Original research

Pathogenesis of multiple sclerosis: genetic, environmental and random mechanisms

Douglas S Goodin [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jnnp-2023-333296).

[1]Neurology, University of California San Francisco, San Francisco, California, USA
[2]Neurology, San Francisco VA Medical Center, San Francisco, California, USA

Correspondence to
Dr Douglas S Goodin; douglas.goodin@ucsf.edu

Received 27 December 2023
Accepted 20 March 2024
Published Online First 2 July 2024

ABSTRACT

Background The pathogenesis of multiple sclerosis (MS) 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, estimated to represent 65–83% of all Canadian patients with MS.

Method 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 many epidemiological parameters, involved in MS pathogenesis, only some of which are directly observable, and establishes 'plausible' value ranges for each parameter. Those parameter value combinations (ie, solutions) that fall within these plausible ranges are then determined.

Results Only a small proportion of the population (≤52%) has any possibility of developing MS, regardless of any environmental conditions that they could experience. Moreover, some of these genetically-susceptible individuals, despite their experiencing a 'sufficient' environmental exposure, will still not develop disease.

Conclusions This analysis explicitly includes all of those genetic factors and environmental events (including their interactions), which are necessary for MS pathogenesis, regardless of whether these factors, events and interactions are known, suspected or as yet unrecognised. Nevertheless, in addition, a 'truly' random mechanism also seems to play a critical role in disease pathogenesis. This observation provides empirical evidence, which undermines the widely-held deterministic view of nature. Moreover, both sexes seem to share a similar genetic and environmental disease basis. If so, then it is this random mechanism, which is primarily responsible for the currently-observed differences in MS disease expression between *susceptible women* and *susceptible men*.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Several epidemiological facts regarding multiple sclerosis (MS) are well-established. These facts include (1) The pathogenesis of MS involves genetic and environmental events; (2) both the prevalence of MS and the proportion of women among MS patients are increasing; (3) women are currently more likely to develop MS than men; and (4) the probability of developing MS for an *MZ*-twin, whose *co-twin* has MS, is substantially greater than this same probability for someone in the general population. However, a unifying concept of how these disparate facts fit together is lacking.

WHAT THIS STUDY ADDS

⇒ This study provides such a unifying concept. It establishes that only a small subset of the general population has *any* non-zero chance of developing MS. Moreover, it finds that, in addition to the necessary genetic and environmental mechanisms, disease pathogenesis also involves '*truly*' random' mechanisms—a finding that undermines the widely-held deterministic view of nature. Finally, it seems likely that these random mechanisms are primarily responsible for the currently-observed differences in disease expression between the sexes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ A better understanding of the precise nature of disease pathogenesis, not only for MS but also for other diseases, can help to guide the development of more specific and more effective therapies.

INTRODUCTION

The pathogenesis of multiple sclerosis (MS) requires both environmental events and genetic factors.[1–4] Considering genetics, the familial aggregation of MS cases is well-established. Thus, compared to the general population, 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 have now been identified.[6] Nevertheless, the genetics of MS is complex. The strongest MS association is with the *HLA* Class-II haplotype, DRB1*15:01~DQB1*06:02, located in the *MHC* region at (6p21.3) on the short arm of chromosome 6. This haplotype has an OR for disease of (~3) in heterozygotes and of (~6) in homozygotes.[1 2 5 6] By contrast, the other MS-associations are quite weak[6]—with a median OR of (1.158) and an IQR of (1.080–1.414). Furthermore, DRB1*15:01~DQB1*06:02 is highly '*selected*', accounting for 12–13% of all DRB1~DQB1 haplotypes—the most frequent such haplotype—among European decedents.[1–8] In addition, everyone (except *MZ*-twins) possesses a unique combination

To cite: Goodin DS. *J Neurol Neurosurg Psychiatry* 2024;**95**:1002–1011.

Table 1 Definition of terms (in rough order of appearance)

Terms	Definitions
(*Z*)	The population—a set consisting of (*N*) individuals.
(*F*), (*M*)	Subsets of women (*F*) and men (*M*) within (*Z*).
(*MS*)	Subset of all individuals within (*Z*) who either have, or will subsequently develop, MS.
(*G*)	Subset of *all* individuals within (*Z*) who have *any* non-zero chance of developing MS under *some* environmental conditions.
G_i	Genotype of the i^{th} susceptible individual (i = 1, 2, …, m)
$\{E_i\}$	A family consisting of *every* set of environmental exposures, each of which is '*sufficient*', by itself, to *cause* MS in the i^{th} susceptible person.
(*MZ*), (*DZ*), (*S*)	Subsets of monozygotic-twins (*MZ*), dizygotic-twins (*DZ*) and non-twin siblings (*S*) within (*Z*).
Proband	An individual, randomly-selected either from (*Z*) or from one of its subsets.
Co-twin, Co-sibling	Either a twin—(*MZ*) or (*DZ*)—or a non-twin sibling (*S*) of the *proband*.
Concordance-Rate (*CR*)	Probability that the *proband* is a member of the (*MS*) subset, given that their *co-twin* or non-twin *co-sibling* is a member of the (*MS*) subset. Also, often referred to as the Recurrence-Rate or Recurrence-Risk.
Penetrance	Probability that a *proband* will develop MS over the course of their life-time.
(*MZ, MS*)	Subset of *MZ*-twin *probands* or *MZ co-twins* within the (*MS*) subset.
(MZ_{MS})	Subset of *MZ co-twins* who are members of the (*MZ, MS*) subset.
DZ_{MS}, S_{MS}	Subsets of *DZ co-twins* (DZ_{MS}) and non-twin *co-siblings* (S_{MS}) within the (*MS*) subset
$P(MS \mid MZ_{MS})$	Concordance Rate of MS for a *proband MZ*-twin, given that their *co-twin* has MS[25]– $P(MS \mid MZ_{MS}) = x''$ Comparable definitions pertain to subsets of *susceptible women* (z_w'') and *susceptible men* (z_m'') (see Methods: *MZ*-twins, *DZ*-twins and siblings; table 3 & online supplemental table S1).
$P(MS \mid IG_{MS})$	Concordance Rate of MS for *MZ-twins*, adjusted because *MZ*-twins, who share identical genotypes–the (*IG*) subset–also share their intrauterine and, probably, some of their other environments. This adjustment isolates the genetic contribution to *MZ-twin* concordance rates (see Methods: *MZ*-twins, *DZ*-twins, and siblings). Notably: (*IG*) = (*MZ*)
(E_T)	Some specific Time-Period (see legend; table 3).
$P(MS \mid E_T)$	Penetrance of MS for the population (*Z*) during (E_T) (see Methods: Genetic susceptibility).
C	Ratio of *MS-penetrance* during *Time-Period #1*, $P(MS)_1$, to that during *Time-Period #2*, $P(MS)_2$.
$P(MS \mid G, E_T)$	Penetrance of MS for the (*G*) subset of the population (*Z*) during (E_T) (see Methods: Genetic susceptibility).

MS, multiple sclerosis.

Table 2 Definition of terms (continued)

Terms	Definitions
Zw	Penetrance of MS for the subset of *susceptible women* (*F, G*) within (*Z*) during (E_T) – Also called the '*failure-probability*' for *susceptible women* during (E_T).
Zm	Penetrance of MS for the subset of *susceptible men* (*M, G*) within (*Z*) during (E_T) – Also called the '*failure-probability*' for *susceptible men* during (E_T).
c, d	Limiting values (constants) for the '*failure- probability*' in *susceptible men* (***c***); and *susceptible women* (***d***)—that is, ($Zm \leq \boldsymbol{c} \leq 1$) and ($Zw \leq \boldsymbol{d} \leq 1$). ***c***=limit of (*Zm*) as: ($a \rightarrow \infty$) & ***d***=limit of (*Zw*) as: ($a \rightarrow \infty$).
p	Proportion of women in the (*G*) subset—that is, $p = P(F \mid G)$.
(*E*)	Event that a randomly-selected member of (*G*)—the *proband*—experiences an environment '*sufficient*' to *cause* MS in them.
$P(E \mid G, E_T)$	Probability that the event (*E*) occurs during (E_T) for a *proband* randomly-selected from (*G*).
u	Variable, which represents the environmental *exposure-level* as measured by the odds that the event (*E*) occurs during any (E_T).
a	Environmental *exposure-level* during some specific (E_T)—that is, when: (u=a).
h(*u*), *k*(*u*)	Unknown (and unspecified) hazard functions for *susceptible men*—*h*(*u*); and for *susceptible women*—*k*(*u*).
H(*a*), *K*(*a*)	Cumulative hazard functions for *susceptible men* – *H*(*a*); and for *susceptible women* – *K*(*a*)—defined as the definite integrals of these unknown and unspecified hazard functions from an *exposure-level* of (u=0) to an *exposure-level* of (u=a).
R>0	Value of the proportionality factor (if the hazards are proportional)—that is, $k(u) = R * h(u)$.
R^{app}	The 'apparent' *value of R*—that is, the value of *R* for proportional hazards whenever: ($\boldsymbol{c}$=$\boldsymbol{d} \leq 1$).
λ_w, λ_m	Environmental exposure thresholds for developing MS in *susceptible women* (λ_w) and *susceptible men* (λ_m) (see Methods: Longitudinal models: General considerations).
λ	Difference in the environmental exposure *threshold* between *susceptible women* and *susceptible men*: that is, $\lambda = \lambda_w - \lambda_m$.

MS, multiple sclerosis.

of these 233 MS-associated genetic traits.[3] Finally, considering the available evidence, the maximum estimate possible for the probability range of *MZ*-twin concordance rates is (0.11–0.46) (see table 4 of reference 3). Consequently, genetics plays only a minor role in determining MS disease expression.

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 concordance rate for dizygotic (*DZ*)-twins (see tables 1 and 2) exceeds that for non-twin siblings[2 3 5]—also implicating intrauterine/perinatal environmental events.[2 3] Third, MS becomes increasingly prevalent in geographical regions farther north or south from the equator.[2 12] Because this gradient is also evident for *MZ*-twin concordance rates (see table 4 of reference 3), environmental factors are likely responsible. Fourth, evidence of a prior Epstein-Barr viral (*EBV*) infection is present in almost all (>99%) current patients with MS.[2 13 14] If these rare *EBV-negative* patients represent *false-negative* tests—either from inherent errors when using any fixed antibody-titre 'cut-off' to determine *EBV-positivity*, or from only determining antibody-responses to some of the *EBV* antigens[2]—then one would conclude that an *EBV* infection is a necessary environmental factor in every causal pathway, which led to MS in these individuals.[2] Regardless, however, it *must* be the case that an *EBV* infection plays an important role 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. For interested readers, the mathematical development of the analytic *Models* is presented in the online supplemental material. This analysis is based on data from the Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis (CCPGSMS) study group[5 8 9 17–23]—a summary of which is provided in the online supplemental material sections 10a,b. The CCPGSMS data set

includes 29 478 patients with MS who were born between 1891 and 1993 and who are estimated to represent 65–83% of Canadian patients with MS.[5 23 24] This cohort is assumed to represent a large random sample of the symptomatic Canadian MS population at the time. Also, this single population provides point estimates and CIs for the MS concordance 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.

(*NB: Generally, publications from the CCPGSMS study group do not distinguish between the different clinical 'subtypes' of MS such as relapsing-remitting MS (RRMS), secondary-progressive MS (SPMS) and primary-progressive MS (PPMS). Nevertheless, 85–90% of diagnosed MS cases have a relapsing onset and all subtypes share similar environmental and genetic determinants.*)[1]

METHODS

Genetic susceptibility

The terms and definitions for the different analytic *Models* described in this manuscript are presented in tables 1 and 2 and in online supplemental table S1 (online supplemental material sections 9a,b).

This analysis considers a population (*Z*), which consists of *N* individuals ($k=1,2,\ldots,N$). The '*genetically-susceptible*' subset of this population (*G*), is defined to include *everyone* who has *any* non-zero life-time chance of developing MS under *some* environmental conditions. Each of the ($m \leq N$) individuals in the (*G*) subset ($i=1,2,\ldots,m$) has a unique genotype (G_i) (*see* online supplemental material sections 1a and 4a). The probability (*P*) of the event that an individual, randomly selected from the population (*Z*)—*the proband*—is a member of the (*G*) subset is: ($P(G)=m/N$). Membership in (*G*)—that is, the *genetic* basis of MS—is assumed to be independent of the environmental conditions during any specific *Time-Period* (E_T)—see the legend of table 3 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 the population (*Z*)—whose relevant exposures occurred during (E_T)—is a member of the (*MS*) subset is called the *MS-penetrance* for the population (*Z*) during (E_T) or $P(MS \mid E_T)$. Similarly, the probability of the event that a *proband* randomly-selected from the (*G*) subset—whose relevant exposures occurred during (E_T)—is a member of the (*MS*) subset, is called the *MS-penetrance* for the (*G*) subset during (E_T), or $P(MS \mid G, E_T)$. Both of these *MS-penetrance* values depend on the environmental conditions during (E_T). Also defined are the subsets of *susceptible women* (*F,G*) and *susceptible men* (*M,G*). The *MS-penetrance* values, during (E_T), for these two subsets are:

$$Zw = P(MS \mid G, F, E_T) \;\&\; Zm = P(MS \mid G, M, E_T)$$

These *MS-penetrance* values, (*Zw*) and (*Zm*), are also called the '*failure-probabilities*' for *susceptible women* and *susceptible men* during (E_T) Because it is assumed (see immediately above) that membership in the (*G*) subset is independent of the environmental conditions of (E_T), the proportion of women (*F*) in the (*G*) subset—that is, $P(F \mid G)$—will also be independent of

Table 3 Parameter-values—point estimates and plausible ranges*

Observed parameters	Definition	Estimate	Estimated range†
Penetrance of MS for the Population (*Z*)	$P(MS) = P(MS \mid Z)$	0.003	0.001–0.006
Proportion of women in the Population (*Z*)	$P(F) = P(F \mid Z)$	0.504	–
Proportion of women in the (*MS*) subset	$P(F \mid MS)$	0.717	0.66–0.78
Time-Period #1 (1941–1945)	$P(F \mid MS)_1$	0.685	0.67–0.71
Time-Period #2 (1976–1980)	$P(F \mid MS)_2$	0.762	0.74–0.78
Concordance Rate (*CR*) for *MZ-twins* (*MZ*)	$x'' = P(MS \mid MZ_{MS})$	0.253	0.18–0.33
(*CR*) for female MZ-twins	$z_w'' = P(MS \mid F, MZ_{MS})$	0.340	0.24–0.44
(*CR*) for male MZ-twins	$z_m'' = P(MS \mid M, MZ_{MS})$	0.065	0.014–0.18
Difference between (*CR*) for females and males	$z_w'' - z_m''$	0.275	0.16–0.39
Ratio of (*CRs*): females to males	z_w'' / z_m''	5.231	1.74–25.1
Concordance Rate for *DZ*-twins (*DZ*)	$P(MS \mid DZ_{MS})$	0.054	0.018–0.09
Concordance Rate for non-twin siblings (*S*)	$P(MS \mid S_{MS})$	0.029	0.017–0.041
(*S:DZ*) Concordance ratio	$P(MS \mid S_{MS}) / P(MS \mid DZ_{MS})$	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 *women* within the (*G*) subset	$p=P(F \mid G)$	–	$0<P(F \mid G) < 1$
Ratio of *P*(*MS*) during *Time-Period #1*, to that during *Time-Period #2*	$C = P(MS)_1 / P(MS)_2$	–	$0.25 \leq C \leq 0.9$

*Estimated values and 'plausible' ranges for observed and non-observed parameters[4] (online 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 $P(F \mid MS)_1$—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 data set.[5 8 9 17–23] The estimate for *P*(*MS*) is based on 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 (≥32%) between the two *Time-Periods* (see online 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 1974 to 2005 whereas, for a person born in 1980, (E_T) would extend from 1979 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% CIs.[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 $P(F \mid MS)$ 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 $P(F \mid G)$ 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 four-fold increase in the prevalence of MS over the last 35–40 years seems implausible based on the available worldwide evidence[2–4]; including the evidence for MS in Canada (see online supplemental material sections 8a and 10a,b; see also Rosati G, Neurol Sci 2001;22:117–39).
CCPGSMS, Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis; MS, multiple sclerosis.

these environmental conditions. Consequently, the '*observed*' (*F:M*) sex-ratio always reflects the ratio of these two failure-probabilities (see online supplemental material section 5d).

Environmental susceptibility

For each member of the (*G*) subset—for example, the i^{th} member of (*G*)—a family of exposures $\{E_i\}$ is defined to include every set of environmental exposures, each of which is '*sufficient*', by itself, to *cause* MS to develop in the i^{th} susceptible individual (see online supplemental material section 1a). Moreover, for *any* susceptible individual to develop MS, that person must experience at least one of the '*sufficient*' *exposure-sets* within their $\{E_i\}$ family. Individuals who share the same $\{E_i\}$ family of '*sufficient*' exposures—although possibly requiring different '*critical exposure-intensities*' [4]—are said to belong to the same '*exposure-group*' (see online supplemental material section 1a).

Certain environmental conditions may be '*sufficient*' to cause MS in anyone but are so unlikely (eg, the intentional inoculation of a person with myelin proteins or other agents) that, effectively, they never occur spontaneously. Nevertheless, even individuals who can *only* develop MS under such improbable (or extreme) conditions, are still members of the (*G*) subset—ie, they can develop MS under *some* environmental conditions.

The term (*E*) is defined as the event that a *proband*, randomly-selected from the (*G*) subset, experiences an environment '*sufficient*' to cause MS in them. The probability of this event for a susceptible *proband*, who has their relevant exposures during (E_T), is represented as: $P(E \mid G,E_T)$. A precise mathematical definition of the event (*E*) is provided in the online 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 and how many of, and in what manner, these exposures need 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*, within the $\{E_i\}$ family, taken together, is '*sufficient*', by itself, to *cause* MS to develop in a specific susceptible individual (ie, the i^{th} susceptible individual) or in susceptible individuals who belong to the same ('*i-type*') *exposure-group* (see online supplemental material section 1a).

MZ-twins, DZ-twins and siblings

The term (*MZ*) represents the event that a *proband*, randomly selected from the population (*Z*), is a member of the (*MZ*) subset or, equivalently, is an *MZ*-twin. This *proband's* twin is called their '*co-twin*' (see table 1). The probability that the *proband* belongs to the (*MS,MZ*) subset, given that their *co-twin* belongs to (*MZ*), is the same as the probability that their *co-twin* belongs to (*MS,MZ*), given that the *proband* belongs to (*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*, given that both *twin* and *co-twin* are members of the (*MZ*) subset. Therefore:

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

The analogous subsets (or events) for *DZ* '*co-twins*' (DZ_{MS}) and non-twin '*co-siblings*' (S_{MS}) are defined similarly (see table 1).

Consequently, $P(MS \mid MZ_{MS})$ 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 that is estimated by the '*observed*' *proband-wise* (or *case-wise*) *MZ*-twin concordance rate.[25]

This *MZ*-twin concordance rate—that is, $P(MS \mid MZ_{MS})$—may require some adjustment because *MZ*-twins, in addition to sharing '*identical*' genotypes (*IG*), also share their intrauterine and, likely, other environments. This adjusted rate—referred to as $P(MS \mid IG_{MS})$—is estimated by multiplying the proband-wise *MZ*-twin concordance rate by the (*S:DZ*) concordance ratio.[4] This estimate isolates the genetic contribution to the observed *MZ*-twin concordance rates (see online supplemental material section 2a). Notably: the subsets (*IG*) and (*MZ*) are identical.

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

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

$$P(G) = P(MS \mid E_T) / P(MS \mid G, E_T) \quad (1)$$

If this ratio is equal to one, 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 one, which indicates 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*'—that is, $P(E \mid G,E_T)$—never reaches 100% under any realistic conditions, if (*Z*) and (*G*) are the same, then this ratio is equal to one during every (E_T). Also, the proportion of women (*F*) among susceptible individuals is expressed as ($p=P(F \mid G)$). For any circumstance, in which this proportion differs from that in the population—ie, ($p \neq P(F)$)—it *must* be the case that ($P(G)<1$).

Data analysis

The Cross-sectional-Models use data from the '*current*' (E_T)—see table 3. The Longitudinal-Models use data regarding changes in MS epidemiology, which have occurred over the past half century[3 4 23] (see also online supplemental material figure S1). The Cross-sectional-Models make the two common assumptions that: (1) *MZ*-twining is independent of genotype and (2) *MS-penetrance* is independent of (*MZ*) subset membership (online supplemental material section 4a). The Longitudinal-Models make neither assumption. Initially, for either Model type, '*plausible*' value-ranges are defined for both '*observed*' and '*non-observed*' epidemiological-parameters (see table 3). Subsequently, incorporating the known (or derived) parameter relationships (see online supplemental material), a '*substitution-analysis*' was used to determine those parameter value combinations (ie, solutions) that fall within the '*plausible*' value ranges for each parameter.[4] For each Model, ($\sim 10^{11}$) possible parameter value combinations were systematically interrogated.

Currently, the *MS-penetrance* for female *probands,* whose *co-twin* belongs to (MZ_{MS}), is ~5-fold greater than the *MS-penetrance* for comparable male *probands* (see table 3; see also online supplemental material section 10b). Moreover, *currently*, both the (*F:M*) sex-ratio and the *MS-penetrance* of the population—i.e., *P*(*MS*)—are known to be increasing, both in Canada and around the world[2–4 23] (see also online supplemental material sections 8a and 10a,b). Under such circumstances, almost certainly, the *current MS-penetrance* in *susceptible women* exceeds that in *susceptible men* (see online supplemental material sections 3a and 7g). Therefore, it is assumed that, *currently*:

$$Zw = P(MS \mid F, G) > P(MS \mid M, G) = Zm$$

No assumptions are made about the relationship between (*Zw*) and (*Zm*) during other *Time-Periods*.

Notably, however, if: ($P(G)=1$); then, during every *Time-Period* it must be that: ($p=P(F)$)—see Methods: Estimating the probability of genetic susceptibility in the population (above). Therefore, in the *current* case, and indeed during any (E_T), whenever: ($P(F|MS,E_T)>P(F|G)=p$)—the relationship of: ($Zw>Zm$) is gurauranteed (see online supplemental material sections 3a and 5d).

Cross-sectional models

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

During any (E_T), the *MS-penetrance* of the population (Z) is $P(MS)$. As demonstrated in the online supplemental material section 4a, during any (E_T), the *MS-penetrance* of the *genetically-susceptible* subset (G) is:

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

Consequently, during any (E_T), the probability of *genetic-susceptibility* in the population ($P(G)$) is estimated by the ratio of these two *MS-penetrance* values (see equation 1; Methods: Estimating the probability of genetic susceptibility in the population).

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 online supplemental material sections 1a and 5a–c). Hazard functions in men, $h(u)$, and women, $k(u)$, are defined in the standard manner[26] and, if these unknown (and unspecified) 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 is being used here as a measure of exposure, not failure.*[4] *Failure is the event that the randomly-selected proband develops MS. The mapping of (u=a) to both H(a) and K(a), if proportional, is 'one-to-one and onto'.*[4] *Therefore, in this case, the two exposure measures—ie, (a) and H(a)—are equivalent. However, the failure-probabilities, (Zw) and (Zm) are exponentially related to cumulative-hazard and, therefore, the exposure-measures of H(a) and K(a) are mathematically tractable, despite the underlying hazard functions being unknown and unspecified—see online supplemental material sections 1a and 5a–c. Moreover, notably, any two points on any exponential response-curve define the entire response-curve completely.*)

In true survival, everyone dies if given a sufficient amount of time. By contrast, as the *exposure-probability*, $P(E|G,E_T)$, approaches unity, the probability of failure (ie, developing MS), either for *susceptible men* (*Zm*) or for *susceptible women* (*Zw*), may not similarly approach 100%. Moreover, the maximum value for this *failure-probability* in *susceptible men* (c) might not be the same as the maximum value for this *failure-probability* in *susceptible women* (d) (see online supplemental material sections 5b–e). Also, the constants (c) and (d) are estimated from the Longitudinal Model, using the parameter values of $P(MS)$ and the (F:M) sex-ratio '*observed*' during any two *Time-Periods* (see Methods: Data analysis; see also online supplemental material section 5e).

By definition, the exposure-level at which the development of MS becomes possible (ie, the *threshold*) must occur at zero for susceptible women, or for susceptible men, or for both. The difference (λ) between the *threshold* in *susceptible women* (λ_w) and that in *susceptible men* (λ_m) is defined as: ($\lambda = \lambda_w - \lambda_m$).

And, therefore:

1. If the *environmental-threshold* in *susceptible women* is greater than that in *susceptible men*
   –that is, if ($\lambda_w > \lambda_m$): then (λ) is positive and ($\lambda_m = 0$)
2. If the *environmental-threshold* in *susceptible men* is greater than that in *susceptible women*
   –that is, if ($\lambda_w < \lambda_m$): then (λ) is negative and ($\lambda_w = 0$)
3. If the *environmental-threshold* in *susceptible women* is the same as that in *susceptible men*
   –that is, if ($\lambda_w = \lambda_m$): then: ($\lambda = \lambda_w = \lambda_m = 0$)

If the hazards are proportional and if: ($H(a)\geq\lambda$)), then the relationship between the cumulative hazard for *susceptible women* and that for *susceptible men* (above) can be generalised (see online supplemental material section 7a) 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 necessary interactions between the two). Therefore, if any member of (G) experiences an environmental 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 *susceptible men* and *susceptible women* in the likelihood of their experiencing a '*sufficient*' exposure, to a difference in the value of the limiting probabilities (c) and (d), or to a difference in both (online supplemental material section 5d). Therefore, by assuming that: ($c=d\leq1$), one also assumes that any difference in the *failure-probability* between *susceptible men* and *susceptible women* is due, exclusively, to a difference in the likelihood of their experiencing a '*sufficient*' environmental-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 different places (see online supplemental material section 6a). In this case, both the proportionality factor (R) and the parameter (λ)—which relates the threshold in *susceptible men* to that in *susceptible women*—are meaningless.

Proportional hazard

An '*apparent*' value of (R), or (R^{app}), can be defined as the value of (R) whenever: ($c=d\leq1$) and, under proportional hazard conditions, with proportionality factor (R)—see online supplemental material section 7c and g—two conditions must hold:

1. if: $R\leq1$; or, if: $R<R^{app}$; or, if: $\lambda\leq0$; then: $c<d$
   Therefore: if: $c=d\leq1$; then, both: $R>1$ and: $\lambda>0$
2. if: $R>1$; then: $\lambda>0$

Condition #1 excludes any possibility that: ($c=d\leq1$) (see figures 1 and 2 and Results).

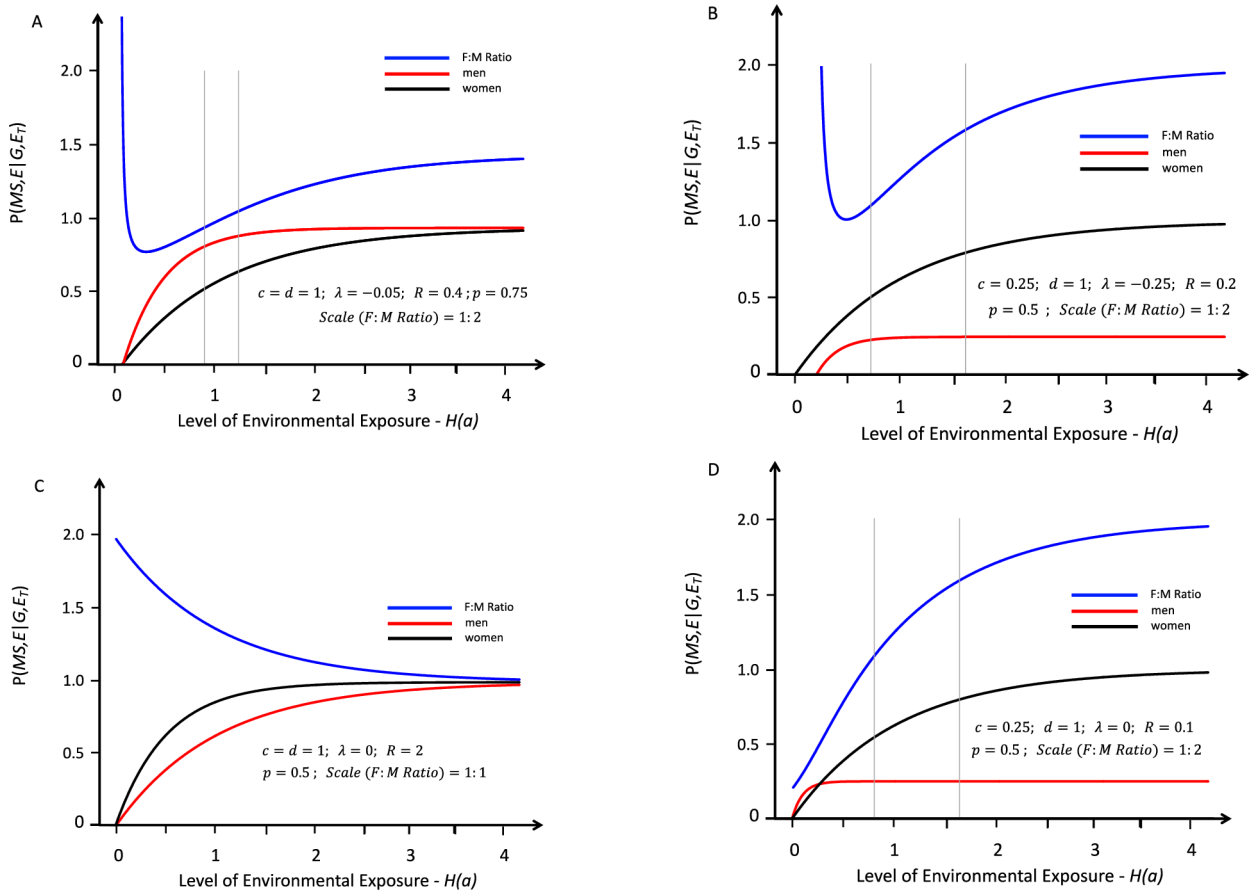

Figure 1 Using the Canadian MS data (online supplemental material 10 a,b), response-curves are depicted for developing MS in *genetically-susceptible women* and men to an increasing probability of sufficient environmental exposure and under conditions, in which the environmental threshold is the same, or greater, in men than it is in women (ie, conditions where: ($\lambda \leq 0$)—see: Longitudinal models; Proportional hazard; & online supplemental material section 1a). Response-curves representing women (black lines) and men (red lines) are depicted separately. 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—that is, under conditions in which: ($\lambda < 0$) (see ethods: Longitudinal models: General considerations. The curves depicted in Panels C and D are 'strictly' proportional, meaning that the environmental threshold is the same for both men and women—that is, under conditions in which: ($\lambda = \lambda_w = \lambda_m = 0$) (see Methods: Longitudinal models: Proportional hazard). The blue lines represent the change in the (*F:M*) sex-ratio with increasing exposure. This ratio is plotted at various scales (indicated in each Panel) so that it can be displayed in the same graph. The thin grey vertical lines represent the narrow portion of the response-curves that covers the change in the (*F:M*) sex-ratio from 2.2 to 3.2 (ie, 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 ($Zw > Zm$) 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 online supplemental material sections 7 c–f). Response curves A and C reflect conditions in which ($\boldsymbol{c} = \boldsymbol{d} = 1$); whereas curves B and D reflect those conditions in which ($\boldsymbol{c} < \boldsymbol{d} = 1$). *F:M*, female-to-male; MS, multiple sclerosis.

Condition #2 (ie, where: $\lambda > 0$), requires that, as the odds of a '*sufficient*' environmental exposure decrease, 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—that is, how can *susceptible women* be less environmentally susceptible than *susceptible men* when the *exposure-probability* is low and, yet, be more environmentally susceptible when the *exposure-probability* is high.

There are two obvious ways to avoid this paradox (see online supplemental material section 7d–h). The first is that the hazards are non-proportional, although this creates other problems. For example, women and men in the same *exposure-group*, necessarily, have proportional hazards (see online 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.

The second is that Condition #1 applies. For example, this condition is compatible with any (λ) 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

Cross-sectional models

Parameter abbreviations (see Methods: Cross-sectional models) are used such that the (*G*) subset consists of all *genetically-susceptible* individuals (see Methods: Genetic susceptibility); the set (*X*) consists of *MS-penetrance* values for all susceptible individuals; the variance of (*X*) is: (σ_X^2) ; *MS-penetrance* for the (*G*) subset is: ($x = P(MS \mid G)$); and

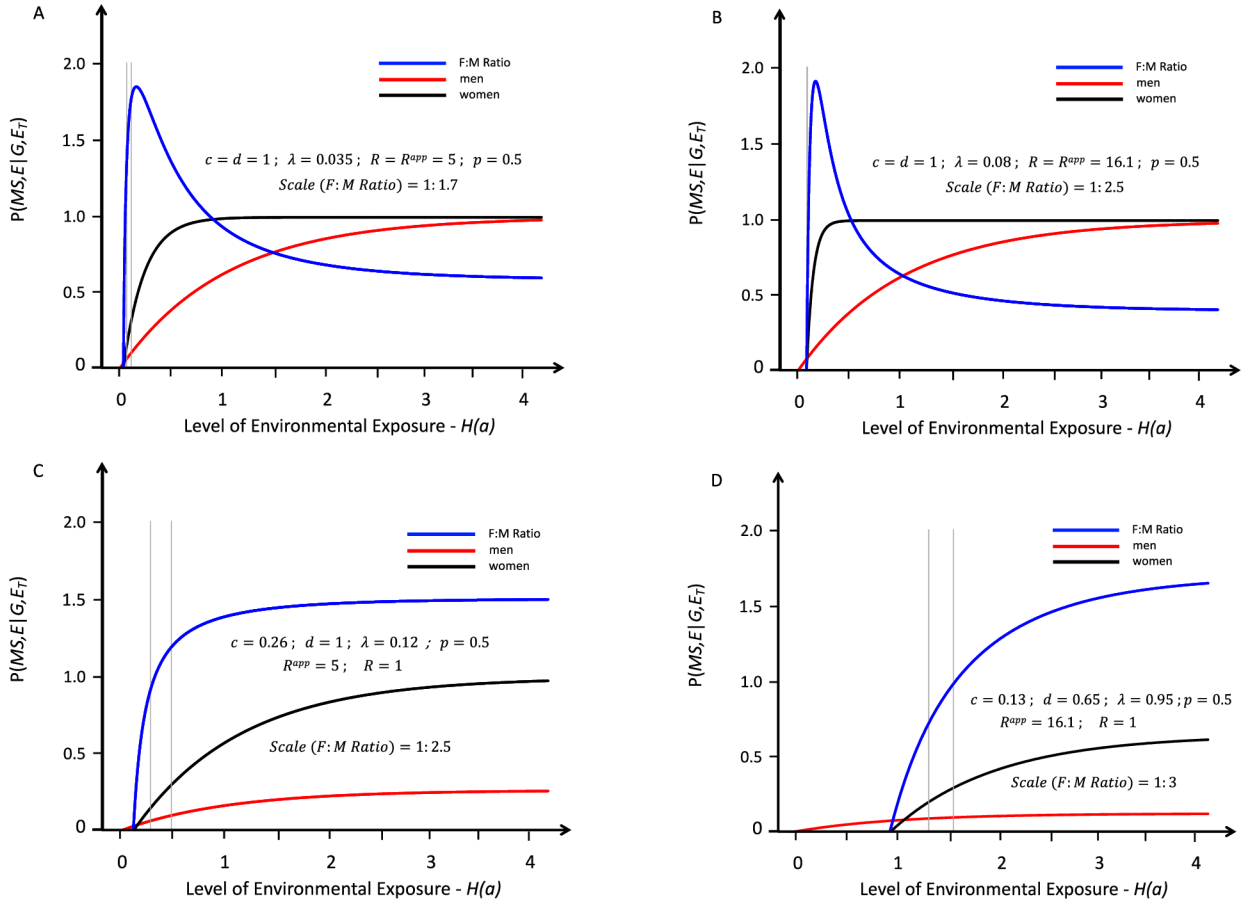

Figure 2 Using the Canadian MS data (online supplemental material section 10a,b), response-curves are depicted for developing MS in *genetically-susceptible women* and men to an increasing probability of sufficient environmental exposure and under conditions, in which the environmental threshold in women is greater than it is in men (ie, conditions where: (λ>0) (see Methods: Longitudinal models; Proportional hazard; & online supplemental material section 1a). Response-curves for women (black lines) and men (red lines) are depicted separately. The curves depicted are proportional, with a proportionality factor (*R*). Also, all of these response curves represent actual solutions. The blue lines represent the change in the (*F:M*) sex-ratio with increasing exposure. This ratio is plotted at various scales (indicated in each Panel) so that it can be displayed in the same graph. 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***≤1), both: ($R=R^{app}$) and the response curves for men and women have the same relationship with each other (see online supplemental material sections 7c–f). By contrast, Panels C and D represent conditions where: (***c***<***d***≤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—that is, 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*. *F:M*, female-to-male; MS, multiple sclerosis.

the '*adjusted*' MZ-twin concordance rate (see Methods: *MZ*-twins, *DZ*-twins, and siblings) is: ($x'=P(MS \mid IG_{MS})$).

For all Cross-sectional Models of the Canadian MS data,[4] the supported range for the probability of being a member of the *genetically-susceptible* subset, *P*(*G*), is:

$0.003 \le P(G) < 0.83$.

From equation 1 (Methods: Estimating the probability of genetic susceptibility in the population), and assuming: ($x \ge x'/2$)—see reference 4—the supported range for *P*(*G*) is:

$0.003 \le P(G) < 0.55$.

Longitudinal models

Parameter abbreviations, again, are used (see Methods: Longitudinal models: General considerations) such that (λ) represents the difference in the *environmental-threshold* between *susceptible women* and that in *susceptible men*; and (*R*) represents the hazard proportionality factor for *susceptible women* compared with *susceptible men*.

For all Longitudinal Models of the Canadian MS data[4]—with either non-proportional or proportional hazards—and, if proportional, with any (*R*)—the supported range for *P*(*G*) is:

$0.001 < P(G) \le 0.52$.

For proportional hazards, whenever: ($\lambda \le 0$)—figure 1—and, thus, when: ($R<1$)—or whenever either: ($R<R^{app}$) or: ($R \le 1$), the condition that: ($c<d$) is established (see Methods: Longitudinal models: Proportional hazard). Considering the alternative that both: ($\lambda>0$) & ($R>1$)—figure 2—it is conceivable that: ($c=d \le 1$). However, in every such circumstance, the conditions required whenever: ($c<d \le 1$) are far less extreme (see figures 5 and S1–S3 in reference[4]; see also Discussion).

Under proportional hazard conditions, when: ($c=d=1$), the supported ranges for the *threshold-difference* between *susceptible women* and *susceptible men* (λ); for the proportionality factor ($R=R^{app}$); and for the probability-ratio of experiencing a '*sufficient*' exposure—that is, ($P(E \mid F, G)$)/($P(E \mid M, G)$)—are:

$0.0005 \leq \lambda \leq 0.13$
$1.3 \leq R = R^{app} \leq 1177$
$1.2 \leq P(E \mid F, G)/P(E \mid M, G) \leq 32.$

Under proportional hazard conditions, when both: (R=1) & (d=1), the supported ranges for (λ) and for the limiting probability of developing MS in *susceptible men* (c) are:

$0.002 < \lambda < 2.4$
$0.002 \leq c \leq 0.786.$

DISCUSSION

There are two principal conclusions derived from this analysis. First, the *MS-penetrance* of the *genetically-susceptible* subset, (G), is greater than that of the population, (Z), and, thus, not everyone in the population is *genetically-susceptible*. Consequently, some members of the population (Z) cannot develop MS regardless of their environmental experiences. And second, at maximum *exposure-levels*, the limiting probability of developing MS in *susceptible men* (c) is less than that for *susceptible women* (d). These two conclusions, stated explicitly, are:

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

Conclusion #1 seems inescapable (see Results). Indeed, given any of the reported *MZ*-twin concordance rates, the notion that the *MS-penetrance* for (G) is the same as that for (Z) is untenable (see table 4 of 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.

However, considering Conclusion #2—ie, that: (c<d)—there are scenarios, in which the condition of: (c=d≤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 (see Methods: Longitudinal models: Proportional hazard; see also online supplemental material section 7h). However, given the genetic and environmental evidence, this possibility, also, seems untenable. For example, all but 1 of the 233 MS-associated loci are autosomal, and the single X-chromosome risk variant is present in both sexes.[6] In this case, any difference between sexes in the genetics of MS is unlikely (see online supplemental material section 7f). In addition, the pattern of the MS association with the different *HLA*-haplotypes is the same for both sexes (see tables 3 & 4 of reference 4). Family studies also suggest a common genetic basis for MS in women and men.[2–5 8 22 27] Thus, both twin and non-twin siblings (male or female) of an *MS-proband* have increased MS risk, regardless of *proband* sex.[5 8 27] Similarly, both sons and daughters of conjugal couples have markedly increased MS risk.[8 27] Also, male and female full-siblings or half-siblings with an *MS-proband* parent (mother or father) have increased MS risk.[2 8 22 27] Each of these observations supports the view that the genetic basis for MS is similar (if not the same) in both sexes.

Moreover, for all non-proportional hazard conditions where: (c=d≤1), the '*current*' condition—that is, where the ratio of: (Zw/Zm) is both greater than one and increasing over time—can only be explained by the fact that, '*currently*', *susceptible women* are more likely to experience a '*sufficient*' environmental-exposure compared with *susceptible men* (see Methods: Longitidinal models: Proportional hazard; see also online supplemental material sections 3a, 5d and 10a). Nevertheless, contrary to this requirement, women do not seem to be more likely than men to experience the various MS-associated environmental events, regardless of whether these events are known or just suspected. In addition, women and men do not seem to require different environmental events. 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 (online supplemental material section 2c); *EBV* infection is equally common and disease associated[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 same 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 must be proportional (Methods: Longitudinal models: Proportional hazard; see also online supplemental material section 7h).

In addition, several lines of evidence indicate that, when the hazards are proportional, the condition of: (c=d≤1) is also unlikely. First, in all circumstances where the proportionality factor (R) is greater than unity—that is, where (R>1)—as it must be whenever: (c=d≤1)—see Methods: Longitudinal models: Proportional hazard—*susceptible women*, compared with *susceptible men*, must be more responsive to the changes in the environmental *exposure-level*, which have taken place over the past 50 years. As discussed in connection with non-proportional hazards (see Discussion, above), there is little current evidence for this. Second, the genetic and environmental observations (described in the Results and Discussion) suggest that: (R^{app}>R≈1), which is impossible whenever: (c=d≤1) (Methods: Longitidinal models: Proportional hazard). Third, as in figure 1, whenever (λ≤0) or whenever (R≤1), the condition that: (c<d) is established (see Methods: Longitudinal models: Proportional hazard; see also online supplemental material section 5d and 7d–g). Fourth, the alternative of: (R>1) & (λ>0) creates a potential paradox (see Methods: Longitudinal models: Proportional hazard). Although there are ways to rationalise this paradox with: (c=d≤1), in every case, the conditions required whenever: (c<d≤1) are far less extreme (see figures 5 and S1–S3 in reference[4]. Finally, the response curves when: (c=d≤1) & (R>1) are steeply ascending and present only a very narrow *exposure-window* to explain the Canadian (*F:M*) sex-ratio data[23] (see figure 2A and B). 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 a 50-year time-span[4 23] (see also online 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 mechanism. This cannot be attributed to other, unidentified, environmental factors (eg, other infections, diseases, nutritional deficiencies, toxic exposures) 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 the '*sufficient*' *exposure-sets* within their $\{E_i\}$ family would include these conditions (see Methods: Environmental susceptibility). 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 on '*observable*' parameter-values (see Methods: Longitidinal models: Proportional hazard). Finally, this cannot be attributed to mild or asymptomatic disease (eg, clinically, or radiographically, isolated syndromes) because this disease-type occurs disproportionately often among women compared with the *current*

(*F:M*) sex-ratio in MS.[4 23] Naturally, invoking '*truly*' random events in MS disease expression requires replication. Nevertheless, any finding that: ($c<d$) indicates that the behaviour of some complex physical systems (eg, organisms) involves '*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 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: ($R\approx1$)—see immediately above—also, support the conclusion that it is this random mechanism of disease pathogenesis, which is primarily responsible for the difference in MS disease expression currently-observed between *susceptible women* and *susceptible men*. Importantly, the fact that a process favours disease development in women over men does not imply that the process must be non-random. For example, when flipping a biased coin compared with a fair coin—if both processes are random—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. It is unclear what characteristics might be implied by these two terms although, perhaps, the general differences in anatomy, physiology and gene or RNA expression, which exist between males and females, create a '*different milieu*' that translates to setting a different bias for each sex. Moreover, these general differences between the sexes are deeply rooted in our evolutionary tree and, presumably, are highly conserved in all animal species that reproduce sexually. Therefore, it seems very likely that these general differences between sexes do not change appreciably from one generation of human beings to the next, so that whatever biases are introduced by them will also be essentially unchanging.

Other authors, modelling immune system function, also invoke random events in MS disease expression (see reference 4 for a review). 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 does not constitute a *test* of whether '*true*' randomness ever occurs. For example, the outcome of a dice roll may be most accurately modelled 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 about the initial conditions of the dice (eg, its orientation and weight) and the direction, location and magnitude of the forces that act on the dice during the roll.[4 28 29]

In 1814, the French polymath and scholar, Pierre-Simon de Laplace, introduced the concept of *causal determinism* based on well-established and strongly confirmed physical laws.[4 29] Following this introduction, deterministic views of nature became increasingly prevalent among scientists and this notion is still current among many (perhaps most) authorities today.[4 29] For example, in 1908, the physicist Henri Poincaré, clearly articulated this *point-of-view*, stating that: 'every phenomenon, however trifling it be, has a cause, and a mind infinitely powerful and infinitely well-informed concerning the laws of nature could have foreseen it from the beginning of the ages. If a being with such a mind existed, we could play no game of chance with him; we should always lose'.[4 29] Similarly, in a 1926 letter to Max Born, Albert Einstein, reflecting on the evolving notions of quantum uncertainty, expressed his belief that '[God] does not play dice'. Nevertheless, to Poincaré's point (above), even if she or he did play dice, likely, the game would not be random. Many contemporary authorities, also, largely agree with such deterministic ideas. For example, the physicist Brian Greene, states that, although 'the quantum equations lay out many possible futures, … they deterministically chisel the likelihood of each in mathematical stone'.[4] The physicist, Stephen Hawking, writes that 'the wave function contains all that one can know of the particle, both its position, and its speed. If you know the wave function at one time, then its values at other times are determined by what is called the Schrödinger equation. Thus, one still has a kind of determinism, but it is not the sort that Laplace envisaged.' Nevertheless, despite agreeing that the quantum equations imply this certain kind of determinism and also envisioning an early universe with minimal entropy, Hawking further argues that the existence of black hole radiations implies that 'the loss of particles and information down black holes [means] that the particles that [come] out [are] random. One [can] calculate probabilities, but one [cannot] make any definite predictions. Thus, the future of the universe is not completely determined by the laws of science.[4]

By contrast, other authorities find it very difficult to rationalise *any* notion that the outcomes of complex biological processes such as evolution by natural selection or immune system function are predetermined, especially considering the fact that each of these processes is so remarkably *adaptive* to contemporary external events.[4 29] Nevertheless, proving that any macroscopic process includes '*truly*' random mechanisms is difficult. This requires an experiment (ie, a test), in which the outcome predicted by determinism differs from that predicted by non-determinism.

The longitudinal MS data from Canada provides an opportunity to apply just such a test. For example, the widely-held deterministic view requires that: ($c=d=1$). By contrast, any observation that either: ($c<d=1$) or: ($c\leq d<1$) indicates that '*true*' randomness must be a component of disease development and undermines the deterministic hypothesis. Thus, 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 their interactions), which are necessary for MS pathogenesis, regardless of whether these factors, events, and interactions are known, suspected, or as yet unrecognised. Nevertheless, in addition to these necessary prerequisites, '*true*' randomness also seems to play a critical role in MS disease pathogenesis. Moreover, both sexes seem to have the same underlying disease. Thus, both sexes seem to have a similar genetic basis and, also, a similar response to the same environmental disease determinants (see Discussion). These observations suggest both that the hazards are proportional (Methods: Longitidinal models: Proportional hazard) and that ($R\approx1$). If correct, this indicates that it is this '*truly*' random mechanism in disease pathogenesis, which is primarily responsible for the currently-observed differences in MS disease expression between *susceptible women* and *susceptible men*.

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 in this project. 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.

Contributors DSG: Conceptualisation; Formal analysis; Methodology; Software; Writing—original draft, review and editing. JP: Critical review of statistical analysis. MJA: Critical review of the manuscript.DSG is the guarantor. The guarantor accepts full responsibility for the finished work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. NA.

ORCID iD

Douglas S Goodin http://orcid.org/0000-0002-5985-0822

REFERENCES

1 Compston A, Confavreux C, Lassman H, *et al*. McAlpine 's Multiple Sclerosis. Elsevier: Churchill Livingstone, 2006:287–346.
2 Goodin DS. The epidemiology of multiple sclerosis: insights to disease pathogenesis. In: Aminoff MJ, Boller F, Swaab DF, eds. *Handbook of Clinical Neurology*. London: Elsevier, 2014: 122. 231–66.
3 Goodin DS, Khankhanian P, Gourraud PA, *et al*. The nature of genetic and environmental susceptibility to multiple sclerosis. *PLoS One* 2021;16:e0246157.
4 Goodin DS, Khankhanian P, Gourraud PA, *et al*. Multiple sclerosis: exploring the limits and implications of genetic and environmental susceptibility. *PLoS One* 2023;18:e0285599.
5 Willer CJ, Dyment DA, Risch NJ, *et al*. 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:eaav7188.
7 Goodin DS, Khankhanian P, Gourraud PA, *et al*. 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, Dyment DA, *et al*. The Canadian collaborative study group. evidence for genetic basis of multiple sclerosis. *Lancet* 1996;347:1728–30.
9 Willer CJ, Dyment DA, Sadovnick AD, *et al*. Timing of birth and risk of multiple sclerosis: population based study. *BMJ* 2005;330:120.
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;267:2815–22.
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;144:2038–46.
13 Kreft KL, Van Nierop GP, Scherbeijn SMJ, *et al*. 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;296:2832–8.
16 Sowah D, Fan X, Dennett L, *et al*. Vitamin D levels and deficiency with different occupations: a systematic review. *BMC Public Health* 2017;17:519.
17 Sadovnick AD, Dyment DA, Ebers GC, *et al*. Evidence for genetic basis of multiple sclerosis. *The Lancet* 1996;347:1728–30.
18 Sadovnick AD, Risch NJ, Ebers GC. Canadian collaborative project on genetic susceptibility to MS, phase 2: rationale and method. *Can J Neurol Sci* 1998;25:216–21.
19 Ebers GC, Yee IML, Sadovnick AD, *et al*. Conjugal multiple sclerosis: population based prevalence and recurrence risks in offspring. *Ann Neurol* 2000;48:927–31.
20 Ebers GC, Sadovnick AD, Dyment DA, *et al*. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. *Lancet* 2004;363:1773–4.
21 Sadovnick AD, Yee IML, Ebers GC, *et al*. Multiple sclerosis and birth order: a longitudinal cohort study. *Lancet Neurol* 2005;4:611–7.
22 Dyment DA, Yee IML, Ebers GC, *et al*. Multiple sclerosis in step siblings: recurrence risk and ascertainment. *J Neurol Neurosurg Psychiatry* 2006;77:258–9.
23 Orton S-M, Herrera BM, Yee IM, *et al*. Sex-ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol* 2006;5:932–6.
24 Canadian Census. 2010. Available: https://www150.statcan.gc.ca/n1/en/pub/89-503-x/2010001/article/11475-eng.pdf?st=WVL9_Ggm [Accessed 15 May 2023].
25 Witte JS, Carlin JB, Hopper JL. Likelihood-based approach to estimating twin concordance for dichotomous traits. *Genet Epidemiol* 1999;16:290–304.
26 Fisher LD, Belle G. Biostatistics: A Methodology for the Health Sciences. New York: John Wiley & Sons, 1993:369–73.
27 Robertson NP, O'Riordan JI, Chataway J, *et al*. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. *Lancet* 1997;349:1587–90.
28 Green B. Until the End of Time. New York, USA: Alfred A Knopf, Penguin Random House, 2020.
29 Layzer D. Why We Are Free: Consciousness, Free Will and Creativity in a Unified Scientific Worldview. Information Publisher, 2021.

Supplemental Material

Mathematical Development of the Susceptibility Models

For Definitions of Model Terms – see Section 9; Tables S1a-b

1. Environmental Susceptibility

1a. Defining Environmental Susceptibility (***E***)

The population (Z) consists of (N) individuals. The "*genetically-susceptible*" subset (G) – i.e., the subset of *everyone* who has *any* non-zero chance of developing MS under *some* environmental circumstances – consists of $(m \leq N)$ individuals $(i = 1,2, \dots, m)$, each having a unique genotype $[G_i]$. Even *MZ*-twins, despite having "*identical*" genotypes (IG), still have subtle genetic differences from one another [4]. For the purpose of this analysis, it is assumed these subtle differences are unrelated to susceptibility [4]. The probability of the event that an individual, randomly selected from (Z), is a member of the (G_i) subset – a subset consisting of a single individual – is: $[P(G_i) = 1/N]$. The *MS-penetrance* for this subset (x_i), during (E_T), is: $[x_i = P(MS \mid G_i, E_T)]$. By the definition of (G), *above*, it *must* be that: $(\forall\, G_i \in (G):\ x_i > 0)$ under *some* environmental conditions.

The family $\{E_i\}$ includes *every* set of environmental exposures, each of which is "*sufficient*", by itself, to *cause* MS to develop in the *i*th susceptible individual (including *any* necessary interactions between genes and environment). Each "*sufficient*" *exposure-set* within the $\{E_i\}$ family must be distinct (in some way) from each other although, otherwise, there can be any degree of overlap between the exposures that comprise these sets. Moreover, the $\{E_i\}$ family can contain an unlimited number of "*sufficient*" *exposure-sets* although, because: $[\forall\, G_i \in (G): x_i > 0]$ under *some* environmental conditions, the family cannot be empty. The event $\{E_i\}$ indicates that, at least, one of these "*sufficient*" *exposure-sets* within the $\{E_i\}$ family occurs. Moreover, it is possible that two or more members of (G) may share the same $\{E_i\}$ family of exposures – although perhaps requiring different "*critical exposure intensities*" [4]. If so, such individuals are said to belong to the same "*i-type*" *exposure-group*.

For the *i*th susceptible individual to develop MS, the events $\{E_i\}$ and (G_i) *must* occur jointly – i.e., the individual (G_i) *must* experience one or more of the $\{E_i\}$ environments. This joint occurrence is reflected by the subset $(\{E_i\}, G_i)$ and the occurrence of $(\{E_i\}, G_i)$ represents the event that an individual, selected randomly from (Z) – the *proband* – is both the *i*th susceptible individual and that they experience an $\{E_i\}$ environment "*sufficient*" to cause MS in them. The probability of this event, given that this person is in (G) and given the environmental conditions of (E_T), is represented as $P(\{E_i\}, G_i \mid G, E_T)$. If the event (G_i) occurs without $\{E_i\}$, then whatever exposure does occur, it is *insufficient*, and the *i*th susceptible individual cannot develop MS.

The event (E) is defined to be the union of the disjoint events, which exhibit the pairing of the (m) susceptible individuals with their "*sufficient*" *exposure-sets*, such that:

$$(E) = (\{E_1\}, G_1) \cup (\{E_2\}, G_2) \cup \dots \cup (\{E_m\}, G_m)$$

in which case: $$P(E \mid G, E_T) = \sum_{i=1}^{m} P(\{E_i\}, G_i \mid G, E_T)$$

or: $$P(E \mid G, E_T) = \sum_{i=1}^{m} P(G_i \mid G, E_T) * P(\{E_i\} \mid G_i, G, E_T)$$

Because genotype is assumed to be independent of the environmental conditions of (E_T):

$$\forall\, G_i \in (G):\ P(G_i \mid G, E_T) = P(G_i \mid G) = \frac{P(G_i)}{P(G)} = \frac{(1/N)}{(m/N)} = 1/m$$

so that: $$P(E \mid G, E_T) = (1/m) * \sum_{i=1}^{m} P(\{E_i\} \mid G_i, G, E_T)$$

In this way, the term $P(E \mid G, E_T)$ represents the probability of the event that an individual, selected randomly from the (G)-subset, and whose relevant-exposure occurs during (E_T), actually experiences an environmental exposure, which is "*sufficient*" to cause MS in them. Furthermore, by definition, the event (E) can only occur in circumstances where the event (G) also occurs. Therefore: $P(E, G) = P(E)$

2. Adjusting the MZ-twin Concordance for the Shared Environment of Twins

2a. Adjustment for the Shared Environment of MZ-twins – $P(MS \mid IG_{MS})$

By definition, anyone with MS *must* belong to (G) and *must* have experienced the event (E). Therefore:

$$P(MS) = P(MS, G) = P(MS, E) = P(MS, E, G)$$

so that: $P(MS \mid MZ_{MS}) = P(MS, E, G \mid MZ_{MS}) = \sum_{i=1}^{m} P(MS, \{E_i\}, G_i \mid MZ_{MS})$

where: $\forall(i)$: $(i = 1,2, \dots, m)$:

$$P(MS, \{E_i\}, G_i \mid MZ_{MS}) = P(MS \mid \{E_i\}, G_i, MZ_{MS}) * P(\{E_i\} \mid G_i, MZ_{MS}) * P(G_i \mid MZ_{MS})$$

In this manner, the probability that the *proband* is a member of the $(MS, \{E_i\}, G_i)$ subset, given the fact that their *co-twin* is a member of the (MZ_{MS}) subset – i.e., $P(MS, \{E_i\}, G_i \mid MZ_{MS})$ – can be deconstructed and re-expressed as the product of three component probabilities – 1) the probability that MS develops in an *MZ-proband* (G_i) who experiences a "*sufficient*" exposure $\{E_i\}$; 2) the probability that this *MZ-proband* experiences an $\{E_i\}$ exposure, which is "*sufficient*" to *cause* MS in them; and 3) the probability that this *MZ-proband* is a member of the (G_i)-subset – where each probability is conditioned on fact that the *proband* has an *MZ co-twin*, who is a member of the (MZ_{MS}) subset within (Z) – *see Main Text*.

For *probands* who are members of (G), but who are otherwise unspecified, the analogous probabilities can be written:

$$P(MS \mid G) = P(MS, E, G \mid G) = \sum_{i=1}^{m} P(MS, \{E_i\}, G_i \mid G)$$

where: $\forall(i)$: $(i = 1,2, \dots, m)$:

$$P(MS, \{E_i\}, G_i \mid G) = P(MS \mid \{E_i\}, G_i) * P(\{E_i\} \mid G_i) * P(G_i \mid G)$$

Therefore, to determine the necessary adjustment, the impact of *MZ*-twins sharing environments needs to be removed while, at the same time, leaving the genetic impact of being *MZ*-twins unchanged. To this end, one can define the term (IG_{MS}) such that:

$$P(MS, \{E_i\}, G_i \mid IG_{MS}) = P(MS \mid \{E_i\}, G_i, IG_{MS}) * P(\{E_i\} \mid G_i, IG_{MS}) * P(G_i \mid IG_{MS})$$

where: $P(\{E_i\} \mid G_i, IG_{MS}) = P(\{E_i\} \mid G_i)$

and: $P(G_i \mid IG_{MS}) = P(G_i \mid MZ_{MS})$

Moreover, the conditioning events $(\{E_i\}, G_i)$ and $(\{E_i\}, G_i, MZ_{MS})$ both represent the same underlying event for the *proband* – i.e., the event that the i^{th} susceptible individual (the *proband*) experiences an environment "*sufficient*" to cause MS in them. In this circumstance, therefore:

$$P(MS \mid \{E_i\}, G_i, MZ_{MS}) = P(MS \mid \{E_i\}, G_i, IG_{MS}) = P(MS \mid \{E_i\}, G_i)$$

Incorporating these equivalences, into the *above* definition of (IG_{MS}), yields:

$$P(MS, G_i \mid IG_{MS}) = P(MS, \{E_i\}, G_i \mid IG_{MS}) = P(MS \mid \{E_i\}, G_i) * P(\{E_i\} \mid G_i) * P(G_i \mid MZ_{MS})$$

or: $$P(MS, G_i \mid IG_{MS}) = P(MS \mid G_i) * P(G_i \mid MZ_{MS}) = (x_i) * P(G_i \mid MZ_{MS})$$

In this manner, the *above* definition for (IG_{MS}) can be re-expressed such that:

$$\forall\, G_i \in (G): \quad \frac{P(MS, G_i \mid IG_{MS})}{P(G_i \mid IG_{MS})} = P(MS \mid G_i, IG_{MS}) = P(MS \mid G_i) = x_i$$

and $\forall\, G_i \in (G)$: $P(G_i \mid IG_{MS}) = P(G_i \mid MZ_{MS})$

And, thus, the appropriate "*adjusted*" probability, $P(MS \mid IG_{MS})$, can be expressed as:

$$P(MS \mid IG_{MS}) = \sum_{i=1}^{m} P(MS, G_i \mid IG_{MS}) = \sum_{i=1}^{m} P(G_i \mid MZ_{MS}) * (x_i)$$

This adjustment, effectively, represents a thought-experiment, in which susceptible *MZ*-twins are separated at conception, and where the *proband* twin is expected to experience the same environmental exposure as would any (G)-subset member, given the environmental conditions of (E_T).

{NB: This definition represents the intended meaning of the "adjusted" proband-wise (or case-wise) recurrence rate [25] for MZ-twins – i.e., $P(MS \mid IG_{MS})$*. The appropriate adjustment can be made such that:*

$$s_a = P(MS \mid DZ_{MS}) / P(MS \mid S_{MS})$$

and: $$P(MS \mid IG_{MS}) = P(MS \mid MZ_{MS}) / s_a$$

as demonstrated in the Supplementary Material of Reference #4.}

2b. Adjustment for the Susceptible Women and Men Considered Together

Assertion: $P(MS \mid IG_{MS}) = 0.136$

Proof: The following point-estimates (*Table 3; Main Text; see also Section 10b; Table S2; below*) from the Canadian twin-study [5] will be used:

$$P(MS \mid MZ_{MS}) = 0.253$$
$$P(MS \mid DZ_{MS}) = 0.054$$
$$P(MS \mid S_{MS}) = 0.029$$

From the *Supplementary Material* (*Reference #4*), one can estimate the *point-value* of $\{P(MS \mid IG_{MS})\}$ as:

$$s_a = P(MS \mid DZ_{MS}) / P(MS \mid S_{MS}) = 0.054/0.029 = 1.86$$

and: $$P(MS \mid IG_{MS}) = P(MS \mid MZ_{MS}) / s_a = 0.253/1.86 = 0.136$$

2c. Adjustments for Susceptible Women and Men Considered Separately

Assertions: $P(MS \mid F, IG_{MS}) = P(MS \mid F, MZ_{MS})/1.95$

$P(MS \mid M, IG_{MS}) = P(MS \mid M, MZ_{MS})/1.63$

Proof: Two parameters $(s_{aw} \geq 1)$ and $(s_{am} \geq 1)$ are defined such that:

$$P(MS \mid F, IG_{MS}) = P(MS \mid F, MZ_{MS})/s_{aw}$$

$$P(MS \mid M, IG_{MS}) = P(MS \mid M, MZ_{MS})/s_{am}$$

From the point-estimates of the Canadian epidemiological data [5,8,17-23] – see *Section 10b; below* – and from *Assertion 4A* (*Section 4a; below*), therefore:

$$P(F \mid MS) = P(F \mid IG_{MS}) = 0.717$$

$$P(F \mid MS, MZ_{MS}) = P(F \mid MS, IG_{MS}) = 0.917$$

$$P(MS \mid F, MZ_{MS}) = 0.340$$

$$P(MS \mid M, MZ_{MS}) = 0.065$$

The term, $P(MS, F \mid IG_{MS})$, can be deconstructed in two different ways:

$$P(MS, F \mid IG_{MS}) = P(F \mid IG_{MS}) * P(MS \mid F, IG_{MS}) = (0.717 * 0.340)/s_{aw}$$

and: $P(MS, F \mid IG_{MS}) = P(MS \mid IG_{MS}) * P(F \mid MS, IG_{MS}) = (0.136 * 0.917)$

Combining these two equations leads to:

$$s_{aw} = (0.717 * 0.340)/(0.136 * 0.917) = 1.95$$

Similarly: $P(MS, M \mid IG_{MS}) = P(M \mid IG_{MS}) * P(MS \mid M, IG_{MS}) = (0.283 * 0.065)/s_{am}$

and: $P(MS, M \mid IG_{MS}) = P(MS \mid IG_{MS}) * P(M \mid MS, IG_{MS}) = (0.136 * 0.083)$

leading to: $s_{am} = (0.283 * 0.065)/(0.136 * 0.083) = 1.63$

Thus, the point estimate for the impact of *MZ-twins* sharing their intrauterine and *some* of their other environments on the likelihood that the *proband* twin is a member of (MS), given the fact that their *co-twin* a member of the (MZ_{MS}), is very similar for both susceptible *women* and *men*.

3. Enrichment of Women among MS-Patients and Concordant MZ- Twins

3a. Enrichment of More Penetrant Genotypes

If the *MS-penetrance* for susceptible *women* exceeds that in susceptible *men* (i.e., $Zw > Zm$) then, from the *Supplementary Material* (*Reference #4*), from the definition of (IG_{MS}) – *see Section 2a; above* – and from *Assertion 4A* (*below*), *women* can be described as being "*enriched*" such that:

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

The terms (G_{i1}) and (G_{i2}) represent the events that *any* pair of *probands*, randomly selected from (G), belong, respectively, to the (G_{i1}) and (G_{i2}) subsets – each subset consisting of a single individual. The probability of each of these events – *see Section 1a; above* – is:

$$P(G_{i1} \mid G) = P(G_{i2} \mid G) = 1/m$$

The *MS-penetrance values* of these two subsets are designated, respectively, as:

$$x_{i1} = P(MS \mid G, G_{i1}) \quad \text{and:} \quad x_{i2} = P(MS \mid G, G_{i2})$$

Moreover, these two subsets can be suitably defined such that: $(x_{i1} \geq x_{i2})$.

For notational simplicity, the following probability terms [including the definition of (IG_{MS}) – *Section 2a* (*above*) *& Table S1*– and from *Assertion 4A; below*] are defined such that:

$$p = P(F \mid G) \;;\; x = P(MS \mid G) \;;\; x' = P(MS \mid IG_{MS}) \;;\; Zw = z_w \;;\; Zm = z_m$$

$$z_w = P(MS \mid F, G) \;;\; z'_w = P(MS \mid F, G, IG_{MS}) \;;\; z_m = P(MS \mid M, G) \;;\; z'_m = P(MS \mid M, G, IG_{MS})$$

$$x'_{i1} = P(MS \mid G_{i1}, IG_{MS}) = x_{i1} \;;\; x'_{i2} = P(MS \mid G_{i2}, IG_{MS}) = x_{i2}$$

Assertion: Almost certainly**:** $Zw = P(MS \mid F, G) > P(MS \mid M, G) = Zm$

Development: With respect to the subsets (G_{i1}) and (G_{i2}), therefore:

$$P(G_{i1} \mid G, MS) = \frac{P(G_{i1}, G, MS)}{P(G, MS)} = \frac{P(G_{i1} \mid G) * P(MS \mid G, G_{i1})}{P(MS \mid G)} = (G_{i1} \mid G) * (x_{i1}/x)$$

and: $$P(G_{i2} \mid G, MS) = \frac{P(G_{i2}, G, MS)}{P(G, MS)} = \frac{P(G_{i2} \mid G) * P(MS \mid G, G_{i2})}{P(MS \mid G)} = (G_{i2} \mid G) * (x_{i2}/x)$$

Therefore: $\forall\, G_{i1} \;\&\; G_{i2} \in (G)$: $P(G_{i1} \mid G, MS) \geq P(G_{i2} \mid G, MS)$

Also: $$P(G_{i1} \mid G, MS, IG_{MS}) = \frac{P(G_{i1}, G, MS, IG_{MS})}{P(G, MS, IG_{MS})} = \frac{P(G_{i1} \mid G, IG_{MS}) * P(MS \mid G, G_{i1}, IG_{MS})}{P(MS \mid G, IG_{MS})}$$

From the definition of (IG_{MS}) – *Section 2a* (*above*) – and *Assertion 4A* (*below*), therefore

$$P(G_{i1} \mid G, MS, IG_{MS}) = P(G_{i1} \mid G, MS) * (x_{i1}/x')$$

and similarly: $$P(G_{i2} \mid G, MS, IG_{MS}) = P(G_{i2} \mid G, MS) * (x_{i2}/x')$$

Thus: $\forall\, G_{i1} \,\&\, G_{i2} \in (G)$: $P(G_{i1} \mid G, MS, IG_{MS}) \geq P(G_{i2} \mid G, MS, IG_{MS})$

Therefore, within the (MS, G) subset, genotypes are "*sorted*" in the sense that the most prevalent genotypes are also the most penetrant for *every pair-wise* comparison. Similarly, within the (MS, G, IG_{MS}) subset, this "*sorting*" is even more extreme for *every pair-wise* comparison and, therefore, there is a continuing "*enrichment*" of more penetrant genotypes such that:

$$\forall\ G_{i1}\ \&\ G_{i2} \in (G):\ \ 1 = \frac{P(G_{i1}\,|\,G)}{P(G_{i2}\,|\,G)} \leq \frac{P(G_{i1}\,|\,G,MS)}{P(G_{i2}\,|\,G,MS)} \leq \frac{P(G_{i1}\,|\,G,MS,IG_{MS})}{P(G_{i2}\,|\,G,MS,IG_{MS})}$$

Moreover, using the terminology of *Section 7h* (*below*) to specify members of the (G) subset, the (mp) members of the $[(F,G) = (G_w)]$ subset are designated such that: $(d = 1,2, \dots, mp)$, each with a unique genotype (G_{dw}), an *MS-penetrance value* of (z_{dw}), and a variance for the set of these *penetrance values* of (σ_w^2). Analogously, the $[m(1-p)]$ members of the $[(M,G) = (G_m)]$ subset can be designated such that: $[d = 1,2, \dots, m(1-p)]$, each with a unique genotype (G_{dm}), an *MS-penetrance* value of (z_{dm}), and a variance for the set of these *penetrance values* of (σ_m^2). In this case:

$$z_w = P(MS\,|\,F,G) = \sum_{dw=1}^{mp} P(MS, G_{dw}\,|\,F,G) = \sum_{dw=1}^{mp} P(G_{dw}\,|\,F,G) * (z_{dw}) = E(z_{dw})$$

and also: $z_m = P(MS\,|\,M,G) = E(z_{dm})$

Similarly: $P(MS\,|\,F,G,IG_{MS}) = \sum_{dw=1}^{mp} P(MS, G_{dw}\,|\,F,G,IG_{MS}) = \sum_{dw=1}^{mp} P(G_{dw}\,|\,F,G,IG_{MS}) * (z_{dw})$

where: $P(G_{dw}\,|\,F,G,IG_{MS}) = P(G_{dw}\,|\,F,G) * (z_{dw})/P(MS\,|\,F,G)$

so that: $z'_w = E[(z_{dw})^2]/z_w = z_w + \sigma_w^2/z_w$

Following the logic of the *Assertion 4B* proof (*below*), therefore: $z_w = (z'_w/2) \pm \sqrt{(z'_w/2)^2 - \sigma_w^2}$

And also: $z'_m = E[(z_{dm})^2]/z_m = z_m + \sigma_m^2/z_m$ so that: $z_m = (z'_m/2) \pm \sqrt{(z'_m/2)^2 - \sigma_m^2}$

Both $P(MS)$ and the (*F:M*) *sex ratio* are currently increasing, both around the world and in Canada [1-4,23] – *see also Table 3* (*Main Text); Sections 8a & 10a-b* (*below*). Therefore, also, *currently*, (Zw) must be increasing at a faster rate than (Zm) – *see Section 7g* (*below*). Moreover, the MS data from Canada [5] – *see Section 10b* (*below*) – indicate that *currently*:

$$P(MS\,|\,F, MZ_{MS})_2 = (5.2) * P(MS\,|\,M, MZ_{MS})_2 \quad \& \quad P(F\,|\,MS)_2 = 0.762$$

Therefore, unless $[P(F\,|\,G) \geq P(F\,|\,MS)_2]$ – or, equivalently, unless: $[p/(1-p)] \geq$ *the current* (*F:M*) *sex ratio – see Equation S5j; below* – and unless susceptible *men* and *women* have markedly different *variance-distributions* for their *MS-penetrance values*, one of which is non-unimodal [2-4], then, *currently*, it ***must*** be that:

$$z_w = Zw = P(MS\,|\,F,G) > P(MS\,|\,M,G) = Zm = z_m$$

Moreover, if susceptible *men* and *women* can both be members of every "*i-type*" *exposure-group* (*see Sections 7g-h; below*), it would be very hard to rationalize such an extreme difference in *variance-distributions*. Consequently, we assume that this relationship pertains during the "*current*" *Time Period*.

{NB: Because the observations regarding (Zw) *and* (Zm)*, presented in the Main Text* (*Table 3*)*, only relate to the "current" Time Period, the circumstances of other Time Periods cannot be determined.}*

4. Cross-sectional Model

4a. Model Development

For notational simplicity, the following probability terms are defined:

$$p = P(F \mid G)\ ;\quad x_i = P(MS \mid G_i);\ x_i'' = P(MS \mid G_i, MZ_{MS})\ ;\ x = P(MS \mid G)\ ;\ \text{and:}\ x' = P(MS \mid IG_{MS})$$

Assertions: ***4A.***

$$\forall\, G_i \in (G):\quad P(G_i, MS \mid MZ) = P(G_i, MS)$$

$$\forall\, G_i \in (G):\quad P(G_i \mid IG_{MS}) = P(G_i \mid MZ_{MS}) = P(G_i \mid MS)$$

$$P(IG_{MS}) = P(MZ_{MS}) = P(MS)$$

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

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

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

4C. $0 \le \sigma_X^2 \le (x'/2)^2$

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

Definitions and Assumptions: The subset (G) is defined (*see Main Text & Section 1a*) and, as noted:

$$\forall\, G_i \in (G):\quad x_i = P(MS \mid G_i)$$

Thus, (x_i) represents the *MS-penetrance* for the *i*th susceptible individual whose exposure occurs during <u>any</u> specific *Time Period* and it is unique to the *i*th individual. The set (X) is defined to include the *penetrance-value* for each of the (m) members of the (G) subset – i.e., $(X) = (x_1, x_2, \ldots, x_m)$ – and its variance is defined to be (σ_X^2). Finally, each of the (k) individuals in the population $(k = 1, 2, \ldots, N)$ has a unique genotype $\{G_k\}$ – including *MZ*-twins who, despite sharing "*identical*" genotypes, still have subtle genetic differences from one another [4].

A random variable (x_G) can be defined to represent any of the $\{x_i\}$ elements within the set (X) and from this, and from *Section 1a* (*above*), the following terms can be defined:

$$P(G) = m/N$$

$$\forall\, G_i \in (G):\ P(G_i \mid G) = 1/m$$

$$E(x_G) = \sum_{i=1}^{m}(x_i) * (1/m) = P(MS \mid G) = x \qquad (\textit{Equation S4a})$$

$$E(x_G^2) = \sum_{i=1}^{m}(x_i^2) * (1/m) = x^2 + \sigma_X^2 \qquad (\textit{Equation S4b})$$

$$x' = P(MS \mid IG_{MS}) = P(MS, G \mid G, IG_{MS}) = \sum_{i=1}^{m} P(MS, G_i \mid G, IG_{MS}) \qquad (\textit{Equation S4c})$$

These *Equations*, and those derived *below*, describe relationships for the subset (G). In a similar manner, analogous relationships can be established and derived for the subsets (F, G) and (M, G) – *see Section 3a; above – see also Supplemental Material; Reference #4.*

Two assumptions are made:

Assumption #1

MZ-twinning is generally thought to be non-hereditary [4]. If so, then every person (i.e., genotype) in the population (Z) has the same chance, *a priori*, of having an *MZ*-twin (i.e., *MZ*-status is independent of genotype). In this circumstance, during any *Time Period*, it will be the case that:

$$\forall\, G_k \in (Z):\ P(MZ \,|\, G_k) = P(MZ)$$

and, thus: $$\forall\, G_i \in (G):\ P(MZ \,|\, G_i) = P(MZ)$$

Even if *MZ*-twinning were thought to be hereditary in some circumstances [4], but where those genetic factors, which relate to *MZ*-twinning, are independent of MS-susceptibility, then the same conclusion would follow. Either this, or the *above* condition, are assumed to pertain.

Assumption #2

The *MS-penetrance* for *any proband MZ*-twin (whose *co-twin* is of unknown status) is assumed to be independent of *MZ*-status. Thus, this *penetrance-value* for any genotype is presumed to be the same regardless of whether that genotype occurs with or without having an *MZ co-twin*. This assumption is equivalent to assuming that experiencing any particular environment together with an *MZ co-twin* has the same impact as experiencing that environment alone. Alternatively, it is presumed that the mere fact of having an *MZ co-twin* does not alter the environment in such a way that the development of MS becomes more or less likely in both the *proband* and the *co-twin*. Specifically, it is assumed, for any *Time Period*, that:

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

Proof of Assertion 4A:

From *Assumption #1*, it follows that:

$$\forall\, G_k \in (Z):\quad P(G_k, MZ) = P(G_k) * P(MZ \,|\, G_k) = P(G_k) * P(MZ)$$

and therefore: $$\forall\, G_k \in (Z):\ P\big(G_k \,|\, MZ\big) = P(G_k, MZ)/P(MZ) = P(G_k)$$

Consequently, also: $$\forall\, G_i \in (G):\ P\big(G_i \,|\, MZ\big) = P(G_i)$$

From this conclusion, from the definitions of (MZ_{MS}) and from *Assumption #2*, it follows that, during any *Time Period*:

$$\forall\, G_i \in (G):\ P(G_i, MZ_{MS}) = P\big(MS, G_i \,\big|\, MZ\big) = P\big(G_i \,\big|\, MZ\big) * P\big(MS \,\big|\, G_i, MZ\big)$$
$$= P(G_i) * P\big(MS \,\big|\, G_i\big) = P(MS, G_i)$$

and: $$P(MZ_{MS}) = P\big(MS \,\big|\, MZ\big) = \sum_{i=1}^{m} P\big(G_i \,\big|\, MZ\big) * P\big(MS \,\big|\, G_i, MZ\big)$$
$$= \sum_{i=1}^{m} P(G_i) * P\big(MS \,\big|\, G_i\big) = P(MS)$$

From the definition of (IG_{MS}) – *see Section 2a* (*above*) – and from these two equivalences, therefore, during any *Time Period*:

$$\forall\, G_i \in (G):\quad P\big(G_i \,\big|\, IG_{MS}\big) = P\big(G_i \,\big|\, MZ_{MS}\big) = \frac{P(G_i, MZ_{MS})}{P(MZ_{MS})} = \frac{P(G_i, MS)}{P(MS)} = P\big(G_i \,\big|\, MS\big)$$

Also, because the subsets (IG) and (MZ) are identical (*see Main Text*), therefore, both:

$$P(IG) = P(MZ) \qquad \text{and:} \qquad P(G_i, IG) = P(G_i, MZ)$$

Consequently, from *above*, it follows that: $P(G_i, IG_{MS}) = P(G_i, MZ_{MS})$ and: $P(IG_{MS}) = P(MZ_{MS})$

Therefore: $P(IG_{MS}) = P(MZ_{MS}) = P(MS)$

Moreover, from the definition of (IG_{MS}) – *see Section 2a; above* – it follows that:

$$P\big(MS \,\big|\, MZ_{MS}\big) = (s_a) * P\big(MS \,\big|\, IG_{MS}\big)$$

Therefore: $$P\big(MS \,\big|\, MZ_{MS}\big) = \sum_{i=1}^{m} (s_a) * P\big(MS, G_i \,\big|\, IG_{MS}\big) = \sum_{i=1}^{m} P\big(G_i \,\big|\, IG_{MS}\big) * [\,(s_a) * (x_i)]$$

and also: $$P\big(MS \,\big|\, MZ_{MS}\big) = \sum_{i=1}^{m} P\big(MS, G_i \,\big|\, MZ_{MS}\big) = \sum_{i=1}^{m} P\big(G_i \,\big|\, MZ_{MS}\big) * (x_i'')$$

Consequently, from *above*: $\forall\, G_i \in (G)$:

$$P\big(G_i \,\big|\, IG_{MS}\big) * [(s_a) * (x_i)] = P\big(G_i \,\big|\, MZ_{MS}\big) * [(s_a) * (x_i)] = P\big(G_i \,\big|\, MZ_{MS}\big) * (x_i'')$$

so that: $$\forall\, G_i \in (G):\ (s_a) * (x_i) = (x_i'') \quad \& \quad (s_a) * P\big(MS, G_i \,\big|\, IG_{MS}\big) = P\big(MS, G_i \,\big|\, MZ_{MS}\big)$$

Therefore: $$P\big(G_i \,\big|\, MS, IG_{MS}\big) = P\big(G_i \,\big|\, MS, MZ_{MS}\big)$$

Using the terminology of *Section 7h* (*below*) to designate the *women* of (G), it follows that each of these (mp) *women* $(d = 1,2, \dots, mp)$, has a unique genotype (G_{dw}) and, therefore:

$$P\big(F \,\big|\, IG_{MS}\big) = \sum_{d=1}^{mp} P\big(G_{dw} \,\big|\, IG_{MS}\big) = \sum_{d=1}^{mp} P\big(G_{dw} \,\big|\, MS\big) = P\big(F \,\big|\, MS\big)$$

and: $$P\big(F \,\big|\, MS, IG_{MS}\big) = \sum_{d=1}^{mp} P\big(G_{dw} \,\big|\, MS, IG_{MS}\big) = \sum_{d=1}^{mp} P\big(G_{dw} \,\big|\, MS, MZ_{MS}\big) = P\big(F \,\big|\, MS, MZ_{MS}\big)$$

similarly: $$P\big(M \,\big|\, IG_{MS}\big) = P\big(M \,\big|\, MS\big) \qquad \text{and:} \qquad P\big(M \,\big|\, MS, IG_{MS}\big) = P\big(M \,\big|\, MS, MZ_{MS}\big)$$

Proof of Assertion 4B:

From the definitions of (G) & (IG_{MS}) – *see Main Text* ; *Section 2a* (*above*) *& Table S1* – it follows that:

$$P(MS \mid G_i, IG_{MS}) = P(MS \mid G_i, G, IG_{MS}) = x_i' = P(MS \mid G_i, G) = P(MS \mid G_i) = x_i$$

Therefore, during any *Time Period*, the probability $P(MS, G_i \mid G, IG_{MS})$ can be re-expressed as:

1. $$P(MS, G_i \mid G, IG_{MS}) = P(G_i \mid G, IG_{MS}) * P(MS \mid G_i, G, IG_{MS})$$
$$= P(G_i \mid G, IG_{MS}) * (x_i)$$

From *Assertion 4A* and from the definitions of (G) & (IG_{MS}) – *see Main Text & Sections 1a & 2a* (*above*) – the term $P(G_i \mid G, IG_{MS})$ can be re-expressed as:

2. $$P(G_i \mid G, IG_{MS}) = P(G_i \mid G, MS) = P(G_i, G, MS)/P(MS, G)$$
$$= P(MS \mid G_i, G) * P(G_i, G)/P(MS, G)$$
$$= (x_i) * P(G_i \mid G)/P(MS \mid G) = (x_i) * (1/m)/x$$

Combining *1 & 2 (above)* yields:

$$P(MS, G_i \mid G, IG_{MS}) = (x_i)^2 * (1/m)/x$$

However, from *Equations S4b-c*, it is the case that:

$$x' = P(MS, G \mid G, IG_{MS}) = \sum_{i=1}^{m} P(MS, G_i \mid G, IG_{MS})$$

where: $$\sum_{i=1}^{m} P(MS, G_i \mid G, IG_{MS}) = \sum_{i=1}^{m} (x_i^2) * (1/m)/x = E(x_G^2)/x$$

Therefore, from *Equation S4b*, it follows that:

$$x' = (x^2 + \sigma_X^2)/x = x + \sigma_X^2/x \qquad \text{(}\textit{Equation S4d}\text{)}$$

Rearrangement of *Equation S4d*, yields a standard-form quadratic *Equation* in (x) such that:

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

which, in turn, can be solved to yield:

$$x = (x'/2) \pm \sqrt{(x'/2)^2 - \sigma_X^2} \qquad \text{(}\textit{Equation S4e}\text{)}$$

Proof of Assertion 4C:

Equation S4e has real solutions only for the range of:

$$0 \leq \sigma_X^2 \leq (x'/2)^2 \qquad \text{(Equation S4f)}$$

Notably, the maximum variance (σ^2) for *any* distribution [*Reference: see footnote #1; below*] on the closed interval $[a, b]$ is:

$$\sigma^2 = [(b-a)/2]^2$$

Consequently, regardless of any *Assumptions* (*see above*), the *variance-range* indicated by *Equation S4f* represents the maximum possible *variance-range* for <u>any</u> distribution on the closed interval of: $[0, x']$.

Also, rearrangement of *Equation S4d* yields:

$$\sigma_X^2 = x(x'-x) \qquad \text{(Equation S4g)}$$

4b. Quadratic Equations for Penetrance in Succeptible Women and Men

For notational simplicity, the following probability terms are defined:

$x = P(MS \mid G)$; $\quad x' = P(MS \mid G, IG_{MS}) = P(MS \mid IG_{MS})$;

$Zw = z_w = P(MS \mid F, G)$; $\quad z'_w = P(MS \mid F, G, IG_{MS}) = P(MS \mid F, IG_{MS})$;

$Zm = z_m = P(MS \mid M, G)$; $\quad z'_m = P(MS \mid M, G, IG_{MS}) = P(MS \mid M, IG_{MS})$;

$p = P(F \mid G)$; and the two ratios: $\quad r = z'_w/z_w \quad$ and: $\quad s = z'_m/z_m$

Assertions: 1. $Zw = z_w = \frac{x+\sqrt{x^2-\{1+(r/s)(1-p)/p)\}\{x^2-xx'(1-p)/s\}}}{p+(r/s)(1-p)}$

2. $Zm = z_m = \frac{x-\sqrt{x^2-\{1+(s/r)(p/(1-p)\}\{x^2-xx'\,p/r\}}}{(1-p)+(s/r)p}$

Proof: $P(MS \mid G) = P(MS, F \mid G) + P(MS, M \mid G)$

$= P(F \mid G) * (P(MS \mid F, G) + P(M \mid G) * (P(MS \mid M, G)$

or: $x = p(z_w) + (1-p)(z_m)$

with re-arrangement, this becomes:

$$z_m = [x - p(z_w)]/(1-p) \qquad \text{(Equation S4h)}$$

Also: $x' = P(MS \mid G, IG_{MS}) = P(MS, F \mid G, IG_{MS}) + P(MS, M \mid G, IG_{MS})$

#1 Jacobson HI. The maximum variance of restricted unimodal distributions. Ann Math Stat. 1969;40:1746–52.

Therefore, from *Assertion 4A* (*above*):

$$P(MS, F \mid G, IG_{MS}) = P(F \mid G, IG_{MS}) * (z'_w) = P(F \mid G, MS) * (z'_w)$$

and: $P(MS, M \mid G, IG_{MS}) = P(M \mid G, IG_{MS}) * (z'_m) = P(M \mid G, MS) * (z'_m)$

where: $P(F \mid G, MS) = P(F, MS \mid G)/P(MS \mid G) = p(z_w)/x$

and, similarly: $P(M \mid G, MS) = (1-p)(z_m)/x$

so that: $xx' = p(z_w)(z'_w) + (1-p)(z_m)(z'_m) = pr * (z_w)^2 + (1-p)s * (z_m)^2$

or: $(z_m)^2 = [xx' - pr(z_w)^2]/[(1-p)s]$ (*Equation S4i*)

Therefore, there are two simultaneous *Equations* for $(z_m)^2$ – i.e., *Equations S4h and S4i, above*. Using these two estimates to eliminate the (z_m) parameter, yields:

$$[\{x - p(z_w)\}/(1-p)]^2 = (z_m)^2 = [xx' - pr(z_w)^2]/[(1-p)s]$$

or: $\{x - p(z_w)\}^2 = \{xx' - pr(z_w)^2\}(1-p)/s = \{xx'(1-p)/s\} - (r/s)p(1-p)(z_w)^2$

and: $x^2 - 2xp(z_w) + p^2(z_w)^2 - xx'(1-p)/s + (r/s)p(1-p)(z_w)^2 = 0$

This last *Equation* can be rearranged to yield a standard-form quadratic *Equation* in (z_w) such that:

$$\{p^2 + (r/s)p(1-p)\}(z_w)^2 - \{2xp\}(z_w) + \{x^2 - xx'(1-p)/s\} = 0 \quad \text{(Equation S4j)}$$

Because: $(z'_w \gg z'_m)$ and because both $P(MS)$ and the (*F:M*) *sex ratio* are "*currently*" known to be increasing [3,4,23] – *see also Sections 8a & 10a-b* (*below*) – it is assumed, during the *current Time Period,* that: $(z_w > z_m)$ – *see Section 3a* (*above*). Therefore, *Equation S4j* is solved for (z_w) as:

$$Zw = z_w = \frac{x + \sqrt{x^2 - \{1 + (r/s)(1-p)/p)\}\{x^2 - xx'(1-p)/s\}}}{p + (r/s)(1-p)} \quad \text{(Equation S4k)}$$

Equation S4h (*above*) can then be solved for (z_m). Alternatively, the *above* arguments can be reframed to eliminate (z_w) instead of (z_m), and the resulting quadratic *Equation* can be solved for (z_m) as:

$$Zm = z_m = \frac{x - \sqrt{x^2 - \{1 + (s/r)(p/(1-p))\}\{x^2 - xx'\,p/r\}}}{(1-p) + (s/r)p} \quad \text{(Equation S4l)}$$

***5. Longitudinal Model**:*

5a. Model Development

Following standard survival analysis methods [26], the cumulative survival $\{S(u)\}$ and failure $\{F(u)\}$ functions where: $F(u) = 1 - S(u)$ can be defined separately for susceptible *men* $\{S_m(u)\ and\ F_m(u)\}$ and for susceptible *women* $\{S_w(u)\ and\ F_w(u)\}$. Also, the (unknown and unspecified) hazard functions for developing MS at different environmental *exposure-levels* (*u*) – i.e., $h(u)$ and $k(u)$ – can be defined for susceptible *men* and susceptible *women*, respectively. These hazard functions for susceptible *women* and *men* may be proportional to each other and, if they are proportional, a hazard proportionality factor $(R > 0)$ can then be defined such that: $[k(u) = R * h(u)]$. Furthermore, from *Section 1a* (*above*), the term, $P(E \mid G, E_T)$, represents the probability of the event that a *proband*, randomly selected from (G), and who has their relevant exposures during (E_T), experiences an environmental exposure "*sufficient*" to *cause* MS in them. The *exposure-level* (*u*) is then defined as the odds that this event occurs such that:

$$u = \frac{P(E \mid G, E_T)}{[1 - P(E \mid G, E_T)]}$$

The cumulative hazard function (for susceptible *men*), $H(a)$, is defined as the definite integral of the hazard function, $h(u)$, from an *exposure-level* of $(u = 0)$ to an *exposure-level* of $(u = a)$ such that:

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

Similarly, the cumulative hazard function (for susceptible *women*), $K(a)$, is defined as the definite integral of the hazard function, $k(u)$, from an *exposure-level* of $(u = 0)$ to an *exposure-level* of $(u = a)$ such that:

$$K(a) = \int_0^a k(u)du$$

If these hazards are proportional, then:

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

For susceptible *men*, using the common definition of the hazard function [26] that:

$$h(u) = f_m(u)/S_m(u)$$

together with the fact that, by definition:

$$f_m(u) = d[F_m(u)]/du = -d[S_m(u)]/du$$

a standard derivation from survival analysis methods [26] demonstrates that, for susceptible *men*, because:

$$h(u)du = -d[S_m(u)]/S_m(u)$$

Therefore, the cumulative hazard function $[H(a)]$ can be re-expressed such that:

$$H(a) = -\int_0^a d[S_m(u)]/S_m(u) = \ln[S_m(0)] - \ln[S_m(a)]$$

where: $H(0) = \ln[S_m(0)] - \ln[S_m(0)] = 0$

Exposure is here being measured as the *odds*, during (E_T), that a (G)-subset member experiences an environmental exposure "*sufficient*" to *cause* MS in them. By definition, when: $[P(E \mid G, E_T) = 0]$, no member of (G) can develop MS [i.e., $S_m(0) = 1$], in which case: $\{\ln[S_m(0)] = \ln(1) = 0\}$. Thus:

$$S_m(a) = e^{-H(a)}$$

This standard derivation from survival methods [26], therefore, demonstrates that the survival function is exponentially related to the integral of the underlying hazard function – i.e., the cumulative hazard function. Consequently, the failure function for susceptible *men* can be stated such that:

$$F_m(a) = 1 - S_m(a) = 1 - e^{-H(a)}$$

5b. Environmental Exposure Levels during Different Time Periods

{NB: In this and the Sections that follow, observations made during the two Time Periods are distinguished by the use of subscripts (1) and (2). For example, $P(MS)_1$ refers to $P(MS)$ during the 1st Time Period whereas $P(MS)_2$ refers to $P(MS)$ during the 2nd Time Period. Also, it is important to note that cumulative hazard is being used as a measure of exposure, not failure – see Main Text & Reference #4.}

The environmental *exposure-level* for susceptible *men* during the *1st Time Period* is defined as $[H(a_1)]$. In turn, the *failure-probability* for a susceptible *man* is defined as: $[F_m(a) = Zm]$, which represents the life-time probability of the event that a susceptible *man*, randomly selected from (M, G), and who has their relevant exposures during (E_T), develops MS. Moreover, if the constant $(\boldsymbol{c})$ is defined as the maximum possible *failure-probability* for susceptible *men*, then:

$$F_m(a) = Zm = P(MS \mid M, G, E_T) = P(MS, E \mid M, G, E_T)$$

and: $$\boldsymbol{c} = \lim_{a \to \infty}(Zm) = P(MS \mid M, G, E) \leq 1$$

In this circumstance, this *failure-probability* during the *1st Time Period* (Zm_1), can be stated as:

$$F_m(a_1) = Zm_1 = P(MS, E \mid M, G)_1 = \boldsymbol{c} * [1 - e^{-H(a_1)}] \qquad (\textit{Equation S5a})$$

If the *exposure-level* for susceptible *men* during the *2nd Time Period* is defined as $[H(a_2)]$, then, because (Zm) is *currently* increasing with time [3,4] – *see also Section 8a* (*below*) – the difference in the *exposure-level* for *men* between the *1st and 2nd Time Periods* can be represented by the parameter (q_m) such that:

$$H(a_2) - H(a_1) = q_m > 0$$

In this case, the *failure-probability* during the *2nd Time Period* (Zm_2), can be stated as:

$$F_m(a_2) = Zm_2 = P(MS, E \mid M, G)_2 = \boldsymbol{c} * [1 - e^{-\{H(a_1)+q_m\}}] \qquad (\textit{Equation S5b})$$

Equations S5a & S5b can be rearranged to yield:

$$1 - Zm_1/\boldsymbol{c} = e^{-H(a_1)}$$

and: $$1 - Zm_2/\boldsymbol{c} = e^{-\{H(a_1)+q_m\}} \qquad (\textit{Equation S5c})$$

Dividing the 1st of these two *Equations* by the 2nd yields:

$$(1 - Zm_1/\boldsymbol{c})/(1 - Zm_2/\boldsymbol{c}) = e^{q_m} \qquad (\textit{Equation S5d})$$

or: $$q_m = \ln(1 - Zm_1/\boldsymbol{c}) - \ln(1 - Zm_2/\boldsymbol{c}) \qquad (\textit{Equation S5e})$$

This unit (q_m) is arbitrary but, nonetheless, depends upon the actual (but unknown) change in the environmental *exposure-level*, which has taken place between the two *Time Periods*. From *Equations S5d–e*, the estimated magnitude of this *exposure-level* change depends upon the value of $(\boldsymbol{c})$, which can range over the interval of: $(1 \geq \boldsymbol{c} > Zm_2)$. The ratio on the *LHS* of *Equation S5d* (*above*) is always greater than unity because (Zm) increases with increasing exposure. Moreover, it increases monotonically as $(\boldsymbol{c})$ varies throughout its range – being at a minimum when: $(\boldsymbol{c} = 1)$ and approaching infinity as: $(\boldsymbol{c} \to Zm_2)$.

Consequently, the term (q_m^{min}) is defined to be the *"minimum" exposure-level* change that is possible for susceptible *men* between these two *Time Periods*. In this case, this *minimum exposure-level* change will occur when:

$$\boldsymbol{c} = P(MS \mid M, E, G) = 1$$

Therefore, from *Equation S5e*: $$q_m^{min} = \ln(1 - Zm_1) - \ln(1 - Zm_2)$$

Nevertheless, this *minimum exposure-level* change (q_m^{min}) may not accurately reflect the actual (but unknown) change in the *exposure-level*, which has taken place between the two *Time Periods*. Therefore, the term (q_m) is called the *"actual" exposure-level* change for susceptible *men*. This may well be different from the "*minimum*" possible *exposure-level* change so that:

$$q_m \geq q_m^{min}$$

In a directly analogous manner, the term $[F_w(a) = Zw]$ is defined to be the *failure-probability* for susceptible *women* during any *Time Period* and the constant $(\boldsymbol{d})$ is defined to be the maximum possible *failure-probability* for susceptible *women* such that:

$$F_w(a) = Zw = P(MS \mid F, G, E_T) = P(MS, E \mid F, G, E_T)$$

and: $$\boldsymbol{d} = \lim_{a \to \infty}(Zw) = P(MS \mid M, F, E) \leq 1$$

Similar to *Equations S5a-b* (*above*), because (Zw) is also increasing with time [3,4], the *failure-probability* in susceptible *women* during the *1st& 2nd Time Periods*, $(Zw_1 \text{ and } Zw_2)$, can be stated as:

$$F_w(a_1) = Zw_1 = P(MS, E \mid F, G)_1 = \boldsymbol{d} * [1 - e^{-K(a_1)}] \qquad (\textit{Equation S5f})$$

and: $$F_w(a_2) = Zw_2 = P(MS, E \mid F, G)_2 = \boldsymbol{d} * [1 - e^{-\{K(a_1) + q_w\}}] \qquad (\textit{Equation S5g})$$

where $\{K(a_1)\}$ indicates the *exposure-level* in *women* during the *1st Time Period* and the term (q_w) is called the *"actual" exposure-level* change for *women* that has occurred between the two *Time Periods*. Therefore:

$$K(a_2) - K(a_1) = q_w > 0$$

Also, in a directly analogous manner to the derivation of *Equation S5e* (*above*):

$$q_w = \ln(1 - Zw_1/\boldsymbol{d}) - \ln(1 - Zw_2/\boldsymbol{d}) \qquad (\textit{Equation S5h})$$

Therefore, similar to those circumstances in susceptible *men*, the *"minimum"* possible value (q_w^{min}) for the *exposure-level* change in susceptible *women* will occur when: $(\boldsymbol{d} = 1)$, so that:

$$q_w^{min} = \ln(1 - Zw_1) - \ln(1 - Zw_2)$$

and: $$q_w \geq q_w^{min}$$

5c. Relationship between Failure to True Survival

In true survival everyone dies if given a sufficient amount of time. By contrast, as the *exposure-probability*, $P(E \mid G, E_T)$, approaches unity, the probability of failure (i.e., developing MS), either for *susceptible-men* (Zm) or for *susceptible-women* (Zw), may not similarly approach 100%. Moreover, the *maximum* possible value for this *failure-probability* for susceptible *men* $(\boldsymbol{c})$ might not be the same as the *maximum* possible *failure-probability* for susceptible *women* $(\boldsymbol{d})$. Although the values of the $(\boldsymbol{c})$ and $(\boldsymbol{d})$ parameters are unknown, they are constants whenever the pathogenesis of disease involves environmental events, and regardless of whether the hazards are proportional. Finally, because exposure is being measured as the *odds* that the *proband* experiences a "*sufficient*" environment, the "*threshold*" exposure (i.e., the *exposure-level* at which MS becomes possible) must occur at: $P(E \mid G, E_T) = 0$; for susceptible *men*, or for susceptible *women*, or for both, provided that this *exposure-level* is possible [3]. If the hazards are proportional, the *threshold- difference* (λ) is defined to be the difference between the threshold in susceptible *women* (λ_w) and that in susceptible *men* (λ_m) – i.e., $(\lambda = \lambda_w - \lambda_m)$. Consequently, if the threshold in susceptible *men* is greater than that in *women*, (λ) will be negative and $(\lambda_w = 0)$; if the threshold in *women* is greater than that in *men*, (λ) will be positive and $(\lambda_m = 0)$; and if the threshold in *women* and *men* is the same, then: $(\lambda = \lambda_w = \lambda_m = 0)$.

Also, in true survival, both the clock and the risk of death begin at time-zero and continue into the future indefinitely. As a consequence, the cumulative probability of death increases monotonically with time. By contrast, for MS, it may be that the prevailing environmental conditions, during some *Time Period* (E_T),

are such that: $P(E \mid G, E_T) = 0$; even for a very extended *Time Period* (e.g., for centuries or millennia). Moreover, unlike the cumulative probability of death, for MS, the *exposure-level* may vary in any direction with time, depending upon the specific environmental conditions during (E_T). Therefore, despite the cumulative probability of failure (i.e., of developing MS) increasing monotonically with increasing *exposure-level*, it may decrease, increase, or stay constant with time.

5d. Relationship of the (F:M) Sex Ratio to Exposure

Regardless of (λ), and regardless of whether the hazards are proportional, the *failure-probability* during any *Time Period* for susceptible *women* (Zw) can be stated as:

$$Zw = P(MS, E \mid G, F, E_T) = P(E \mid G, F, E_T) * P(MS \mid E, G, F)$$

or: $$Zw = P(E \mid G, F, E_T) * \boldsymbol{d}$$

and, similarly, the *failure-probability* for susceptible *men* (Zm) can be stated as:

$$Zm = P(MS, E \mid G, M, E_T) = P(E \mid G, M, E_T) * \boldsymbol{c}$$

Dividing the *1st* of these two *Equations* by the *2nd*, during any *Time Period*, yields:

$$Zw/Zm = [P(E \mid G, F, E_T)/P(E \mid G, M, E_T)] * [\boldsymbol{d}/\boldsymbol{c}] \qquad (\textit{Equation S5i})$$

Consequently, during any *Time Period*, any disparity observed between (Zw) and (Zm), must be due to a difference between *men* and *women* in the likelihood of their experiencing a "*sufficient*" environmental exposure, to a difference in the values of constants $(\boldsymbol{c})$ and $(\boldsymbol{d})$, or to a difference in both.

Therefore, by assuming that: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$, one is also assuming that any difference observed in disease expression between susceptible *women* and *men* is due entirely to a difference between susceptible *men* and *women* in the likelihood of their experiencing a "*sufficient*" exposure, despite the fact that, for every (i), the exposure $\{E_i\}$ is both fixed and *population-wide* during any (E_T). Thus, this exposure is "*available*" to everyone, so that, if the "*sufficient*" *exposure-level* differs between sexes, one possible explanation might be a systematic behavioral difference between susceptible *women* and *men* – i.e., to an increased exposure to, or avoidance of, susceptible environments by one or the other sex (perhaps consciously or unconsciously; or perhaps as a result of differing recreational activities, differing occupations, differing gender-roles, etc.). Nevertheless, the fact that *most men* behave differently from *women* does not indicate that *all men* do so, which makes a difference in *threshold* difficult to rationalize. Notably, also, if a finding of $(\lambda \neq 0)$ were to be explained by a systematic behavioral difference, then the finding of $(\lambda > 0)$ would suggest that the behavior of *men* leads to a greater exposure than the behavior of *women*. Any general conclusion in this regard, however, cannot be easily rationalized with the *current* observation that: $(Zw_2 > Zm_2)$. – *see Section 3a* (*above*)*; see also Supplemental Material; Reference #4*.

Another possible explanation for $(\lambda > 0)$, is that there may be distributions of so-called "*critical exposure intensity*" levels (i.e., "*thresholds*") that differ between susceptible *men* and *women* who are members

of the same "*i-type*" *exposure-group* (*see Supplemental Material; Reference #4*). In such a case, perhaps, despite the fact that the same "*exposure-level*" is experienced equally by the two sexes, the "*intensity*" of this exposure might be disproportionately "*sufficient*" for susceptible *women* or susceptible *men* [4].

Membership in (G) is assumed to be independent of (E_T). In this case, the proportion of *women* among susceptible individuals $[p = P(F \mid G)]$ is also independent of (E_T). Following the logic and notation leading to *Equation S4h* (*Section 4b; above*), therefore, regardless of whether the hazards are proportional, for *any* solution, the observed (*F:M*) *sex ratio* during *any Time Period* is proportional to the observed (Zw/Zm) ratio. Thus:

$$(F:M)\ sex\ ratio = \frac{P(MS,F \mid E_T)}{P(MS,M \mid E_T)} = \left(\frac{Zw}{Zm}\right) * \left(\frac{p}{1-p}\right) \qquad (Equation\ S5j)$$

5e. Response-Curves to Increasing Exposure

From *Section 5a* (*above*) the response curves for both susceptible *men* and *women* are exponential. Importantly, <u>*any*</u> two points on <u>*any*</u> exponential curve completely defines the entire response curve. Thus, the values of Zw, Zm, $P(MS)$, and the (*F:M*) *sex ratio*, during *any* two *Time Periods,* completely defines these response curves for both susceptible *men* and susceptible *women* – *see Equations S5a & S5b* and *S5f & S5g* (*above*). Moreover, if these response curves for both sexes can be plotted on the same *x-axis* (i.e., if both sexes are responding to the same environmental events), the hazards are always proportional (*see Section 7h; below*). Also, in this circumstance, the values of $(R = q_w/q_m)$ and (λ) are determined from *Equations S7f-g* (*below*).

6. Non-proportional Hazard Models

6a. General Considerations

If the hazard functions for susceptible *men* and *women* are not proportional, the "*actual*" *exposure-level* changes for susceptible *men* and *women* could each be at their *"minimums"* – i.e., (q_m^{min}) and (q_w^{min}). Such a circumstance, however, occurs when, and only when: $(\boldsymbol{c} = \boldsymbol{d} = 1)$ – *see Section 5b* (*above*).

Also, in this circumstance, although the "*plausible*" *parameter-value-ranges* for both observed and non-observed epidemiological parameters (*see Table 3; Main Text*) still limit possible solutions and, although $(\boldsymbol{c} \leq 1)$ and $(\boldsymbol{d} \leq 1)$ will be constants, nothing about them or about their relationship to each other can be inferred from the changes that take place in the (*F:M*) *sex ratio* and $P(MS)$ over time. Thus, any differences in the *values* that these *parameters* take during different *Time-Periods* could be attributed, both potentially and plausibly, to the differing environmental circumstances of different times and different places. In this circumstance, both the hazard proportionality factor (R) and the parameter (λ) – which relates the threshold in susceptible *women* to that in susceptible *men* – are meaningless.

Nevertheless, during *any Time Period*, the ratio of (Zw/Zm) will still be proportional to the observed (*F:M*) *sex ratio* (*see Equation S5j*) and, if: $\boldsymbol{c} = \boldsymbol{d} \leq 1$, then any observed difference between (Zw) and (Zm), <u>*must*</u> be the result of a difference between susceptible *women* and susceptible *men* in the likelihood that they have experienced a "*sufficient*" environmental exposure during that *Time Period* (*see Equation S5i*).

7. Proportional Hazard Models

7a. General Considerations

If the hazards for susceptible *women* and *men* are proportional with the proportionality factor (R), the situation is altered. First, because $(R > 0)$, the *penetrance-values* of $P(MS \mid F, G)$ and $P(MS \mid M, G)$, if they change over time, ***must*** have the same directionality. Indeed, the epidemiological observation that *MS-prevalence* has been increasing for both *women* and *men* over the past several decades, accords with this requirement [3,4,23]. Second, the proportional hazard *Model* (*see Section 5a; above*), including the possibility of a difference in the "*threshold*" *exposure-level* between the sexes, can be generalized such that:

$$\forall\, H(a) \geq \lambda: \quad K(a) = R * \{H(a) - \lambda\} \geq 0 \qquad \textit{(Equation S7a)}$$

In this circumstance, *Equations S5f & S5g*, which represent the *failure-probabilities* during the *1st & 2nd Time Periods* for susceptible *women*, can be re-stated as:

$$Zw_1 = \boldsymbol{d} * \left[1 - e^{-K(a_1)}\right] = \boldsymbol{d} * [1 - e^{-R*\{(H(a_1)-\lambda\}}] \qquad \textit{(Equation S7b)}$$

and: $$Zw_2 = \boldsymbol{d} * \left[1 - e^{-K(a_2)}\right] = \boldsymbol{d} * [1 - e^{-R*\{H(a_1)+q_m-\lambda\}}] \qquad \textit{(Equation S7c)}$$

Equations S5a & S7b can be rearranged for ***any*** *Time Period* to yield:

$$1 - Zw/\boldsymbol{d} = e^{-K(a)} = e^{-R*\{H(a)-\lambda\}} \qquad \textit{(Equation S7d)}$$

and: $$1 - Zm/\boldsymbol{c} = e^{-H(a)} \qquad \textit{(Equation S7e)}$$

Dividing *Equation S7d* by *S7e*, this result can be rearranged to yield:

$$\lambda = \{\ln [1 - Zw/\boldsymbol{d}] - \ln [1 - Zm/\boldsymbol{c}]\}/R + [(R-1)/R] * H(a) \qquad \textit{(Equation S7f)}$$

Then *Equation S7f* can be applied to the *exposure-levels* $H(a_1)$ and $H(a_2)$ and one can subtract the 2nd of the resulting two *Equations* from the 1st. Then, applying *Equations S5e & S5h*, together with the defining *Equations* for (q_m) and (q_w) from *Section 5b* (*above*), this result can be rearranged to yield:

$$(R-1) * (q_m) = (q_w - q_m)$$

or: $$R = q_w/q_m \qquad \textit{(Equation S7g)}$$

In addition, under circumstances where: $(R = 1)$, *Equation S7f* becomes:

$$\lambda = \ln [1 - Zw/\boldsymbol{d}] - \ln [1 - Zm/\boldsymbol{c}] \qquad \textit{(Equation S7h)}$$

At any specific *exposure-level* $[H(a) \geq \lambda]$, the values of (Zw) and (Zm) are unknown. However, if a proportional hazard *Model* is appropriate for the disease being considered, the parameters $(\boldsymbol{c}, \boldsymbol{d}, R, \& \lambda)$ are constants (albeit unknown), so that, from *Equations S7d & S7e,* the probabilities of (Zm) and (Zw) are also fixed at any specific *exposure-level* $[H(a)]$.

7b. Defining an "Apparent" Proportionality Factor

An "*apparent*" hazard proportionality factor (R^{app}) can be defined such that: $R^{app} = (q_w^{min}/(q_m^{min})$, which represents the value (R) when: $(\boldsymbol{c} = \boldsymbol{d} = 1)$ – see *Section 6a*; *above*. Potentially, this value incorporates

two different processes. First, it may reflect the increased level of "*sufficient*" exposure experienced by one sex compared to the other. Indeed, from *Equation S5i*, this is the only possible interpretation for circumstances in which: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$. Second, however, if: $(\boldsymbol{c} < \boldsymbol{d} \leq 1)$ is admitted as a possibility, then a portion of (R^{app}) will be due to the difference of $(\boldsymbol{c})$ from unity.

{NB: The possibility that: $(\boldsymbol{d} < \boldsymbol{c})$*, is directly analogous to that of:* $(\boldsymbol{c} < \boldsymbol{d})$*, and, thus, is not considered further.}*

Considering those circumstances in which: $(\boldsymbol{d} = 1)$ & $(R \geq 1)$, from *Sections 5b* (*above*) and *Section 8a* (*below*), the *"actual" exposure-level* change in susceptible *men* (q_m) has a limited range such that:

$$\forall(R^{app} \geq R \geq 1): \quad q_m^{min} \leq q_m \leq q_w^{min}$$

$$\text{where:} \quad \boldsymbol{c} = (Zm_2) * \{e^{q_m} - [P(M,MS)_1/P(M,MS)_2]\}/(e^{q_m} - 1) \leq 1$$

From this, the *"actual"* hazard proportionality factor $(R^{app} \geq R \geq 1)$, at $(\boldsymbol{d} = 1)$, can be defined such that:

$$R^{app} \geq R = q_w^{min} \,/\, q_m$$

In this manner, if $(q_m > q_m^{min})$, some of the "*apparent*" value (R^{app}) will be accounted for by the fact that, in this case, $(\boldsymbol{c} < 1)$. Furthermore, if a reduction of $(\boldsymbol{c})$ from unity is possible in susceptible *men*, then, clearly, it is also possible for the value of $(\boldsymbol{d})$ in susceptible *women* to be less than unity. For example, when: $(\boldsymbol{c} < \boldsymbol{d} < 1)$, the *"actual" exposure-level* in *women* (q_w) will be greater than its *minimum* value (q_w^{min}) such that:

$$R = \; q_w/q_m > q_w^{min}/q_m$$

As a result, in each of these cases, the *"actual"* (R) value may differ from its *"apparent"* value (R^{app}).

7c. Implications that the Values of* $(\boldsymbol{R})$*,* $(\boldsymbol{\lambda})$*,* $(\boldsymbol{c})$ *and* $(\boldsymbol{d})$ *have for Each Other

Assertions:

1. $\forall(R \geq 1): \; \lambda > 0$
2. $\forall(\lambda \leq 0): \; \boldsymbol{c} < \boldsymbol{d} \leq 1$
3. $\forall(R \leq 1) \;\&\; \forall(R < R^{app}): \; \boldsymbol{c} < \boldsymbol{d} \leq 1$
4. $\forall(\boldsymbol{c} = \boldsymbol{d} \leq 1): \; both \;\; (R > 1) \;\; and \;\; (\lambda > 0)$

Proof: The ratios $(C_F \,\&\, C_M)$ are defined in *Section 8a* (*below)* and, because both $P(MS)$ and the (*F:M*) *sex ratio* are *currently* increasing [3,4,23] – *see also Sections 8a & 10a; Figure S1* (*below*) – therefore:

$$C_F = P(F,MS)_1/P(F,MS)_2 < P(M,MS)_1/P(M,MS)_2 = C_M$$

From *Equation S5j*, during *any Time Period*:

$$(F:M) \; sex \; ratio = (Zw/Zm) * \{p/(1-p)\}$$

and, as noted earlier, $[p = P(F \mid G)]$ is independent of the environmental conditions during (E_T). Therefore, for all solutions, the (Zw/Zm) ratio mirrors the (*F:M*) *sex ratio – see Section 5d* (*above*).

1. *For those Conditions in which:* $(R = 1)$:

From *Section 7a* (*above*) for circumstances where: $\{R = (q_w/q_m) = 1\}$, it must be that:

$$q_m = q_w \geq q_w^{min}$$

When: $(\lambda = 0)$, from *Equation S7h* (*above*):

$$Zm/\boldsymbol{c} = Zw/\boldsymbol{d}$$

or: $$Zw/Zm = \boldsymbol{d}/\boldsymbol{c} \qquad \text{(Equation S7i)}$$

Therefore, the (*F:M*) *sex ratio* will remain constant in this case, regardless of the *exposure-level*.

However, when: $(R = 1)$, then: $[q_w = q_m]$ – *see above*. Therefore, from *Section 8; Equations S8c-d* (*below*):

$$\boldsymbol{d}/\boldsymbol{c} = \{Zw/Zm\} * \{(e^{q_w} - C_F)/(e^{q_w} - C_M)\} > Zw/Zm$$

or, with rearrangement: $$Zm/\boldsymbol{c} > Zw/\boldsymbol{d}$$

Therefore, from *Equation S7h*: $\lambda > 0$

Consequently, if $(R = 1)$, and if both the (*F:M*) *sex ratio* and $P(MS)$ are *currently* increasing, then the threshold for susceptible *women must* be greater than that it is for susceptible *men*.

2. *For those Conditions in which:* $(\lambda \leq 0)$ & $(R > 1)$:

For $\{H(a) \geq 0\}$, from *Equation S7f*, during any (E_T), under these conditions:

$$\{\ln(1 - Zw/\boldsymbol{d}) - \ln(1 - Zm/\boldsymbol{c})\}/R = \lambda - [(R-1)/R] * H(a) \leq 0$$

or: $$\ln(1 - Zw/\boldsymbol{d}) - \ln(1 - Zm/\boldsymbol{c}) \leq 0 \qquad \text{(Equation S7j)}$$

In turn, under these conditions, *Equation S7j* requires that:

$$Zm/\boldsymbol{c} \leq Zw/\boldsymbol{d}$$

or: $$Zw/Zm \geq \boldsymbol{d}/\boldsymbol{c} \qquad \text{(Equation S7k)}$$

Also, regardless of the value of (R), from the definitions of $(\boldsymbol{c})$, and $(\boldsymbol{d})$ – *Section 5b* – from the definition of (E) – *Section 5b* – and from *Equation S5i*:

$$\lim_{a \to \infty}(Zw/Zm) = \boldsymbol{d}/\boldsymbol{c} \qquad \text{(Equation S7l)}$$

Because, with increasing exposure, both (Zw) and (Zm) increase monotonically (*see Section 4a*), and because $(R > 1)$, and because $(\lambda \leq 0)$, and because $\{H(a) \geq 0\}$, the condition that:

$$Zw/Zm \geq \boldsymbol{d}/\boldsymbol{c}$$

requires that: $Zw_1/Zm_1 \geq Zw_2/Zm_2 \geq \boldsymbol{d}/\boldsymbol{c}$:

Thus, under these conditions, the (Zw/Zm) ratio either decreases or remains constant with increasing exposure. Because the (Zw/Zm) ratio mirrors the (*F:M*) *sex ratio*, therefore, the (*F:M*) *sex ratio* will also decrease or remain constant (e.g., *Figure 1C; Main Text*) – a conclusion, which is inconsistent with the evidence [1-4,23]. Thus, the conditions: $(R > 1)$ & $(\lambda \leq 0)$ are not plausible, given the Canadian data [23].

Combining these two conclusions (i.e., *Conditions 1 & 2* ; *above*), it *must* be the case that:

$$\forall(R \geq 1):\ \lambda > 0$$

From the Canadian MS data [23], both $P(MS)$ and the (*F:M*) *sex ratio* are *currently* increasing when the "*current*" epoch is compared to any of the previous 5-year epochs from the same study – *see Section 10a, Figure S1* (*below*). An increasing MS-prevalence disproportionately affecting *women* is also reported from other parts of the world [1-4]. Therefore, based exclusively on the increasing $P(MS)$ and (*F:M*) *sex ratio*, and on purely theoretical grounds, one can conclude, that, if the hazards in susceptible *men* and *women* are proportional and if: $(R \geq 1)$, then susceptible *women must* have a higher threshold than susceptible *men*.

3. *For those Conditions, in which:* $(\lambda \geq 0)$ & $(R \leq 1)$

If: $(\lambda \geq 0)$ & $(R \leq 1)$ & $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$; then the *failure-probability* for susceptible *men* would be as great (or a greater) than the *failure-probability* for *women* (i.e., $Zm \geq Zw$) at every *exposure-level* (*see Figure 2B; Reference #4*). Because: $(Zw_2 > Zm_2)$, these conditions are impossible. Therefore, whenever: $(\lambda \geq 0)$ & $(R \leq 1)$, then: $(\boldsymbol{c} < \boldsymbol{d} \leq 1)$ – e.g., *Figure 1B* (*Main Text*).

4. *For those Conditions, in which:* $(\lambda < 0)$ & $(R \leq 1)$:

In these conditions, *Equation S7k* still applies and, thus, if: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$, following the intersection of the response curves for susceptible *men* and *women*, then $(Zm > Zw)$ at every *exposure-level* (e.g., *Figure 1; Main Text*). Because an increasing (*F:M*) *sex ratio* only takes place after this intersection, the condition that both: $(Zw_2 > Zm_2)$ & $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$, is not possible. Nevertheless, the condition that: $(\boldsymbol{c} < \boldsymbol{d} \leq 1)$ is still possible – e.g., *Figure 1D* (*Main Text*). Therefore, combining *Conditions 2 & 4* (*above*), it *must* be the case that:

$$\forall(\lambda \leq 0):\ \boldsymbol{c} < \boldsymbol{d} \leq 1$$

5. *For those Conditions, in which:* $(R^{app} > 1)$ *or* $(R^{app} > R)$:

The value of (R) is related to how quickly the response curves for susceptible *men* and *women* go from onset to their maximums. Thus, this value is independent of (λ). Rather, it depends only upon how quickly this transition occurs. Consequently, for comparison, one is free to choose *any* (λ) value. Therefore, when $(\boldsymbol{c} = \boldsymbol{d})$ & $(\lambda = 0)$, for any (E_T), *Equations S5a & S5f* can be multiplied by the scaling factor of: $(1/\boldsymbol{c})$, and then restated as:

$$Zm/\boldsymbol{c} = (1 - e^{\{H(a)\}})$$

and: $$Zw/\boldsymbol{c} = (1 - e^{R*\{H(a)\}})$$

The *RHS* of both *Equations* is independent of scale. Also, the relationship *between* the *LHS* of two *Equations* is also independent of scale. Therefore, the relationship *between* these two *Equations*, when $(\boldsymbol{c} = \boldsymbol{d})$, is independent of scale. In his case, when: $(\boldsymbol{c} = \boldsymbol{d})$ the value of (R) is constant for all: $(Zm_2 < \boldsymbol{c} \leq 1)$ and therefore:

$$\forall(\boldsymbol{c} = \boldsymbol{d}):\ R^{app} = q_w^{min}/q_m^{min} = q_w/q_m = R$$

However, whenever: $(R \leq 1)$, then also, $(q_w \leq q_m)$.

Consequently, whenever: $(R^{app} > 1)$, then:

$$R^{app} = q_w^{min}/q_m^{min} > 1 \geq q_w/q_m = R$$

Any circumstance in which: $(R^{app} > 1)$, therefore, implies that:

$$\forall(R \leq 1): \ \boldsymbol{c} < \boldsymbol{d}$$

Combining the three conclusions from *Conditions 3–5* (*above*), it is clear that:

$$\forall(R \leq 1): \ \boldsymbol{c} < \boldsymbol{d} \leq 1$$

Indeed, following a logic directly analogous to that *above*, it must also be that:

$$\forall(R^{app} > R): \ \boldsymbol{c} < \boldsymbol{d}$$

6. Finally: Combining each of the conclusions from *Conditions 1–5* (*above*), one can further conclude, based on purely theoretical grounds, that whenever: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$, it *must* also be the case that both: $(R > 1)$ and: $(\lambda > 0)$.

7d. Strictly Proportional Hazard: $(\boldsymbol{\lambda = 0})$

If the condition of "strictly" proportional hazards in susceptible *men* and *women* were to apply, then, by definition:$(\lambda = 0)$. Consequently, whenever: $(\lambda > 0)$, as it *must* be when $(R \geq 1)$, the hazards cannot be "strictly" proportional to each other. In fact, for those cases in which $(R \geq 1)$ and $(\lambda = 0)$, the observed (*F:M*) *sex ratio* either decreases or remains constant with increasing exposure (*see Equations S7j–l; above*), regardless of the values that $(\boldsymbol{c})$ and $(\boldsymbol{d})$ parameters take – e.g., *Figure 1C* (*Main Text*). Therefore, the only possible "strictly" proportional conditions, are those in which the hazard in susceptible *men* is greater than that in susceptible *women* – i.e., $(R < 1)$. Importantly, if the hazard in susceptible *men* is greater than that in *women*, then, as noted in *Section 7c;* (*above*), the simultaneous conditions of: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$ & $(\lambda = 0)$ are excluded.

Consequently, the only "strictly" proportional conditions possible are those, in which both $(R < 1)$ and $(\boldsymbol{c} < \boldsymbol{d} \leq 1)$ – e.g., *Figure 1D* (*Main Text*).

{NB: In the Figures presented in the Main Text, all response curves serving as examples for conditions in which: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$*, are depicted for the condition* $(\boldsymbol{c} = \boldsymbol{d} = 1)$*. Nevertheless, for all conditions (and, therefore, for all Figures) in which the condition of* $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$ *applies, the depicted response curves differ only in so far as the scale of the y-axis is different. Thus, any response curve, depicted at:*$(\boldsymbol{c} = \boldsymbol{d} = 1)$*, is representative of all curves for conditions in which* $(\boldsymbol{c} = \boldsymbol{d})$ *– see Section 7c; Condition 5 (above).}*

7e. Intermediate Proportional Hazard: $(\boldsymbol{\lambda < 0})$

It is possible that a different *Model*, the so-called "*intermediate*" *Model*, is more appropriate than the "strictly" proportional *Model* considered *above*. In this *Model*, the hazards in susceptible *women* and *men* are

still held to be proportional to one another but the onset of the response curves in susceptible *women* and *men* are offset from each other by an amount $(\lambda \neq 0)$. As noted previously: $\forall\ (R \geq 1)$: $\lambda > 0$. Consequently, whenever: $(\lambda < 0)$, it *must* be that the hazard in susceptible *men* is greater than it is in *women*. In addition, under conditions, where $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$ & $(R < 1)$ & $(\lambda < 0)$, the (*F:M*) *sex ratio* initially decreases with increasing exposure until the two response curves intersect at a point below $[p/(1-p)]$ on the *y-axis* (e.g., *Figure 1A; Main Text*). Following this intersection, the (*F:M*) *sex ratio* increases steadily, ultimately reaching a level of $[p/(1-p)]$ on the *y-axis* and, notably, never exceeds this level. In addition, after this intersection (i.e., after the nadir), the response curves maintain a relationship such that: $(Zm > Zw)$, throughout the remainder of response curve until the (*F:M*) *sex ratio* reaches the level of: $[p/(1-p)]$ *on the y-axis* (e.g., *Figure 1A; Main Text*). Moreover, defining the term: $[(p') = P(F \mid MS)]$; it follows from *Equation S5j* (*above*) and the condition that: $(Zw_2 > Zm_2)$ requires both of the conditions:

$$(p')_2 > p \quad \& \quad [p'/(1-p')]_2 > p/(1-p)$$

Therefore, the condition of: $(\lambda < 0)$ is only possible, when: $(\boldsymbol{c} < \boldsymbol{d})$ – e.g., *Figure 1B* (*Main Text*).

7f. Intermediate Proportional Hazard: $(\boldsymbol{\lambda > 0})$ *& Autosomal Genotypes*

By contrast, when $(\lambda > 0)$, there are no constraints on the relationship that the hazards can take in susceptible *women* compared to susceptible *men*. Thus, both the conditions of: $(R < 1)$ & $(\lambda \geq 0)$ and the conditions of: $(R \geq 1)$ & $(\lambda > 0)$ lead to similar conclusions (*see Figures 3 & 4; Reference #4*).

In this case, it is useful to define a so-called "*susceptibility genotype*", (G_{is}), for the *i*th susceptible individual. This genotype includes only those genetic factors (located on any chromosome), which are related to MS susceptibility. Because (G_{is}) includes the specification of fewer genetic factors than does the complete genotype of the *i*th individual (G_i), it is possible for more than one person in the population to belong to the same *susceptibility-genotype*. For example, because *MZ*-twins have "*identical genotypes*", therefore, based on our assumption (*see Section 1a, above*), they necessarily have the same *susceptibility-genotype*. The group of individuals, who have the same *susceptibility-genotype* as the *i*th individual is referred to as the (G_{is}) subset within (Z). The occurrence of (G_{is}) represents the event that a person, randomly selected from (Z), belongs to the (G_{is}) subset. The probability of this event is represented as $P(G_{is})$. Because some members of (G) are *MZ*-twins, therefore, the total number of these *susceptibility-genotypes* in the population (m_{is}) is less than (m) – i.e., $(m_{is} < m)$. The subset (G_s) includes *all* of the *susceptibility genotypes* within (Z). The occurrence of (G_s) represents the event that an individual, selected randomly from (Z), is member of the (G_s) subset.

Also, it is possible for two or more individuals (perhaps, each with a different *susceptibility genotype*) to share the same family of "*sufficient*" environmental exposures $\{E_i\}$ with the *i*th individual (*see Section 1a*). Therefore, the "*i-type*" *exposure-group* (G_{it}) – or the "*i-type*" group – is defined to include *all* individuals (possibly with different "*susceptibility genotypes*") who share the same $\{E_i\}$ family. The probability: $P(G_{it})$

represents the probability of the event, (G_{it}), that an individual, randomly selected from (Z), belongs to the (G_{it}) *exposure-group*. Also, from *above,* the total number of "*i-type*" *exposure-groups* in the population (m_{it}) must be less than (m) – i.e., $(m_{it} \leq m_{is} < m)$. The family $\{G_t\}$ is defined to include <u>*all*</u> of the "*i-type*" *exposure-groups*, (G_{it}), within (Z).

The "*autosomal susceptibility genotype*" of the *i*th susceptible individual, (G_{ia}), is defined to include all of genetic factors (located on autosomal chromosomes) that are related to MS susceptibility. The occurrence of (G_{ia}) represents the event that an individual, randomly selected from (Z), is a member is a member of the (G_{ia}) subset – a subset consisting of a single *autosomal susceptibility genotype*. The subset (G_a) is defined to include all of these *autosomal susceptibility genotypes* within the (G) subset. In a similar manner, the occurrence of (G_a) represents the event that an individual, randomly selected from (Z), is a member of the (G_a) subset.

Because the genotypes within (G_a) are exclusively autosomal, it is anticipated that:

$$\forall\, G_{ia} \in (G_a): \quad P(G_{ia} \,|\, M) = P(G_{ia} \,|\, F)$$

$$\forall\, G_{ia} \in (G_a): \quad P(G_{ia}, F, G, G_{is}) = P(F, G_{is})$$

and: $$\forall\, G_{ia} \in (G_a): \quad P(G_{ia}, M, G, G_{is}) = P(M, G_{is})$$

Certainly, it is possible for susceptible *women* and *men* may be members of the same (G_{ia}) subset, but not be members of the same (G_{is}) subset. Consequently, these anticipated equivalences do not, necessarily, imply either that:

$$\forall\, G_{is} \in (G_s): \quad P(F, G_{is}) = P(M, G_{is})$$

or that both: $\forall\, G_{is} \in (G_s): \quad P(F, G_{is}) > 0$ and: $\forall\, G_{is} \in (G_s): \quad P(M, G_{is}) > 0$

However, all but one of the 233 MS-associated genetic loci, reported by the *International Multiple Sclerosis Genetics Consortium*, are located on autosomal chromosomes [6]. Moreover, even for the single locus found on the X-chromosome, *men* and *women* both carried the risk-variant [6]. In such a circumstance, therefore, it seems very likely that:

$$\forall\, G_{is} \in (G_s): \quad P(F, G_{is}) \approx P(M, G_{is})$$

And that the same conclusion will hold for all "*i-type*" *exposure-groups* (G_{it}). Therefore, likely:

$$\forall\, G_{it} \in \{G_t\}: \quad P(F, G_{it}) \approx P(M, G_{it})$$

As a result, likely, both *men* and *women* (at least potentially) could belong to any of the "*i-type*" *exposure-groups* – in which case they will be referred to as "*i-type*" individuals. The same conclusion is suggested by the evidence from the occurrence of MS within families (*see Main Text*). In this context, those environmental factors, which comprise each of the "*sufficient*" *exposure-sets* within the $\{E_i\}$ family, are envisioned to be the same regardless of whether the "*i-type*" individual is a *woman* or a *man*. However, it may be that the "*sufficient*" exposure for an "*i-type*" *woman* needs to be more or less "*intense*" than it is for an "*i-type*" *man* [4].

7g Considerations of Exposure "Intensity"

Before considering notions of "*exposure-intensity*", it is notable that there seem to be four well-established conclusions. First, for *every* proportional hazard solution, which was identified by this analysis (*see Results; Main Text*), it was found that:

$$(R^{app} > 1)$$

Second, on theoretical grounds, from *Section 7c* (*above*), it ***must*** be the case that:

$$\forall(R \leq 1)\ \&\ \forall(R < R^{app})\ \&\ \forall(\lambda \leq 0):\ \boldsymbol{c < d}$$

Third, from *Section 7c* (*above*), under those conditions where both $P(MS)$ and $P(F \mid MS)$ are increasing, then it ***must*** be the case that:

$$\forall(R \geq 1):\ \lambda > 0$$

And fourth, from the Canadian MS-data [23], as the probability of a "*sufficient*" environmental exposure has increased over the last several decades, so too has the (*F:M*) *sex ratio – see Sections 8a & 10a-b; see also Figure S1* (*below*). From these two observations, one can conclude that, over this period of time, the probability $[Zw = (MS \mid F, G)]$ ***must*** have increased at a faster rate than has the probability $[Zm = (MS \mid M, G)]$ and, therefore, almost certainly, it is *currently* the case that: $(Zw > Zm)$ *– see Section 3a* (*above*).

From these four conclusions, if susceptible *men* and *women* have proportional hazards, it follows (*see Section 7c; above*) that following two conditions ***must*** also hold.

1) if: $R \leq 1$; or, if: $R < R^{app}$; or, if: $\lambda \leq 0$; then: $\boldsymbol{c < d}$

   Therefore: if: $\boldsymbol{c = d} \leq 1$; then, both: $R > 1$ and: $\lambda > 0$

2) if: $R > 1$; then: $\lambda > 0$

Condition *#1*, clearly, excludes *any* possibility that: $\boldsymbol{c = d} = 1$

Considering condition *#2*, notably, both of the exposure measures used in this analysis– i.e., (a) and $H(a)$ – are directly related to the parameter $P(E \mid G)$, which represents the probability of the event that an individual, randomly selected from the (G) subset, experiences an environmental exposure "*sufficient*" to cause MS in them. Consequently, this condition – i.e., where: $\lambda > 0$ – indicates 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 *exposure-level*, $(R = 0)$. As a result, the additional requirement that: $(R > 1)$ poses a potential paradox in that, if both of these conditions were true, susceptible *women* would be more likely than *men* to experience a "*sufficient*" exposure when the probability $[P(E \mid G)]$ is high and, yet, susceptible *men* would be considerably more likely than susceptible *women* to experience a "*sufficient*" exposure when this probability is low.

There are two obvious ways to avoid this paradox. Principal among them is for one to conclude that the hazards are not proportional. Despite this possibility, however, such a conclusion creates other problems (*see Main Text*). For example, susceptible *women* and *men* who are members of the same "*i-type*" *exposure-group*

necessarily have proportional hazards (*see Section 7h; below*). Therefore, in this case, one would also have to conclude further that susceptible *women* and *men* can never be in the same "*i-type*" *exposure-group* and, consequently, that the "*sufficient"* exposure sets are different for the two sexes. In such a circumstance, MS in *women* would represent a different disease from MS in *men*. Alternatively, if it were possible that both *women* and *men* could be members of some "*i-type*" *exposure-groups* but not others, one would conclude that MS represents three distinct diseases (one in *women*, one in *men*, and a third in both). Neither conclusion is supported by the available genetic and the epidemiological evidence (*see Main Text*).

The second way to avoid the paradox is to accept *Condition #1*, which is compatible with any (λ). However, if: $(\lambda > 0)$ and $(R \leq 1)$, then, at every population *exposure-level* (a), the probability of the event that a *susceptible-man*, randomly-selected, will experience a *sufficient-exposure* is as great, or greater, than the same probability for a *susceptible-woman*. Thus, although developing a notion of a so-called "*critical exposure-intensity*" may be necessary to rationalize any threshold difference between susceptible *women* and *men* [4], it is not necessary to resolve any paradox. Nevertheless, accepting the conclusion that $(\lambda > 0)$ and $(R \leq 1)$, does require also accepting the fact that $(\boldsymbol{c < d})$ and therefore that some susceptible *men* will never develop MS, even when the correct genetic background occurs together with an environmental exposure "*sufficient*" to cause MS in a person with that genetic background (*see Section 7c; above*).

7h. Variability in the Values of $(\boldsymbol{R_i})$ *and* $(\boldsymbol{\lambda_i})$ *between "i-type" Groups*

In the circumstance where both *men* and *women* are (or potentially could be) members of some specific "*i-type*" *exposure-group* $\{G_{it}\}$, by definition, such *men* and *women* each will have *some* non-zero probability of developing MS in response to *every* "*sufficient*" *exposure-set* within the $\{E_i\}$ family.

For notational clarity, a subset (G_w) will be defined to include of <u>*all*</u> *female* members of the (G) subset {i.e., $(G_w) = (F, G)$}. As in previous *Sections*, the proportion of *women* in the (G) subset is defined as: $[p = P(F \mid G)]$. In this case, each of the $(m * p)$ *women* in the (G_w) subset $(d = 1,2, \dots, mp)$ has a unique genotype (G_{dw}). The occurrence of (G_{dw}) represents the event that an individual, selected randomly from the population (Z), belongs to the (G_{dw}) subset – a subset consisting of only single individual (i.e., the d^{th} susceptible *woman*) – and the probability of this event is represented as: $\{P(G_{dw}) = 1/N\}$. Also, the probability of the event that an individual, selected randomly from the population (Z), belongs to the (G_w) subset is represented as: $\{P(G_w) = P(F, G) = mp/N\}$.

{NB: The use of (G_w) *and* (G_{dw}) *terminology is used only when the listing of individual susceptible genotypes for women is important to the argument being made.}*

In the circumstances where both *men* and *women* are (or, potentially, could be) members of *every* "*i-type*" *exposure-group* and where every *exposure-group* as the same threshold difference (λ), then, at every *exposure-level* for a *man* $\{H(a) \geq \lambda\}$, a proportionality constant ($R_i > 0$) is defined, so that the *exposure-level* for any *i-type* susceptible *woman* $\{K_i(a) \geq 0\}$ can be stated as:

$$\forall\ G_{dw} \in (F, G_{it}):\ \ K_i(a) = R_i * \{H(a) - \lambda\}$$

{NB: In this case, one doesn't need to consider the "i-type" specific exposure for men, $H_i(a)$*, because, by definition, if each exposure-group has the same threshold difference* $(\lambda > 0)$ *then, for all* $\{H(a) \geq \lambda\}$ *and for all* (i)*, it will be true that, for all* (a)*, both* $\{H(a) - \lambda \geq 0]\}$ *and* $\{K_i(a) \geq 0\}$*. Consequently, in this case, there will be some constant* $(R_i > 0)$ *that permits this statement to be true for each* (i)*. The impact of different "i-type" exposure-groups having different thresholds is considered below.}*

Because each susceptible *woman* (G_{dw}) is a member of <u>*some*</u> "*i-type*" *exposure-group* (G_{it}), an *exposure-level* $[K_{dw}(a)]$ and a proportionality factor $[R_{dw}]$ can be defined for each susceptible *woman* so that:

$$\forall\ G_{dw} \in (G_w):\ \ K_{dw}(a) = R_{dw} * (H(a) - \lambda)$$

where: $\forall\ G_{dw} \in (F, G_{it}):\ K_{dw}(a) = K_i(a)\ \ and:\ \ R_{dw} = R_i$

In this circumstance, the expected *exposure-level* for susceptible *women* can be stated as:

$$K(a) = E\{K_{dw}(a)\} = \sum_{d=1}^{mp}\ R_{dw} * \{H(a) - \lambda\}/mp = R * \{H(a) - \lambda\}$$

where: $R = E(R_{dw})$

Consequently, if *women* and *men* can (at least potentially) be members of every "*i-type*" *exposure-group*, the hazards for *women* and *men* will always be proportional. However, the hazard proportionality factor (R_i) may be different for different "*i-type*" *exposure-groups*.

It is also possible that the threshold-difference between suscitpible *women* and *men* (λ_i) varies between the different "*i-type*" *exposure-groups*. Initially, the circumstances where $(\lambda > 0)$ are considered. The "*i-type*" *exposure group* (j) with the smallest threshold (λ_{jm}), for *men* of any "*i-type*" *group*, can be defined such that:

$$[\lambda_{jm} = \min(\lambda_m)]$$

By definition: $(\lambda_{jm} = 0)$ – *see Section 5c; above*. Similarly, in this case, the *i-type*" *exposure-group* (k) with the smallest threshold (λ_{kw}), for *women* of any "*i-type*" *group* can be defined such that:

$$[\lambda_{kw} = \min(\lambda_w) > 0]$$

In this case, from the definition of threshold, some *men* and some *women* will begin to develop MS at these *exposure-levels* so that, in this circumstance:

$$\lambda = \lambda_{kw} - \lambda_{jm} = \lambda_{kw}$$

Moreover, it is possible that the *men* and *women* who develop MS at these *exposure-levels* are not members of the same "*i-type exposure-group* and, therefore, it is not necessarily the case that $(j = k)$. Regardless,

however, a difference in threshold can then be defined between (λ_{dw}) for each susceptible *woman* and $(\lambda_{jm} = 0)$. In this circumstance, therefore, one can define one can define $(\lambda_i > 0)$ such that:

$$\forall\, G_{it} \in \{G_t\}\ \&\ \forall\, G_{dw} \in (F, G_{it}):\ \ \lambda_{dw} = \lambda_i$$

In this way, the proportionality constants for each "*i-type*" $(R_i > 0)$ and each *woman* $(R_{dw} > 0)$ can be replaced by a *"adjusted"* proportionality constants $(R'_i > 0)$ and $(R'_{dw} > 0)$ such that:

$$\forall\, G_{dw} \in (G_w):\ \ K_{dw}(a) = R_{dw} * (H(a) - \lambda_{dw}) = R'_{dw} * \{H(a) - \lambda\}$$

where: $\forall\, G_{dw} \in (F, G_{it}):\ K_{dw}(a) = K_i(a)\,;\ \ R_{dw} = R_i\ ;\ \ R'_{dw} = R'_i\ ;$ and: $\ \lambda_{dw} = \lambda_i$

Thus, in this circumstance, the expected *exposure-level* for susceptible *woman* can be stated as:

$$K(a) = E\{K_{dw}(a)\} = \sum_{d=1}^{mp}\ R'_{dw} * \{H(a) - \lambda\}/mp = R * \{H(a) - \lambda\}$$

where: $\ R'_{dw} = R_{dw} * \{(H(a) - \lambda_{dw})/(H(a) - \lambda)\} \leq R_{dw}$

and where now: $\quad R = E(R'_{dw})$

When: $(\lambda < 0)$, this analysis is only changed in that the roles of susceptible *men* and *women* are interchanged for all of the above arguments and conditions. Thus, in both cases, the hazards will be proportional. Moreover, because *failure-probability* is described only as a function of the *probability* of a "*sufficient*" exposure, given the environmental conditions of the time (*see Section 5a; above*), and because it is posited that *women* and *men* can (at least potentially) be members of every "*i-type*" *exposure-group*, it is unnecessary to specify the composition of the "*sufficient*" *exposure-sets*, within each $\{E_i\}$, which have resulted in the observed *failure-probability* change between *Time Period #1* and *Time Period #2*.

By contrast, if *men* and *women* each require distinct "*sufficient*" *exposure-sets*, the hazards will not be proportional and *women* and *men* would require their response curves plotted separately; each graph having its own *x-axis* scale. In this case, one would also need to envision *men* and *women* with MS as each having different underlying diseases.

{NB: One might also imagine the possibility that (R_i) *or* (λ_i) *or both varied between the different exposure-sets within* $\{E_i\}$*. In such a circumstance, susceptible-men and susceptible-women (considered separately) would still have an exponential relationship between their failure-probability and their exposure as measured by the odds that a proband (either male or female) experiences an exposure "sufficient" to cause MS in them (see Section 5a; above). However, if this variability were large enough, the relationship between "i-type" men and "i-type" women could become non-proportional and effectively equivalent to those circumstances, in which these men and women were actually members of distinct "i-type" exposure-groups. In this case, for such "i-type" individuals, as is also the case in other non-proportional circumstances (see above), female-MS and male-MS would represent distinct diseases.}*

8. Summary Equations for the Longitudinal Model

8a. Derivations

For notational simplicity, three related ratios are defined:

$$C = P(MS)_1/P(MS)_2 \quad \text{or:} \quad P(MS)_1 = C * P(MS)_2$$

$$C_F = P(F,MS)_1/P(F,MS)_2 = C * [P(F \mid MS)_1/P(F \mid MS)_2]$$

$$C_M = P(M,MS)_1/P(M,MS)_2 = C * [P(M \mid MS)_1/P(M \mid MS)_2]$$

The following *Summary Equations* can be derived using these definitions:

1. First, one can re-express (Zw_2) & (Zw_1) such that:

$$Zw_2 = P(MS,E \mid G,F)_2 = P(MS \mid G,F)_2 = P(F \mid MS)_2 * \left(\frac{P(MS)_2}{P(G,F)}\right)$$

$$Zw_1 = P(MS \mid G,F)_1 = \frac{P(MS)_1 * P(F \mid MS)_1}{P(G,F)} = C * P(F \mid MS)_1 * \left(\frac{P(MS)_2}{P(G,F)}\right)$$

Therefore: $Zw_2/P(F \mid MS)_2 = Zw_1/\{C * P(F \mid MS)_1\}$

so that: $$Zw_1 = Zw_2 * C * \left(\frac{P(F \mid MS)_1}{P(F \mid MS)_2}\right) = Zw_2 * \left(\frac{P(F,MS)_1}{P(F,MS)_2}\right) = Zw_2 * C_F$$ *Equation S8a*

and similarly: $$Zm_1 = Zm_2 * C * \left(\frac{P(M \mid MS)_1}{P(M \mid MS)_2}\right) = Zm_2 * \left(\frac{P(M,MS)_1}{P(M,MS)_2}\right) = Zm_2 * C_M$$ *Equation S8b*

Equation S5d (*see Section 5b; above*) for *men* can then be rearranged to yield:

$$\boldsymbol{c} = \{e^{q_m} * Zm_2 - Zm_1\}/(e^{q_m} - 1)$$

Substituting in this equation for (Zm_1) from *Equation S8b* yields:

$$\boldsymbol{c} = Zm_2(e^{q_m} - C_M)/(e^{q_m} - 1)$$ *Equation S8c*

and similarly: $$\boldsymbol{d} = Zw_2(e^{q_w} - C_F)/(e^{q_w} - 1)$$ *Equation S8d*

2. Also, notably, both: $(Zm_2 < \boldsymbol{c})$; and: $(Zw_2 < \boldsymbol{d})$. Therefore, from *Equation S8c* and from the definition of the ratio (C_M) – *see above* – it must be the case that:

$$Zm_2 < Zm_2 * \{e^{q_m} - C * \{P(M \mid MS)_1/P(M \mid MS)_2\}/(e^{q_m} - 1)$$

Dividing both sides of this inequality by (Zm_2) and, with rearrangement, yields:

$$C < P(M \mid MS)_2/P(M \mid MS)_1$$ *Equation S8e*

and similarly: $$C < P(F \mid MS)_2/P(F \mid MS)_1$$ *Equation S8f*

One can use the point estimates from *Section 10b* (*below*) – i.e., $P(M \mid MS)_2 = 0.238$ & $P(M \mid MS)_1 = 0.315$ – and, inserting these estimates into *Equation S8e*, yields:

$$C < P(M \mid MS)_2 / P(M \mid MS)_1 = 0.238/0.315 = 0.756$$

Therefore, the observations from the *CCPGSMS* dataset [23] translate to a *minimum* increase in *MS-penetrance* by more than 32% between *Time Period #1* and *Time Period #2* – or, equivalently, to an increase in the prevalence of MS in Canada by more than 32% between the two *Time Periods*.

3. And, finally, because: $P(MS \mid E, G, M) = \boldsymbol{c}$ and: $P(MS \mid E, G, F) = \boldsymbol{d}$; during any *Time Period*, then:

$$Zm_2 = P(MS, E \mid G, M)_2 = P(E \mid G, M)_2 * P(MS \mid E, G, M)$$

or: $$Zm_2 = P(E \mid G, M)_2 * (\boldsymbol{c})$$

with rearrangement, this becomes:

$$P(E \mid G, M)_2 = Zm_2/\boldsymbol{c} \qquad \textit{Equation S8g}$$

and, similarly: $$P(E \mid G, F)_2 = Zw_2/\boldsymbol{d} \qquad \textit{Equation S8h}$$

8b. Limits on the Value of the Parameters: $P(MS \mid E)$, $(\boldsymbol{c})$ and $(\boldsymbol{d})$

As noted earlier (*see Section 2a*), the observed *MZ*-twin concordance rate [i.e., $P(MS \mid MZ_{MS}, E_T)$] may need to be converted into an adjusted rate [i.e., $P(MS \mid IG_{MS}, E_T)$] because the observed rate may reflect, in part, the fact that *MZ*-twin *probands* share both their intrauterine and <u>*some*</u> of their other environments with their *co-twin*. If this *co-twin* either has, or will subsequently develop, MS then, potentially, these shared environmental experiences may also make MS more likely in the *proband*. In this case, to isolate the genetic contribution, the impact of these environmental similarities needs to be removed (*see Section 2a*).

By definition, an *exposure-level* can never be greater than its *maximum* value so that:

$$[P(E \mid MZ_{MS})_2 \leq 1].$$

Moreover, if *any* susceptible *MZ-proband* (G_i) is known to have experienced $\{E_i\}$, then both the environmental experience of their *co-twin*, and the *Time Period*, become irrelevant such that:

$$P(MS \mid E, MZ_{MS})_2 = P(MS \mid E)_2 = P(MS \mid E)$$

Therefore: $$P(MS \mid MZ_{MS})_2 = P(MS, E \mid MZ_{MS})_2 = P(MS \mid E, MZ_{MS})_2 * P(E \mid MZ_{MS})_2$$

or: $$P(MS \mid MZ_{MS})_2 = P(MS \mid E) * P(E \mid MZ_{MS})_2$$

so that: $$P(MS \mid E) \geq P(MS \mid MZ_{MS})_2 \qquad \textit{Equation S8i}$$

Thus, the value of the parameter $[P(MS \mid E)]$ must be, at least, as large as the *currently* observed *MZ*-twin concordance rate. And similarly:

$$\boldsymbol{c} = P(MS \mid E, M) \geq P(MS \mid M, MZ_{MS})_2 \qquad \textit{Equation S8j}$$

and: $$\boldsymbol{d} = P(MS \mid E, F) \geq P(MS \mid F, MZ_{MS})_2 \qquad \textit{Equation S8k}$$

Table S1a. Definitions for Terms used in the Mathematical Development – *see also Tables 1 &2; Main Text*

Terms	Definitions
(Z)	The population – a set consisting of (N) individuals – *see Main Text*
G_k	The unique genotype of the k^{th} individual within the population (Z): $(k = 1,2,\dots,N)$ – *see Main Text & Section 4a*
(F) , (M)	Subsets of *women* (F) and *men* (M) within (Z) – *see Main Text*
(MS)	Subset of *all* individuals within (Z) who either have, or will subsequently develop, MS; or, equivalently, *all* individuals who develop MS over the course of their *life-time – see Main Text*;
(G)	Subset of individuals within (Z) who have *any* non-zero *life-time* chance of developing MS under *some* environmental conditions – *see Main Text & Section 1a*
G_i	The unique genotype of the i^{th} susceptible individual within (G): $(i = 1,2,\dots,m)$ – *see Main Text*
p	Proportion of *women* in the (G) subset – i.e., $p = P(F \mid G)$ – *see Main Text*
(E_T)	Environmental conditions of *some* specific *Time-Period – see legend; Table 3; Main Text*
Subscripts (1) & (2)	Designations for *parameter-values* during *Time Period #1* (*1941-1945*) and *Time Period #2* (*1976-1980*) – e.g., $P(MS)_2$ represents $P(MS)$ during *Time Period #2 – see Section 5b*
$P(MS \mid E_T)$	*Penetrance* of MS for the population (Z) during (E_T) – *see Main Text*
$x = P(MS \mid G, E_T)$	*Penetrance* of MS for the (G) subset of the population (Z) during (E_T) – *see Main Text*
$x_i = P(MS \mid G_i, E_T)$	*Penetrance* of MS for the i^{th} individual in the (G) subset of (Z) during (E_T) – *see Section 1a* By the definition of (G) – *above* – it *must* be that, during *some* (E_T): $\forall\, G_i \in (G)$: $x_i > 0$
$Zw = z_w$	*Penetrance* of MS for the subset of *susceptible women* (F, G) within (Z) during (E_T) – Also called the *failure probability* for *susceptible women* during (E_T) – *see Sections 3a & 5b*
$Zm = z_m$	*Penetrance* of MS for the subset of *susceptible men* (M, G) within (Z) during (E_T) . – Also called the *failure probability* for *susceptible men* during (E_T) – *see Sections 3a & 5b*
$\boldsymbol{c}$, $\boldsymbol{d}$	Limiting values (constants) for the *maximum failure probability* in *susceptible-men* $(\boldsymbol{c})$; and *susceptible women* $(\boldsymbol{d})$ – i.e., $(Zm \le \boldsymbol{c} \le 1)$ and $(Zw \le \boldsymbol{d} \le 1)$ – *see Sections 5b-c*
s_a , s_{aw} , s_{am}	The ratio of: $[P(MS \mid DZ_{MS})/P(MS \mid S_{MS})]$; used to adjust the *MZ*-twin concordance for the environments shared by *MZ*-twins; considered collectively (s_a), or the comparable ratios for *women* (s_{aw}) and *men* (s_{am}); considered separately – *see Main Text & Sections 2b-c*
x'', z''_w , z''_m , x''_i	*MZ-twin Concordance* (*penetrance*) values for members of the (G, MZ_{MS}) subset, (x''), for the subsets (G, F, MZ_{MS}) – (z''_w) – and (G, M, MZ_{MS}) – (z''_m) – and for the subset (G_i, MZ_{MS}) – (x''_i) – considered separately – e.g., $x'' = P(MS \mid MZ_{MS})$ – *see Main Text & Sections 4a-b & 10b*
x', z'_w , z'_m , x'_i	"Adjusted" *MZ-twin Concordance* (*penetrance*) values for members of the (G, MZ_{MS}) subset, (x'), for members of the subsets (G, F, MZ_{MS}) – (z'_w) – and (G, M, MZ_{MS}) – (z'_m) – and (G_i, MZ_{MS}) – (x'_i) – considered separately – e.g., $x' = P(MS \mid MZ_{MS})/s_a = P(MS \mid IG_{MS})$ – *see Main Text & Sections 2a, 3a & 4a-b* By the definition of the *adjusted MZ-twin Concordance*, $P(MS \mid IG_{MS})$: $(x'_i = x_i)$ – *see Section 2a*
r , $\boldsymbol{s}$	Ratios of the adjusted *MZ-twin Concordance* to the *MS penetrance* in *susceptible women, i.e.,* $(r = z'_w/z_w)$; and *susceptible men, i.e.,* $(s = z'_m/z_m)$ – *see Section 4b*
(X)	Set of *MS-penetrance* values for *all* (m) members of the "*genetically-susceptible*" subset (G) – i.e., $(X) = (x_1, x_2, \dots, x_m)$ – *see Main Text & Section 4a*
σ_X^2 , σ_w^2 , σ_m^2	Variance of the *MS-penetrance* values for *all* susceptible individuals (σ_X^2) and for *susceptible women*, (σ_w^2), and *susceptible men*, (σ_m^2), considered separately – *see Sections 3a & 4a*
(G_w) , (G_m)	Alternative designations for subsets of all *susceptible women* – i.e., $(G_w) = (F, G)$ – and all susceptible *men* – i.e., $(G_m) = (M, G)$ – *see Sections 3a, 4a & 7h*
G_{dw} , G_{dm}	Alternative designations for the genotypes of the (mp) *women* in the (F, G) subset – $(d = 1,2,\dots,mp)$ – and for the genotypes of the $[m(1-p)]$ *men* in the (M, G) subset – $[d = 1,2,\dots,m(1-p)]$ – *see Sections 3a, 4a & 7h*
z_{dw} , z_{dm}	*MS-Penetrance* values for the d^{th} *susceptible woman* in (G_w): $(d = 1,2,\dots,mp)$; and for the d^{th} *susceptible man* in (G_m): $[d = 1,2,\dots,m(1-p)]$ – *see Section 3a*

Table S1b. Definitions for Terms used in the Mathematical Development – *Continued*

Terms	Definitions
G_{i1} , G_{i2}	Any pair of *susceptible* individuals, randomly-selected from (G) – *see Section 3a*
x_{i1} , x_{i2}	*MS-penetrance* values, respectively, for the individuals G_{i1} and G_{i2} – *see Section 3a*
x'_{i1} , x'_{i2}	*Adjusted MS-Penetrance* values, respectively, for members of the (G_{i1}, MZ_{MS}) and (G_{i2}, MZ_{MS}) subsets By the definition of the *adjusted MZ-twin Concordance*, $P(MS \mid IG_{MS})$: $(x'_{i1} = x_{i2})$ and: $(x'_{i2} = x_{i2})$ – *see Sections 2a & 3a*
(G_{is})	Subset of susceptible individuals who share the same "*susceptibility genotype*" with the i^{th} susceptible individual – i.e., the genotype considering *only* those genetic factors related to "*genetic susceptibility*" – *see Sections 7f-h*
(G_s)	Subset of all "*susceptibility genotypes*" within (Z) – *see Sections 7f-h*
(G_{ia})	Subset of susceptible individuals who share the same "*autosomal susceptibility genotype*" with the i^{th} susceptible individual – i.e., the genotype considering *only* those *autosomal* genetic factors related to "*genetic susceptibility*" – *see Sections 7f-h*
(G_a)	Subset of all "*autosomal susceptibility genotypes*" within (Z) – *see Sections 7f-h*
(G_{it})	Subset of susceptible individuals (possibly with different *susceptibility genotypes*) who are in the same "*i-type*" *exposure-group* – i.e., individuals who share the same $\{E_i\}$ family of "*sufficient*" *environmental-exposures* – *see Sections 7f-h*
$\{G_t\}$	Family of all "*i-type*" *exposure-groups* within (Z) – *see Sections 7f-h*
$\{E_i\}$	Family of <u>*every*</u> set of *environmental-exposures*, each of which is "*sufficient*", by itself, to *cause* MS in the i^{th} susceptible individual within (G): $(i = 1,2, \dots, m)$ – *see Section 1a*
(E)	Event that a randomly selected member of (G) – the *proband* – experiences an environment *sufficient* to *cause* MS in them – *see Section 1a*
$P(E \mid G, E_T)$	Probability that the event (E) occurs during (E_T) – *see Section 1a*
u	Variable representing the level of *environmental-exposure*, as measured by the odds that the event (E) occurs during *any* (E_T) – *see Section 5a*
a	Level of *environmental-exposure* during *some* specific (E_T) – i.e., when: $(u = a)$
$h(u)$, $k(u)$	Unknown (and unspecified) hazard functions for *susceptible men* – $h(u)$ – and for *susceptible women* – $k(u)$ – *see Section 5a*
$H(a)$, $K(a)$	Cumulative hazard functions for *susceptible-men* – $H(a)$; and *susceptible-women* – $K(a)$ – Defined as the definite integrals of these unknown and unspecified hazard functions from an *exposure-level* of: $(u = 0)$ to an *exposure-level* of: $(u = a)$ – *see Section 5a*
q_w , q_m	Actual *exposure-level* change between *Time Periods* for *women* (q_w) and *men* (q_m) – *see Section 5b*
q_w^{min} , q_m^{min}	Minimum *exposure-level* change possible between *Time Periods* for *women* (q_w^{min}) and *men* (q_m^{min}) – *see Section 5b*
$R > 0$	Value of the proportionality-factor (*if the hazards are proportional*) – i.e., $k(u) = R * h(u)$ – *see Main Text & Sections 5a & 7a*
R^{app}	The "apparent" value of R – i.e., the value of R for proportional hazards when: $(\boldsymbol{c} = \boldsymbol{d} \leq 1)$ – *see Section 7b*
C	Ratio of the *MS-penetrance* during *Time Period #1*, $[P(MS)_1]$, to that during *Time Period #2*, $[P(MS)_2]$ – i.e., $C = [P(MS)_1 / P(MS)_2]$ – *see Section 8a*
C_F , C_M	Analogous ratios to (C) considering *women* (C_F) and *men* (C_M) separtately. – i.e., $C_F = [P(MS,F)_1 / P(MS,F)_2]$ & $C_M = [P(MS,M)_1 / P(MS,M)_2]$ – *see Section 8a*
λ_w , λ_m	Environmental *exposure-thresholds* for developing MS in *susceptible women* (λ_w) and *susceptible men* (λ_m) – *see Main Text & Section 5c*
λ	Difference in the environmental *exposure-threshold* between *susceptible women* and *susceptible men* – i.e., $\lambda = \lambda_w - \lambda_m$ – *see Main Text & Section 5c*

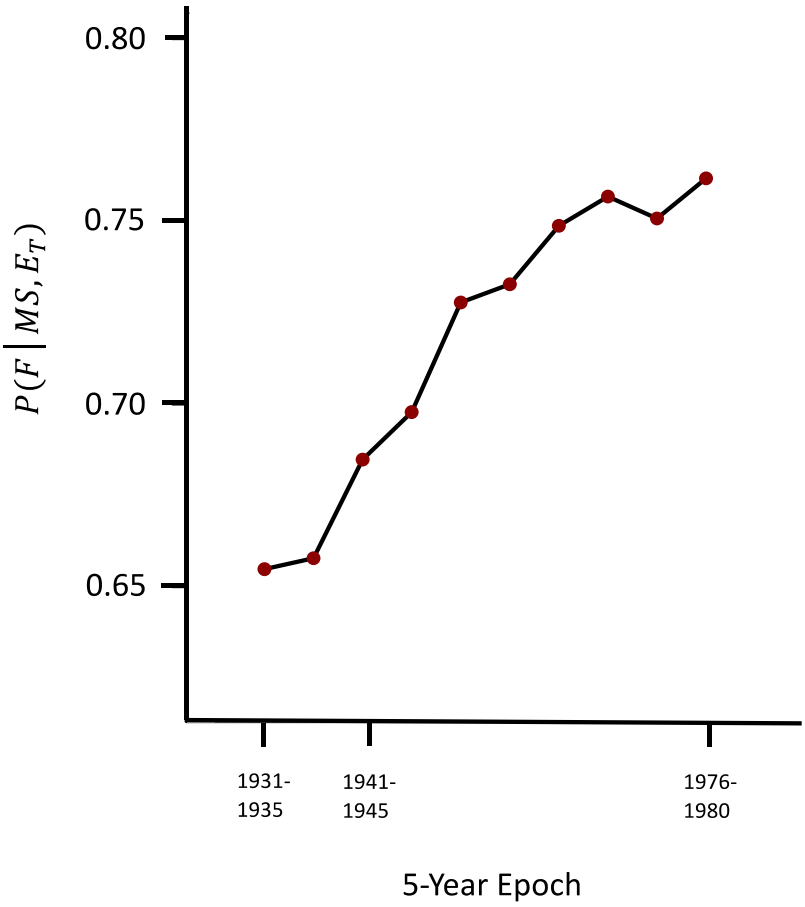

Figure S1. Reported change [23] in the proportion of *women* among MS patients (*y-axis*) – i.e., $[P(F \mid MS, E_T)]$ – over the time of birth-year date (*x-axis)* for persons born in Canada from 1931 until 1980. Each data point in the *Figure* represents a sequential 5-year epoch beginning with $(1931 - 1935)$ and ending with $(1976 - 1980)$. In the *CCPGSMS* dataset there were total of (29,748) identified *MS-cases*, of whom (27,074) were born during this *date-range* and who were included in the analysis [23]. Each 5-year epoch from $(1931 - 1980)$ contained a minimum of (500) identified patients and, of the total number of patients identified in this *date-range*, (19,417) were *women* and (7,657) were *men*. In addition, there were reported to be an average of (2,400) patients identified in each of the ten 5-year epochs, for an average of (480) patients in each birth-year [23]. *{NB: It is unclear from Reference #23 why these last two numbers are not reported as* (2,707.4) *and* (541.48)*, respectively}*. For purposes of the present analysis, the epoch of $(1941 - 1945)$ was chosen as *Time Period #1* because it was the earliest epoch with a very small *confidence-interval* [23]. The epoch of $(1976 - 1980)$ was chosen as *Time Period #2* because it represents the most recent of the reported Canadian epochs [23]. Nevertheless, choosing *any* 5-year epoch from $(1931 - 1975)$ as *Time Period #1* still demonstrates and increasing proportion of *women* between *Time Period #1* and *Time Period #2.*

Table S2. Epidemiological Data regarding Multiple Sclerosis in Canada circa (*2000 –2015*)

Canadian Data for MZ Twin-Pairs*	***Women***	***Men***	***Totals***
Concordant for MS	22	2	24
Discordant for MS	66	43	109
Totals	88	45	133
*Proband-wise Concordance***	0.340	0.065	0.253
Concordance Ratio (F/M)	5.231	–	–
Proportion of Concordant Twins	0.917	0.083	1.000
Proportion of Discordant Twins	0.606	0.394	1.000

Population Data for Canada in 2001-2010:

$Total\ Population = 34{,}108{,}800\ individuals$ -- *from the 2010 Canadian census* [24]

$P(F) = 0.504$ -- *from the 2010 Canadian census* [24]

MS-prevalence (*~2001*) $= (100 - 153)\ cases\ per\ (100{,}000)\ population$ -- *from Reference* [5]

Case Ascertainment in the CCPGSMS:

Estimated using: Twin-rate $= (0.0091)$ *twins per birth;* <u>*and*</u>*: MS-prevalence* $= 100\ cases\ per\ 10^5\ persons$

-- $(454)\ Indentified\ Cases\ /\ (547)\ Expected\ Cases = 83.0\%$ -- *from Reference* [5]

-- *Expected Number of Concordant MZ-Twins* $= (2 * 24)/0.83 = 57.8$ -- *from both above and Table*

-- $(37)\ Ascertained\ /\ (57.8)\ Expected\ = 64.0\%$ -- *from Reference* [5]

Estimated from the Double Ascertainment Rate for Concordant MZ-Twins – i.e., $(13)\ out\ of\ (24)\ Total$

-- $(13)\ Doubly\ Ascertained\ /\ (24 - 5.5)\ = 70.3\%$ -- *from Table, above; References* [5,25]

{NB: This "double ascertainment" estimate is independent of the twin-rate and the MS-prevalence}

Summary Data for MS-Concordance among DZ-Twins and Non-twin Siblings in Canada:

$P(MS \mid DZ_{MS}) = 0.054$ -- *from Reference* [5]

$P(MS \mid S_{MS}) = 0.029$ -- *from Reference* [5]

Summary Data for the Preponderance of Women among MS Patients in Canada

$P(F \mid MS) = 19{,}417/27{,}074 = 0.717$ -- *from Reference* [23]

$P(F \mid MS, MZ_{MS}) = 22/24 = 0.917$ -- *from Table, above; Reference* [5]

During Time Period #1 (*1941-1945*): $P(F \mid MS)_1 = 0.685$ -- *from Figure S1, above; Reference* [23]

During Time Period #2 (*1976-1980*): $P(F \mid MS)_2 = 0.762$ -- *from Figure S1, above; Reference* [23]

---

* Data drawn from the MS-patients in the *CCPGSMS* database as of (~ 2001) – *Reference* [5]

** *Proband-wise* (or *case-wise*) concordance calculated according to [5,25] -- adjusted for double ascertainments $(13/24 = 54\%)$

-- *Proband-wise Concordance* in *men* $= P(MS \mid M, MZ_{MS})$

-- *Proband-wise Concordance* in *women* $= P(MS \mid F, MZ_{MS})$