

### ***Manuscript Information***

***Title: Pathogenesis of Multiple Sclerosis: Genetic, Environmental, and Random Mechanisms***

***Author:*** Douglas S. Goodin, MD

Department of Neurology, *University of California, San Francisco* & the *San Francisco VA Medical Center*, San Francisco, CA, USA

### ***Address for Correspondence:***

Douglas S. Goodin, MD

Department of Neurology; *University of California, San Francisco*

675 Nelson Rising Lane, Suite #221D

San Francisco, CA 94158

Phone: (415) 514 2464

Fax: (415) 514 2470

E mail: [douglas.goodin@ucsf.edu](mailto:douglas.goodin@ucsf.edu)

### ***Manuscript Subject Areas:***

*Multiple Sclerosis, pathogenesis, environmental factors, genetic factors, randomness, determinism*

***Author Contributions:*** Douglas S. Goodin:

Conceptualization; Formal analysis; Methodology; Software; Writing – original draft, review & editing

***Funding:*** None

***Conflicts of Interest:*** None

### ***Acknowledgements:***

I am especially indebted to John Petkau, PhD, Professor Emeritus, Department of Statistics, University of British Columbia, Canada, for enormous help with this project. He devoted many hours of his time to critically reviewing early versions of this analysis and contributed immensely both to the clarity and to the logical development of the mathematical and statistical arguments presented herein. I am also indebted to my mentor, Michael J. Aminoff, MD, Professor Emeritus, Department of Neurology, University of California, San Francisco, USA, for his invaluable help with this project. He critically, and thoughtfully, reviewed many drafts of this manuscript and contributed enormously to the logic and clarity of its presentation.

## Abstract

**BACKGROUND:** MS-pathogenesis requires both genetic factors and environmental events. The question remains, however, whether these factors and events *completely* describe the MS disease-process. This question was addressed using the Canadian MS-data, which includes 29,478 individuals, representing 65-83% of all Canadian MS-patients.

**METHODS:** The “*genetically-susceptible*” subset of the population, (*G*), includes *everyone* who has *any non-zero* life-time chance of developing MS, under *some* environmental-conditions. A “*sufficient*” environmental-exposure, for *any* “*genetically-susceptible*” individual, includes *every* set of environmental conditions, each of which is *sufficient*, by itself, to *cause* MS in that person. This analysis incorporates several different *epidemiologic-parameters*, involved in MS-pathogenesis, only some of which are directly-observable, and establishes “*plausible-value-ranges*” for each parameter. Those *parameter-value* combinations (solutions) that fall within these *plausible-ranges* are then determined.

**RESULTS:** Only a fraction of the population can possibly be “*genetically-susceptible*”. Thus, many individuals have *no possibility* of developing MS under *any* environmental conditions. Moreover, *some* “*genetically-susceptible*” individuals, despite their experiencing a “*sufficient*” environmental-exposure, *will never* develop disease.

**CONCLUSIONS:** This analysis *explicitly includes* all of those genetic factors and environmental events (including interactions), which are necessary for MS-pathogenesis, regardless of whether these are known, suspected, or as yet unrecognized. Nevertheless, in addition, “*true*” randomness seems to play a critical role in disease-pathogenesis. This observation provides empirical evidence that undermines the widely-held deterministic view of nature. Moreover, both sexes seem to have a similar genetic and environmental disease-basis. If so, this indicates that this random element is primarily responsible for the currently-observed differences in disease-expression between *susceptible-women* and *susceptible-men*.

## Introduction

Multiple sclerosis (MS)-pathogenesis requires both environmental-events and genetic-factors [1-4]. Considering genetics, familial-aggregation of MS-cases is well-established. MS-risk is increased ~30-fold in non-twin siblings and ~250-fold in monozygotic (*MZ*)-twins of an *MS-proband* [1,2,5]. Moreover, 233 MS-associated genetic-traits are now identified [6]. Nevertheless, MS-genetics is complex. The strongest MS-association is with the *Class-II* haplotype, *HLA-DRB1\*15:01~DQB1\*06:02*, located at (6p21), having an odds-ratio (*OR*) of (~3) in heterozygotes and of (~6) in homozygotes [1,2,5,6]. Other MS-associations are quite weak [6] – (median – *OR* = 1.158; interquartile – range = 1.080 – 1.414). Also, *DRB1\*15:01~DQB1\*06:02* is highly “selected”, accounting for (~13%) of *DRB1~DQB1*-haplotypes – the most frequent such haplotype – among European-decedents [1-8]. Moreover, *everyone* (except *MZ*-twins) possesses a unique combination of the 233 MS-associated genetic-traits [3]. Finally, the observed *probability-range* for *MZ*-twin-concordance is: (0.11 – 0.46) and, consequently, genetics plays only a minor role in determining disease-expression (*Table-4; Reference:[3]*).

MS is also linked to environmental-events. First, a well-documented *month-of-birth* effect, linking MS-risk to the solar cycle, likely implicates intrauterine/perinatal environmental-events in MS-pathogenesis [2,9-11]. Second, given an *MS-proband*, the MS-recurrence-rate for dizygotic (*DZ*)-twins exceeds that for non-twin siblings [2,3,5] – also implicating intrauterine/perinatal environmental-events [2,3]. Third, MS becomes increasingly prevalent farther north or south from equatorial-regions [2,12]. Because this gradient is also evident for *MZ*-twin-recurrence-rates (*Table-4; Reference:[3]*), environmental-factors are likely responsible. Fourth, a prior Epstein-Barr viral (*EBV*) infection is found in almost all (> 99%) *current* MS patients [2,13,14]. If these rare *EBV-negative* patients represent *false-negative-tests* – either from inherent-errors when using *any* fixed antibody-titer “*cut-off*” to determine *EBV-positivity*, or from *only* determining antibody-responses to *some EBV-antigens* [2] – then one can conclude that *EBV-infection* is a necessary-factor in *every* causal-pathway, which led to MS in these individuals [2]. Regardless, an *EBV-infection must* somehow be involved in MS-pathogenesis [2,13,14]. Lastly, smoking and vitamin-D deficiency are implicated in MS-pathogenesis [2,15,16].

This manuscript presents an analysis regarding genetic and environmental susceptibility to MS [4] in a relatively non-mathematical format to make its conclusions accessible. The terms and definitions used for this analysis are presented in *Table-1*. For interested readers, the mathematical-development is presented in the *Supplemental-Material*. This analysis is based on the *Canadian-Collaborative-Projection-Genetic-Susceptibility-to-Multiple-Sclerosis (CCPGSMS)*-dataset [5,8,9,17-23] – a summary of which is provided in the *Supplemental-Material-Sections:10a-b*. The *CCPGSMS*-dataset includes 29,478 MS-patients (*born: 1891 – 1993*) – estimated to represent (65 – 83%) of Canadian MS-patients [5,23,24]. This cohort is assumed to represent a large random sample of the Canadian MS-population. Also, this

single population provides point-estimates and confidence-intervals for the MS-recurrence-rates in *MZ*-twins, *DZ*-twins, and non-twin siblings (*S*), and for the time-dependent changes in the *female-to-male* (*F:M*)-sex-ratio.

## Methods

### 1. Genetic-Susceptibility

A population (*Z*), consists of *N* individuals. The “genetically-susceptible” (*G*)-subset includes *everyone* who has *any* non-zero chance of developing MS under *some* environmental-conditions. Each of the ( $m \leq N$ ) individuals in the (*G*)-subset ( $i = 1, 2, \dots, m$ ) has a unique-genotype [ $G_i$ ]. The probability of the event that a *proband*, randomly-selected from (*Z*), is a (*G*)-subset member is: [ $P(G) = m/N$ ]. Membership in (*G*) is assumed independent of the environmental-conditions during *any* specific *Time-Period* ( $E_T$ ) – see legend Table-2; considering the definition of ( $E_T$ ).

The (*MS*)-subset includes *everyone* who either has, or will subsequently develop, MS. The probability of the event that a *proband*, randomly-selected from (*Z*), is both an (*MS*)-subset member and whose relevant-exposures occurred during ( $E_T$ ), is called the *MS-penetrance* for the population (*Z*) during ( $E_T$ ), or  $P(MS | E_T)$ . Also, during ( $E_T$ ), the probability of the event that a *proband*, randomly-selected from (*G*), is an (*MS*)-subset member, is called the *MS-penetrance* for the (*G*)-subset during ( $E_T$ ), or  $P(MS | G, E_T)$ . Both *penetrance-values* depend upon the environmental-conditions during ( $E_T$ ). Also defined are the subsets of *susceptible-women* (*F, G*) and *susceptible-men* (*M, G*). Their *MS-penetrance-values*, during ( $E_T$ ), are:

$$Z_w = P(MS | G, F, E_T)$$

$$\text{and: } Z_m = P(MS | G, M, E_T).$$

These *MS-penetrance-values*, ( $Z_w$ ) and ( $Z_m$ ), are also called the “*failure-probabilities*” for *susceptible-women* and *susceptible-men* during ( $E_T$ ). During *any* ( $E_T$ ), because the proportion of *women* in (*G*) is independent of environmental-conditions, the (*F:M*)-sex-ratio always reflects the ratio of these two *failure-probabilities* (see *Supplemental-Material-Section:5d*).

### 2. Environmental-Susceptibility

For each (*G*)-subset member, a family of exposures is defined that includes *every* set of environmental-exposures, each of which is “*sufficient*”, by itself, to *cause* MS in that person. Moreover, for *any* susceptible-individual to develop MS, that person must experience at least one of the “*sufficient-exposure-sets*” within *their* family. Individuals sharing the same family of *sufficient-exposures* – although possibly requiring different “*critical-exposure-intensities*” [4] – belong to the same “*exposure-group*”.

Certain environments may be *sufficient* to cause MS in *anyone* but are so improbable (e.g., intentional inoculation of someone with myelin proteins or other agents) that, effectively, they never occur spontaneously. Nevertheless, even individuals who can *only* develop MS under such extreme (or unlikely) conditions, are still ( $G$ )-subset members – i.e., they *can* develop MS under *some* environmental-conditions.

The probability of the event ( $E$ ) – i.e., that a randomly-selected member of ( $G$ ), during ( $E_T$ ), experiences an environment *sufficient* to cause MS in them – is represented as:  $P(E | G, E_T)$ . A mathematical definition for the ( $E$ )-event is provided in the *Supplemental-Material-Section:1a*.

Each set of *sufficient-exposures* is *completely* undefined and agnostic regarding: 1) how many environmental-exposures are involved; 2) when, during life, and in what order, these exposures need to occur; 3) the intensity and duration of the required exposures; 4) what these exposures are; 5) whether *any* of these exposures needs to interact with *any* genetic-factors; and 6) whether certain exposures need to be present or absent. The *only* requirement is that each *exposure-set*, taken together, is *sufficient*, by itself, to cause MS in a specific susceptible-individual or in susceptible-individuals belonging to the same *exposure-group*.

### 3. *MZ-Twins, DZ-Twins, and Siblings*

The term ( $MZ$ ) represents the event that a *proband*, randomly-selected from ( $Z$ ), is an ( $MZ$ )-subset member or, equivalently, is an *MZ-twin*. This *proband*'s twin is called their "*co-twin*". The probability that the *proband* belongs to the ( $MS, MZ$ )-subset and their *co-twin* belongs to ( $MZ$ ) is the same as the probability that the *proband* belongs to ( $MZ$ ) and their *co-twin* belongs to ( $MS, MZ$ ). Therefore, for clarity, ( $MS, MZ$ ) indicates this subset (or event) for the *proband*, whereas ( $MZ_{MS}$ ) indicates the same subset (or event) for their *co-twin*. The analogous subsets (or events) for *DZ* "*co-twins*" ( $DZ_{MS}$ ) and non-twin "*co-siblings*" ( $S_{MS}$ ) are defined similarly (*Table-1*).

Consequently,  $P(MS | MZ_{MS})$  represents the life-time probability that a randomly-selected *proband* belongs to ( $MS, MZ$ ), given that their *co-twin* belongs to ( $MZ_{MS}$ ) – a probability estimated by the *proband-wise* (or *case-wise*) *MZ-twin-concordance-rate* [25].

The term  $P(MS | IG_{MS})$  represents this concordance-rate – i.e.,  $P(MS | MZ_{MS})$  – adjusted because *MZ-twins*, in addition to sharing "identical" genotypes ( $IG$ ), also share intrauterine and, probably, other environments. This adjustment – made by multiplying the *proband-wise MZ-twin-concordance-rate* by the ( $S: DZ$ ) concordance-ratio [4] – isolates the genetic-contribution to the observed *MZ-twin concordance-rates* (*see Supplemental-Material-Section:2a*).

#### 4. Estimating $P(G)$

If the population ( $Z$ ) and the ( $G$ )-subset are identical, then, during any ( $E_T$ ), the *MS-penetrance* of ( $Z$ ) and ( $G$ ) are also identical. Consequently, the ratio of these two *penetrance-values* [4] estimates  $P(G)$  such that:

$$P(G) = P(MS | E_T) / P(MS | G, E_T) \quad \text{Equation-1}$$

If this ratio is (1), then *everyone* in the population can develop MS under some environmental-conditions. However, if the *MS-penetrance* of ( $G$ ) exceeds that of ( $Z$ ), then this ratio is less than (1), indicating that only *some* members of ( $Z$ ) have any possibility of developing MS, regardless of any exposure they either have had or could have had. Even if the “*exposure-probability*”,  $P(E | G, E_T)$ , never reaches 100% under any realistic conditions, if ( $Z$ ) and ( $G$ ) are the same, then this ratio is (1) during every ( $E_T$ ). Moreover, in any circumstance where:  $[p = P(F | G) \neq P(F)]$ , it must be that:  $(P(G) < 1)$ .

#### 5. Data-Analysis

*Cross-sectional-Models* use data from the “*current*” ( $E_T$ ) – Table-2. *Longitudinal-Models* use data regarding changes in MS-epidemiology, which have occurred over the last century [3,4,23] – see also *Supplemental-Material; Figure-S1*. *Cross-sectional-Models* make the two common assumptions that: 1) *MZ-twinning* is independent of genotype and: 2) *MS-penetrance* is independent of (*MZ*)-subset membership (*Supplemental-Material-Section:4a*). *Longitudinal-Models* make neither assumption. Initially, “*plausible-value-ranges*” are defined for both “*observed*” and “*non-observed*” *parameter-values* (Table-2). Subsequently, a “*substitution-analysis*” determined those *parameter-value-combinations* (solutions) that fall within the “*plausible-value-ranges*” for each *parameter*. For each *Model*, ( $\sim 10^{11}$ ) possible *parameter-value-combinations* were systematically-interrogated.

Currently, *MS-penetrance* for *female-probands*, whose *co-twin* belongs to ( $MZ_{MS}$ ), is (5.7)-fold greater than *MS-penetrance* for comparable *male-probands* (Table-2). Moreover, currently, both the ( $F:M$ )-sex-ratio and  $P(MS)$  are increasing [2-4,23]. Under such circumstances, almost certainly, the *current MS-penetrance* in *susceptible-women* exceeds that in *susceptible-men* (see *Supplemental-Material-Sections:3a&7g*). Therefore, it is assumed that, currently:

$$Z_w = P(MS | F, G) > P(MS | M, G) = Z_m$$

No assumptions are made regarding the circumstances of other *Time-Periods*

#### 6. Cross-sectional Models:

For notational simplicity, *parameter-abbreviations* are used: *MS-penetrance* for the  $i^{\text{th}}$  susceptible individual is:  $\{x_i = P(MS | G_i)\}$ ; the set  $(X)$  consists of *MS-penetrance* values for all susceptible-individuals – i.e.,  $(X) = (x_1, x_2, \dots, x_m)$ ; the variance of  $(X)$  is:  $\sigma_X^2$ ; *MS-penetrance* for the  $(G)$ -subset is:  $x = P(MS | G)$ ; the adjusted *MZ-twin concordance-rate* is:  $x' = P(MS | IG_{MS})$ .

During *any*  $(E_T)$ , the *MS-penetrance* for  $(Z)$  is  $P(MS)$ . As demonstrated in the *Supplemental-Material-Section:4a*, during *any*  $(E_T)$ , the *MS-penetrance* for  $(G)$  is:

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

Consequently, during *any*  $(E_T)$ , the probability  $P(G)$  is estimated by the ratio of these *penetrance-values* (*Methods #4*).

## 7. Longitudinal Models:

### General Considerations

Using standard survival-analysis methods [26], the exposure  $(u)$  is defined as the *odds* that the event  $(E)$  occurs for a randomly-selected member of the  $(G)$ -subset during *any Time-Period* (see *Supplemental-Material-Sections:5a-c*). Hazard-functions in *men*,  $h(u)$ , and *women*,  $k(u)$ , are defined in the standard manner [26] and, if these unknown hazard-functions are proportional, a proportionality-factor  $(R > 0)$  is defined such that:

$$k(u) = R * h(u).$$

The *exposure-level*  $(u = a)$ , during *some Time-Period*, is then converted into “*cumulative-hazard-functions*”,  $H(a)$  and  $K(a)$ , which represent definite integrals of these unspecified hazard-functions from an *exposure-level* of:  $(u = 0)$  to an *exposure-level* of:  $(u = a)$ .

{NB: *Cumulative-hazard measures exposure, not failure. Failure is the event that the proband develops MS. The mapping of  $(u = a)$  to  $H(a)$  or  $K(a)$ , if proportional, is “one-to-one and onto” [4]. Therefore, both exposure-measures are equivalent. However, the failure-probabilities  $(Zw$  and  $Zm)$  are exponentially related to cumulative-hazard and, therefore, are mathematically-tractable, despite the underlying hazard-functions being unspecified. Moreover, any two points on any exponential curve defines the entire curve.}*

Unlike true-survival, for MS, the *failure-probability* may not approach 100% as the *exposure-probability* approaches unity (see *Supplemental-Material-Sections:5b-e*). Moreover, the limiting-value for this *failure-probability* in *susceptible-men* (**c**) and *susceptible-women* (**d**) may not be the same. Also, (**c**) and (**d**) are constants, estimated from the *parameter-values* of  $(Zw)$ ,  $(Zm)$ ,  $P(MS)$ , and the  $(F:M)$ -



*sex-ratio* “observed” during any two *Time-Periods*.

The *exposure-level* at which MS becomes possible (i.e., the threshold) must be zero for *susceptible-women* or *susceptible-men* or both. The difference between the threshold in *women* ( $\lambda_w$ ) and that in *men* ( $\lambda_m$ ) is defined as:

$$(\lambda = \lambda_w - \lambda_m).$$

And, therefore:      if: ( $\lambda_w > \lambda_m$ ); then ( $\lambda$ ) is positive and ( $\lambda_m = 0$ )

                         if: ( $\lambda_w < \lambda_m$ ); then ( $\lambda$ ) is negative and ( $\lambda_w = 0$ )

                         if: ( $\lambda_w = \lambda_m$ ); then: ( $\lambda = \lambda_w = \lambda_m = 0$ )

As demonstrated in the *Supplemental-Material-Section:7a*, if hazards are proportional and if: [ $H(a) \geq \lambda$ ], the *cumulative-hazards* in *men* and *women* are related such that:

$$K(a) = R * [H(a) - \lambda]$$

Moreover, any *causal-chain* leading to disease can *only* include genetic-factors, environmental-events, or both (including any interactions). Therefore, if any member of (*G*) experiences an exposure *sufficient* to cause MS in them, and if, in this circumstance, this person’s probability of developing (*MS*) is less than (100%); then their outcome, in part, must be due to a “*truly*” random mechanism.

Consequently, if randomness plays *no role* in MS-pathogenesis, then: ( $c = d = 1$ ) – see *Discussion*.

Also, regardless of proportionality, any disparity between *women* and *men* in their likelihood of developing MS, during any *Time-Period*, must be due to a difference between (*c*) and (*d*), between *susceptible-men* and *women* in the likelihood of their experiencing a *sufficient-exposure*, or between both (*Supplemental-Material-Section:5d*). Therefore, assuming that: ( $c = d \leq 1$ ), also assumes that any difference between *men* and *women* in their *failure-probability* is due, exclusively, to a difference in their likelihood of experiencing a *sufficient-exposure*.

#### *Non-proportional Hazard*

If hazards in *women* and *men* are not proportional, the “*plausible-parameter-value*” ranges still limit possible solutions. However, any difference that these *values* take during different *Time-Periods* could be attributed, both potentially and plausibly, to the different environmental-circumstances of different times and places.

#### *Proportional Hazard*



The “*apparent*” value of  $(R)$ , or  $(R^{app})$ , is defined as *the* value of  $(R)$  when:  $(c = d \leq 1)$ . As demonstrated in *Supplemental-Material-Sections:7c&7g*, for proportional-hazards with proportionality-factor  $(R)$ , three conditions must hold:

- 1) if:  $R \leq 1$  ; or, if:  $R < R^{app}$  ; or, if:  $\lambda \leq 0$  ; then:  $c < d$
- 2) if:  $c = d \leq 1$  ; then, both:  $R > 1$  and:  $\lambda > 0$
- 3) if:  $R > 1$  ; then:  $\lambda > 0$

*Condition #1* excludes any possibility that:  $(c = d \leq 1)$  – see *Figures-1&2 and Results*.

*Conditions #2&3* (where:  $\lambda > 0$ ), require that, as the *odds* of a *sufficient-exposure* decreases, there must come a point where only *susceptible-men* can develop MS. This implies that, at (or below) this *sufficient-exposure-level*,  $(R = 0)$ . Consequently, the additional requirement that:  $(R > 1)$  poses a potential paradox.

There are two obvious ways to avoid this paradox (see *Supplemental-Material-Sections:7d-h*). First, if the hazards are non-proportional, although this creates other problems. For example, *women* and *men* in the same *exposure-group*, necessarily, have proportional-hazards (*Supplemental-Material-Section:7h*). Therefore, if *women* and *men* are never in the same *exposure-group*, each sex *must* develop MS in response to distinct  $\{E_i\}$  families, in which case *female-MS* and *male-MS* would represent different diseases.

Second, *Condition #1* is compatible with any  $(\lambda)$  so that, if:  $(\lambda > 0)$  and  $(R \leq 1)$ , then, at every *sufficient-exposure-level*  $(u = a)$ , the probability that a *susceptible-man*, randomly-selected, will experience a *sufficient-exposure* is as great, or greater, than this probability for a *susceptible-woman*.

## Results

### 1. Cross-sectional Models

For all *Cross-sectional* analyses [4], the supported-range for  $P(G)$  is:

$$0.003 \leq P(G) < 0.83$$

From *Equation-1*, assuming:  $(x \geq x'/2)$ ; the supported-range for  $P(G)$  is:

$$0.003 \leq P(G) < 0.55$$

### 2. Longitudinal Models

For either non-proportional or proportional-hazards and, if proportional, any  $(R)$ , the supported-range for  $P(G)$  is:

$$0.001 < P(G) \leq 0.52$$

For proportional-hazards and:  $(c = d = 1)$ , the supported ranges for the threshold-difference between *susceptible-women* and *susceptible-men*  $(\lambda)$ ; for the proportionality-factor  $(R = R^{app})$ ; and the *probability-ratio* for receiving a *sufficient-exposure* – i.e.,  $P(E | F, G)/P(E | M, G)$  – are:

$$\begin{aligned}0.0005 &\leq \lambda \leq 0.13 \\1.3 &\leq R = R^{app} \leq 1177 \\1.2 &\leq P(E | F, G) / P(E | M, G) \leq 32\end{aligned}$$

For proportional-hazards and both ( $R = 1$ ) & ( $d = 1$ ), the supported-ranges for ( $\lambda$ ) and for the limiting probability of developing MS in *susceptible-men* ( $c$ ) are:

$$\begin{aligned}0.002 &< \lambda < 2.4 \\0.002 &\leq c \leq 0.786\end{aligned}$$

## Discussion

The two principal conclusions from this analysis are: 1) the *penetrance* of ( $G$ ) is greater than that of ( $Z$ ) and, thus, *not everyone* in the population is susceptible to developing MS and: 2) at *maximum exposure-levels*, the limiting probability for developing MS in *susceptible-men* ( $c$ ) is less than that for *susceptible-women* ( $d$ ). These two conclusions, stated explicitly, are:

1.  $P(G) \leq 0.52$
- and:    2.  $c < d \leq 1$

*Conclusion #1* seems inescapable (*see Results*). Indeed, given *any* of the reported *MZ*-twin-concordance-rates, the notion that *MS-penetrance* for ( $G$ ) is the same as that for ( $Z$ ) is untenable (*Table-4; Reference:[3]*). Therefore, a large proportion of the population ( $Z$ ) must be impervious to developing MS, regardless of *any* environmental-events they either have experienced or could have experienced.

Regarding *Conclusion #2*, however, scenarios exist where: ( $c = d \leq 1$ ) might be possible. Principal among these is the possibility of non-proportional-hazards, which requires *female-MS* and *male-MS* to be different diseases (*Methods #7*). However, given the genetic and environmental evidence, this too seems untenable. For example, all but one of the 233 MS-associated loci are *autosomal*, and the single *X-chromosome* risk-variant is present in both sexes [6] – *see also Supplemental-Material-Section:7f*. Moreover, the MS-association with different *HLA*-haplotypes is the same for both sexes – *see Tables-3&4; Reference:[4]*. Family studies also suggest a common genetic-basis for MS in *women* and *men* [2-5,8,22,27,28]. Thus, both twin and non-twin siblings (*male* or *female*) of an *MS-proband* have increased MS-risk, regardless of *proband* sex [5,27,28]. Similarly, both *sons* and *daughters* of conjugal couples have markedly increased MS-risk [8,27]. Also, *male* and *female* full- or half-siblings with an *MS-proband* parent (*mother* or *father*) have increased MS-risk [2,8,22,27]. Each of these observations, supports a similar (if not the same) genetic-basis for MS in both sexes.

Also, contrary to those circumstances required whenever the proportionality-factor ( $R$ ) is greater than (1) – i.e., when ( $R > 1$ ) – *women* don't seem more likely than *men* to experience the MS-associated environmental-events (either known or suspected). Thus, for both sexes, the *month-of-birth* effect is

equally evident [2,4,9-11]; the latitude gradient is the same [2,4,12]; the impact of intrauterine/perinatal environments is similar (*Supplemental-Material-Section:2c*); EBV infection is equally common [2,4,13,14]; vitamin-D levels are the same [2,4,15,16]; and smoking tobacco is actually less common among *women* [2,4]. Collectively, these observations suggest that, *currently*, each sex experiences the relevant environmental-events in an approximately equivalent manner. Taken together, this genetic and environmental evidence implies that *female-MS* and *male-MS* represent the same underlying disease-process and, therefore, that the hazards are proportional (*Methods #7*).

Also, several lines of evidence indicate that the proportional-hazard condition of: ( $c = d \leq 1$ ) is unlikely. First, the environmental-observations (*described above*) suggest that: ( $R^{app} > R \approx 1$ ), which is impossible whenever: ( $c = d \leq 1$ ) – see *Methods #7 & Results*. Second, as in *Figure-1*, whenever ( $\lambda \leq 0$ ) or whenever ( $R \leq 1$ ), the condition that: ( $c < d$ ) is established (*Methods #7*). Third, the alternative of: ( $R > 1$ ) & ( $\lambda > 0$ ) creates a potential paradox (*Methods #7*). Although there are ways to rationalize this potential paradox with: ( $c = d \leq 1$ ), in every case, the conditions required *whenever*: ( $c < d \leq 1$ ) are far less extreme [4]. Finally, the response-curves when: ( $c = d \leq 1$ ) & ( $R > 1$ ) are steeply ascending and present only a very narrow exposure-window to explain the (*F:M*)-sex-ratio data [23] – see *Figures-2A&2B*. Moreover, following this narrow-window, the (*F:M*)-sex-ratio decreases with increasing exposure. By contrast, the Canadian MS-data documents a steadily-progressive rise in the (*F:M*)-sex-ratio over the past century [4,23] – see also *Supplemental-Material; Figure-S1*.

Nevertheless, whenever ( $c < d$ ), some *susceptible-men* will *never* develop MS, even when a susceptible-genotype co-occurs with a *sufficient-exposure*. Thus, the Canadian MS-data [5,8,9,17-23] seems to indicate that MS-pathogenesis involves a “*truly*” random element. This cannot be attributed to other, unidentified, environmental-factors (e.g., other infections, diseases, nutritional deficiencies, toxic-exposures, etc.) because each set of environmental-exposures is defined to be *sufficient*, by itself, to *cause* MS in a specific susceptible-individual. If other conditions were necessary for this individual to develop MS, then one (or more) of their *sufficient-exposure sets* would include these conditions (*Methods #2*). This also cannot be attributed to the possibility that *some* individuals can *only* develop MS under improbable conditions. Thus, the estimates for ( $c$ ) and ( $d$ ) are based solely upon observable *parameter-values* (*Methods #7*). Finally, this cannot be attributed to mild or asymptomatic-disease because this *disease-type* occurs disproportionately among *women* compared to the *current (F:M)-sex-ratio* in MS [4,23]. Naturally, invoking “*truly*” random-events in disease-expression requires replication. Nevertheless, *any* finding that: ( $c < d$ ) indicates that the behavior of some complex physical-systems (e.g., organisms) involve “*truly*” random-mechanisms.

Moreover, considering those circumstances where: ( $R = 1$ ) & ( $d = 1$ ) and, also, considering a

*man*, randomly-selected from the ( $M, G$ )-subset, who also experiences a *sufficient-environment*, the chance that he *will not* develop MS is (21 – 99%) – *see Results*. Consequently, both the genetic and environmental data, which support the conclusion that: ( $R \approx 1$ ) – *see above* – also, support the conclusion that it is this “random-element” of disease-pathogenesis, which is primarily responsible for the difference in disease-expression currently-observed between *susceptible-women* and *men*. Importantly, the fact that a process favors disease-development in *women* over *men* does *not* imply that the process *must be* non-random. For example, when flipping a biased-coin compared to a fair-coin – if both are random-processes – the only difference is that, for the biased-coin, the two possible outcomes are not equally likely. In the context of MS-pathogenesis, the characteristics of “*female-ness*” and “*male-ness*” would each simply be envisioned as biasing the coin differently (whatever characteristics are implied by these two terms).

Other authors, modeling immune-system function, also invoke random-events in MS-disease-expression [4]. In these cases, however, randomness is incorporated into their *Models* to reproduce the MS-disease-process more faithfully. However, the fact that including randomness improves a *model's* performance doesn't constitute a *test* of whether “*true*” randomness ever occurs. For example, the outcome of a dice-roll may be most accurately *modeled* by treating this outcome as a random-variable with a well-defined probability-distribution. Nevertheless, the question remains whether this probability distribution represents a *complete* description of the process, or whether this distribution is merely a convenience, compensating for our ignorance regarding the initial orientation of the dice and the direction, location, and magnitude of the forces that act on the dice during the roll [4,29,30].

It is hard to imagine that the outcomes of complex-biological-processes such as evolution and immune-function are pre-determined events, especially considering the fact that both processes are so remarkably adaptive to contemporary external-events [4,30]. Nevertheless, *proving* that any macroscopic-process is “*truly*” random is difficult. This requires an experiment (i.e., a *test*), in which the outcome predicted by determinism differs from that predicted by non-determinism.

The Canadian MS-data presents an opportunity to apply just such a *test*. Thus, the widely-held deterministic-hypothesis [4,30] *requires* that: ( $c = d = 1$ ). Any observation that either: ( $c < d = 1$ ) or: ( $c \leq d < 1$ ) indicates that “*true*” randomness is a component of disease-development and undermines the deterministic-view. Therefore, the Canadian MS-data [5,8,9,17-23], which strongly implies that: ( $c < d$ ), provides empirical evidence in support of the non-deterministic hypothesis. Importantly, this analysis *explicitly includes* all those genetic-factors and environmental-events (including interactions), which are necessary for MS-pathogenesis, regardless of whether these factors, events, and interactions are known, suspected, or as yet unrecognized. Nevertheless, in addition to these necessary prerequisites, “*true*” randomness also seems critical to disease-pathogenesis. Moreover, both sexes seem to have the same

underlying-disease. Thus, both seem to have a similar genetic-basis and, also, a similar response to the same environmental disease-determinants (*see above*). These observations suggest both that the hazards are proportional (*Methods #7*) and that ( $R \approx 1$ ). If correct, this indicates that it is this “*truly*” random-element in disease-pathogenesis, which is primarily responsible for the currently-observed differences in disease-expression between *susceptible-women* and *susceptible-men*.

## References

1. Compston A, Confavreux C, Lassman H, McDonald I, Miller D, Noseworthy J, Smith K, Wekerly H. *McAlpine 's Multiple Sclerosis*. Churchill Livingstone, Elsevier. 2006: pp.287-346.
2. Goodin DS. The Epidemiology of multiple sclerosis: Insights to disease pathogenesis. In. *Handbook of Clinical Neurology* Aminoff MJ, Boller F, Swaab DF (eds). Elsevier, London. 2014;122:231-266.
3. Goodin DS, Khankhanian P, Gourraud PA, Vince N. The nature of genetic and environmental susceptibility to multiple sclerosis. *PLoS One*. 2021;16(3): e0246157.
4. Goodin DS, Khankhanian P, Gourraud PA, Vince N. Multiple Sclerosis: Exploring the limits and implications of genetic and environmental susceptibility. *PLoS. One*. 2023;18(6): e0285599.
5. Willer CJ, Dymant DA, Rusch NJ, et al. and 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.
6. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019;365 (6460): doi:10.1126/science.aav7188.
7. Goodin DS, Khankhanian P, Gourraud PA, Vince N. Genetic susceptibility to multiple sclerosis: interactions between conserved extended haplotypes of the MHC and other susceptibility regions. *BMC Med Genomics*. 2021;14:183
8. Sadovnick AD, Ebers GC, Dymant DA, Risch NJ. The Canadian Collaborative Study Group. Evidence for genetic basis of multiple sclerosis. *Lancet*. 1996;347(9017):1728-1730.
9. Willer CJ, Dymant DA, Sadovnick AD, et al. Timing of birth and risk of multiple sclerosis: population based study. *Br Med J*. 2005;330:120-124.
10. Staples J, Ponsonby AL, Lim L. Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: longitudinal analysis. *Br Med J*. 2010;340:c1640.
11. Pantavou KG, Bagos PG. Season of birth and multiple sclerosis: a systematic review and multivariate meta-analysis *J Neurol*. 2020;267:2815-2822.
12. Sabel CE, Pearson JF, Mason DF, et al. The latitude gradient for multiple sclerosis prevalence is established in the early life course. *Brain*. 2021;144:2038-2046.
13. Kreft KL, VanNierop, GP, Scherbeijn SMJ, et al. Elevated EBNA-1 IgG in MS is associated with genetic MS risk variants. *Neurol Neuroimmunol Neuroinflamm* 2017;4:e406
14. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*. 2022;375:296-301.

15. Munger KL, Levin LI, Hollis BW, et al. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA*. 2006;296:2832-2838
16. Sowah D, Fan X, Dennett L, et al. Vitamin D levels and deficiency with different occupations: a systematic review. *BMC Public Health*. 2017;22:17:519.
17. Sadovnick AD, Ebers GC, Dyment DA, Risch NJ, and the Canadian Collaborative Study Group. Evidence for genetic basis of multiple sclerosis. *Lancet*. 1996;347:1728–1730
18. Sadovnick AD, Risch NJ, Ebers GC. Canadian collaborative project on genetic susceptibility to MS, phase 2: Rationale and method. Canadian Collaborative Study Group. *Can J Neurol Sci*. 1998;25:216-21.
19. Ebers GC, Yee IML, Sadovnick AD, Duquette P, and the Canadian Collaborative Study Group. Conjugal multiple sclerosis: Population based prevalence and recurrence risks in offspring. *Ann Neurol*. 2000;48:927–931.
20. Ebers GC, Sadovnick AD, Dyment DA, et al. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. *Lancet*. 2004;363:1773–1774.
21. Sadovnick AD, Yee IML, Ebers GC, and the Canadian Collaborative Study Group. Multiple sclerosis and birth order: A longitudinal cohort study. *Lancet Neurol*. 2005;4:611–617.
22. Dyment DA, Yee IML, Ebers GC, Sadovnick AD, and the Canadian Collaborative Study Group. Multiple sclerosis in step siblings: Recurrence risk and ascertainment. *J Neurol Neurosurg Psychiatry*. 2006;77:258–259
23. Orton SM, Herrera BM, Yee IM, et al., and the Canadian Collaborative Study Group. Sex-ratio of multiple sclerosis in Canada: A longitudinal study. *Lancet Neurol*. 2006 ;5 :932–936.
24. Canadian Census. 2010. Available from: [https://www150.statcan.gc.ca/n1/en/pub/89-503-x/2010001/article/11475-eng.pdf?st=WVL9\\_Ggm](https://www150.statcan.gc.ca/n1/en/pub/89-503-x/2010001/article/11475-eng.pdf?st=WVL9_Ggm) (last accessed 15 May 2023)
25. Witte JS, Carlin JB, Hopper JL. Likelihood-based approach to estimating twin concordance for dichotomous traits. *Genetic Epidemiol*. 1999;16:290–304.
26. Fisher LD, van Belle G. Biostatistics: A Methodology for the Health Sciences, John Wiley & Sons, New York, 1993; pp. 369-373 & 786-829.
27. Robertson NP, O’Riordan JI, Chataway J, et al. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. *Lancet*. 1997;349:1587-1590.
28. Sadovnick AD, Ebers GC, Dyment DA, Risch NJ. The Canadian Collaborative Study Group. Evidence for genetic basis of multiple sclerosis. *Lancet*. 1996;347:1728-1730.
29. Green B. Until the end of time. Alfred A Knopf, Penguin Random House, New York, USA. (2020).
30. Layzer D. Why we are free: Consciousness, free will and creativity in a unified scientific worldview. Information Publisher. 2021;ISBN-10 0983580251.



## Figure Legends

**Figure 1.** Response-curves representing the likelihood of developing MS in genetically-susceptible *women* (black lines) and *men* (red lines) with an increasing probability of a *sufficient* environmental exposure – see *Methods #2 & Supplemental-Material-Section:1a*. The curves depicted in *Panels A and B* are proportional, with a proportionality-factor ( $R$ ), although the environmental threshold is greater for *men* than for *women* – i.e., under conditions in which: ( $\lambda < 0$ ) – see *Text*. The curves depicted in *Panels C and D* are “strictly” proportional, meaning that the environmental threshold is the same for both *men* and *women* – i.e., under conditions in which: ( $\lambda = \lambda_w = \lambda_m = 0$ ) – see *Text*. The blue lines represent the change in the ( $F:M$ )-sex-ratio (plotted at various scales; indicated in each *Panel*) with increasing exposure. The thin grey vertical lines represent the portion of the response-curve that covers the change in the ( $F:M$ )-sex-ratio from 2.2 to 3.2 (i.e., the actual change observed in Canada [23] between *Time-Periods #1 & #2*). The grey lines are omitted in *Panel C* because the observed ( $F:M$ )-sex-ratio change is not possible under these conditions. In *Panel A*, although the ( $F:M$ )-sex-ratio change is possible, the condition ( $Z_w > Z_m$ ) is never possible throughout the entire response-curve. Response-curves *A, B, and D* reflect conditions in which ( $R < 1$ ); whereas curve *C* reflects conditions in which ( $R > 1$ ). If ( $R = 1$ ), the blue line in *Panel C* would be flat (see *Supplemental-Material-Sections;7c-f*). Response-curves *A and C* reflect conditions in which ( $c = d = 1$ ); whereas curves *B and D* reflect those conditions in which ( $c < d = 1$ ).

**Figure 2.** Response-curves for the likelihood of developing MS in genetically *susceptible-women* (black lines) and *men* (red lines) with an increasing probability of a *sufficient* environmental exposure – see *Methods #2 & Supplemental-Material-Section:1a*. The curves depicted are proportional, with a proportionality-factor ( $R$ ), although the environmental threshold is greater for *women* than that it is in *men* – i.e., these are conditions in which: ( $\lambda > 0$ ). Also, all of these response-curves represent actual solutions. The blue lines represent the change in the ( $F:M$ )-sex-ratio (plotted at various scales; indicated in each *Panel*) with increasing exposure. *Panels A and B* are for conditions where: ( $c = d = 1$ ). The value of ( $R$ ), specific for this condition, is termed ( $R^{app}$ ). Indeed, for *every* condition in which: ( $c = d \leq 1$ ), both: ( $R = R^{app}$ ) and the response curves for *men* and *women* have the same relationship to each other (see *Supplemental-Material-Sections:7c-f*). By contrast, *Panels C and D* represent conditions where: ( $c < d \leq 1$ ) and, in these circumstances: ( $R < R^{app}$ ). To account for the observed increase in the ( $F:M$ )-sex-ratio, the response-curves in *Panels A and B* require that the Canadian observations [23] were made within a very narrow window – i.e., for most of these response-curves, the ( $F:M$ )-sex-ratio is actually decreasing. By contrast, the response-curves in *Panels C and D* demonstrate an increasing ( $F:M$ )-sex-ratio for *every* two-point interval of exposure along the entire response-curves for *women* and *men*. The thin grey vertical lines represent the portion of these response-curves (for the depicted solution), which represents the actual change in the ( $F:M$ )-sex-ratio for specific “solutions” between *Time-Periods #1 & #2*.

**Table 1.** Definition of Terms

Terms	Definitions
$(Z)$	The population – a set consisting of $(N)$ individuals
$(F), (M)$	Subsets of <i>women</i> $(F)$ and <i>men</i> $(M)$ within $(Z)$
$(MS)$	Subset of <i>all</i> individuals within $(Z)$ who either have, or will subsequently develop, MS
$(G)$	Subset of <i>all</i> individuals within $(Z)$ who have <i>any</i> non-zero chance of developing MS under <i>some</i> environmental conditions
$(MZ), (DZ), (S)$	Subsets of monozygotic-twins $(MZ)$ , dizygotic-twins $(DZ)$ , and non-twin siblings $(S)$ within $(Z)$
<i>Proband</i>	An individual, randomly-selected either from $(Z)$ or from one of its subsets
<i>Co-twin, Co-sibling</i>	Either a twin – $(MZ)$ or $(DZ)$ – or a non-twin sibling $(S)$ of the <i>proband</i>
<i>Recurrence Rate</i>	Probability that the <i>proband</i> is a member of the $(MS)$ -subset, given that their <i>co-twin</i> or non-twin <i>co-sibling</i> is a member of the $(MS)$ -subset
<i>Penetrance</i>	Probability that a <i>proband</i> will develop MS over the course of their <i>life-time</i>
$(MZ, MS)$	Subset of <i>MZ-twin probands</i> within the $(MS)$ subset
$(MZ_{MS})$	Subset of <i>MZ co-twins</i> within the $(MS)$ subset
$(DZ_{MS}), (S_{MS})$	Subsets of <i>DZ co-twins</i> $(DZ_{MS})$ and non-twin <i>co-siblings</i> $(S_{MS})$ within the $(MS)$ subset
$P(MS   IG_{MS})$	Recurrence-rate for <i>MZ-twins</i> , adjusted because <i>MZ-twins</i> , who share identical genotypes $(IG)$ , also share their intrauterine and <i>some</i> of their childhood environments – this adjustment isolates the genetic contribution to <i>MZ-twin</i> recurrence-rates – <i>see Text</i>
$(E_T)$	Some specific <i>Time-Period</i> – <i>see legend; Table 2</i>
$P(MS   E_T)$	<i>Penetrance</i> of MS for the population $(Z)$ during $(E_T)$ – <i>see Text</i>
$C$	Ratio of <i>MS-penetrance</i> during <i>Time-Period #1</i> , $P(MS)_1$ , to that during <i>Time-Period #2</i> , $P(MS)_2$
$P(MS   G, E_T)$	<i>Penetrance</i> of MS for the $(G)$ subset of the population $(Z)$ during $(E_T)$ – <i>see Text</i>
$Z_w$	<i>Penetrance</i> of MS for the subset of <i>susceptible-women</i> $(F, G)$ within $(Z)$ during $(E_T)$ – Also called the <i>failure probability</i> for <i>susceptible-women</i> during $(E_T)$
$Z_m$	<i>Penetrance</i> of MS for the subset of <i>susceptible-men</i> $(M, G)$ within $(Z)$ during $(E_T)$ – Also called the <i>failure probability</i> for <i>susceptible-men</i> during $(E_T)$
$c, d$	Limiting values (constants) for the <i>failure probability</i> in <i>susceptible-men</i> $(c)$ ; and <i>susceptible-women</i> $(d)$ – i.e., $(Z_m \leq c \leq 1)$ and $(Z_w \leq d \leq 1)$
$p$	Proportion of women in the $(G)$ subset – i.e., $p = P(F   G)$
$(E)$	Event that a randomly-selected member of $(G)$ – the <i>proband</i> – experiences an environment <i>sufficient to cause</i> MS in them
$P(E   G, E_T)$	Probability that the event $(E)$ occurs during $(E_T)$ for a <i>proband</i> randomly-selected from $(G)$
$u$	Variable, which represents the level of environmental-exposure as measured by the odds that the event $(E)$ occurs during <i>any</i> $(E_T)$
$a$	Level of environmental-exposure during <i>some</i> specific $(E_T)$ – i.e., when: $(u = a)$
$h(u), k(u)$	Unknown (and unspecified) hazard functions for <i>susceptible-men</i> – $h(u)$ ; and for <i>susceptible-women</i> – $k(u)$
$H(a), K(a)$	Cumulative hazard functions for <i>susceptible-men</i> – $H(a)$ ; and for <i>susceptible-women</i> – $K(a)$ – Defined as the definite integrals of these unknown and unspecified hazard functions from an <i>exposure-level</i> of: $(u = 0)$ to an <i>exposure-level</i> of: $(u = a)$ .
$R > 0$	Value of the proportionality-factor (if the hazards are proportional) – i.e., $k(u) = R * h(u)$
$R^{app}$	The “apparent” value of $R$ – i.e., the value of $R$ for proportional hazards when: $(c = d \leq 1)$
$\lambda_w, \lambda_m$	Environmental-exposure thresholds for developing MS in <i>susceptible women</i> $(\lambda_w)$ and <i>susceptible men</i> $(\lambda_m)$ – <i>see Text</i>
$\lambda$	Difference in the environmental-exposure threshold between <i>susceptible women</i> and <i>susceptible men</i> : i.e., $\lambda = \lambda_w - \lambda_m$

**Table 2.** *Parameter-values* – Point Estimates and Plausible Ranges \*

Observed Parameters	Definition	Estimate	Estimated Range †
Penetrance of MS for the Population ( $Z$ )	$P(MS) = P(MS   Z)$	0.0035	0.001 – 0.006
Proportion of <i>women</i> in the Population ( $Z$ )	$P(F) = P(F   Z)$	0.504	–
Proportion of <i>women</i> in the ( $MS$ ) subset	$P(F   MS)$	0.717	0.66 – 0.78
<i>Time-Period #1</i> (1941 – 1945)	$P(F   MS)_1$	0.685	0.67 – 0.71
<i>Time-Period #2</i> (1976 – 1980)	$P(F   MS)_2$	0.762	0.74 – 0.78
Recurrence-Rate for $MZ$ -twins ( $MZ$ )	$P(MS   MZ_{MS})$	0.253	0.18 – 0.33
<i>female</i> $MZ$ -twins	$Z_w = P(MS   F, MZ_{MS})$	0.340	0.24 – 0.44
<i>male</i> $MZ$ -twins	$Z_m = P(MS   M, MZ_{MS})$	0.065	0.014 – 0.18
Difference between <i>females</i> and <i>males</i>	$Z_w - Z_m$	0.275	0.16 – 0.39
Recurrence-Rate for $DZ$ -twins ( $DZ$ )	$P(MS   DZ_{MS})$	0.054	0.018 – 0.09
Recurrence-Rate for non-twin siblings ( $S$ )	$P(MS   S_{MS})$	0.029	0.017 – 0.041
( $S: DZ$ ) Concordance-ratio	$P(MS   S_{MS}) / P(MS   DZ_{MS})$	0.537	0.12 – 1.0
Non-observed Parameters	Definition	Estimate	Plausible Range ††
Proportion of Population in the ( $G$ ) subset	$P(G)$	–	$0 < P(G) \leq 1$
Proportion of <i>women</i> in the ( $G$ ) subset	$p = P(F   G)$	–	$0 < P(F   G) < 1$
Ratio of [ $P(MS)$ ] during <i>Time-Period #1</i> , to that during <i>Time-Period #2</i>	$C = [P(MS)_1 / P(MS)_2]$	–	$0.25 \leq C \leq 0.9$

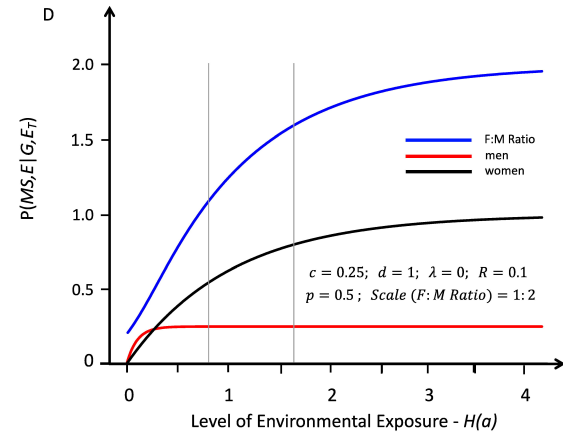
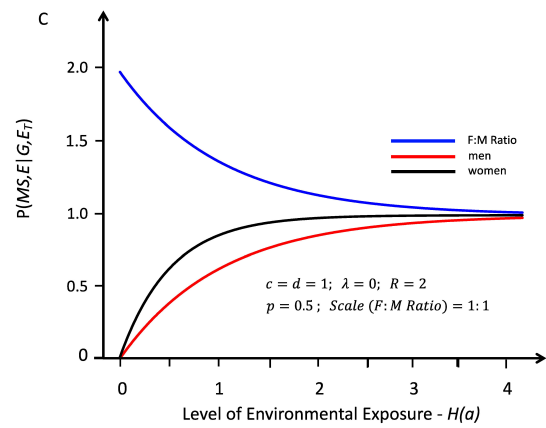
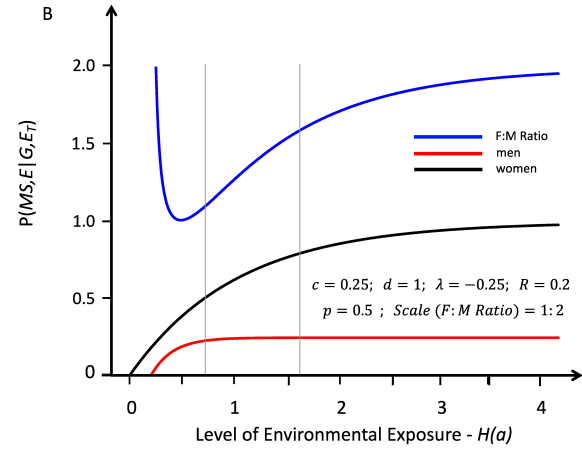
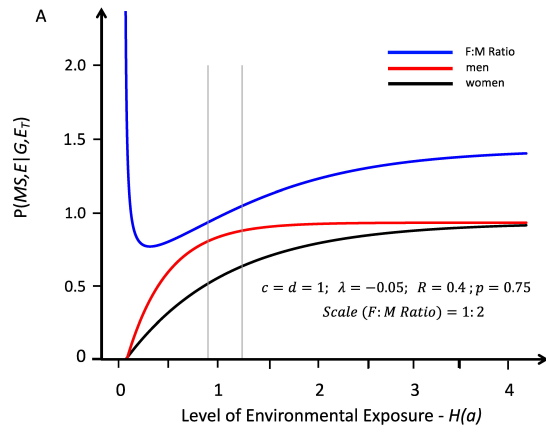
\* Estimated values and “plausible” ranges for observed and non-observed *parameters* [4] – *Supplemental Material* (Sections 10a-b). Because the MS-status of individuals born during *Time-Period #2* (1976–1980), cannot be determined until 25-35 years later, all parameter estimates – except  $P(F | MS)_1$  – are for the “current” *Time-Period* (2001–2015). Estimates for all observed *parameter-values* – except  $P(MS)$  and  $P(F)$  – are exclusively from the *CCPGSMS*-dataset [5,8,9,17-23]. The estimate for  $P(MS)$  is based upon three measures: 1) the population-prevalence of MS; 2) the age-specific prevalence of MS in the age-band of 45-54 years; and 3) the proportion of death certificates mentioning MS [3]. The parameter  $P(F)$  is taken from the 2010 Canadian census [24]. Also,  $P(MS)$  has been increasing in many regions around the world – especially among *women* [3,4]. In Canada, based on the point-estimates provided in this *Table*, it has increased by ( $\geq 32\%$ ) between the two *Time-Periods* – see *Supplemental-Material-Section:8a*). If all of the environmental events, relevant to *MS*-pathogenesis, take place prior to the age of 30 years, then, for an individual born in 1975, ( $E_T$ ) would extend from 1975 to 2005 whereas, for a person born in 1980, ( $E_T$ ) would extend from 1980 to 2010. If the relevant age-window is different than 30 years, then the definition of ( $E_T$ ) would change accordingly.

† Ranges represent the 95% confidence intervals [4]. To include a broader range of possible solutions, the range for  $P(MS)$  was expanded beyond the range of:  $\{0.0025 \leq P(MS) \leq 0.0046\}$ , which was supported by the three *above* methods [3]. The range for  $P(F | MS)$  was similarly expanded [4], as was the range for the ( $S: DZ$ ) concordance-ratio, considering the

theoretical constraint [4] that:  $[(S: DZ) \leq 1]$ . Because  $P(F)$  is taken from a census of the entire “current” Canadian population at the time (2010), there is no estimated range [24].

†† Ranges represent the “plausible” parameter-value range for each parameter. For example, because, currently, both men and women can (and do) develop MS,  $P(G)$  cannot be (0) and  $P(F | G)$  cannot be either (0) or (1). Also, the theoretical upper-limit for the value of the ratio ( $C$ ) is 0.9 [2-4]. In addition, a greater than 4-fold increase in the prevalence of (MS) over the last 35 years seems implausible based upon the available world-wide evidence [2-4]; including the evidence for MS in Canada – see *Supplemental Material (Sections 8a & 10a-b)*; see also Rosati G, *Neurol Sci* 2001;22:117-39.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).



It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/) .

