019-2026.
17. Ramagopalan, SV, Anderson, C, Sadovnick, AD, Ebers, GC. Genomewide study of multiple sclerosis. *N. Engl. J. Med*. 2007;357, 2199–2200.
18. Link J, Kockum I, Lorentzen AR, et al. Importance of Human Leukocyte Antigen (HLA) Class I and II Alleles on the Risk of Multiple Sclerosis. *PLoS One.* 2012;7(5):e36779.

19. Patsopoulos NA, Barcellos LF, Hintzen RQ, et al. Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet.* 2014;9(11):e1003926.
20. Chao MJ, Barnardo MC, Lincoln MR, et al. HLA class I alleles tag HLA-DRB1*1501 haplotypes for differential risk in multiple sclerosis susceptibility. *Proc Natl Acad Sci USA.* 2008;105:13069-74.
21. Lincoln MR, Ramagopalan SV, Chao MJ, et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc Natl Acad Sci USA.* 2009;106:7542-7.
22. Sadovnick AD, Ebers GC, Dyment DA, et al., and the Canadian Collaborative Study Group. Evidence for genetic basis of multiple sclerosis. *Lancet* 1996;347:1728-1730.
23. Ebers GC, Yee IML, Sandovnick AD, Duquette P, and the Canadian Collaborative Study Group. Conjugal multiple sclerosis: Population-based prevalence and recurrence risks in offspring. *Ann Neurol.* 2000;48:927-931.
24. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, the Canadian Collaborative Study Group. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc Natl Acad Sci U S A.* 2003;100:12877–82.
25. Dyment DA, Yee IML, Ebers GC, SadovnickAD, and the Canadian Collaborative Study Group. Multiple sclerosis in stepsiblings: Recurrence risk and ascertainment. *J Neurol Neurosurg Psychiatry.* 2006;77:258–259.
26. Orton SM, Herrera BM, Yee IM, et al., and the Canadian Collaborative Study Group. Sex ratio of multiple sclerosis in Canada: A longitudinal study. *Lancet Neurol*. 2006;5:932–6.
27. Sadovnick AD, Yee IML, Ebers GC, and the Canadian Collaborative Study Group. Multiple sclerosis and birth order: A longitudinal cohort study. *Lancet Neurol.* 2005;4:611–617.
28. Ebers GC, Sadovnick AD, Dyment DA, et al. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. *Lancet.* 2004;363: 1773–1774.
29. Bager P, Nielsen NM, Bihrmann K, et al. Sibship characteristics and risk of multiple sclerosis: A nationwide cohort study in Denmark. *Am J Epidemiol.* 2006;163:1112–1117.
30. Kantarci OH, Barcellos LF, Atkinson EJ, et al. Men transmit MS more often to their children vs women: The Carter effect *Neurology.* 2006;67:305–310
31. Herrera BM, Ramagopalan SV, Orton S, et al. Parental transmission of MS in a population-based Canadian cohort. *Neurology.* 2007;69:1208–1212.
32. Islam T, Gauderman WJ, Cozen W, et al. Differential twin concordance for multiple sclerosis by latitude of birthplace *Ann Neurol.* 2006;60:56–64.
33. Robertson NP, Fraser M, Deans J, et al. Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain.* 1996;119, 449-455.
34. Compston A, Coles A. Multiple sclerosis. *Lancet.* 2002;359:1221-31.
35. Willer CJ, Dyment DA, Sadovnick AD, et al. Timing of birth and risk of multiple sclerosis: population based study. *Br Med J.* 2005;330:120-124.
36. Templer DI, Trent NH, Spencer DA, et al. Season of birth in multiple sclerosis. *Acta Neurol Scand*. 1992;85:107-109.

37. Staples J, Ponsonby AL, Lim L. Low maternal exposure to ultraviolet radiation in pregnancy, month of birth and risk of multiple sclerosis in offspring: a longitudinal analysis. *Br Med J.* 2010;340:c1640.
38. Torkildsen Ø, Aarseth J, Benjaminsen E, et al. Month of birth and risk of multiple sclerosis: confounding and adjustments. *Ann Clin Transl Neurol*. 2014;1:141–144.
39. Pantavou KG, Bagos PG. Season of birth and multiple sclerosis: a systematic review and multivariate meta-analysis *J Neurol.* 2020;267:2815–2822.
40. Fiddes B, Wason J, Kemppinen A, et al. Confounding underlies the apparent month of birth effect in multiple sclerosis. *Ann Neurol.* 2013;73:714-270.
41. Fiddes B, Wason J, Sawcer S. Confounding in association studies: month of birth and multiple sclerosis *J Neurol.* 2014;261:1851-1856.
42. Dean G, Kurtzke JF. On the risk of multiple sclerosis according to age at immigration to South Africa. *Br Med J.* 1971;3:725-729.
43. Alter M, Kahana E, Loewenson R. Migration and risk of multiple sclerosis. *Neurology.* 1978;28:1089-1093.
44. Elian M, Nightingale S, Dean G. Multiple sclerosis among United Kingdom-born children of immigrants from the Indian subcontinent, Africa, and the West Indies. *J Neurol Neurosurg Psychiatr*. 1960;53:906-911.
45. Kahana E, Zilber N, Abramson JH, et al. Multiple sclerosis: Genetic versus environmental aetiology: Epidemiology in Israel updated. *J Neurol.* 1994;241:341-346.81
46. Cabre P. Signate A. Olindo S, et al. Role of return migration in the emergence of multiple sclerosis in the French West Indies *Brain.* 2005;128:2899–2910.
47. Compston A, Confavreux C, Lassmann H, et al., eds. McAlpine's Multiple Sclerosis, 4th ed. London, UK: Churchill Livingston; 2006
48. Sumaya CV, Myers LW, Ellison GW. Epstein-Barr virus antibodies in multiple sclerosis. *Arch Neurol.* 1980;37:94–96.
49. Sumaya CV, Myers LW, Ellison GW, et al. Increased prevalence and titer of Epstein-Barr virus antibodies in patients with multiple sclerosis. *Ann Neurol.* 1985;17:371–377.
50. Ascherio A, Munger KL, Lennette ET, et al. Epstein-Barr virus antibodies and the risk of multiple sclerosis: A prospective study. *Am Med Assoc J.* 2001;286:3083-3088.
51. Goldacre MJ, Wotton CJ, Seagroatt V, et al. Multiple sclerosis after infectious mononucleosis: record linkage study. *J Epidemiol Community Health*. 2004;58:1032-1035.
52. Sundström P, Juto G, Wadell G, et al. An altered immune response to Epstein-Barr virus in multiple sclerosis: A prospective study. *Neurology.* 2004;62;2277-2282.
53. Haahr S, Plesner AM, Vestergaard BF, et al. A role of late Epstein-Barr virus infection in multiple sclerosis. *Acta Neurol Scand.* 2004;109: 270–275.
54. Ponsonby AL, van der Mei, Dwyer T, et al. Exposure to infant siblings during early life and risk of multiple sclerosis. *Am Med Assoc J*. 2005;293:463-9.
55. Farrell PJ. Role for HLA in susceptibility to infectious mononucleosis. *J Clin Invest.* 2007;117:2756-2758.
56. Nielsen TR, Rostgaard K, Nielsen NM, et al. Multiple sclerosis after infectious mononucleosis. *Arch Neurol.* 2007;64:72-75.
57. Serafini B, Roiscarelli B, Franciotta D, et al. Dysregulated Epstein-Barr virus infection in multiple sclerosis. *J Exp Med.* 2007;204:2899-2912.

58. Ascherio A, Munger KL. Epidemiology of Multiple Sclerosis: From Risk Factors to Prevention—An Update. *Semin Neurol.* 2016;36:103-114.
59. Abrahamyan S, Eberspächer B, Hoshi MM, et al. Complete Epstein-Barr virus seropositivity in a large cohort of patients with early multiple *sclerosis J Neurol Neurosurg Psychiatry.* 2020;91:681-686.
60. Henle W, Henle G, Andersson J, et al. Antibody responses to Epstein-Barr virus-determined nuclear antigen (EBNA)-1 and EBNA-2 in acute and chronic Epstein-Barr virus infection. *Proc Natl Acad Sci (USA).* 1987;84:570-574.
61. Holick MF. Vitamin D requirements for humans of all ages: new increased requirements for women and men 50 years and older. *Osteoporos Int.* 1998;8(Suppl 2):S24-29.
62. Hayes CE, Nashold FE Spach KM, et al. The immunological functions of the vitamin D endocrine system. *Cell Molec Biol.* 2003;49:277-300.
63. Nagpal S, Na S, Rathnachalam R. Noncalcemic actions of vitamin D receptor ligands. *Endocr Rev.* 2005;26:662-687.
64. Lips P. Vitamin D physiology. *Prog Biophys Molec Biol.* 2006;92:4-8.
65. Sinclair H. Polyunsaturated fatty acids in multiple sclerosis. *Br Med J.* 1977;2:1217.
66. Koch-Henderson N. Multiple sclerosis in Scandinavia and Finland. *Acta Neurol Scand.* 1995;161:55-59.
67. Grønlie SA, Myrvoll E, Hansen G, et al. Multiple sclerosis in North Norway, and first appearance in an indigenous population. *J Neurol.* 2000;247:129–133.
68. Gillie O. A new government policy is needed for sunlight and vitamin D. *Br J Dermatol.* 2006;154:1052–1061.
69. Ramagopalan SV, Maugeri NJ, Handunnetthi L, et al. Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1*1501 is regulated by vitamin D. *PLoS Genet.* 2009;5(2):e1000369
70. Jablonski NG, Chaplin G. The evolution of human skin coloration. *J Hum Evol.* 2000;39:57-106.
71. Jablonski NG, Chaplin G. Skin deep. *Sci Am.* 2002;287:74-81.
72. Adams CWM. A Colour Atlas of Multiple Sclerosis & Other Myelin Disorders. Ipswich, Suffolk, UK: Wolfe Medical Publications; 1989, p. 101.
73. Kimlin MG, Olds WJ, Moore MR. Location and vitamin D synthesis: Is the hypothesis validated by geophysical data? *J Photochem Photobiol B: Biology.* 2007;86:234–239.
74. Cantorna MT. Vitamin D and autoimmunity: Is vitamin D status an environmental factor affecting autoimmune disease prevalence? *Proc Soc Exp Biol Med.* 2000;223:230-233.
75. Deluca HF, Cantorna MT. Vitamin D: Its role and uses in immunology. *FASEB J.* 2001;15:2579-2585.
76. Griffin MD, Lutz W, Phan VA, et al. Dendritic cell modulation by 1-alpha,25 dihydroxyvitamin D3 and its analogs: A vitamin D receptor-dependent pathway that promotes a persistent state of immaturity in vitro and in vivo. *Proc Natl Acad Sci.* 2001;98:6800-6805.
77. Tavera-Mendoza LE, White JH. Cell defenses and the sunshine vitamin. *Sci Am.* 2007;297(5):62-72. doi:10.1038/scientificamerican1107-62.
78. Cantorna MT, Mahon BD. Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence. *Exp Biol Med.* 2004;229:1136-1142.

79. Van der Mei IA, Ponsonby AL, Dwyer T, et al. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: Case-control study. *Br Med J.* 2003;327:316-322.
80. Van der Mei IA, Ponsonby AL, Dwyer T, et al. Vitamin D levels in people with multiple sclerosis and community controls in Tasmania, Australia. *J Neurol.* 2007;254:581-90.
81. Munger KL, Levin LI, Hollis BW, et al. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *Am Med Assoc J.* 2006;296:2832-2838.
82. Munger KL, Zhang SM, O'Reilly E, et al. Vitamin D intake and incidence of multiple sclerosis. *Neurology.* 2004;62:60-65.
83. Hernán MA, Olek MJ, Ascherio A. Geographic variation of MS incidence in two prospective studies of US women. *Neurology.* 1999;53:1711–1718.
84. Koch-Henriksen N. The Danish Multiple Sclerosis Registry: a 50-year follow-up. *Mult Scler.* 1999;5:293-296.
85. Celius EG, Vandvik B. Multiple sclerosis in Oslo, Norway: prevalence on 1 January 1995 and incidence over a 25-year period. *Eur J Neurol.* 2001;8:463-469.
86. Barnett MH, William DB, Day S, et al. Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: a 35-year study. *J Neurol Sci.* 2003;213:1–6.
87. Ranzato F, Perini P, Tzintzeva E, et al. Increasing frequency of multiple sclerosis in Padova, Italy: a 30 year epidemiological survey. *Mult Scler.* 2003;9:387-392.
88. Sundström P, Nyström L, Forsgren L. Incidence (1988-97) and prevalence (1997) of multiple sclerosis in Västerbotten County in northern Sweden. *J Neurol Neurosurg Psychiatry.* 2003;74:29-32.
89. Sarasoja T, Wikström J, Paltamaa J, et al. Occurrence of multiple sclerosis in central Finland: a regional and temporal comparison during 30 years. *Acta Neurol Scand.* 2004;110:331-6.
90. Freedman DM, Mustafa M, and Alavanja MCR. Mortality from multiple sclerosis and exposure to residential and occupational solar radiation: A case-control study based on death certificates. *Occup Environ Med.* 2000;57:418–421.
91. Emmons KM, Colditz GA. Preventing excess sun exposure: It is time for a national policy. *J Natl Cancer Inst.* 1999;91:1269-1270.
92. Goodin DS. The causal cascade to multiple sclerosis: a model for MS pathogenesis. *PLoS One.* 2009;4:e4565.
93. Mumford CJ, Wood NW, Kellar-Wood H, et al. The British Isles survey of multiple sclerosis in twins. *Neurology.* 1994;44:11-15.
94. Nielsen NM, Westergaard T, Rostgaard K, et al. Familial risk of multiple sclerosis: A nationwide cohort study *Am J Epidemiol* 2005;162:774–778.
95. Kuusisto H, Kaprio J, Kinnunen E, et al. Concordance and heritability of multiple sclerosis in Finland: Study on a nationwide series of twins. *Eur J Neurol.* 2008;15:1106-1110.
96. Hansen T, Skytthe A, Stenager E, et al. Risk for multiple sclerosis in dizygotic and monozygotic twins. *Mult Scler.* 2005;11:500-503.
97. Hansen T, Skytthe A, Stenager E, et al. Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. *Mult Scler.* 2005;11:504-510.

98. Ristori G, Cannoni S, Stazi MA, et al., and the Italian Study Group on MS in Twins. Multiple sclerosis in twins from continental Italy and Sardinia: A Nationwide Study *Ann Neurol.* 2006;59:27–34.
99. Goodin DS, Khankhanian P. Single Nucleotide Polymorphism (SNP)-Strings: An Alternative Method for Assessing Genetic Associations. *PLoS One.* 2014;9(4):e90034.
100. Goodin DS, Khankhanian P, Gourraud, PA, Vince N. Highly conserved extended haplotypes of the major histocompatibility complex and their relationship to multiple sclerosis susceptibility. *PLoS One.* 2018;13(2):e0190043.
101. Goodin DS, Khankhanian P, Gourraud, PA, Vince N. Genetic susceptibility to multiple sclerosis: Interactions between conserved extended haplotypes of the MHC and other susceptibility regions. (*submitted*).
102. Goodin DS, Khankhanian P, Gourraud, PA, Vince N. The nature of genetic and environmental susceptibility to multiple sclerosis. *PLoS One.* 2020;16(3): e0246157.
103. Harrison's Principles of Internal Medicine, 18th Edition. Longo DL, Kasper, DL, Jameson JL, Fauci AS, Hauser SL, Loscalzo JL., eds. New York, NY: McGraw Hill Medical; 2012.
104. Witte JS, Carlin JB, Hopper JL. Likelihood-based approach to estimating twin concordance for dichotomous traits. *Genetic Epidemiol.* 1999;16:290–304
105. Wallin MT, Culpepper WJ, Campbell JD, et al. US Multiple Sclerosis Prevalence Workgroup. The prevalence of MS in the United States: A population-based estimate using health claims data *Neurology*. 2019;92:e1029-e1040.
106. Harding K, Zhu F, Alotaibi MD, et al. Causes that contribute to deaths due to multiple sclerosis: analysis of population-based multiple-cause-death data. *Presentation 144*. ECTRIMS 2018, Berlin.
107. Jacobson HI. The maximum variance of restricted unimodal distributions. *Ann Math Stat.* 1969;40:1746–52.
108. Freeman JB, Dale R. Assessing bimodality to detect the presence of a dual cognitive process. *Behav Res.* 2013;45:83–97.
109. Fisher LD, van Belle G. Biostatistics: A Methodology for the Health Sciences, New York, NY: John Wiley & Sons; 1993, pp. 786-829.
110. Okuda DT, Mowery EM, Cree BAC, et al. Asymptomatic spinal cord lesions predict disease progression in radiologically isolated syndrome. *Neurology.* 2011;76:686-692.

