BMC Neurology

The nature of genetic susceptibility to multiple sclerosis: constraining the possibilities

Douglas S. Goodin

Abstract

Background: Epidemiological observations regarding certain population-wide parameters (e.g., disease-prevalence, recurrence-risk in relatives, gender predilections, and the distribution of common genetic-variants) place important constraints on the possibilities for the genetic-basis underlying susceptibility to multiple sclerosis (MS).

Methods: Using very broad range-estimates for the different population-wide epidemiological parameters, a mathematical model can help elucidate the nature and the magnitude of these constraints.

Results: For MS no more than 8.5 % of the population can possibly be in the "genetically-susceptible" subset (defined as having a life-time MS-probability at least as high as the overall population average). Indeed, the expected MS-probability for this subset is more than 12 times that for every other person of the population who is not in this subset. Moreover, provided that those genetically susceptible persons (genotypes), who carry the well-established MS susceptibility allele (DRB1*1501), are equally or more likely to get MS than those susceptible persons, who don't carry this allele, then at least 84 % of MS-cases must come from this "genetically susceptible" subset. Finally, because men, compared to women, are at least as likely (and possibly more likely) to be susceptible, it can be demonstrated that women are more responsive to the environmental factors that are involved in MS-pathogenesis (whatever these are) and, thus, susceptible women are more likely actually to develop MS than susceptible men. Finally, in contrast to genetic susceptibility, more than 70 % of men (and likely also women) must have an environmental experience (including all of the necessary factors), which is sufficient to produce MS in a susceptible individual.

Conclusions: As a result, because of these constraints, it is possible to distinguish two classes of persons, indicating either that MS can be caused by two fundamentally different pathophysiological mechanisms or that the large majority of the population is at no risk of the developing this disease regardless of their environmental experience. Moreover, although environmental-factors would play a critical role in both mechanisms (if both exist), there is no reason to expect that these factors are the same (or even similar) between the two.

Keywords: Multiple sclerosis, Genetic susceptibility, Heritability, Pathogenesis, Causation, Epidemiology, Environment, Complex disease, Complex, Twin studies

Correspondence: douglas.goodin@ucsf.edu
Department of Neurology, UCSF MS Center, University of California, San Francisco, 675 Nelson Rising Lane, Suite #221D, San Francisco, CA 94158, USA

Background

Introduction

Complex genetic disorders are those that are caused by the interaction of multiple genetic and environmental factors [1]. Many human diseases are examples of such disorders, including multiple sclerosis (MS) – a common neurological condition, in which recurrent immune-mediated injuries occur to the central nervous system [2, 3]. Epidemiological evidence has implicated the involvement of multiple environmental factors, including vitamin D deficiency and Epstein-Barr viral infections – see [3] for a review. Nevertheless, it is on the genetic side that most of the recent progress has come. The associations of MS with the human leukocyte antigens (HLA) on the short arm of chromosome 6 have been known for decades [2–9]. More recently, from several genome-wide association studies (GWAS) of single-nucleotide polymorphisms (SNPs) in MS, disease-associations have been identified in more than 150 different non-HLA locations scattered throughout the genome [4–7]. However, the translation of these associations into a clinically useful assessment of an individual's disease-risk has been limited. This is due to the fact that a large proportion of the heritability for many complex diseases, including MS, remains unexplained. Indeed, in MS the 110 genes so far identified (in addition to the HLA associations) only account for only 28 % of the known heredity [4, 10]. Although a good deal of effort is currently being made to narrow this so-called "heritability gap", it is unclear how likely these efforts are to succeed.

Much will depend upon the underlying basis of genetic susceptibility. For example, suppose that the individuals in a population exist on a continuum of susceptibility (i.e., anyone can develop the illness under the proper circumstances and a person's individual genetic make-up only serves to make this outcome more or less likely to occur). In this case, although individuals at especially high-risk could, perhaps, be identified, the development of a sensitive and specific genetic test for susceptibility, which could be applied to the population as a whole, will likely not be possible. By contrast, if only a small segment of the population is genetically susceptible and only these susceptible individuals can develop disease, then the task of developing such a test, should, in theory, be much more likely to succeed.

The epidemiological observations regarding the various population-wide parameters such as the disease prevalence, the recurrence risk in relatives, gender predilections, and the distribution of common genetic variants (observations which have been made for years in many different parts of the world) place important constraints on the possibilities for the underlying genetic basis of susceptibility. Although applicable to any complex genetic disease, for illustrative purposes, this paper considers the nature and magnitude of these constraints as they apply to the study of MS.

Model overview and implications

Because much of this model development is technical, an overview of the basic ideas behind the model (and their implications) is here provided for purposes of clarity. Thus, in this model, directly-measurable epidemiological data are used to estimate the likelihood that an individual from the general population is "genetically-susceptible" to getting MS. This probability is defined as $P(G)$. The definition used for this parameter is provided both below and in Table 1. In order to estimate the value of this parameter, we define certain other quantities such as the conditional probability that a "genetically-susceptible" individual will get MS $\{P(MS|G)\}$ and the probability that an individual

Table 1 Definitions for estimating the probability of genetic susceptibility – $P(G)$

Assume a population (P) of (n) individuals: ($i = 1,2,…,n$)

$P(MS)$	=	The life-time probability of developing multiple sclerosis (MS) in the population
(G_i)	=	Genotype of the (i^{th}) individual in the population
$P(MS \mid G_i) = z_i$	=	Expected life-time probability of MS in the (i^{th}) individual (genotype)
$(G-)$	=	The subset of "non-susceptible" individuals for whom: $P(MS \mid G_i) = 0$
(G_{min})	=	The subset of "minimally susceptible" individuals for whom: $0 < P(MS \mid G_i) < P(MS)$
(G)	=	The subset of "genetically susceptible" individuals for whom: $P(MS \mid G_i) \geq P(MS)$
(G_T-)	=	the combined subset: $(G_{min}) \cup (G-)$
(G_T)	=	the combined subset: $(G_{min}) \cup (G)$
Z, X, Y, W, V	=	sets of: $\{z_i\}$; in the entire population (Z); in the (G) subset (X); in the ($G-$) subset (Y), in the (G_T) subset (W), and in the (G_T-) subset (V)
$P(MS \mid G-)$, $P(MS \mid G_{min})$, $P(MS \mid G)$,	=	Expected life-time probability of MS for individuals in the subsets ($G-$), (G_{min}), or (G). By definition: $P(MS \mid G) > P(MS \mid G_{min}) > P(MS \mid G-) = 0$
p, q	=	$p = P(G)$; $q = P(G \mid MS) = P(G \mid IG_{MS})$
x, x', x_i	=	$x = P(MS \mid G)$; $x' = P(MS \mid G, IG_{MS})$; $x_i = P(MS \mid G_i)$ given $z_i \varepsilon X$
y, y'	=	$y = P(MS \mid G_{min})$; $y' = P(MS \mid G_{min}, IG_{MS})$
r	=	The largest value of $P(MS \mid G_i)$ in the population
$P(MS \mid MZ_{MS})$	=	The conditional life-time probability of an individual developing MS, given that their monozygotic (MZ)-twin either has or will develop MS. This is equal to the proband-wise concordance rate for MZ twins.
$P(MS \mid DZ_{MS}), P(MS \mid S_{MS})$	=	The equivalent definition as for $P(MS \mid MZ_{MS})$ except for the individual having either a dizygotic (DZ) twin or sibling (S) with MS
$P(MS \mid IG_{MS}) = b$	=	$P(MS \mid MZ_{MS})$ adjusted for the impact of an identical genotype (IG) sharing the same childhood and intrauterine micro-environments

with MS is also "genetically-susceptible" {$P(MS, G)$}. By the rules of conditional probability we know that:

$$P(G) = P(MS, G)/P(MS|G)$$

Therefore, in order to estimate $P(G)$, we just need to know (approximately) what these other two probabilities are. Fortunately, these other probabilities can be estimated from directly-observed epidemiological data. For example, it must be the case that $P(MS, G)$ is less than the probability of MS in the population $P(MS)$ and this probability, in turn, can be approximated by the measured prevalence of MS in the population. Also, we will define the term $P(MS|MZ_{MS})$ as the probability that a monozygotic (MZ) twin will get MS given that their co-twin already has (or will develop) MS. This also is a measureable epidemiological parameter – the proband-wise (or case-wise) MZ-twin concordance rate for MS [10]. Moreover, because, this observed rate must be less than the concordance rate in "genetically-susceptible" MZ-twins, the observed rate can be used to approximate the term $P(MS|G)$. Thus, using these two approximations for the populations of North America and northern Europe (where these two measured parameters are reasonably well-established), our estimate for $P(G)$ becomes:

$$P(G) \approx P(MS)/P(MS|MZ_{MS}) \approx (0.001)/(0.25) = 0.004$$

Consequently, this simple "back of the envelope" calculation suggests that the occurrence of "genetic-susceptibility" in the population must be extremely rare (~0.4 %). The present manuscript refines this calculation, providing an estimate for the maximum possible value that this probability (G) can take, given the uncertainties in the estimates for these two measurable parameters and given the possible relationships that these measureable parameters have to the actual parameters of interest {i.e., $P(MS, G)$ and $P(MS|G)$}. In addition, this markedly asymmetric division of individuals into those who are "genetically-susceptible" and those who are not has important implications for the nature of genetic susceptibility to MS. Thus, in contrast to a log-normal model (in which the odds of MS are increased by each independent genetic risk factor that a person possesses), the epidemiological data strongly suggests that most (possibly all) individuals with MS come from the "genetically-susceptible" group and that the population is markedly bimodal with respect to the likelihood (risk) that individual members of the population will develop MS.

Moreover, there seem to be differences in the nature of the pathogenesis of MS between men and women. Thus, genetic risk factors are critical to each. However, because women comprise almost three quarters of the MS population, it is perhaps surprising that men are, if anything, more likely than women to be "genetically-susceptible" to getting MS. Therefore, because of this fact, the final gender distribution must be related to environmental factors (either from differences in exposure between men and women or from differences in the response by women to a given exposure). The environment is known to play a critical role in MS-pathogenesis and at least three separate environmental factors (events) are implicated. Each of these events occurs in the large majority of individuals within the population (i.e., they are population-wide environmental events). One event occurs near birth (either *in utero* or in the immediate post-natal period), another occurs during adolescence, and a third (and possibly more) occurs thereafter. Based on the observed epidemiological data, it can be shown that the basis for the final gender distribution in MS, and for the increasing proportion of women in contemporary MS cohorts, is that susceptible women are much more likely to develop MS than susceptible men under similar environmental conditions.

The remainder of this manuscript (together with the Additional file 1) is devoted to developing these ideas in a more rigorous manner.

Methods
Model definitions for determining genetic susceptibility in MS

1. The definitions used for establishing an upper limit for the probability of being genetically susceptible to MS in the population are listed in Table 1.

Consider a population (P_0) of (n) individuals ($i = 1,2,...,n$), each with their own unique genotype (G_i). Let the term $P(MS)$ be defined as the expected life-time probability that a member of the population will develop MS. This probability is related to a directly-observable population parameter – the disease prevalence. Let the expected life-time probability of getting MS for a specific individual (i.e., for their unique genotype) be defined as the conditional probability $P(MS|G_i)$. Let (Z) be the set of all these individual probability values within the population. Thus:

$$(Z) = \{z_i\};$$

where:

$$\forall G_i \in (P_0) : z_i = P(MS|G_i)$$

Further, let the population be partitioned into three mutually exclusive subsets of individuals based on their individual expected life-time probability values. These three subsets (G), (G_{min}), and (G-) are defined in the following manner:

$(G) = \{G_i \in (P_0) | P(MS|G_i) \geq P(MS)\};$ $P(MS|G) = x$

$(G_{min}) = \{G_i \in (P_0) | P(MS) > P(MS|G_i) > 0\}; \ P(MS|G_{min}) = y$

and:

$(G-) = \{G_i \in (P_0) | P(MS|G_i) = 0\}; \ \ P(MS|G-) = 0$

By these definitions:

$x \geq P(MS) > y > 0$

and:

$P(G) + P(G_{min}) + P(G-) = 1$

The subset $(G-)$, members of which have no chance of getting MS, will be referred to as "non-susceptible"; the subset (G_{min}), members of which have a very small chance of getting MS. will be referred to as "minimally-susceptible"; and the subset (G) will be referred to as "genetically-susceptible". Let the sets (X) and (Y) be the sets of the individual life-time probability values for members of the (G) and (G_{min}) subsets respectively. Thus:

$(X) = \{x_i\};$

where:

$\forall G_i \in (G) : x_i = P(MS|G_i)$

and:

$(Y) = \{y_i\};$

where:

$\forall G_i \in (G_{min}) : y_i = P(MS|G_i)$

Let the combined subset of all genotypes, which are not in the "genetically-susceptible" subset, be defined as:

$(G_T-) = (G_{min}) \cup (G-)$

Let the set (V) be the set of the individual life-time probability values for members of the (G_T-) subset. Thus:

$(V) = \{v_i\};$

where:

$\forall G_i \in (G_T-) : v_i = P(MS|G_i)$

And finally, let the combined subset of all genotypes, which have a non-zero probability of developing MS (G_T), be defined as:

$(G_T) = (G) \cup (G_{min})$

Let the set (W) be the set of the individual life-time probability values for members of the (G_T) subset. Thus:

$(W) = \{w_i\};$

where:

$\forall G_i \in (G_T) : w_i = P(MS|G_i)$

Whether defining genetic susceptibility in this manner or creating these different categories has any utility is not known. Therefore, the value (if any) of these constructs needs to be established. Nevertheless, any population can be partitioned in this manner and such a division makes no assumptions about the underlying distribution of the individual expected life-time probability values within the population. For example, if everyone has the exact same expected life-time probability, $P(MS)$, then everyone will belong to the subset (G) and the subsets (G_{min}) and $(G-)$ will be empty. If everyone has a non-zero expected life-time probability of MS, then the subset $(G-)$ will be empty. If the distribution of individual expected life-time probability values $\{w_i\}$ within the (G_T) subset is normal and centered at $P(MS)$, then the two subsets (G) and (G_{min}) will be symmetrical to each other, $P(G) = P(G_{min})$, and each subset will have the half-normal distribution. If the distribution of individual expected life-time probability values is something else, then individuals will be assigned to the three subsets accordingly. For MS, the subset (G) cannot be empty; neither can it encompass the entire population [2–9].

Results
Estimating the probability of genetic susceptibility – *P(G)*

2. Therefore, from the definitions provided in (1) above, the proportion (p) of "genetically susceptible" individuals in the population (P_0) is:

$p = P(G|P_0) = P(G) > 0$
$1 - p - P(G-) = P(G_{min}) \geq 0$
$1 - p = P(G_T-) > 0$

so that:

$P(MS) = P(G)x + P(G_{min})y = px + \{1 - p - P(G-)\}y$

And also the proportion (q) of "genetically susceptible" individuals in the subset (MS) is:

$q = P(G|MS) > 0$

and:

$(1-q) = P(G_{\min}|MS) \geq 0$

As previously, we will let the term $P(MS|MZ_{MS})$ be defined as the conditional life-time probability that an MZ-twin will develop MS, given that his or her co-twin either already has, or will develop, MS. This probability is related to a directly-observable population parameter – the proband-wise (or case-wise) concordance rate for MZ-twins [10]. Let the purely hypothetical term, $P(MS|IG_{MS})$, be introduced to represent the MZ-concordance rate, which has had the impact of the environment (shared by the twins) removed. Thus, this term envisions what the expected concordance rate would be if the MZ-twins, with their identical genotypes (IG), were to be separated at conception and to grow up independently in different environments (both intrauterine and childhood).

Let the term (b) be defined such that:

$b = P(MS|IG_{MS})$

Because MZ-twins are genetically identical, and because MZ-twining is thought to be non-genetic, therefore, it is assumed (see Additional file 1) that:

$q = P(G|MS) = P(G|MZ_{MS}) = P(G|IG_{MS})$

Let the two quantities (x') and (y') be defined such that:

$x' = P(MS|G, IG_{MS})$

and:

$y' = P(MS|G_{\min}, IG_{MS})$

so that:

$b = qx' + (1-q)y'$

3. From the Additional file 1:

$x' \geq x \geq P(MS) > y' \geq y > 0$

Moreover, because ($0 < q \leq 1$), the final equation in (2) above can easily be rearranged to yield:

$0 < q = (b-y')/(x'-y') \leq 1$

and, therefore:

$x' \geq b$

4. Let the maximum probability of MS (r) within the set (X) be defined such that:

$\forall x_i, x_j \in (X) : x_j = r$; if it is true that : $\forall G_i \in (G)$
$: x_i \leq x_j$

By definition, (r) is also the maximum probability within the set (Z). Moreover, because there must be at least one person in (X), for whom: $x_i \geq b$
therefore, it must also be the case that: $r \geq b$

5. From the Additional file 1:

$x' = x + \sigma_X^2/x$

or:

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

which, because of the constraint that when:

$\sigma_X^2 = 0$;

then:

$x = x'$

and because, by definition:

$x > 0$

this has the unique solution:

$x = x'/2 + \left(\sqrt{(x')^2 - 4\sigma_X^2} \right)/2$

From (3) above, under any circumstance, the following limits must apply:

$x \geq x'/2 \geq b/2$;

and:

$\sigma_X^2 \leq (x')^2/4$

Also, when

$\{x' = b\}$;

then:

$\sigma_X^2 \leq b^2/4$

This theoretical limit for (σ_X^2) is only slightly greater than the maximum possible variance for the set (X), which occurs for a bimodal population [11], in which:

$$P\{x_i = b\} = 0.5;$$

and:

$$P\{x_i = P(MS)\} = 0.5$$

so that:

$$b = r;$$

and:

$$E(X) = x = \{b + P(MS)\}/2$$

and:

$$\sigma_X^2 = E(x_i - x)^2 = \{r - P(MS)\}^2/4 = \{b - P(MS)\}^2/4$$

However, these circumstances describe a bimodal population. Therefore, if these circumstances pertain, the case for a bimodal distribution is already made. Consequently, for any unimodal distribution, (σ_X^2) must be lower than this upper-bound and (x) must be greater than this lower-bound.

6. For example, the uniform distribution is an example of unimodal distribution, which is evenly spread out [11]. The uniform distribution is defined such that:

$$\forall x_i, x_j \in (X) : P(x_i) = P(x_j)$$

where:

$$\sigma_X^2 = \{r - P(MS)\}^2/12$$

and:

$$x = \{r + P(MS)\}/2$$

In this circumstance, from this and from the Additional file 1:

$$\sigma_X^2 = x(x' - x) = \{r - P(MS)\}^2/12$$

which, when $(x' = b)$, can be rewritten to become:

$$\{r + P(MS)\}b/2 - \{r + P(MS)\}^2/4 = \{r - P(MS)\}^2/12$$

This last expression is a quadratic in (r), which otherwise includes only the population parameters $\{b$ and $P(MS)\}$ and, thus, (r) in this circumstance can be estimated based on direct epidemiological observations.

7. For a population with a similar age-structure to the US, the quantity $P(MS)$ will be between one and two times the population prevalence [2]. The prevalence of MS can be very broadly estimated to be

50–250/100,000 in northern populations [2, 3, 12–15], which yields the range-estimate for $P(MS)$ of:

$$0.0005 \leq P(MS) \leq 0.005$$

To estimate the quantity $\{b = P(MS|IG_{MS})\}$ requires an understanding of the impact that the shared childhood and intrauterine micro-environments have on the likelihood that MS will develop. From multiple observations, the shared childhood micro-environment seems to make little difference [15–22]. The shared intrauterine (IU) micro-environment, by contrast, may be important [13, 14, 23–29]. Thus, the Canadian data [23] regarding the recurrence risk for dizygotic (DZ) twins and siblings (S) suggests that this IU effect may be as large as:

$$P(MS|DZ_{MS})/P(MS|S_{MS}) = 0.054/0.029 = 1.86$$

A similar disparity has been noted in a review of the available epidemiological data [14]. Nevertheless, in a recent population-based study from Sweden, DZ-twins and siblings seemed to have the same risk [13, 14].

Assuming the truth lies between these two extremes then, using a very broad range-estimate for the proband-wise concordance rates for MZ twins {i.e., $P(MS|MZ_{MS})$} in northern populations of between 0.15 and 0.40, inclusive [2, 3, 14, 23–29] yields the range-estimate for (b) of:

$$0.15/1.86 = 0.081 \leq \{b = P(MS|IG_{MS})\} \leq 0.4$$

8. Substituting: $\{b = 0.081 ; \text{ and} : P(MS) = 0.005\}$ into the Equations from (6) above, yields the minimum possible estimate for (x) of:

$$x \geq b/2 \geq 0.081/2 = 0.0405$$

Because:

$$P(G) = P(MS, G)/x \leq 2*P(MS, G)/b \leq 2*P(MS)/b$$

therefore:

$$P(G) \leq 2*P(MS)/b \leq 0.01/0.081 = 0.123$$

Consequently, under any circumstance, the maximum possible percentage of the population (P_0) that members of the (G) subset could comprise is 12.3 %.

A lower bound for $P(G)$ can also be established by noting that:

$$x' = P(MS, G|IG_{MS})/P(G|IG_{MS})$$
$$= P(MS, G|IG_{MS})/P(G|MS)$$

Because:

$$b = P(MS|IG_{MS}) \geq P(MS, G|IG_{MS});$$

and because:

$$x' \geq x$$

Therefore:

$$P(G) = P(MS, G)/x \geq P(MS, G)/x' \geq \{P(G|MS)\}^2 \{P(MS)/b\}$$

And, using the definition of (g) from (Additional file 1: Table S1) as:

$$g = P(G|MS)$$

the maximum possible range for $P(G)$ can be expressed as:

$$(g^2)\{P(MS)/b\} \leq P(G) \leq (2)\{P(MS)/b\}$$

9. Nevertheless, as noted in (5) above, this particular upper-bound is that for a bimodal population. Substituting these same values into the Equations from (6) above for the uniform distribution {at: $(x' = b)$} yields the minimum (x) and the maximum variance of (X) for this situation of:

$$r = 0.121; \quad x = 0.063; \quad \sigma_X^2 = 0.0011;$$

and:

$$\sigma_X = 0.034$$

Notably, however, the largest possible variance for any unimodal population [11], has been shown to be:

$$\sigma_X^2 = x(x'-x) = \{r-P(MS)\}^2/9$$

Therefore, using the same estimates for {b and $P(MS)$} as above, yields

$$r = 0.113; \quad x = 0.059; \quad \sigma_X^2 = 0.0013;$$

and:

$$\sigma_X = 0.036$$

From (5) above, as either (b) increases or as (x') increases relative to (b), both the minimum (x) and the maximum variance of (X) increase.

Despite this slightly larger upper-bound for the variance of any unimodal distribution compared to the uniform distribution, several conditions (e.g., the distribution is symmetrical, the median is equal to x, or the mode is equal to x) are sufficient to make the maximum variance be that of the uniform distribution [11]. Nevertheless, the larger estimate for (σ_X^2) – i.e., that for any unimodal distribution – and the smaller estimate for the minimum (x) will be used in the remainder of the calculations.

10. Finally, because the probability of the part cannot exceed the probability of the whole, it follows that:

$$P(MS)/P(G) \geq P(MS, G)/P(G) = P(MS|G) = x$$

Using the result from (9) above that ($x \geq 0.059$), therefore:

$$P(MS)/P(G) \geq x \geq 0.059$$

With simple rearrangement, this condition means that the maximum possible estimate for the upper-bound of $P(G)$, given any unimodal distribution of the set (X), is:

$$P(G) \leq P(MS)/x \leq 0.005/0.059 = 0.085$$

11. Recall that the subset of all genotypes that have a non-zero probability of developing MS was defined as:

$$(G_T) = (G) \cup (G_{min})$$

It is noteworthy, however, that the difference in the expected life-time probability of MS between these two subsets of (G_T) is substantial.
Thus, because:

$$P(MS|G_T-) \leq \{y = P(MS|G_{min})\} < P(MS) \leq 0.005$$

and because:

$$x = P(MS|G) \leq 0.059$$

Therefore:

$$x/y = P(MS|G)/P(MS|G_{min}) > P(MS|G)/P(MS)$$
$$\geq 0.059/0.005 = 12$$

so that:

$$x = P(MS|G) > 12 * P(MS) > 12 * P(MS|G_{min})$$
$$\geq 12 * P(MS|G_T-)$$

In fact, by the definition of the (G_{min}) and (G-) subsets, the expected probability of MS in the set (X) is also more than 12 times the likelihood of MS developing for every other individual member of the population who is not in the (G) subset.

Contrast this with the very small difference in subset means permitted for any symmetric distribution centered on $P(MS)$. Thus, by the definition of symmetry:

$$x-P(MS) = P(MS)-y$$

Because, by definition:

$$0 < y < P(MS)$$

Therefore, in this situation:

$$0 < y < P(MS) \leq x < 2{*}P(MS)$$

12. From (10) above, the fact that:

$$P(MS|G) = x \geq 0.059$$

and:

$$P(G) \leq 0.085$$

and, the definition that:

$$0 < P(MS|G_{min}) = y < P(MS)$$

together with the extreme separation of the subset mean $\{P(MS|G)\}$ from the means of the subsets $\{P(MS|G_{min}), P(MS|GT-), \text{ and } P(MS)\}$, constrain, in important ways, the possibilities for the distribution of (Z), which is the set of individual life-time MS probabilities in the population.

For example, the observation that no more than 8.5 % of the population can possibly be members of a unimodal subset (G) indicates that (Z) can't have a symmetric distribution centered on $P(MS)$. This is because, in such a circumstance:

$$\forall z_i \in (Z) : P(z_i \in X) = P(G) = P(G_T-) = P(z_i \in V)$$

whereas, in fact:

$$P(G) \leq 0.085 << 0.915 \leq P(G_T-)$$

Indeed, this disparity is so large that it precludes even a roughly symmetric distribution for (Z), which is centered on $P(MS)$. Moreover, the extreme separation of both the subset means $\{P(MS|G_T-) \text{ and } P(MS|G)\}$, as well as the means for the set (G) and the (P_0) population {i.e., $P(MS|G)$ and $P(MS)$}, together with the very restricted range for the MS probabilities $\{v_i\}$ in the set (V) – i.e., for individuals in the (G_T-) subset – indicates that the distribution of MS probabilities for the whole population (Z) must be, at least, bimodal [30]. This is illustrated in Fig. 1a for the circumstances in which $\{P(W) \approx P(Z) = 1\}$. The means and variances for the distribution used in this illustration are based on considerations developed in (9) above and, for illustrative purposes only, the two distributions of the bimodal population have been each represented as normal. Moreover, this distribution could also be trimodal – i.e., the set (V) could itself be bimodal – if the both of its subsets $\{(G_{min}) \text{ and } (G-)\}$ are non-empty (e.g., Fig. 1b).

Nevertheless, there are other (unimodal) distributions, which can also be markedly asymmetric. The question, therefore, naturally arises as to how confident we are that the distribution of the odds of MS in the (P_0) population can be distinguished from these unimodal alternatives. As noted above, the extreme separation of the sub-set means suggests that the distribution is bimodal [30]. Nevertheless, the possibility that the distribution conforms to a log-normal model needs to be considered carefully. In the first place, the log-normal distribution is both unimodal and asymmetric and, moreover, this asymmetry can be of any specified degree. In the second place, the log-normal model has considerable theoretical appeal, particularly in the setting of a complex disease such as MS, which is associated with multiple genetic risk factors [6, 7]. Thus, if these multiple risk factors are independent of each other (and there is little doubt that they are sorted independently), then (by the central limit theorem) the resulting probability distribution for the odds of MS will follow a log-normal probability density function [31]. And, indeed, Clayton and colleagues recently concluded, based on experimental evidence, that a log-normal model was appropriate for another complex genetic disease, which is comparable epidemiologically to MS – type I diabetes [31].

Similarly, in MS, a log-normal distribution of the odds could account either for the minimum asymmetry of 91.5 % in (G_{min}) and 8.5 % in (G), or for an even more asymmetric split. Nevertheless, for a log-normal distribution having such a split (i.e., 91.5 %/8.5 %), the mean (t) for that portion of a log-normal population, which is at or above the mean for the entire distribution {i.e., $P(MS)$}, is more than 4-fold less than the minimum mean for the odds at $P(MS|G)$– see Additional file 1. Thus, in this circumstance:

$$P(MS|G) > 4.6{*}t$$

As such, $P(MS|G)$ cannot be the mean for this portion of a log-normal distribution (see Additional file 1). This situation is changed only slightly with even much more asymmetric splits. Thus, even when the distribution is severely asymmetric and $P(G)$ is truly tiny (e.g., 10^{-14}), there is still a more than 3-fold difference between $P(MS|G)$ and the mean of that portion of the log-normal distribution, which is at or above $P(MS)$. Indeed, even in these extreme circumstances:

$$P(MS|G) > 3.1{*}t$$

Consequently, having a mean for the odds of getting MS in the (G) population {i.e., $P(MS|G)$}, which is more than $12{*}P(MS)$, is not compatible, under any circumstance, with the subset (G) simply being a part of a unimodal log-normal distribution (Additional file 1: Figure S2B).

It might be argued that such a bimodal structure implies that evidence of either strong interactions or linkage should be present – neither of which has been found. However, with so many genes involved (>150) and such a small
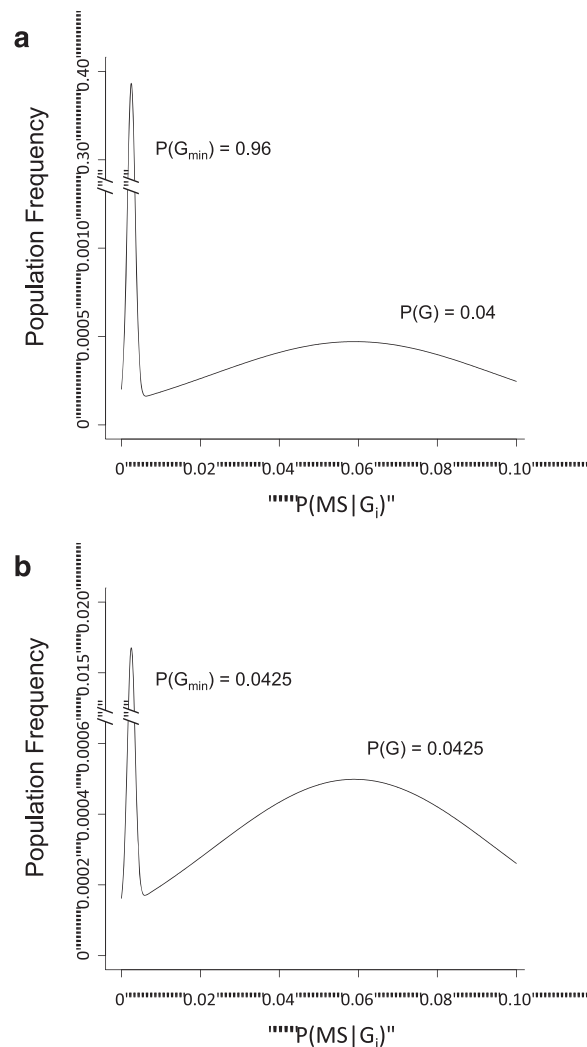
Fig. 1 Plots of the distribution of MS-probabilities in the (*W*) set − i.e., from the combined set (*G_T*) − are shown in two hypothetical circumstances, chosen to match the conditions derived from the Model. Because, as noted in the text, {*PMS*|*G*) > 12 * *P*(*MS*)}, the minimum asymmetry of the distribution is far too great to be explained by a log-normal distribution (see Additional file 1). For illustrative purposes only, the distriburtion of the MS-probabilities in both of the (*G_T*) subsets are plotted as normal. However, because of the large difference in sub-subset means (see text and below), this arbitrary choice makes no difference to the final conclusions [29]. The means (μ) and standard deviations (σ) for the two distributions have been chosen to fit with the following four derived or defined relationships (see text and Additional file 1): 1. $E(X) = x = P(MS|G) \geq 0.059$; 2. $0 < y = P(MS|G \text{ min}) < P(MS) \leq 0.005$; 3. $P(MS) = P(G)x + P(G \text{ min})y$ and: 4. $\sigma_X^2 = E(xi - x)^2 \leq 0.0013$; $\therefore$ $\sigma_X \leq 0.036$. Thus, for the distribution surrounding $P(MS|G)$, these values have been taken to be (μ1 = 0.055 ; σ1 = 0.036) and for the distribution surrounding $P(MS|G \text{ min})$, they have been taken to be (μ2 = 0.0025 ; σ2 = 0.001). The total area underneath the entire (*W*) distribution (depicted) is equal to $P(G_T)$. The areas underneath each sub-distribution {i.e., splitting the two distributions at $P(MS)$}, are equal to $P(G)$ and $P(G_{min})$, respectively. For the first distribution, (μ1) is slightly less than {$x = P(MS|G)$} because a small portion of the left-hand tail of this distribution belongs to the (*G* min) subset. The first plot (panel A), considers the distribution of (*W*) in a circumstance where the distribution of MS probabilities in the combined (*G_T*) set is unlikely to be bimodal {i.e., where $P(G-) = 0$}. In this circumstance, the area underneath (and, thus, the height of) the curve representing the (*X*) distribution (i.e., for the "genetically susceptible" subset) depends upon the value of $P(MS,G_{min})$. Thus: 5. when: $P(MS, G_{min}) \approx 0$; then: $P(G) \approx P(MS)/x$; 6. when: $P(MS, G_{min}) \approx 0.5 * P(MS)$; then: $P(G) \approx 0.5 * P(MS)/x$; 7. when: $P(MS, G_{min}) \approx P(MS)$; then: $P(G) \approx 0$. However, because {$P(MS, G_{min}) \leq 0.56 * P(MS)$}, circumstance #7 is impossible (see Additional file 1). Panel A represents circumstance #6. By contrast, the second plot (panel B), considers the distribution of (*W*) in a circumstance where the distribution of (*Z*) is definitely bimodal. Specifically this figure considers the circumstsnce, in which: $P(G-) > 0.83$; and: $P(G_{min}) \leq P(G)$. In the particular case illustrated − i.e., where $P(G \text{ min}) \approx P(G) \approx 0.0425$ − the distribution is actually trimodal although the zero probability of MS for the subset (*G-*) subset is not depicted in the graph, despite the fact that this subset constitutes the large majority of the population. If the distribution for the (*G* min) subset were assumed to have a uniform distribution, on the plot in Panel B, the plateau of the (*G* min) distribution would be above the population frequency of (0.0008) and, thus, would still be clearly bimodal. This same pattern persists regardless of the actual value chosen for $P(G) \approx P(G_{min})$. Both panels A and B demonstrate the severly bimodal character for the (*W*) distribution that results under any circumstance. Indeed, such severe bimodality will exist, regardless of the acutal shape of the distribution of MS probabilities for members of each of the two subsets of (*G_T*). This is both because of the extreme separation of the subset means and because of the very restricted range for the MS probabilities within the (*G* min) subset [29]

fraction of the population being in the subset (G), this is not the case. Indeed, considering only susceptibility genes, it seems very likely that almost all MS patients will have a unique genotype (Additional file 1) and, empirically, this seems to be true. Thus, using the first 95 MS-associated SNPs identified in the WTCCC data set [6, 7], 105 of the genotypes (at these SNP locations) were identical in, at least, 1 pair of MS cases. Nevertheless, regardless of the basis for these apparent duplications, for all of the other 10,643 MS cases in this dataset, their genotypes (at these SNP locations) were unique. Moreover, none of these apparently duplicated genotypes bore any obvious resemblance to each other – sharing identity at only 43 (on average) and 59 (at most) of the 95 SNPs. Under such circumstances, almost certainly, there will be no linkage and no strong interactions, even if the population (P_0) is bimodal.

13. There are two further possibilities. First, it could be that $\{P(G\,min) \leq P(G)\}$. In this case, the distribution of MS probabilities within (W) – i.e., for members of the (G_T) subset – could either be symmetric or not. This relationship necessarily pertains for any symmetric distribution of (W) because, by definition, and from the above considerations:

$$P(MS|G_T) = P(MS, G_T)/P(G_T)$$
$$= P(MS)/P(G_T) \geq P(MS)$$

From this it follows that, any symmetric distribution for (W) must be centered on some probability value (μ) such that:

$$x \geq \mu = PMS|G_T) \geq P(MS)$$

and, thus, that:

$$P(G) \geq P(G_{min})$$

However, regardless of the nature of the distribution of (W), if this relationship holds, then it must also be the case that:

$$P(G-) = 1 - P(G) - P(G_{min}) \geq 1 - 2 * P(G) \geq 0.83$$

In addition, the actual values that $P(G_{min})$ can take will depend, in part, upon value of $P(G)$.

For example, when $P(G)$ is at its upper-bound of:

$$P(G) = P(MS)/x$$

then:

$$P(G_{min}) = 0$$

From this point, as the quantity $P(G_{min})$ becomes larger, the quantity $P(G)$ will have to become smaller in order to maintain the relationship:

$$P(MS) = P(G)x + P(G_{min})y$$

Therefore, we conclude that, if $P(G min) \leq P(G)$, then it must also be the case that the large majority of the population (P_0) must be in the (G-) subset and that the distribution of (Z) is, at least, bimodal.

14. Second, it could be that: $\{P(G_{min}) > P(G)\}$. In this case, the extreme separation of the subset means within (W) – i.e., the separation of $P(MS|G)$ from $P(MS|G_{min})$ – together with the very restricted range for the MS probabilities in the (G_{min}) subset, again indicates that the distribution of $\{w_i\}$ within the set (W) must be bimodal [30]. This is illustrated in Fig. 1b, in which the resulting extreme bimodality is demonstrated even for the circumstance where:

$$P(G_{min}) \approx P(G) \approx 0.0425$$

Again, the means and variances for the distributions used in this illustration are based on considerations developed in (9) above. This same pattern (as illustrated) persists regardless of the value chosen for $P(G) \approx P(G_{min})$.

Discussion

Using very broad range-estimates for the basic epidemiological parameters of MS-prevalence and the recurrence risk of MS in MZ-twins, this analysis indicates that no more than 8.5 % of individuals in the general population can possibly be "genetically susceptible" to developing MS as herein defined. In all likelihood, this percentage is actually much smaller [2, 3]. In addition, as demonstrated in Additional file 1, more than 43 % (and, likely, more than 84 %) of MS cases develop through this genetic pathway. Importantly, each of these estimates are based on directly-observable population parameters, which have been repeatedly verified in different parts of the world.

The implications of these conclusions are substantial. Recall that the subset (G_{min}) is defined as consisting of only those individuals who have very low individual expected life-time probability values of more than zero but less than $P(MS)$. By contrast, individuals in the (G) subset, collectively, have an expected life-time probability of MS at least 12 times the maximum possible either for that of the (G_{min}) subset as a whole or for that of any individual member of this subset.

Consequently, even though this analysis does not assume that a group of individuals who are "genetically susceptible" is distinct from other individuals in the population, it does, in fact, establish that these two groups can be so distinguished. Thus, based on the

considerations developed in (13) and (14) above, there are two possibilities.
If:

$$P(G_{min}) \leq P(G)$$

Then:

$$P(G-) = 1 - P(G_{min}) - P(G) \geq 1 - 2 * P(G) > 0.83$$

And, in this situation, the two groups are distinguished by the fact that members of the "non-susceptible" subset $(G-)$, which represents the overwhelming majority of the population, are at no risk of developing MS, regardless of their environmental exposure.

Conversely, if: $P(G_{min}) > P(G)$

Then, from (14) above, even considering only those individuals who belong to the combined (G_T) subset, the extreme separation of the means of $P(MS)$ and $P(MS|G)$, together with the very narrow range of MS probabilities within the (G_{min}) subset, requires the distribution of individual expected life-time MS probabilities in the set (W) to be bimodal [30] and, thus, to reflect the existence of two distinct groups of MS patients (see Fig. 1a; Additional file 1: Figure S2B). In this circumstance, the two classes of MS patients are distinguished by the fact that the MS, which develops, seems to be caused by two, fundamentally different, pathophysiological mechanisms. If the subset (G_{min}) is, indeed, non-empty, then, in the first mechanism, MS is very improbable, the genetic contribution seems to be minor, and, thus, environmental factors are likely to be primary. By contrast, in the second mechanism, MS is comparatively much more likely to occur and the combination of both genetic and environmental events are each critical determinants of disease. Importantly, if a second group of individuals with non-zero probabilities of MS actually exists, then, despite the fact that environmental factors would be involved in both mechanisms, there is no reason to expect that the environmental events involved in the first pathway are the same as (or even similar to) those involved in the second pathway.

In most cases of MS, the genetic route seems to dominate (Additional file 1). Indeed, more than 94 % of concordant MZ-twins {i.e., individuals in the (MS, IG_{MS}) sub-subset} come from the (G) subset (Additional file 1). However, these observations do not mean that the genetics primarily determines the disease, even in these cases. In fact, the increasing prevalence of MS world-wide [32–40], its increasing prevalence in women [32, 34, 36, 39, 40], and the change in prevalence and MZ-twin concordance based on latitude [3, 12] are better explained by differences in environmental exposure and by a women's greater physiological responsiveness to environmental events (Additional file 1) than by any differences in genetic

susceptibility between groups or regions [2, 3]. Also, given the wide disparity between the probability of developing MS between men and women, the set (G) must itself be bimodal (Additional file 1).

These conclusions also have important implications with regard to our ability, ultimately, to determine a person's risk through genetic analysis. Indeed, if more than 84 % of MS occurs in the genetically susceptible population (G) and less than 8.5 % of the population is susceptible, it should be possible, in theory, to characterize a person's individual risk with high sensitivity and specificity. The fact that it has proven difficult to do this so far probably relates, in part, to the fact that the genetic associations have been defined on the basis of SNPs [4–7] rather than on the basis of more extended SNP-haplotypes [41, 42].

Conclusions

It is possible to distinguish two classes of persons in the general population, indicating either that MS can be caused by two fundamentally different pathophysiological mechanisms or that the large majority of the population is at no risk of the developing this disease regardless of their environmental experience. Moreover, although environmental-factors would play a critical role in both mechanisms (if both exist), there is no reason to expect that these factors are the same (or even similar) between the two.

The definitions for the parameters used in the model are presented in Table 1.

Ethics approval and consent to participate
Ethical approval and consent from patients were not required because neither human subjects nor animals were used. All supporting data is available to researchers in the Additional file 1 provided.

Consent for publication
This manuscript contains does not contain any person's individual data.

Availability of data and materials
All of the data contained in this manuscript is publically available.

Additional file

Additional file 1: Contains several derivations related to the points made in the main text. It also contains 3 **Table S1** and **Figure S1**. (PDF 2039 kb)

References
1. Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. Biochim Biophys Acta. 2014;1842(10): 1889–95.
2. Goodin DS. The genetic and environmental bases of complex human disease: Extending the utility of twin-studies. PLoS One. 2012;7(12), e47875.
3. Goodin DS. The epidemiology of multiple sclerosis: insights to disease pathogenesis. Handb Clin Neurol. 2014;122:231–66.
4. Ramagopalan SV, Anderson C, Sadovnick AD, Ebers GC. Genomewide study of multiple sclerosis. N Engl J Med. 2007;357:2199–200.
5. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, Piccio L, Raychaudhuri S, Tran D, Aubin C, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009;41:776–82.
6. The International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell mediated immune mechanisms in multiple sclerosis. Nature. 2011;476:214–9.
7. International Multiple Sclerosis Genetics Consortium. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2014;45:1353–60.
8. Dyment DA, Herrera BM, Cader MZ, Willer CJ, Lincoln MR, Sadovnick AD, Risch N, Ebers GC. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. Hum Mol Genet. 2005;14:2019–26.
9. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Baker PI, Gabriel SB, Mirel DB, Ivinson AJ, et al., and International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007;357:851–62.
10. Witte JS, Carlin JB, Hopper JL. Likelihood-Based Approach to Estimating Twin concordance for dichotomous traits. Genetic Epidemiol. 1999;16:290–304.
11. Jacobson HI. The maximum variance of restricted unimodal distributions. Ann Math Stat. 1969;40:1746–52.
12. Rosati G. The prevalence of multiple sclerosis in the world: an update. Neurol Sci. 2001;22:117–39.
13. Hansen T, Skytthe A, Stenager E, Petersen HC, Kyvik KO, Brønnum-Hansen H. Risk for multiple sclerosis in dizygotic and monozygotic twins. Mult Scler. 2005;11:500–3.
14. Hansen T, Skytthe A, Stenager E, Petersen HC, Brønnum-Hansen H, Kyvik KO. Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. Mult Scler. 2005;11:504–10.
15. O'Gorman C, Lin R, Stankovich J, Broadley SA. Modelling genetic susceptibility to multiple sclerosis with family data. Neuroepidemiology. 2013;40:1–12.
16. Bager P, Nielsen NM, Bihrmann K, Frisch M, Wohlfart J, Koch-Henriksen N, Melbye M, Westergaard T. Sibship characteristics and risk of multiple sclerosis: A nationwide cohort study in Denmark. Am J Epidemiol. 2006;163:1112–7.
17. Compston A, Coles A. Multiple sclerosis. Lancet. 2002;359:1221–31.
18. Dyment DA, Yee IML, Ebers GC, Sadovnick AD, the Canadian Collaborative Study Group. Multiple sclerosis in stepsiblings: Recurrence risk and ascertainment. J Neurol Neurosurg Psychiatry. 2006;77:258–9.
19. Ebers GC, Sadovnick AD, Dyment DA, Yee IM, Willer CJ, Risch N. Parent of-origin effect in multiple sclerosis: observations in half-siblings. Lancet. 2004;363:1773–4.
20. Ebers GC, Yee IML, Sadovnick AD, Duquette P, the Canadian Collaborative Study Group. Conjugal multiple sclerosis: Population based prevalence and recurrence risks in offspring. Ann Neurol. 2000;48:927–31.
21. Sadovnick AD, Yee IML, Ebers GC, the Canadian Collaborative Study Group. Multiple sclerosis and birth order: A longitudinal cohort study. Lancet Neurol. 2005;4:611–7.
22. Sadovnick AD, Ebers GC, Dyment DA, Risch NJ, the Canadian Collaborative Study Group. Evidence for genetic basis of multiple sclerosis. Lancet. 1996; 347:1728–30.
23. Willer CJ, Dyment DA, Rusch 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.
24. French Research Group on Multiple Sclerosis. Multiple sclerosis in 54 twinships: Concordance rate is independent of zygosity. Ann Neurol. 1992;32:724–7.
25. Islam T, Gauderman WJ, Cozen W, Hamilton AS, Burnett ME, Mack TM. Differential twin concordance for multiple sclerosis by latitude of birthplace. Ann Neurol. 2006;60:56–64.
26. Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. Neurology. 1994;44:11–5.
27. Ristori G, Cannoni S, Stazi MA, Vanacore N, Cotichini R, Alfò M, Pugliatti M, Sotgiu S, Solaro C, Bomprezzi R,et al. Multiple sclerosis in twins from continental Italy and Sardinia: A Nationwide Study. Ann Neurol. 2006;59:27–34.
28. Kuusisto H, Kaprio J, Kinnunen E, Luukkaala T, Koskenvuo M, Elovaara I. Concordance and heritability of multiple sclerosis in Finland: Study on a nationwide series of twins. Eur J Neurol. 2008;15:1106–10.
29. Fagnani C, Neale MC, Nisticò L, et al. Twin studies in multiple sclerosis: A meta-estimation of heritability and environmentality. Mult Scler. 2015; 21(11):1404–13.
30. Freeman JB, Dale R. Assessing bimodality to detect the presence of a dual cognitive process. Behav Res. 2013;45:83–97.
31. Clayton DC. Prediction and interaction in complex disease genetics: Experience in Type 1 diabetes. PLoS Genet. 2009;5(7), e1000540.
32. Hernán MA, Olek MJ, Ascherio A. Geographic variation of MS incidence in two prospective studies of US women. Neurology. 1999;53:1711–8.
33. Koch-Henriksen N. The Danish Multiple Sclerosis Registry: a 50-year follow-up. Mult Scler. 1999;5:293–6.
34. Freedman DM, Dosemeci M, Alavanja MC. Mortality from multiple sclerosis and exposure to residential and occupational solar radiation: A case control study based on death certificates. Occup Environ Med. 2000;57:418–21.
35. 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;8:463–9.
36. Barnett MH, William DB, Day S, Macaskill P, McLeod JG. Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: a 35-year study. J Neurol Sci. 2003;213:1–6.
37. 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.
38. Ranzato F, Perini P, Tzintzeva E, Tiberio M, Calabrese M, Emani M, Davetag F, De Zanche L, Garbin E, Verdelli F, et al. Increasing frequency of multiple sclerosis in Padova, Italy: a 30 year epidemiological survey. Mult Scler. 2003;9:387–92.
39. Sarasoja T, Wikström J, Paltamaa J, Hakama M, Sumelahti ML. Occurrence of multiple sclerosis in central Finland: a regional and temporal comparison during 30 years. Acta Neurol Scand. 2004;110:331–6.
40. Orton SM, Herrera BM, Yee IM, Valdar W, Dyment DA, Ramagopalan SV, Sadovnick AD, Ebers GC, and the Canadian Collaborative Study Group. Sex ratio of multiple sclerosis in Canada: A longitudinal study. Lancet Neurol. 2006;5:932–6.
41. Goodin DS, Khankhanian P. Single Nucleotide Polymorphism (SNP)-Strings: An Alternative Method for Assessing Genetic Associations. PLoS One. 2014; 9(4), e90034.
42. Khankhanian P, Gourraud PA, Lizee A, Goodin DS. Haplotype-based approach to known MS-associated regions increases the amount of explained risk. J Med Genet. 2015;52(9):587–94.