OPEN ACCESS Freely available online

PLOS ONE

The Genetic and Environmental Bases of Complex Human-Disease: Extending the Utility of Twin-Studies

Douglas S. Goodin*

Department of Neurology, University of California San Francisco, San Francisco, California, United States of America

Abstract

Making only the assumption that twins are representative of the population from which they are drawn, we here develop a simple mathematical model (using widely available epidemiological information) that sheds considerable light on the pathogenesis of complex human diseases. Specifically, for the case of multiple sclerosis (MS), we demonstrate that the vast majority of patients (≥94%), possibly all, require genetic susceptibility in order to get MS. Nevertheless, only a tiny fraction of the population (≤2.2%) is actually susceptible to getting this disease; a finding which is highly consistent in all of the studied populations across both North America and Europe. Men are more likely to be susceptible than women although susceptible women are more than twice as likely to actually develop MS compared to susceptible men (i.e., they have a greater disease penetrance). This is because women are more responsive to the environmental factors involved in MS pathogenesis than men. These differences account for the current gender-ratio (3:1, favoring women) and also for the increasing incidence of MS in women around the world. By contrast, the most important genetic marker for MS susceptibility (DRB1*1501) influences the likelihood of susceptibility but not the penetrance of the disease. Nevertheless, even for this major susceptibility allele, only a very small fraction of DRB1*1501carriers (<5%) are susceptible to getting MS and for only a minority of MS patients (~41%) does this allele contribute to their susceptibility. Moreover, each copy of this allele seems to make an independent contribution to susceptibility. Finally, at least three environmental events are necessary for MS pathogenesis and, during the course of their lives, the large majority of the population (≥69%) experiences an environmental exposure, which is sufficient to produce MS in, at least, some susceptible genotypes. Also, susceptible men (compared to susceptible women) have a lower threshold, a greater hazard-rate, or both in response to the environmental factors involved in MS pathogenesis.

Citation: Goodin DS (2012) The Genetic and Environmental Bases of Complex Human-Disease: Extending the Utility of Twin-Studies. PLoS ONE 7(12): e47875. doi:10.1371/journal.pone.0047875

Editor: Frank Emmert-Streib, Queen's University Belfast, United Kingdom

Received February 7, 2012; **Accepted** September 24, 2012; **Published** December 18, 2012

Funding: The author has no support or funding to report.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: douglas.goodin@ucsf.edu

Introduction

The etiologies of many chronic human-diseases are complex and their basis often includes both the individual's genotype and their environmental experiences [1]. Recurrence-risk data for disease in monozygotic (MZ)-twins, dizygotic (DZ)-twins, and siblings (S) of an affected-proband, provides insight to the nature of disease-susceptibility [2]. For example, if the disease-risk in MZ-twins of the affected-proband is substantially greater than the risk in DZ-twins, this suggests the importance of genetics to disease pathogenesis. Similarly, if this disease-risk is considerably less than 100% in MZ-twins, this suggests of the importance of environmental-factors. In fact, by assuming that both MZ- and DZ-twins have similarly "shared environments" and that twins are representative (genetically) of the general population, the difference in disease-risk between MZ-twins and DZ-twins can estimate the proportion of the variance in disease-occurrence that can be attributed to heritable-factors, shared environmental-factors, or non-shared environmental-factors [2].

While these approaches offer a broad outline of disease pathogenesis, epidemiological data could potentially provide more quantitative information. Here we develop a simple mathematical model, using concordance (recurrence-risk) data from twin and familial studies, to elucidate the nature and frequency of genetic-susceptibility to complex human-diseases. This is not to downplay the importance of environmental factors in disease pathogenesis, which, as noted both here and elsewhere [1–4], is considerable. Neither is this an exploration of the genes themselves. Rather, it is an attempt to understand the nature of genetic-susceptibility, the importance of environmental-risk, and to delineate the constraints on the genetic and environmental bases of these complex diseases, which are imposed by certain epidemiological observations or facts.

Although broadly applicable to many chronic complex diseases, these principals are here applied specifically to the example of multiple sclerosis (MS), because of the ready-availability of both familial-recurrence data and world-wide epidemiological information [3–21]. For example, it is well-established that the prevalence of MS in the northern regions of either Europe or North America is approximately 0.1–0.2% [3]. For individuals with an affected family member, the MS-risk increases roughly in proportion to the amount of shared genetic-information between the affected-relative and the individual [3,6,9,13,17,20]. Although at least three environmental-factors, each acting at specific periods during a person's life, seem critical to disease pathogenesis [4], genetic-factors are also, unquestionably, part of a causal pathway leading to MS [3–21].

The earliest and the best-established genetic-association with MS-susceptibility is the HLA-DRB1 locus on the short-arm of chromosome six [21–26]. Within this locus, the DRB1*1501 allele has the strongest and most consistent association with MS in both northern European and North American populations [21–26]. Nevertheless, despite its importance, only about half of MS patients are DRB1*1501 carriers and only a small percentage of carriers (<1%) will ever develop the disease [21–26]. These observations indicate that other genes, at different locations, are necessary and/or sufficient to produce MS-susceptibility [26,27].

The Genetic Model

The definitions for the principle model terms are presented in Table 1. Further model definitions, assumptions, and explanatory tables are presented in Appendix S1; Section A. In addition, Appendix S1 presents both the conceptualizations of genetic-susceptibility and environmental-risk used for the model (Section B) as well as a rigorous presentation of model development (Sections C–F). The basic epidemiological and familial-concordance data used for quantitative analysis of model implications are presented in Tables 2,3,4, and the detailed MZ-twin data regarding DRB1*1501 and gender are presented in Tables 5&6. The principle conclusions and range-estimates derived for the model are summarized in Table 7.

We define disease-penetrance as the conditional life-time probability of disease given the specific genotype for a member of the general population (see Appendix S1; Section B). We can also partition the general population into the mutually exclusive sets of carriers *(HLA+)* and non-carriers *(HLA−)* of at least one copy of the DRB1*1501 allele.

In MS, it is well established [21–26] that:

$$P(HLA+|MS) > P(HLA+)$$

Therefore, it must also be the case that:

$$P(MS|HLA+) > P(MS) > P(MS|HLA-)$$

This last statement indicates, unequivocally, that some genotypes have a greater penetrance than others and, therefore, that at least one genotype must have the least penetrance of any. Consequently, the individual genotypes can be partitioned into two subsets, *(G)* and *(G−)*, where the term $P(MS|G-)$ or, more generally, $P(D|G-)$, represents the disease-penetrance of the least-penetrant genotype in the population (see Appendix S1; Section B).

It could be the case that: $P(D|G-)=0$

However, if so, and if we define (Table 1) disjoint sets of individual environmental experiences or exposures that either are *(E)* or are not *(E−)* sufficient to produce disease environmentally (see Appendix S1; Sections A&B), then this circumstance requires that:

$$P(D|G-) = P(D,E|G-) + P(D,E-|G-) = 0$$

and, thus, that: $P(D,E|G-)=0$

Consequently, the circumstance in which: $P(D|G-)=0$; implies that "purely environmental" disease does not occur (see Appendix S1; Sections A&B).

Conversely, if "purely environmental" disease is possible, then:

$$P(D|G-) > 0$$

Members of the subset *(G)* are said to be "genetically susceptible" whereas members of the subset *(G−)* are said to be "genetically non-susceptible". In this conceptualization, genetic-susceptibility is, by definition, binary (quantitatively) although the subset of susceptible individuals *(G)* could, at least theoretically, encompass virtually the entire population (i.e., all but one genotype) and the penetrance of the different susceptible genotypes within *(G)* could range from nearly zero to one (Appendix S1; Section B). The terms $P(D|G)$ and $(\mathbf{z})$ are used interchangeably and represent the expected disease-penetrance in genetically susceptible individuals. In the model, we imagine that, within the population of all susceptible-individuals *(G)*, each individual has their own individual-specific susceptibility-genotype, and each genotype has its own genotype-specific penetrance-value. The penetrance of disease for the (i^{th}) genotype (G_i) within *(G)* is represented as either $P(D|G_i)$ or $(\mathbf{z}_i)$. The term $P(D)$ represent the probability that a random member of the general population will develop the disease within their life-time. The set *(D,G−)* represents those cases of disease, which occur in individuals who are not genetically susceptible. We also consider the different circumstances that exist for men *(M)* and women *(F)* and, in addition, we partition the *(HLA+)* subset into those individuals who carry either one *(1HB+)* or two *(2HB+)* copies of the DRB1*1501 allele.

Without making any assumptions, two definitional statements can be made:

$$1.\ P(D) = P(D,G) + P(D,G-)$$

$$\text{and}: 2.\ P(G) = P(D,G)/P(D|G)$$

$$\text{or, in the case of MS}: P(MS) = P(MS,G) + P(MS,G-) \quad (1)$$

$$\text{and}: P(G) = P(MS,G)/P(MS|G) = P(MS,G)/\mathbf{z} \quad (2)$$

From Equation (1), there must be some constant $(0 \le g \le 1)$ such that:

$$P(MS,G) = g * P(MS) \le P(MS) \quad (3)$$

and that: $P(G) = g * P(MS)/P(MS|G)$; and also: $g = P(G|MS)$

Moreover, because some MS-cases involve genetic-factors [3,6,9,13,17,20], then it also must be the case that: $P(G-|MS) < 1$; or, equivalently: $g > 0$

The purpose of the model is to use directly observable epidemiological information of the type presented in Tables 2,3,4,5,6 in order to estimate a variety of unknown quantities including:

$P(G)$, $P(G|MS)$, $P(F|G)$, $P(G|HLA+)$, $P(G|HLA-)$, $P(G|2HB+)$, $P(G|1HB+)$, $P(MS|G)$, $P(MS|G-)$, $P(MS|G,HLA+)$, $P(MS|G,HLA-)$, $P(MS|G,F)$, $P(MS|G,M)$, $P(MS|G,2HB+)$, $P(E)$, $P(MS|E,G,F)$, and $P(MS|E,G,M)$.

Table 1. Model definitions*

$P(MS)$	=	The life-time probability of developing MS in the general population. [equated to the prevalence of the disease]
(G) , $(G–)$	=	Sets of persons who either are (G) or are not $(G–)$ genetically susceptible to MS
$(G1)$, $(G2)$	=	Two mutually exclusive subsets of (G); one consisting of high-penetrance genotypes $(G1)$ and the other consisting of low-penetrance genotypes $(G2)$. $(G1) + (G2) = (G)$
$(G0)$, $(G3)$	=	Mutually exclusive sets of genetically susceptible individuals who depend upon $(G0)$ or don't depend upon $(G3)$ environmental events to get MS. $(G0) + (G3) = (G)$
$P(MS\|G–)$	=	Penetrance of the least penetrant genotype in the population
$P(MS\|G_i)$	=	Penetrance of the i^{th} genotype in the set (G)
$P(MS\|G)$	=	Expected penetrance of the set (G) ; $P(MS\|G) = E\{P(MS\|G_i)\}$
σ_{zi}^2	=	Penetrance Variance within the set (G) ; $\sigma_{zi}^2 = Var(G_i)$
$P(MS\|IG_{MS})$	=	**b** = the conditional life-time probability of developing a MS, given that the person's MZ-twin has MS; adjusted to exclude the impact of twins sharing intra-uterine (IU) and childhood (CH) environments.
(MZ_{MS}) , (DZ_{MS}) , (S_{MS})	=	Sets of persons with a monozygotic (MZ)-twin, a dizygotic (DZ)-twin, or a sibling (S) who either has or will develop MS.
(IU) , (CH)	=	Sets of persons who share, with an MS-proband, either the same intra-uterine (IU) or a similar childhood (CH) environment
(E) , $(E–)$	=	Sets of persons who either do (E) or do not $(E–)$ experience a sufficient environmental exposure to produce MS (see Section B)
(FT) , (ST)	=	The sets of first (FT) or second (ST) twins of an MZ-twin pair
$(Gx+)$, $(Gx–)$	=	The set of persons who either possess $(Gx+)$ or don't possess $(Gx–)$ the particular genetic characteristic (Gx).
$(HLA+)$, $(HLA–)$	=	The set of persons who either carry $(HLA+)$ or don't carry $(HLA–)$ at least one HLA DRB1*1501 allele. $(HLA+) = (2HB+) + (1HB+)$
$(1HB+)$, $(2HB+)$	=	The sets of persons who carry one $(1HB+)$ or two $(2HB+)$ copies of the DRB1*1501 allele.
$(1HB–)$	=	The set of persons who carry one copy of a non-DRB1*1501 allele $P(1HB-,1HB-)=P(HLA-)$; $P(1HB+,1HB-)=P(1HB+)=P(1HB-)$
(F) , (M)	=	Sets consisting of either women (F) or men (M)
a , **a′**	=	$P(MS, G) / P(G1)$ = **a** ; and: $P(MS, G) / P(G2)$ = **a′**
b , **b′**	=	$P(MS\|IG_{MS})$ = **b** ; and: $P(MS\|G, IG_{MS})$ = **b′**
x	=	$P(MS\|G1)$ = Expected Penetrance of the high-penetrance subset
y	=	$P(MS\|G2)$ = Expected Penetrance of the low-penetrance subset
z	=	$P(MS\|G)$ = Expected Penetrance for the entire set (G)
$\mathbf{z_t}$, $\mathbf{z_s}$	=	$P(MS\|G, Gx+)$ = $\mathbf{z_t}$; and: $P(MS\|G, Gx–)$ = $\mathbf{z_s}$
t	=	$P(MS\|Gx+, IG_{MS}) = P(MS, G\|Gx+, IG_{MS})$
t′	=	$P(MS\|G, Gx+, IG_{MS})$
s	=	$P(MS\|Gx–, IG_{MS}) = P(MS, G\|Gx–, IG_{MS})$
s′	=	$P(MS\|G, Gx–, IG_{MS})$
p	=	$P(G1\|G) = P(G1, G) / P(G) = P(G1) / P(G)$; $(G1) \subset (G)$
g	=	$P(G\|MS) = P(G\|IG_{MS})$
g_1	=	$P(G\|Gx+, MS) = P(G\|Gx+, IG_{MS})$
g_2	=	$P(G\|Gx–, MS) = P(G\|Gx–, IG_{MS})$
A_0	=	$P(Gx+)$
A	=	$P(Gx+\|MS) = P(Gx+\|IG_{MS})$
MAF	=	Mean allelic frequency – defined as the frequency of an "allelic state" {e.g., the "$(HLA–)$ allele" at the DRB1 gene = one "non-1501" allele}
HWE	=	Hardy-Weinberg Equilibrium

*See Appendix S1 (Section A) for additional model definitions.

doi:10.1371/journal.pone.0034034.t001

In addition, we also use this information to provide other insight to the nature and basis of genetic-susceptibility in different sub-populations.

Basic model assumptions and derivations

To begin, we define $P(MZ_{MS})$ as the life-time probability that, for an individual from an MZ-twinship, their co-twin either has or will develop MS, independent from whatever has happened or will happen to them. Because there is no known genetic-predilection for having MZ-twins, the genetic composition of the MZ-twin population is assumed to be "representative" of the general population. The definition of "representative" is made explicit by Assumptions (A5)&(A6) – Appendix S1 (Section A). Thus, it is assumed (Appendix S1; Section A; Assumption A5) that $P(MZ_{MS})$ for the first twin (FT) is the same as it is for the second twin (ST), and that the genetic-composition of the sets (MS) and (MZ_{MS}) are the same. In this case:

Table 2. Epidemiological data used in the model[#]

	Population	Women	Men
$P(MS) = P(IG_{MS})$ *	0.0015	0.00204	0.00096
$P(F\|MS) = P(F\|IG_{MS})$ *	0.68		
$P(F\|MS, MZ_{MS}) = P(F\|MS, IG_{MS})$ *	0.92		
$P(F\|HLA+, MZ_{MS}) = P(F\|HLA+, IG_{MS})$ *	0.74		
$P(F\|MS, HLA+, MZ_{MS}) = P(F\|MS, HLA+, IG_{MS})$ *	$> 0.82^{\dagger}$		
Raw MZ-twin Concordance = $P(MS\|MZ_{MS})$ *	0.25	0.34	0.067
Adjusted MZ-twin Concordance = $P(MS\|IG_{MS})$ = b **	0.134	0.183	0.036
Raw DZ-twin Concordance = $P(MS\|DZ_{MS})$ *	0.054	0.051	0.057
Raw Sibling Concordance = $P(MS\|S_{MS})$ *	0.029	0.039	0.019
UCSF (#1) – $P(HLA+\|MS) = P(HLA+\|IG_{MS})$ – Cases [††]	0.56	0.57	0.52
Canadian – $P(HLA+\|MS) = P(HLA+\|IG_{MS})$ – Cases [##]	0.55	0.60	0.52
Canadian – $P(HLA+)$ – Controls [##]	0.24	$\sim$ 0.24	$\sim$ 0.24
UCSF (#2) – $P(HLA+\|MS) = P(HLA+\|IG_{MS})$ – Cases [††]	0.46	0.49	0.39
UCSF (#2) – $P(HLA+)$ – Controls [††]	0.20	0.18	0.22

[#]HLA+ = carrier of $\geq$ 1 copy of the DRB1*1501 allele
*From Canadian Data [21], based on a prevalence of 150 per 10^5 population and split into men and women according to [15]. Concordance rates presented as "proband-wise" rates [30].
[†]Data unavailable on the 2 male patients [21]. The worst case is: 9/11 = 0.82
**See: Prop. (1.4) of Appendix S1 (Section C)
[##]Canadian HLA data: D Sadovnick (personal communication). Based on $\sim$ 3,000 cases and $\sim$ 400 Controls (% women not available). Control rates confirmed in a much larger transplant database.
[††]UCSF Databases: J Oksenberg (personal communication)
UCSF #1 (GeneMSA) - 485 cases (68% women) and 431 Controls (66% women)
UCSF #2 (IMSGC) - 779 cases (76% women)
doi:10.1371/journal.pone.0034034.t002

$$P(MS|FT) = P(MS|ST) = P(MZ_{MS}) = P(MS)$$

Moreover, it is assumed (Appendix S1; Section A; Assumption A6) that the genetic-composition of the sets *(G,FT)*, *(G,ST)*, and *(G)* are the same. Under these conditions:

$$P(G|FT) = P(G|ST) = P(G)$$

Importantly, for MS, the direct observational data supports the validity of the assumption that twins are "representative". Thus, both the twin-rates in an MS-population and the probability of MS in twins are as expected for the population as a whole [21]. These same assumptions also underlie the "classical" twin methods discussed earlier [2].

In addition, we assume that *P(MS)* is approximately equal to the observed prevalence of MS in the general population (Appendix S1; Section A; Assumption A1). Nevertheless, because most prevalence-estimates use, as their denominator, the total population in the region and, because almost all MS cases begin (clinically) between the ages of 15 and 45 years [3] and most survive at least into late middle-age [28], Assumption (A1), almost certainly, underestimates *P(MS)*. A better estimator of *P(MS)* – the life-time risk of MS – will be derived from the prevalence in those aged 45–55 years (Appendix S1; Section A). In this age-bracket, new incident-cases are unlikely to occur [3] and substantial early mortality from MS is unlikely to have yet happened [28]. If so, the true *P(MS)* could, potentially, be double the estimate derived from the population-prevalence (e.g., [29]). The impact of this possibility is considered further in Appendix S1 (Section B) and also, subsequently, as a part of our sensitivity analyses.

MZ-twins, in addition to sharing the same nuclear and mitochondrial genes, also share the same intra-uterine *(IU)* and similar childhood *(CH)* environments. We further assume (Appendix S1; Section A; Assumption A2) that, of these, the *(IU)* environment has a far greater impact on the development of MS than does the shared *(CH)* environment. Once again, for MS, the direct observational data supports the validity of this assumption. Thus, studies in adopted individuals, in siblings and half-siblings raised together or apart, in conjugal couples, and in brothers and sisters of different birth order have generally indicated that MS-risk is unaffected by the *(CH)* micro-environment [4–7,9,10,19,20]. Regardless, however, the shared environmental experiences of MZ-twins, above and beyond the effect of the shared *(CH)* environment of siblings, potentially, could increase the proband-wise concordance rate [30]. As a result, the directly-observed MZ-twin concordance rates (Table 2) need to be adjusted to exclude the impact of these environmental similarities (Appendix S1; Section C; Prop. 1.4). These adjusted concordance rates, therefore, will reflect only the impact of an individual sharing an identical genotype *(IG)* with their MZ-twin who has MS. Two adjustments are envisioned. The first represents the total penetrance of the complex genetic trait (including both purely environmental and genetic cases) is referred to as:

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

This penetrance (**b**) is estimated to be 0.134 (Appendix S1; Section C; Prop. 1.4). The second adjusted rate represents the penetrance of the complex genetic trait exclusively in the set of genetically susceptible individuals and is referred to as:

Table 3. HLA data used in the model[#].

	2HB+		1HB+	HLA−
Canadian Data				
Observed Frequency – Cases (HLA+ and HLA−)[##]		0.55		0.45
Observed Frequency – Controls (HLA+ and HLA−)[##]		0.24		0.76
OR – (2HB+ & 1HB+) vs. (HLA−)*		3.9		
UCSF #1				
Observed Frequency – Cases[†]	0.10		0.46	0.44
Predicted HWE frequencies – Cases[††]	0.11		0.45	0.44
Predicted Controls – HWE at: P(HLA+) = 0.24	0.016		0.224	0.76
OR – (2HB+) vs. (HLA−) & (1HB+) vs. (HLA−) *	10.4		3.6	
OR – (2HB+ & 1HB+) vs. (HLA−)*		4.0		
UCSF #2				
Observed Frequency – Cases[†]	0.07		0.39	0.54
Predicted HWE frequencies – Cases[††]	0.07		0.39	0.54
Observed Frequency – Controls[†]	0.012		0.186	0.80
Predicted HWE frequencies – Controls[††]	0.011		0.186	0.80
OR – (2HB+) vs. (HLA−) & (1HB+) vs. (HLA−) *	9.3		3.1	
OR – (2HB+ & 1HB+) vs. (HLA−)*		3.5		

[#]Numbers listed are genotype frequencies.
2HB+ = carrier of 2 copies of the DRB1*1501 allele (homozygous carrier).
1HB+ = carrier of 1 copies of the DRB1*1501 allele (heterozygous carrier).
HLA− = carrier of 0 DRB1*1501 alleles.
(HLA+) = (2HB+)+(1HB+).
[##]Canadian HLA data: D Sadovnick (personal communication).
Based on ~3,000 cases and ~400 Controls (% women not available). Control rates confirmed in a much larger transplant database.
*Odds ratio (OR) versus controls. Calculated as odds of genotype in cases divided by odds of the same genotype in controls.
[†]UCSF Databases: J Oksenberg (personal communication).
UCSF #1 (IMSGC) – 779 cases (76% women); No observed controls.
UCSF #2 (GeneMSA) – 485 cases (68% women) and 431 Controls (66% women).
[††]Hardy Weinberg Equilibrium (HWE) values predicted based on the observed P(2HB+) in Cases or Controls.
doi:10.1371/journal.pone.0047875.t003

$$P(MS|G,IG_{MS}) = \mathbf{b}'$$

From Prop. (1.6) of Appendix S1 (Section C):

$$P(MS|G,IG_{MS}) = \mathbf{b}' = \mathbf{b}/\mathrm{g} = P(MS|IG_{MS})/\mathrm{g} \quad (4)$$

$$\text{Because : } \mathrm{g} \le 1; \text{therefore : } \mathbf{b}' \ge \mathbf{b} \quad (5)$$

Estimating proportion of population that is genetically susceptible to getting MS

We can partition *(G)* into two mutually-exclusive subsets (*G1* and *G2*) based on their disease-penetrance. The subset *(G1)* is defined as the high-penetrance subgroup (i.e., consisting of genotypes with a penetrance-value as high or higher than the expected penetrance for the entire susceptible-population) whereas *(G2)* is defined as the low-penetrance subgroup (i.e., genotypes having a penetrance-value as low or lower than this expected penetrance). Genotypes with a penetrance-value exactly equal to the expectation are divided evenly (and randomly) between the *(G1)* and *(G2)* subsets (to ensure that the subsets are mutually-exclusive). We define the expected disease-penetrance of these different sets as:

$$P(MS|G1) = \mathbf{x};\ P(MS|G2) = \mathbf{y};\ \text{and : } P(MS|G) = \mathbf{z}$$

By these definitions: $\mathbf{x} \ge \mathbf{z} \ge \mathbf{y}$

From Prop. (2.1) of Appendix S1 (Section C):

$$P(MS|G,IG_{MS}) = \mathbf{b}' \ge \mathbf{z} = P(MS|G) \quad (6)$$

So that, from Equations (2–6):

$$P(G) = P(MS,G)/\mathbf{z} \ge P(MS,G)/\mathbf{b}' = (\mathrm{g}^2) * P(MS)/\mathbf{b} \quad (7)$$

As demonstrated below and in Prop. (5.2b) of Appendix S1 (Section C), we estimate that $(\mathrm{g} \ge 0.94)$. Therefore, using this estimate, together with the values presented in Table 2, yields the estimate of:

$$P(G) \ge (\mathrm{g}^2) * P(MS)/\mathbf{b} \ge (0.0015)(0.94)^2/(0.134) = 0.010 \quad (8)$$

This provides a lower-bound for the probability of being genetically susceptible to MS.

To provide an upper-bound for *P(G)*, we define three quantities (p, **a**, and **a**′) such that:

Table 4. HLA data by gender used in the model[#].

	2HB+		1HB+	HLA−
UCSF #1 (Women)				
Observed Frequency – Cases[†]	0.11		0.46	0.43
Predicted HWE frequencies – Cases[††]	0.11		0.44	0.45
OR – (2HB+ & 1HB+) vs. (HLA−)*		4.3		
UCSF #1 (Men)				
Observed Frequency – Cases[†]	0.06		0.45	0.48
Predicted HWE frequencies – Cases[††]	0.09		0.42	0.48
OR – (2HB+ & 1HB+) vs. (HLA−)*		3.4		
UCSF #2 (Women)				
Observed Frequency – Cases[†]	0.08		0.41	0.51
Predicted HWE frequencies – Cases[††]	0.08		0.41	0.50
Observed Frequency – Controls[†]	0.01		0.17	0.82
Predicted HWE frequencies – Controls[††]	0.01		0.21	0.78
OR – (2HB+) vs. (HLA−) & (1HB+) vs. (HLA−) *	9.7		3.9	
OR – (2HB+ & 1HB+) vs. (HLA−)*		4.4		
UCSF #2 (Men)				
Observed Frequency – Cases[†]	0.05		0.35	0.61
Predicted HWE frequencies – Cases[††]	0.05		0.34	0.61
Observed Frequency – Controls[†]	0.01		0.22	0.78
Predicted HWE frequencies – Controls[††]	0.01		0.21	0.78
OR – (2HB+) vs. (HLA−) & (1HB+) vs. (HLA−) *	8.6		2.0	
OR – (2HB+ & 1HB+) vs. (HLA−)*		2.2		

[#]Numbers listed are genotype frequencies.
2HB+ = carrier of 2 copies of the DRB1*1501 allele (homozygous carrier).
1HB+ = carrier of 1 copies of the DRB1*1501 allele (heterozygous carrier).
HLA− = carrier of 0 DRB1*1501 alleles;
(HLA+) = (2HB+)+(1HB+).
[†]UCSF Databases: J Oksenberg (personal communication).
UCSF #1 (IMSGC) – 779 cases (76% women).
UCSF #2 (GeneMSA) – 485 cases (68% women) and 431 Controls (66% women).
[††]Hardy Weinberg Equilibrium (HWE) values predicted based on the observed P(2HB+) in Cases or Controls. Because of the small number of men in these samples, the number of males with 2 copies of HLA DRB1*1501 was tiny. Therefore, in men, HWE was estimated from the observed P(HLA−).
*Odds ratio (OR) versus controls. Calculated as odds of genotype in cases divided by odds of the same genotype in controls.
doi:10.1371/journal.pone.0047875.t004

Table 5. MS concordance rates in monozygotic twins of DRB1*1501 carrier *(HLA+)* and DRB1*1501 non-carrier *(HLA–)* probands[*].

	Monozygotic Twins of MS Probands		
	HLA+	*HLA–*	**Totals**
Concordant for MS (C)	9	11	20
Discordant for MS (D)	31	42	73
Totals	40	53	93
Pair-wise Concordance[†]	(9/40) = 0.225	(11/53) = 0.207	0.215
Proband-wise Concordance[††]	0.309	0.287	0.297
Proband-wise Concordance (Adjusted) [†††]	**t** = 0.166	**s** = 0.154	**b** = 0.160
Proband-wise Concordance (Adjusted) [††††]	**t** = 0.139	**s** = 0.129	**b** = 0.134
***P(HLA+|MS, IG*$_{MS}$**) **(Adjusted)**[#]	0.57		

*(HLA+) = carrier of ≥1 copy of the DRB1*1501 allele; Data from: Reference [21].
[†]Pair-wise rates (Z_1) calculated as: $Z_1 = C/(C+D)$; see Reference [30]
[††]Proband-wise concordance rates (Z_2) calculated as: $Z_2 = 2C/(2C+D)$; adjusted [30] for the overall probability of doubly ascertaining concordant twin-pairs $(13/24 = 54\%)$in the study of Willer, et al. [21]
[†††]For adjustment: See: Prop. (1.4a) & (1.4b) of Appendix S1 (Section C)
[††††]Further adjusted to the requirement that: **b** = 0.134
[#]Adjusted to the condition where: $P(HLA+|MS) = P(HLA+|IG_{MS}) = 0.55$
doi:10.1371/journal.pone.0034034.t005

Table 6. MS concordance rates in monozygotic twins of female *(F)* and male *(M)* probands*.

	Monozygotic Twins of MS Probands		
	F	***M***	**Totals**
Concordant for MS (C)	22	2	24
Discordant for MS (D)	66	43	109
Totals	88	45	133
Pair-wise Concordance†	(22/88) = 0.25	(2/45) = 0.044	0.18
Proband-wise Concordance††	0.34	0.067	0.25
Proband-wise Concordance (Adjusted) †††	**t** = 0.183	**s** = 0.036	**b** = 0.134
***P(F|MS, IG*$_{MS}$**) (Adjusted)**#	0.92		

*(F) = Women ; (M) = Men ; Data from: Reference [21]
†Pair-wise rates (Z_1) calculated as: $Z_1 = C/(C+D)$; see Ref. [30]
††Proband-wise concordance rates (Z_2) calculated as: $Z_2 = 2C/(2C+D)$; adjusted [30] for the overall probability of doubly ascertaining concordant twin-pairs $(13/24 = 54\%)$ in the study of Willer, et al. [21]
†††For adjustment: See: Prop. (1.4a) & (1.4b) of Appendix S1 (Section C)
#Adjusted to the condition where: $P(F|MS) = P(F|IG_{MS}) = 0.68$
doi:10.1371/journal.pone.0047875.t006

$$0 \le \mathrm{p} = P(G1|G) \le 1$$

so that, also : $\mathrm{p} * P(G) = P(G1)$; and : $(1-\mathrm{p}) * P(G) = P(G2)$

and, therefore, that:

$$\mathbf{a} = P(MS,G)/P(G1) = P(MS,G)/\{\mathrm{p} * P(G)\}$$
$$= P(MS|G)/\mathrm{p} \ge \mathbf{z} = P(MS|G)$$

$$\mathbf{a}' = P(MS,G)/P(G2) = P(MS,G)/\{(1-\mathrm{p}) * P(G)\}$$
$$= P(MS|G)/(1-\mathrm{p}) \ge \mathbf{z} = P(MS|G)$$

From Prop. (3.4) of Appendix S1 (Section C):

Table 7. Summary of conclusions regarding MS pathogenesis derived from the model*

Conclusions about genetic susceptibility (in general)	(see Section C; Props. 4−5)
$0.016 \le P(G) \le 0.022$	$0.94 \le P(G\|MS) \le 1$
$0 \le P(MS\|G-) \le 0.000092$	$P(MS\|G) \ge 728 * P(MS\|G-)$
$0.067 \le P(MS\|G) \le 0.089$	$0.0040 \le {\sigma_{zi}}^2 \le 0.0051$
Conclusions about DRB1*1501 status and genetic susceptibility	(see Sections D&E; Props. 6.3&7.1a)
$0.012 \le P(G\|HLA-) \le 0.014$	$0.036 \le P(G\|1HB+) \le 0.045$
$0.044 \le P(G\|HLA+) \le 0.049$	$0.110 \le P(G\|2HB+) \le 0.140$
$0.54 \le P(HLA+\|G) \le 0.55$	$P(2HB+\|G) \approx 0.10$
$P(MS\|G,2HB+) \approx P(MS\|G,1HB+) \approx P(MS\|G,HLA+) \approx P(MS\|G,HLA-) \approx P(MS\|G)$	
Conclusions about gender status and genetic susceptibility	(see Sections D&E; Props. 6.2&7.1a)
$0.010 \le P(G\|F) \le 0.021$	$0.28 \le P(F\|G) \le 0.48$
$0.023 \le P(G\|M) \le 0.032$	$0.52 \le P(M\|G) \le 0.72$
$0.030 \le P(MS\|M,G) \le 0.040$	$0.096 \le P(MS\|F,G) \le 0.191$
$2.4 \le P(MS\|F,G)/P(MS\|M,G) \le 5.4$	
Conclusions about other relationships regarding genetic susceptibility	(see Section E; Prop. 7.2)
$P(G3\|G) \approx 0$; where, by definition: $P(MS\|G3,E) = P(MS\|G3,E-) = P(MS\|G3)$	
Conclusions about environmental susceptibility (see Section F; Eqs. 24−31)	
$0.114 \le P(MS\|G,E,F) \le 0.277$	$0.030 \le P(MS\|G,E,M) \le 0.056$
$2.5 \le P(MS\|G,E,F)/P(MS\|G,E,M) \le 7.5$	$0.69 \le P(E) \le 1$
$0.100 \le \lambda \le 2.87$; λ = threshold difference between women and men – (see Section F)	
$0.54 \le \mathrm{r} \le 1.6$; r = proportional hazard for MS – women to men – (see Section F)	

*See Table 1 for term definitions; In Table "Section" refers to Sections of Appendix S1
doi:10.1371/journal.pone.0047875.t007

Table 8. Estimated prevalence (probability) of genetic susceptibility in different geographic regions.

	MS Prevalence ‡	MZ-Twin Concordance *	% Susceptible †
	P(MS)	$P(MS\|MZ_{MS})$	*P(G)*
North America			
Canada [21]	68 – 248	25.3%	0.4 – 3.6%
Northern US [12]	100 – 160	31.4%	0.5 – 1.9%
Southern US [12]	22 – 112	17.4%	0.2 – 2.4%
Europe			
Finland [31]	52 – 93	46.2%	0.2 – 0.7%
Denmark [32]	110	24%	0.7 – 1.7%
British Isles [13]	74 – 193	40.0%	0.3 – 1.8%
France [11]	32 – 65	11.1%	0.5 – 2.2%
Sardinia [16]	144 – 152	22.2%	1.1 – 2.5%
Italy [16]	38 – 90	14.5%	0.4 – 2.3%

‡Per 10^5 population. The prevalence of MS in each region is from data provided in Reference [33]. A range is given because, often, a range of estimates are available for a particular region.

*Studies [11–13] reported pair-wise MZ-twin concordance-rates, which have been converted into proband-wise rates assuming a random sampling of twin-pairs [30]. Study [12], however, reported no double ascertainment of twin-pairs and, therefore, almost certainly violates this assumption [30].

†*P(G)* calculated according to Eq. (**6**); Prop. (4.2a); Appendix S1 (Section C); that: $(g^2)*(1.86)*\{P(MS)/P(MS|MZ_{MS})\} \le P(G) \le (3.72)*\{P(MS)/P(MS|MZ_{MS})\}$
This equation assumes that (g) for each geographic region is: $0.94 \le g \le 1$; Appendix S1; Section C; Prop. (5.2b)
Moreover, as noted in Prop. (4.2b), Eq. (**6**) also assumes that: $\mathbf{z}_{\max} = \mathbf{b}'$
A narrower range-estimate could be provided by Eq. (**13**); Prop. (4.2b) of Appendix S1 (Section C) . However, regardless of which range-estimate for ($\mathbf{z}_{max}$) is used, this only impacts the lower-bound estimates for *P(G)*. The upper-bound estimates remain the same.

doi:10.1371/journal.pone.0047875.t008

$$\mathbf{a} \ge \mathbf{b}'; \text{and} : \mathbf{a}' \ge \mathbf{b}' \tag{9}$$

Therefore: $\mathbf{a} = P(MS,G)/P(G1) = P(MS,G)/\{p * P(G)\} \ge \mathbf{b}'$ so that, with rearrangement:

$$P(G) \le \{P(MS,G)/\mathbf{b}'\}/p$$

Moreover, because: $P(MS,G) \le P(MS)$; and, by Equation (5): $\mathbf{b}' \ge \mathbf{b}$
Therefore:

$$P(G) \le \{P(MS)/\mathbf{b}\}/p \tag{10}$$

Similarly:

$$\mathbf{a}' = P(MS,G)/P(G2) = P(MS,G)/\{(1-p) * P(G)\} \ge \mathbf{b}'$$

so that, with rearrangement:

$$P(G) \le \{P(MS,G)/\mathbf{b}'\}/(1-p)$$

And, therefore, also:

$$P(G) \le \{P(MS)/\mathbf{b}\}/(1-p) \tag{11}$$

Because one of the following three statements must be true:

$$p > 0.5; (1-p) > 0.5; \text{or} : p = 0.5$$

Therefore, making only Assumptions (A2–A4) from Appendix S1 (Section A), the Equations (10)&(11), place two simultaneous constraints on *P(G)* and, together with Equations (6–8), require that:

$$(g^2) * P(MS)/\mathbf{b} \le P(G) \le 2 * P(MS)/\mathbf{b} \tag{12}$$

which can be rewritten equivalently as:

$$(g^2) * (1.86) * \{P(MS)/P(MS|MZ_{MS})\} \le P(G) \le 2 * (1.86) * \{P(MS)/P(MS|MZ_{MS})\}$$

Because the quantities *P(MS)* and $P(MS|MZ_{MS})$ are directly observable parameters (Table 2), we can substitute, into Equation (12), the values of:

$$P(MS) = 0.0015; \text{and} : \mathbf{b} = P(MS|MZ_{MS})/(1.86) = 0.134$$

Doing this, together with Equation (8), yields the estimate of:

$$0.010 \le P(G) \le 0.022 \tag{13}$$

Thus, making Assumptions (A1–A4) from Appendix S1 (Section A), no more than 2.2% of the general population is genetically susceptible to getting MS (Appendix S1; Section C; Prop. 4.2). A very similar range-estimate for *P(G)* is derived from epidemiological data obtained from different populations throughout North America and Europe (Table 8).

Estimating the proportion of MS patients who are genetically susceptible

In order to estimate the quantity (g), we can partition the general population into two subsets, *(Gx+)* and *(Gx−)*, based on the presence or absence of some genetic factor *(Gx)* related to susceptibility (Appendix S1; Section C; Props. 1.7&5.2a). In Table 1, as before with (**b**&**b′**), we define two adjusted penetrance-values for each subset, either based on all cases (**s**&**t**) or based on just the genetic cases (**s′**&**t′**). Additionally, as in Table 1, we define two sets of parameters (A_0&A) and (g_1&g_2) such that:

$$A_0 = P(Gx+); \text{and}: A = P(Gx+|MS)$$

and: $g_1 = P(G|MS,Gx+)$; and: $g_2 = P(G|MS,Gx-)$

Using, in part, the result of Equation (13) and, as demonstrated in Prop. (5.1) of Appendix S1 (Section C), four relationships must hold:

$$\#1.\ g = Ag_1 + (1-A)g_2 \tag{14}$$

$$\#2.\ g_1/g_2 \le \mathbf{t}/\mathbf{s} \tag{15}$$

$$\#3.\ 1 \ge P(G-|Gx+) \ge (A_0 - 0.022)/A_0$$

$$\#4.\ 1 \ge P(G-|Gx-) \ge (0.978 - A_0)/(1-A_0)$$

From this, we define the parameter (B) such that:

$$B = (1-g_1)/1-g_2) = P(G-|MS,Gx+)/P(G-|MS,Gx-)$$

which, from Prop. (5.2a) of Appendix S1 (Section C), is equivalent to:

$$B = \{(A_0/A) * P(G-|Gx+)\}/ \{(1-A_0)/(1-A) * P(G-|Gx-)\} \tag{16}$$

$$\text{and}: g_1 = Bg_2 + (1-B) \tag{17}$$

Using Equations (13)&(16) together with the above relationships (#3)&(#4), yields:

$$[(A_0 - 0.022)/A] * [(1-A)/(1-A_0)] \le B \le (A_0/A) * [(1-A)/(0.978 - A_0)] \tag{18}$$

Because the quantities A, A_0, **t**, and **s** are either directly observable (or derived-directly from observations) for any partition of *(G)*, therefore, we can use Equations (14–18) to estimate the unknown values of B, g, g_1, and g_2 using experimental-data (Prop. 5.2a; Appendix S1; Section C).

In MS, from the gender-partition, our estimate is: $0.42 \le g \le 1$

and, from the HLA-partition, our estimate is: $0.94 \le g \le 1$

Notably: $g = P(G|MS) = P(G,Gx+|MS) + P(G,Gx-|MS)$

Therefore, the estimated value of (g) will be the same regardless of which partition is chosen for its estimation (as long as *Gx* is associated with susceptibility – see Props. (1.7)&(5.2a) of Appendix S1 (Section C). Thus, in order to satisfy both the gender and the HLA estimates of (g), we conclude that, for MS, more than 94% of the cases occur in genetically susceptible individuals (Prop. 5.2b; Appendix S1; Section C). The conclusion that the proportion of genetically susceptible cases is very high, is also reached in Prop. (5.3) of Appendix S1 (Section C) using the population-based epidemiological data reported from Finland [31,32].

HLA-DRB1 Subgroup differences in disease-penetrance

There are two possible mechanisms whereby *Gx+* individuals could be enriched in the MS-population compared to the general population (Appendix S1; Sections C&D; Props. (1.7)&(6). These are:

Mechanism (1) $P(Gx+|G) > P(Gx+)$
or, equivalently: $P(G|Gx+) > P(G|Gx-)$
{a difference in "allelic" frequency}

Mechanism (2) $P(MS|G,Gx+) > P(MS|G,Gx-)$
{a difference in penetrance}

In addition, there are three (potential) enrichment-stages for *(Gx+)*, which take place in MZ-twins (Appendix S1; Section D; Prop. 6.1a). The first stage occurs when moving from the set *(Gx+)* to the set *(Gx+,G)*; the second occurs when moving from the set *(Gx+,G)* to the set *(Gx+,G,MS)*, or equivalently to the set *(Gx+,G,IG_{MS})*; and the third occurs when moving from the set *(Gx+,G,IG_{MS})* to the set *(Gx+,G,MS,IG_{MS})*. As discussed in Prop. (6.1a) of Appendix S1 (Section D), the first stage can only involve Mechanism (1) whereas, the second and third stages can only involve Mechanism (2).

Moreover, the ratio (**s′**/**b′**) provides an estimate of the extent to which these two mechanisms operate (Appendix S1; Section D; Props. 6.1&6.2). If only Mechanism (1) is responsible for the enrichment, then:

$$\mathbf{s}'/\mathbf{b}' = 1; \text{otherwise}: \mathbf{s}'/\mathbf{b}' < 1$$

Unfortunately, the quantities (**s′**) and (**b′**), unlike the quantities (**s**) and (**b**), are not derived from direct-observations. However, from Props. (5.1&5.2b) of Appendix S1 (Section C) for MS and for the HLA partition, it is the case that:

$$\mathbf{s}/\mathbf{b} \le \mathbf{s}'/\mathbf{b}' = (g_1/g_2)(\mathbf{s}/\mathbf{b}) \le (0.94/0.90)(\mathbf{s}/\mathbf{b}) = 1.04(\mathbf{s}/\mathbf{b}) \tag{19}$$

So that, for the HLA-partition, this yields: $0.97 \le \mathbf{s}'/\mathbf{b}' \le 1$

and, therefore, it follows that Mechanism (1) accounts almost entirely for the enrichment of DRB1*1501 in an MS-population. Consequently, from Props. (2.3b,6.3b,&7.1a) of Appendix S1 (Sections C&D), the following relationships can be demonstrated:

$$(3.72) * P(G|HLA-) \le P(G|HLA+) \le (3.87) * P(G|HLA-)$$

$$\text{and}: 0.044 \le P(G|HLA+) \le 0.049$$

$$\text{and}: 0.012 \le P(G|HLA-) \le 0.014$$

$$\text{and, finally}: 0.54 \le P(HLA+|G) \le 0.55$$

$$\text{In addition}: 1 \le P(MS|G,HLA+)/P(MS|G,HLA-) \le 1.06$$

so that : $P(MS|G,HLA+)\approx P(MS|G,HLA-)\approx P(MS|G)$

Consequently, despite the importance of DRB1*1501 for genetic-susceptibility, only a very small fraction of carriers (<5%) are even genetically susceptible to getting MS. Also, the conclusion that, for HLA-status, Mechanism (1) operates almost exclusively is supported by the observed lack of any continued HLA-enrichment in moving from the general population, to the *(MS)* population, and then to the *(MS, MZ_{MS})* population. Thus, from Tables 2 and 5:

$$P(HLA+)=0.24<P(HLA+|MS)$$
$$=0.55\approx 0.57=P(HLA+|MS,MZ_{MS})$$

The enrichment of homozygous DRB1*1501 *(2HB+)* is approximately 3-fold greater than for single-allele carrier-status (Prop. 6.3c; Appendix S1; Section D). Nevertheless, even in this circumstance, Mechanism (1) still seems to account (almost entirely) for the enrichment of 2HB+ (Prop. 6.3c; Appendix S1; Section D). This suggests that neither heterozygous nor homozygous carrier-status affects disease-penetrance (Appendix S1; Sections C&D; Props. 5.3a,5.3c,6.3b,&6.3c).

In addition, it is a notable fact that all of these MS-populations seem to be at or near the Hardy-Weinberg equilibrium (HWE) state (Tables 3 and 4). From Prop. (6.4b) of Appendix S1 (Section D), this observation indicates that the relative normalized selection pressure for two DRB1*1501 alleles (w^2) is equal to the square of that for one allele (w>1). In this sense the two DRB1*1501 alleles are said to be independently selected. Thus, the weighting for the homozygous-lack, and for the heterozygous- and homozygous-presence, of the risk allele is geometric (1,w,w^2). This is analogous to the joint probability of two events being the product of the individual probabilities; and it contrasts to the weighting scheme for recessive and dominant traits (assuming a non-zero risk for non-carriers), which would be (1,1,w) and (1,w,w), respectively. This suggests the possibility that each DRB1*1501 allele contributes equally to the total number of susceptibility alleles required (Appendix S1; Section B & Section E; Prop. 6.4b). For example, if susceptible "non-DRB1*1501" genotypes have (on average) ten susceptibility alleles, perhaps susceptible genotypes with one DRB1*1501 allele have only nine, whereas genotypes with two such alleles might have only eight [27].

Finally, susceptible women (compared to susceptible men) have a higher mean allelic frequency (MAF) for the DRB1*1501 allele, a difference which is consistently reflected in MS-populations (Tables 2,3,4 & Appendix S1; Section E; Prop. 6.4d). This imbalance is due primarily to a gender difference in the composition of the subset *(G)* of susceptible individuals (Appendix S1; Section E; Prop. 6.4d).

As noted above, one of the features of susceptible genotypes that include the DRB1*1501 allele seems to be that they have a (slightly) reduced number of susceptibility alleles present (on average) compared to other susceptible genotypes [27]. In this circumstance, the observed MAF gender-difference would be expected if this reduction (for DRB1*1501 genotypes) were somewhat greater in women than in men.

Gender Subgroup differences in disease-penetrance

For MS and for gender, from Prop. (6.1c) of Appendix S1 (Section D), we can also write Equation (19) as:

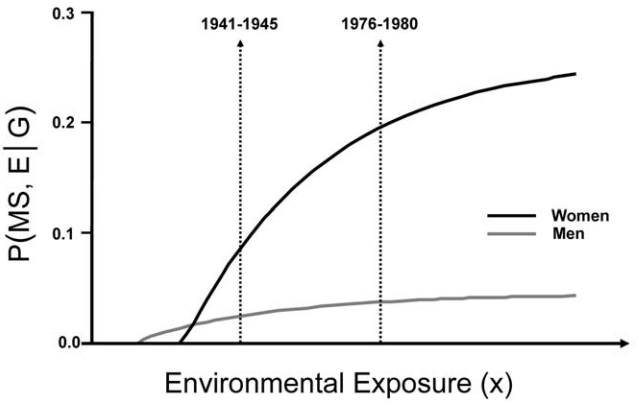

Figure 1. Response-curves for developing MS in susceptible men *(M)* and women *(F)* to an increasing likelihood of a "sufficient" environmental exposure *(E)*. Proportionate hazard is assumed for the two genders (see Appendix S1; Section F). The probability of developing MS – *P(MS, E|G)* – is shown on y-axis and the transformed environmental exposure (x) is shown on the x-axis {NB: (x) increases with *(E)* but not necessarily linearly – see Appendix S1; Section F}. The maximum y-axis excursions have been set to the high-point of the predicted ranges for *P(MS|G, E, M)* & *P(MS|G, E, F)* given by Eqs. (**14**) & (**16**) – Appendix S1; Section E; Prop. (7.1c). The proportionality constants, (C) and (r), are taken to be 0.5 and 1, respectively. One "environmental unit" has been defined arbitrarily as the change in the level of a sufficient environmental exposure *(E)*, which has taken place between the time-periods of (1941–1945) and (1976–1980). Based on the increase in the gender-ratio of MS patients over this interval, together with the proband-wise MZ-twin concordance-rates for MS in men and women from Canada [15,21], two conclusions follow directly. First, there has been more than a 32% increase in the prevalence of MS in Canada between these two time-periods and second, compared to women, men begin to develop MS at a lower level of environmental exposure (x) or they have a greater hazard-rate (see Appendix S1; Section F). In either case, women are more responsive to the environmental changes that are taking place than men (regardless of what these changes actually are). Presumably, this explains the observation that the prevalence of MS is increasing, especially among women [4]. Each of these conclusions is apparent in the Figure. The response curve for men starts at a lower value of (x) than women but their response curve is almost at its plateau in (1941–1945). By contrast, women are nowhere near their (much higher) plateau in (1941–1945) and, compared to men, have a much steeper rise of *P(MS|G, E)* in response to the environmental changes, which have taken place during the interval. {NB: the x-axis is **not** a time-axis. The x-axis represents increasing levels of environmental exposure (x) from whatever cause and over whatever period of time it has taken place.}
doi:10.1371/journal.pone.0047875.g001

$$\mathbf{s}/\mathbf{b}\le \mathbf{s}'/\mathbf{b}'=(g_1/g_2)(\mathbf{s}/\mathbf{b})\le(0.94/0.90)(\mathbf{s}/\mathbf{b})=1.04(\mathbf{s}/\mathbf{b})$$

so that, from Table 6, for the gender partition, this becomes:

$$0.27\le \mathbf{s}'/\mathbf{b}'\le 0.28$$

It turns out that this implies (Appendix S1; Sections D&E; Props. 6.3b&7.1a) that both Mechanisms (1) and (2) operate and, thus, that:

$$(1.08)*P(G|F)\le P(G|M)\le(2.57)*P(G|F)$$

$$\text{and}: 0.010 \le P(G|F) \le 0.021$$

$$\text{and}: 0.023 \le P(G|M) \le 0.032$$

$$\text{and}: 0.28 \le P(F|G) \le 0.48$$

$$\text{and also that}: 2.4 \le P(MS|G,F)/P(MS|G,M) \le 5.4$$

Thus, men are more likely to be genetically susceptible to MS compared with women although susceptible men are less likely to get MS than susceptible women. This same conclusion was suggested earlier [4] and, in fact, the actual response-curves demonstrating the greater responsiveness of women to increasing environmental exposures (and, thus, the greater penetrance of MS in women) can also be derived quantitatively (assuming proportionate hazard for MS in men and women) from the same epidemiological data (Figure 1; & Appendix S1; Section F). Notably, increasing the likelihood of a sufficient environmental exposure *(E)* in susceptible individuals, $P(E|G)$, does not increase the likelihood of MS developing beyond 28% in women and beyond 6% in men (Figure 1; & Appendix S1; Section F). This must be due to the fact that certain genetic backgrounds are only (or more) responsive to certain sufficient environmental experiences (Appendix S1; Section F). For example, even if all susceptible genotypes required a particular environmental stimulus (e.g., vitamin D deficiency), some susceptible genotypes might require a longer duration or a greater intensity of exposure to produce MS than others (Appendix S1; Section F). Also, assuming a proportional hazard for men and women, susceptible men (compared to susceptible women) must have a lower threshold, a greater hazard-rate, or both in response to the environmental factors involved in MS pathogenesis (Figure 1 & Appendix S1; Section F).

In addition, the greater penetrance of MS in susceptible women is also reflected by the continued enrichment of women in going from the general population to the *(MS)* population and then to *(MS, MZ_{MS})* population. Thus, from Tables (2,3,4,5), and as discussed in Prop. (5.3) of Appendix S1 (Section C):

$$P(F) \approx 0.5 < P(F|MS) = 0.68 < 0.92 = P(F|MS, MZ_{MS})$$

As a result, we conclude that gender has a marked impact on both disease penetrance and disease susceptibility (Appendix S1; Sections C&D; Props. 5.3b,5.3c,&6.3a).

Estimating the penetrance of susceptible and non-susceptible genotypes

Rearranging Equations (4) and (11) yields:

$$\mathbf{b}' \ge P(MS|G) = \mathbf{z} \ge \mathbf{b}'/2 \ge \mathbf{b}/2$$

From Prop. (5.2b) and substituting into this equation the value of:

$$\mathbf{b}' \ge \mathbf{b} = 0.134$$

yields the estimate of:

$$(0.134/2) = 0.067 \le \mathbf{z} \le (0.134/0.94) = 0.143 \qquad (20)$$

This range-estimate can be narrowed considerably (see Appendix S1; Section E; Prop. 7.1a) by recognizing that:

$$P(MS|M,G) = \mathbf{z_s} \le P(MS|M,G,IG_{MS}) \le (0.036/0.90) = 0.040$$

$$\text{and}: P(MS|F,G) = \mathbf{z_t} \le P(MS|F,G,IG_{MS}) \le (0.183/0.96) = 0.191$$

$$\begin{aligned}\text{Also}: P(MS|G) &= \mathbf{z} \\ &= P(M|G) * P(MS|M,G) \\ &+ P(F|G) * P(MS|F,G)\end{aligned} \qquad (21)$$

Therefore, the predicted ranges from Prop. (6.2b) of Appendix S1 (Section D) lead to the boundaries:

lower-bound: $P(MS|G,F) = \mathbf{z_t} = (2.3) * P(MS|G,M)$; and: $P(M|G) = 0.51$

upper-bound: $P(MS|G,F) = \mathbf{z_t} = (5.4) * P(MS|G,M)$; and: $P(M|G) = 0.72$

From Prop. (7.1) of Appendix S1 (Section E), substituting these values into Equation (21) yields the boundary estimates of:

$$\mathbf{z} \ge (0.51) * (0.040) + (0.49) * (5.4) * (0.040) = 0.065 \qquad (22)$$

$$\text{and}: \mathbf{z} \ge (0.72) * (0.040) + (0.28) * (5.4) * (0.040) = 0.089$$

However, the lower-boundary of Equation (22) is slightly inconsistent with the most straight-forward lower bound condition that: $\mathbf{z} \ge \mathbf{b}/2 = 0.067$ # see Eq.(20)

For MS, obviously, this discrepancy is quite minor. In other disease states, by contrast, it may be greater. Therefore, we provide a method for making the Equation (20) & (22) estimates "coherent" with each other (Appendix S1; Section E; Prop. 7.1a). For MS, solution of the two simultaneous equations yields the minimally modified lower boundary estimates of:

$$a_1 = P(MS|G,F)/P(MS|G,M) \ge 2.4 \qquad (23)$$

$$\text{and}: P(M|G) \ge 0.52 \qquad (24)$$

so that:

$$\mathbf{z} \ge P(M|G) * (0.040) + \{1 - P(M|G)\} * (a_1) * (0.040) = 0.067$$

And, consequently, this yields the revised range-estimate for ($\mathbf{z}$) of:

$$\mathbf{b}/2 = 0.067 \le \mathbf{z} \le 0.089 \qquad (25)$$

From Equation (25), it also follows (Appendix S1; Section E; Prop. 7.1) that:

$$0.016 \le P(G) \le 0.022$$

$$0.0040 \le \sigma_{zi}{}^2 \le 0.0051$$

$$0.030 \le P(MS|M,G) \le 0.040 \quad (26)$$

$$\text{and, finally : } 0.096 \le P(MS|F,G) \le 0.191 \quad (27)$$

Also, from Props. (4.2&5.2b) of Appendix S1 (Section C):

$$P(MS,G-) \le (0.06) * P(MS) = (0.06)(0.0015) = 0.00009$$

$$P(MS|G-) = P(MS,G-)/P(G-) \le (0.00009/0.978) = 0.000092$$

$$\text{so that : } 0 \le P(MS|G-) \le 0.000092$$

$$\text{and : } P(MS|G) \ge (0.067/0.000092) * P(MS|G-) = 728 * P(MS|G-)$$

Estimating the proportion of "purely genetic" MS

Because "purely genetic" MS is defined to be independent of the environment (see Appendix S1; Section B), its penetrance is expected to very high (i.e., near unity). Thus, we anticipate both that:

$$P(MS|G3) \approx 1; \text{and that : } P(G1|G3) = 1 \quad (28)$$

If these conditions were not met, it would raise the question of what factors determined the lower penetrance. If these factors were potentially identifiable and non-hereditary, then they would constitute environmental events and, thus, these genotypes would be in *(G0)* and not in *(G3)*. Although a purely stochastic mechanism might lower the penetrance somewhat, this seems unlikely to reduce the penetrance markedly.

As shown in Prop. (7.2) of Appendix S1 (Section E), even if we make the extreme assumptions that:

$$P(G3|G) = P(G1|G) = \text{p}; P(MS|G3) = \mathbf{x} \approx 1; \text{and : } P(MS|G2) = \mathbf{y}$$

and assume that the variances of the of the ($\mathbf{x_i}$) and ($\mathbf{y_i}$) terms are zero;

and, finally assume that all values: $P(MS|G3) > 0.8$; satisfy the conditions of Equation (28);

then, even in these extreme conditions, we still estimate that:

$$0 \le P(G3|G) < 0.010$$

However, these conditions seem too extreme for any actual distribution and, notably, less extreme assumptions lead to even smaller estimates for *P(G3|G)*. Therefore, this derived upper limit for the range of *P(G3|G)* is, almost certainly, too large.

And, consequently, it must be that: $P(G3|G) \approx 0$

And, thus, for all practical purposes, "purely genetic" MS does not exist.

Sensitivity considerations

Naturally, all of the range-estimates provided here are dependent upon the accuracy of the underlying epidemiological data in Tables 2,3,4,5,6. To illustrate this, we will use our Equation (13) estimate for P(G) where we estimated that:

$$0.010 \le P(G) \le 0.022$$

For example, if we consider the prevalence of MS in the 45–55 year age-range (e.g., Appendix S1; Section B) to be a better estimator of *P(MS)* then, potentially, the estimate of (0.0015) used here could double [29]. In this case {i.e., if: $P(MS) = 0.0030$; and all else is equal}, then our Equation (13) range-estimate for *P(G)* would be increased to:

$$0.020 \le P(G) \le 0.045$$

By contrast, even though the estimate for (B) changes slightly using this upper bound, the estimate for (g) derived from the HLA partition in Prop. (5.2a2) of Appendix S1 (Section C), remains unchanged at:

$$0.94 \le \text{g} \le 1$$

Similarly, if the proband-wise MZ-twin concordance in northern populations is 35% rather than the 25% used here [3], then this would lead to:

$$\mathbf{b} = 0.188$$

and our Equation (13) estimate would become:

$$0.007 \le P(G) \le 0.016$$

Also, if $P(MS|S_{MS})$ is actually 3.5% instead of 2.9% then:

$$\mathbf{b} = 0.162$$

and the Equation (13) estimate would become:

$$0.008 \le P(G) \le 0.019$$

Finally, if all of these modifications were accepted, then the Equation (13) estimate would become:

$$0.012 \le P(G) \le 0.026$$

Thus, there is an additional level of uncertainty implicit in each of the range-estimates for the different parameters provided here.

Assumption Violations

It is also important to consider what the impact might be if one or more of the Assumptions underlying the model were to be

Table 9. Estimated prevalence (probability) of genetic susceptibility in rheumatoid arthritis, ankylosing spondylitis, and systemic lupus erythematosus

	Prevalence ‡	MZ-Twin Concordance *	% Susceptible
	P(D)	$P(D\|MZ_D)$	*P(G)*
Rheumatoid Arthritis	1 – 2%	~ 35%	5.7 – 11.4%
Ankylosing Spondylitis	0.4 – 4%	~ 53%	1.4 – 15%
Systemic Lupus Erythematosus	~ 0.025%	~ 39%	~ 0.13%

‡The prevalence of diseases {*P(D)*} is from data provided in Reference [35].
*Studies [36-38] report pair-wise MZ-twin concordance-rates. These have been converted into proband-wise rates $\{P(D|MZ_D)\}$ assuming a random sampling of twin-pairs [30]. Also, the IU environment has been assumed to have no impact on the disease. A violation of either of these assumptions will make the estimate of *P(G)* too low.
doi:10.1371/journal.pone.0047875.t009

violated (Appendix S1; Section A). The most basic assumption of the model is that the twin populations are "representative" of the general population (see Assumptions A5&A6; Appendix S1; Section A). This assumption is critical and were it to be violated, the entire model would be invalid. Fortunately, as noted earlier, the direct observational data in MS support the validity of this assumption (e.g., [21]). Moreover, this assumption also underlies the "classical twin study" approach that has been used (and validated) for decades to elucidate the genetic and environmental bases of many human illnesses (e.g., [2]).

The second critical assumption of the model is that the *(CH)* micro-environment does not contribute to disease occurrence (Appendix S1; Section A; Assumption A2). Fortunately, as noted earlier, there is considerable observational data in MS (from numerous studies in adopted individuals, in siblings and half-siblings raised together or apart, in conjugal couples, and in brothers and sisters of different birth order) to support the notion that MS-risk is not impacted by the *(CH)* environment [4–7,9,10,19,20]. Nevertheless, if this assumption were to be violated, it would have a major impact on our ultimate conclusions.

For example, in Parkinson's disease *(PD)*, it has also been observed that siblings of a PD-proband carry a significantly greater risk of disease compared to unrelated controls (33). However, by contrast to MS, the MZ-twins of a PD-proband seem not to be at greater risk compared to DZ-twins, especially if the onset of illness is over age 50 [34]. In such a circumstance, the lack of any difference between the MZ-risk and DZ-risk, most likely reflects the fact that:

$$P(G|PD)\approx 0$$

and, thus, that genetics are only minimally (or not) involved in disease pathogenesis. In this case, the increased-risk in siblings is presumably due to the similar *(CH)* environment, which siblings share, and, therefore that:

$$P(PD|G-,S_{PD})=P(PD|G-,CH)>P(PD|G-)$$

Even if, unlike the situation in PD, both the genetic make-up and the *(CH)* environment contribute to the increased disease *(D)* risk, then it would still be the case that:

$$P(D|G-,S_D)=P(D|G-,CH)>P(D|G-)$$

In this circumstance, however, the relationship between $\{P(D|G-, CH)\}$ and $\{P(D|G, CH)\}$ cannot be deduced. Therefore, this violation would invalidate the conclusion that:

$$P(D,G-|IG_D)<P(D)$$
$$\text{see Appendix S1; Section C; Prop.(1.5)}$$

which would invalidate the further conclusion that:

$$P(D,G|IG_D)>\mathbf{b}-P(D)$$

This, in turn, would invalidate the conclusion that:

$$\mathbf{b}/g=\mathbf{b}'\geq\mathbf{b}$$
$$\text{see Appendix S1; Section C; Prop.(1.6)}$$

which would invalidate most of the Prop. 4&5 conclusions (Appendix S1; Section C).

Despite these consequences, however, a violation of Assumption (A2) would not be fatal to the model. Rather, it would mean that the model would need minor revision and that the *(CH)* impact would need to be estimated from experimental data, for example, by studying siblings raised separately or adopted children raised together with an MS-proband.

Assumption (A4); Appendix S1 (Section A), is crucial to conclusions about the relative prevalence of genetic susceptibility in the *(Gx+)* and the *(Gx−)* subsets. For example, for the gender partition ($Gx+=F$), from Table 2 & Prop. (1.4b) of Appendix S1 (Section C), it seems that:

$$\mathrm{m}_1=(0.051/0.039)=1.31<3.0=(0.057/0.019)=\mathrm{m}_2$$

If these experimental observations are correct, then the impact of this violation would be that the true separation between men and women in the percentage of genetically susceptible individuals (Appendix S1; Sections C & D; Props. 1.4b&6.2a) would be underestimated. Naturally, the impact of the opposite violation (i.e., where: $\mathrm{m}_1>\mathrm{m}_2$), would be to overestimate this separation. However, from the available data, this seems not to be the case.

Other assumption violations would, in general only impact the specific propositions involved. Each of these assumptions, and the propositions they impact, are listed in Appendix S1 (Section A).

Discussion

Both the mathematical model and the data presented here suggest that detailed study of MZ- and DZ-twin concordance data, combined with general epidemiological information regarding the disease from the same population as the twin data, are capable of providing quantitative estimates for many parameters associated with disease pathogenesis, which can't be directly-observed or easily measured. Thus, making only a few very simple (and quite plausible) assumptions about the genetic make-up of MZ- and DZ-twins, quantities such as $P(G)$, $P(E)$, $P(G3|G)$, $P(G|MS)$, $P(MS|G)$, $P(MS|G-)$, $P(F|G)$, $P(MS|E,G,F)$, $P(MS|E,G,M)$, and (σ_{zi}^2) can be estimated from directly observable data (Table 7). Also this model can provide these parameter estimates for other complex genetic disorders (e.g., Table 9). Finally, the model can provide insight to the mechanisms of disease pathogenesis. For example, in MS, this analysis indicates that the basis for the association of DRB1*1501 with MS is due to the fact that persons who carry this allele have a greater likelihood of being genetically susceptible compared to persons who lack this allele. In addition, each DRB1*1501 allele seem to affect susceptibility independently. By contrast, carrier status does not seem to affect the likelihood of developing the disease in the susceptible population. Moreover, despite the strong association of DRB1*1501 with MS, the majority (~59%) of genetically susceptible individuals are susceptible based on genotypes that do not include this allele and, indeed, for the 25% of these individuals who, nonetheless, still carry this allele, the presence of DRB1*1501 seems not to contribute to their susceptibility (Prop. 8.1; Appendix S1; Section E). In addition, among carriers of this allele, fewer than 5% are even susceptible to getting MS in the first place (Appendix S1; Section D; Prop. 6.3b).

In the case of gender, however, the disease association turns out to result from a combination of effects. Thus, despite men having a greater likelihood than women of being genetically susceptible, women who are susceptible are considerably more likely to develop the disease than susceptible men. Although, the distinction between men and women is (in some sense) genetic, the principal anatomic and physiological differences between genders are likely not to be linked to specific allelic variations but, rather, are almost certainly based on differences in the regulation of developmental programs that are shared by all same-sexed individuals. Because the observed gender differences in disease penetrance seem to be the result of an increased physiological responsiveness of women to common environmental events (see Appendix S1; Section F), therefore, the genetic basis of this particular influence is unlikely to be uncovered through approaches such as genome-wide association studies (GWAS). By contrast, the genetic basis for the gender-related differences in the likelihood of susceptibility could arise from either allelic or epigenetic differences between the sexes and might, potentially, be detected using GWAS or other genetic methods, particularly if men and women were to be analyzed separately. Alternatively, if the lower likelihood of susceptibility in women were due to an increase in the average number of susceptibility-genes necessary to produce susceptibility in women, this, also, would likely not be evident using a GWAS approach. Moreover, because of the huge number of anticipated susceptibility-genotypes (Appendix S1; Section B), few MS patients are likely to share the exact same combination of susceptibility genes. Therefore, as discussed in Appendix S1 (Section B), novel approaches to the analysis of these large datasets [26] are almost certainly going to be necessary in order to clarify the genetic underpinnings of MS.

These considerations also have implications for some of the gene-disease associations, which have been occasionally suggested in the literature. For example, recently, Gregory and co-workers, reported genetic evidence that implicated the single nucleotide polymorphism (SNP), rs1800693, as the variant within the TNFRSF1A gene, which is associated with MS-susceptibility by genome wide association studies [39]. This is the gene, which encodes tumor necrosis factor (TNF) receptor-1. These authors further suggested that this particular genetic variant was "causal" for MS-susceptibility by demonstrating that the MS risk-allele results in expression of a novel and soluble form of TNF receptor-1. The novel transcript produced by this mutation skips Exon 6 and results in the formation of a substantially truncated protein, which functions as a TNF-blocker [39]. However, despite the seeming plausibility of this proposed mechanism for MS-susceptibility associated with this SNP, the offered explanation is, at best, incomplete – a conclusion based solely on relationships derived for the proposed model. Thus, because, only a tiny fraction ($\leq$2.2%) of the population is genetically susceptible to getting MS, and because the risk-allele frequency (MAF) for this "causative" SNP-variant is 40% [39], the maximum percentage of "risk-allele" carriers who could possibly be genetically susceptible is only 3.4% (2.2/64). Even if the risk were assumed to be carried exclusively by homozygotes for the risk-allele, this maximum percentage rises to just 13.8% (2.2/16). Consequently, this risk-allele, by itself, is insufficient to produce susceptibility – rather, it is only in combination with other susceptibility alleles that this particular variant can lead to genetic susceptibility to MS [27]. Moreover, the fact that many MS patients are not carriers (~36%) is indicated by the small odds ratio (1.12) for the association of this risk-allele with MS [39]. In such circumstances, this particular SNP-variant can hardly be described as "causative" for MS-susceptibility.

In conclusion, the mathematical model for disease pathogenesis, here developed, is capable of providing considerable insight to the nature and basis of genetic susceptibility to chronic human diseases in different groups of individuals.

Supporting Information

Appendix S1
(PDF)

Acknowledgments

Brian C. Healy, PhD (*Department of Neurology, Brigham and Women's Hospital; Harvard University*) provided invaluable assistance in the development of this mathematical model.

Author Contributions

Conceived and designed the experiments: DSG. Performed the experiments: DSG. Analyzed the data: DSG. Contributed reagents/materials/analysis tools: DSG. Wrote the paper: DSG.

References

1. Rothman KJ, Greenland S (1998) Modern Epidemiology, 2nd Edition, Lippincott Williams & Wilkins, Philadelphia.
2. Boomsma D, Busjahn A, Peltonen L (2002) Classical Twin Studies and Beyond. Nat Rev Genet 3:872–82.
3. Compston A, Confavreux C, Lassmann H, McDonald I, Miller D, et al. (2006), McAlpine's Multiple Sclerosis, 4th Edition, Churchill Livingston, London.
4. Goodin DS (2009) The causal cascade to multiple sclerosis: A model for MS pathogenesis. PLoS One 4:e4565. {hyperlinked and remarked PDF version,

complete with Supplementary Material and Correspondence, is available from corresponding author upon request}
5. Bager P, Nielsen NM, Bihrmann K, Frisch M, Wohlfart J, et al. (2006) Sibship characteristics and risk of multiple sclerosis: A nationwide cohort study in Denmark. Am J Epidemiol 163:1112–1117.
6. Compston A, Coles A (2002) Multiple sclerosis. Lancet 359:1221–31.
7. Dyment DA, Yee IML, Ebers GC, Sadovnick AD, and the Canadian Collaborative Study Group (2006) Multiple sclerosis in stepsiblings: Recurrence risk and ascertainment. J Neurol Neurosurg Psychiatry 77:258–259.
8. Ebers GC, Sadovnick, Risch NJ, and the Canadian Collaborative Study Group (1995) A genetic basis for familial aggregation in multiple sclerosis. Nature 377:150–151.
9. Ebers GC, Sadovnick AD, Dyment DA, Yee IM, Willer CJ, et al. (2004) Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet 363: 1773–1774.
10. Ebers GC, Yee IML, Sadovnick AD, Duquette P, and the Canadian Collaborative Study Group (2000) Conjugal multiple sclerosis: Population-based prevalence and recurrence risks in offspring. Ann Neurol 48:927–931.
11. French Research Group on Multiple Sclerosis (1992) Multiple sclerosis in 54 twinships: Concordance rate is independent of zygosity. Ann Neurol 32:724–727.
12. Islam T, Gauderman WJ, Cozen W, Hamilton AS, Burnett ME, et al. (2006) Differential twin concordance for multiple sclerosis by latitude of birthplace. Ann Neurol 60:56–64.
13. Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, et al. (1994) The British Isles survey of multiple sclerosis in twins. Neurology 44:11–15.
14. Nielsen NM, Westergaard T, Rostgaard K, Frisch M, Hjalgrim H, et al. (2005) Familial risk of multiple sclerosis: A nationwide cohort study Am J Epidemiol 162:774–778.
15. Orton SM, Herrera BM, Yee IM, Valdar W, Ramopgopalan SV, et al (2006) Sex ratio of multiple sclerosis in Canada: A longitudinal study. Lancet Neurol 5:932–936.
16. Ristori G, Cannoni S, Stazi MA, Vanacore N, Cotichini R, et al (2006) Multiple sclerosis in twins from continental Italy and Sardinia: A Nationwide Study. Ann Neurol 59:27–34.
17. Robertson NP, Fraser M, Deans J, Clayton D, Walker N, et al. (1996) Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. Brain 119:449–455.
18. Sadovnick AD, Dircks A, Ebers GC (1999) Genetic counselling in multiple sclerosis: risks to sibs and children of affected individuals. Clin Genet 56:118–122.
19. Sadovnick AD, Yee IML, Ebers GC, and the Canadian Collaborative Study Group (2005) Multiple sclerosis and birth order: A longitudinal cohort study. Lancet Neurol 4:611–617.
20. Sadovnick AD, Ebers GC, Dyment DA, Risch NJ, and the Canadian Collaborative Study Group (1996) Evidence for genetic basis of multiple sclerosis. Lancet 347:1728–1730.
21. Willer CJ, Dyment DA, Rusch NJ, Sadovnick AD, Ebers GC, and the Canadian Collaborative Study Group (2003) Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci (USA) 100:12877–12882.
22. Dyment DA, Herrera BM, Cader Z, Willer CJ, Lincoln MR, et al. (2005) Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. Hum Mol Genet 14:2019–2026.
23. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al (2007) Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 357, 851–862.
24. Ramagopalan SV, Anderson C, Sadovnick AD, Ebers GC (2007) Genomewide study of multiple sclerosis. N Engl J Med 357, 2199–2200.
25. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, et al. (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nature Genetics 41:776–782.
26. The International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium 2 (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476:214–219.
27. Goodin DS (2010) The genetic basis of multiple sclerosis: A model for MS susceptibility. BMC Neurol 10:101. {hyperlinked and remarked PDF version, complete with Supplementary Material, is available from corresponding author upon request}
28. Torkildsen GN, Lie SA, Aarseth JH, Nyland H, Myhr KM (2008) Survival and cause of death in multiple sclerosis: results from a 50-year follow-up in Western Norway. Mult Scler 14:1191–1198.
29. Sundström P, Nyström L, Forsgren L (2003) Incidence (1988–97) and prevalence (1997) of multiple sclerosis in Västerbotten County in northern Sweden. J Neurol Neurosurg Psychiatry 74:29–32.
30. Witte JS, Carlin JB, Hopper JL (1999) Likelihood-Based Approach to Estimating Twin Concordance for Dichotomous Traits. Genetic Epidemiol 16:290–304.
31. Kuusisto H, Kaprio J, Kinnunen E, Luukkaala T, Koskenvuo M. et al. (2008) Concordance and heritability of multiple sclerosis in Finland: Study on a nationwide series of twins. Eur J Neurol 15:1106–1110.
32. Rosati G (2001) The prevalence of multiple sclerosis in the world: an update. Neurol Sci 22:117–139.
33. Payami H, Larsen K, Bernard S, Nutt J (1994) Increased risk of Parkinson's disease in parents and siblings of patients. Ann Neurol 36:659–661.
34. Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, et al. (1999) Parkinson disease in twins: An etiologic study. JAMA 281:341–346.
35. Sundquist K, Martineus JC, Li X, Hemminki K, Sundquist J (2008) Concordant and discordant associations between rheumatoid arthritis, systemic lupus erythematosus and ankylosing spondylitis based on all hospitalizations in Sweden between 1973 and 2004. Rheumatol 47:1199–1202.
36. Höhler T, Hug R, Schneider PM, Krummenauer F, Gripenberg-Lerche C, et al. (1999) Ankylosing spondylitis in monozygotic twins: studies on immunological parameters. (1999) Ann Rheum Dis 58:435–440.
37. Bellamy N, Duffy D, Martin N, Mathews J (1992) Rheumatoid arthritis in twins: a study of aetiopathogenesis based on the Australian Twin Registry. Ann Rheum Dis 51:588–593.
38. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, et al. (1992) A revised estimate of twin concordance in systemic lupus erthymatosus. Arthritis Rheumatism, 35:313–318.
39. Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, et al. (2012) TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. Nature 488:508–11.