



OPEN ACCESS

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>).

¹Neurology, University of California San Francisco, San Francisco, California, USA

²Neurology, 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 ($\leq 52\%$) 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*.

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

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.

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 weak⁶—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



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

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 <i>all</i> individuals within (Z) who have <i>any</i> non-zero chance of developing MS under <i>some</i> environmental conditions. |
| G_i | Genotype of the i^{th} susceptible individual ($i = 1, 2, \dots, m$) |
| $\{E_i\}$ | A family consisting of <i>every</i> set of environmental exposures, each of which is 'sufficient', by itself, to <i>cause</i> 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 <i>proband</i> . |
| Concordance-Rate (CR) | Probability that the <i>proband</i> is a member of the (MS) subset, given that their <i>co-twin</i> or non-twin <i>co-sibling</i> is a member of the (MS) subset. Also, often referred to as the Recurrence-Rate or Recurrence-Risk. |
| Penetrance | Probability that a <i>proband</i> will develop MS over the course of their life-time. |
| (MZ, MS) | Subset of MZ-twin <i>probands</i> or MZ <i>co-twins</i> within the (MS) subset. |
| (MZ _{MS}) | Subset of MZ <i>co-twins</i> who are members of the (MZ, MS) subset. |
| DZ _{MS} , S _{MS} | Subsets of DZ <i>co-twins</i> (DZ _{MS}) and non-twin <i>co-siblings</i> (S _{MS}) within the (MS) subset |
| $P(MS MZ_{MS})$ | Concordance Rate of MS for a <i>proband</i> MZ-twin, given that their <i>co-twin</i> has MS ²⁵ — $P(MS MZ_{MS}) = x''$ Comparable definitions pertain to subsets of <i>susceptible women</i> (z'') and <i>susceptible men</i> (z_m'') (see Methods: MZ-twins, DZ-twins and siblings; table 3 & online supplemental table S1). |
| $P(MS IG_{MS})$ | 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 _i) | Some specific Time-Period (see legend; table 3). |
| $P(MS E_i)$ | Penetrance of MS for the population (Z) during (E _i) (see Methods: Genetic susceptibility). |
| C | Ratio of MS-penetrance during Time-Period #1, $P(MS)_1$, to that during Time-Period #2, $P(MS)_2$. |
| $P(MS G, E_i)$ | Penetrance of MS for the (G) subset of the population (Z) during (E _i) (see Methods: Genetic susceptibility). |
| 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

Table 2 Definition of terms (continued)

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

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 antigens²—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 MS⁴ 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 CCPGMS data set

Multiple sclerosis

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 ($k=1,2,\dots,N$). 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 ($i=1,2,\dots,m$) 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 $P(MS|E_T)$. 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 $P(MS|G,E_T)$. 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:

$$Z_w = P(MS|G, F, E_T) \ \& \ Z_m = P(MS|G, M, E_T)$$

These MS-penetrance values, (Z_w) and (Z_m), are also called the ‘failure-probabilities’ for susceptible women and susceptible men during (E_T). 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, $P(F|G)$ —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) | $P(MS) = P(MS Z)$ | 0.003 | 0.001–0.006 |
| Proportion of women in the Population (Z) | $P(F) = P(F Z)$ | 0.504 | – |
| Proportion of women in the (MS) subset | $P(F MS)$ | 0.717 | 0.66–0.78 |
| Time-Period #1 (1941–1945) | $P(F MS)_1$ | 0.685 | 0.67–0.71 |
| Time-Period #2 (1976–1980) | $P(F MS)_2$ | 0.762 | 0.74–0.78 |
| Concordance Rate (CR) for MZ-twins (MZ) | $x'' = P(MS MZ_{MS})$ | 0.253 | 0.18–0.33 |
| (CR) for female MZ-twins | $z_w'' = P(MS F, MZ_{MS})$ | 0.340 | 0.24–0.44 |
| (CR) for male MZ-twins | $z_m'' = P(MS M, MZ_{MS})$ | 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) | $P(MS DZ_{MS})$ | 0.054 | 0.018–0.09 |
| Concordance Rate for non-twin siblings (S) | $P(MS S_{MS})$ | 0.029 | 0.017–0.041 |
| (S:DZ) Concordance ratio | $P(MS S_{MS}) / P(MS DZ_{MS})$ | 0.537 | 0.12–1.0 |
| Non-observed Parameters | | | |
| Definition | Estimate | Plausible Range‡ | |
| Proportion of Population in the (G) subset | $P(G)$ | – | $0 < P(G) \leq 1$ |
| Proportion of women within the (G) subset | $p = P(F G)$ | – | $0 < P(F G) < 1$ |
| 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 parameters⁴ (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 $P(F|MS)$ —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 ($\geq 32\%$) 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 $P(F|MS)$ was similarly expanded,⁴ as was the range for the (S:DZ) concordance ratio, considering the theoretical constraint⁴ that $((S:DZ) \leq 1)$. Because $P(F)$ is taken from a census of the entire ‘current’ Canadian population at the time (2010), there is no estimated range.²⁴

‡Ranges represent the ‘plausible’ parameter value range for each parameter. For example, because, currently, both men and women can (and do) develop MS, $P(G)$ cannot be (0) and $P(F|G)$ cannot be either (0) or (1). Also, the theoretical upper limit for the value of the ratio (C) is 0.9.^{2–4} In addition, a greater than 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: $P(E|G, E_T)$. 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:

$$P(MZ_{MS}) = P(MS, MZ | MZ) = P(MS | MZ)$$

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, $P(MS | MZ_{MS})$ 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, $P(MS | MZ_{MS})$ —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 $P(MS | IG_{MS})$ —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* values⁴ estimates $P(G)$ such that:

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

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, $P(E | G, E_T)$ —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 ($p = P(F | G)$). 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-twinning 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—ie., $P(MS)$ —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*:

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

Multiple sclerosis

No assumptions are made about the relationship between (Z_w) and (Z_m) 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: ($Z_w > Z_m$) is guaranteed (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, ($X = (x_1, x_2, \dots, x_m)$); 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 manner²⁶ 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, (Z_w) and (Z_m) 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* (Z_m) or for *susceptible women* (Z_w), 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: ($c=d \leq 1$), 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: ($c=d \leq 1$) and, under proportional hazard conditions, with proportionality factor (R)—see online supplemental material section 7c and g—two conditions must hold:

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

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

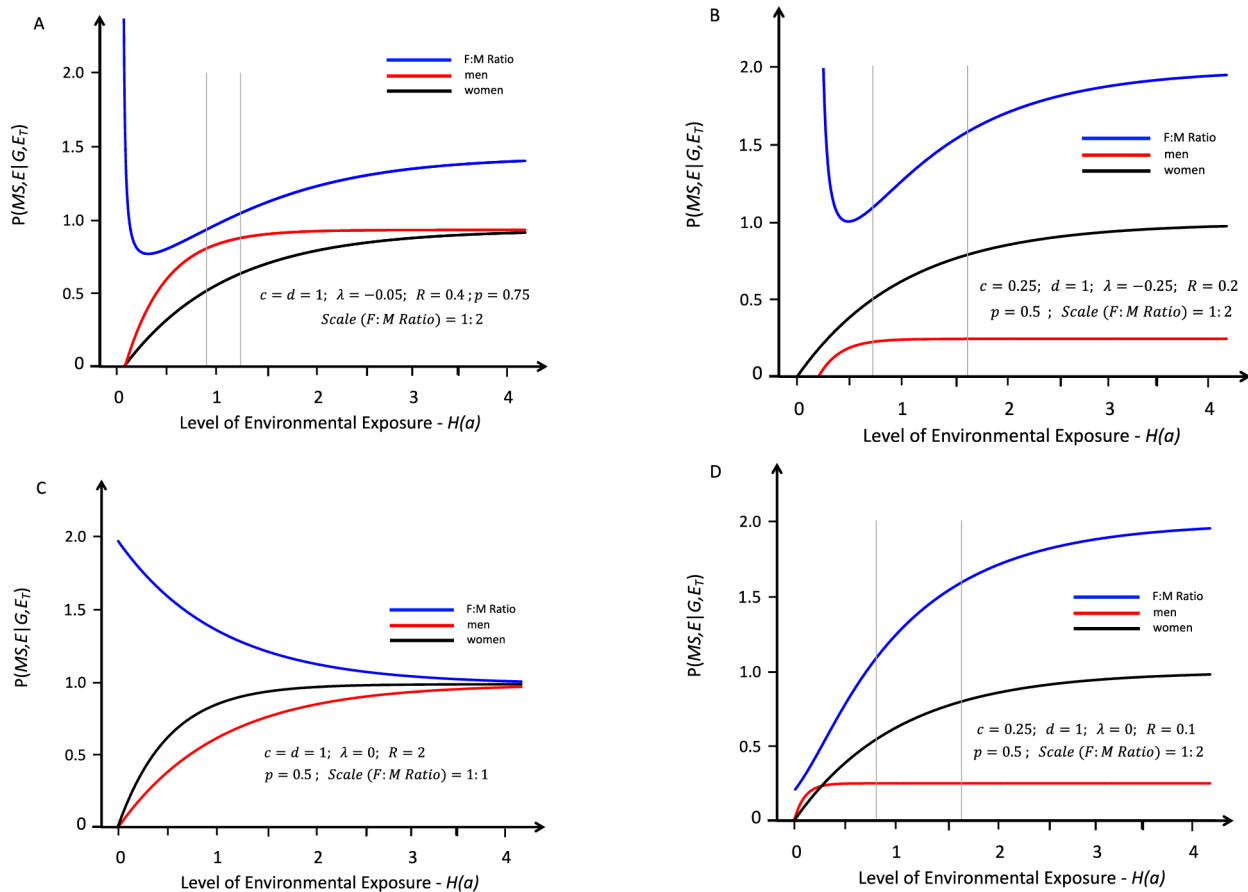


Figure 1 Using the Canadian MS data (online supplemental material 10 a,b), response-curves are depicted for developing MS in *genetically-susceptible women* and men to an increasing probability of sufficient environmental exposure and under conditions, in which the environmental threshold is the same, 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 methods: 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 Canada²³ between *Time-Periods* #1 & #2). The grey lines are omitted in Panel C because the observed ($F:M$) sex-ratio change is not possible under these conditions. In Panel A, although the ($F:M$) sex-ratio change is possible, the condition ($Z_w > Z_m$) is never possible throughout the entire response curve. Response curves A, B, and D reflect conditions in which ($R < 1$); whereas curve C reflects conditions in which ($R > 1$). If ($R = 1$), the blue line in Panel C would be flat (see online supplemental material sections 7 c–f). Response curves A and C reflect conditions in which ($c = d = 1$); whereas curves B and D reflect those conditions in which ($c < d = 1$). $F:M$, 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: ($x = P(MS|G)$); and

Multiple sclerosis

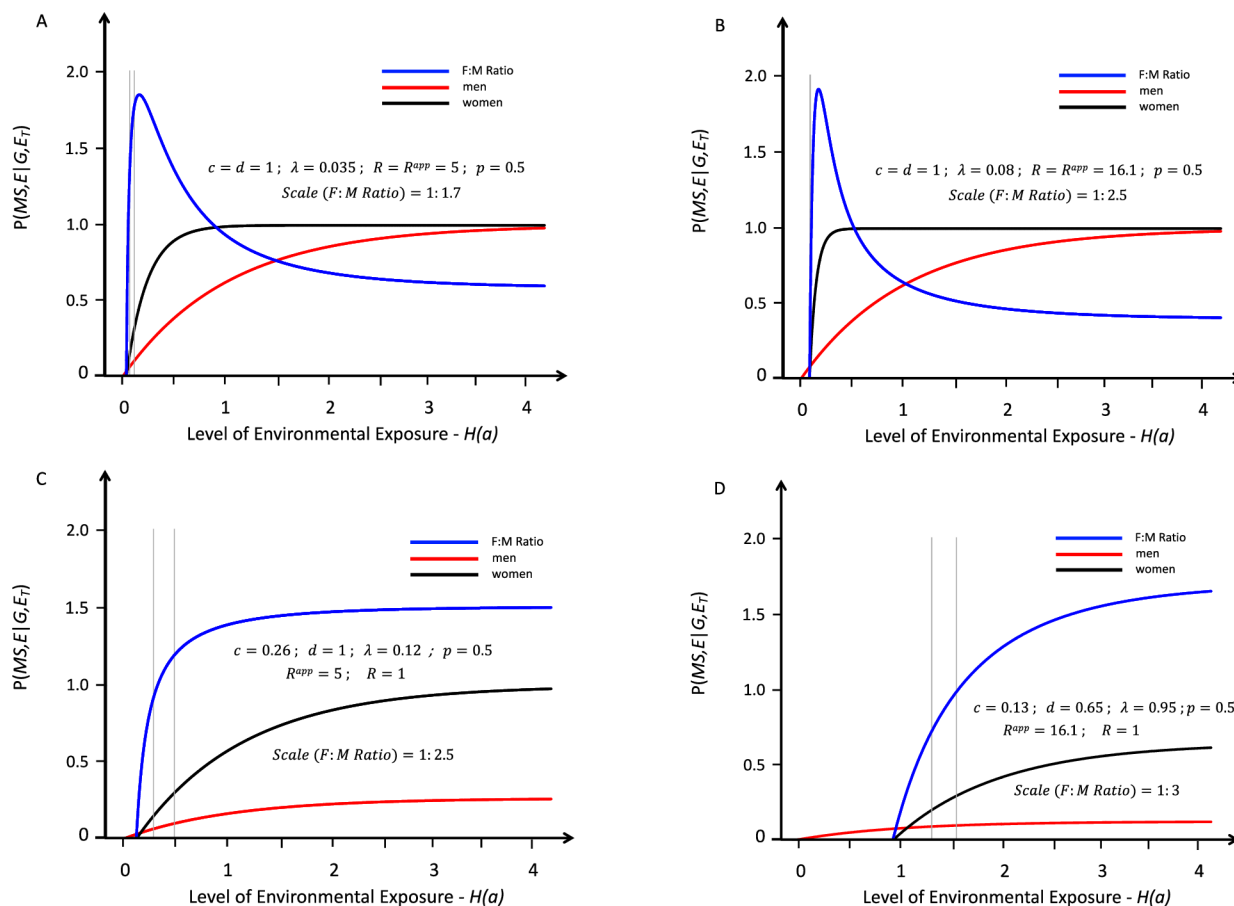


Figure 2 Using the Canadian MS data (online supplemental material section 10a,b), response-curves are depicted for developing MS in *genetically-susceptible women* and men to an increasing probability of sufficient environmental exposure and under conditions, in which the environmental threshold in women is greater than it is in men (ie, conditions where: $(\lambda > 0)$) (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: $(c=d \leq 1)$, 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: $(c < d \leq 1)$ and, in these circumstances: $(R < R^{app})$. To account for the observed increase in the ($F:M$) sex-ratio, the response curves in Panels A and B require that the Canadian observations²³ 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: $(x' = P(MS | IG_{MS}))$.

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:

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

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

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

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 data⁴—with either non-proportional or proportional hazards—and, if proportional, with any (R)—the supported range for $P(G)$ is:

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

For proportional hazards, whenever: $(\lambda \leq 0)$ —figure 1—and, thus, when: $(R < 1)$ —or whenever either: $(R < R^{app})$ or: $(R \leq 1)$, 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: $(c = d \leq 1)$. However, in every such circumstance, the conditions required whenever: $(c < d \leq 1)$ 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, $(P(E | F, G)) / (P(E | M, G))$ —are:

$$0.0005 \leq \lambda \leq 0.13$$

$$1.3 \leq R = R^{app} \leq 1177$$

$$1.2 \leq P(E|F, G)/P(E|M, G) \leq 32.$$

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: ($c = d \leq 1$) might be possible. Principal among these is the possibility of non-proportional hazards, which requires female-MS and male-MS to be different diseases (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: ($c = d \leq 1$), the ‘*current*’ condition—that is, where the ratio of: (Z_w/Z_m) 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: Longitudinal 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: ($c = d \leq 1$) 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: ($c = d \leq 1$)—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: ($R^{app} > R \approx 1$), which is impossible whenever: ($c = d \leq 1$) (Methods: Longitudinal models: Proportional hazard). Third, as in figure 1, whenever ($\lambda \leq 0$) or whenever ($R \leq 1$), 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: ($R > 1$) & ($\lambda > 0$) creates a potential paradox (see Methods: Longitudinal models: Proportional hazard). Although there are ways to rationalise this paradox with: ($c = d \leq 1$), in every case, the conditions required whenever: ($c < d \leq 1$) are far less extreme (see figures 5 and S1–S3 in reference⁴). Finally, the response curves when: ($c = d \leq 1$) & ($R > 1$) are steeply ascending and present only a very narrow *exposure-window* to explain the Canadian ($F:M$) sex-ratio data²³ (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: Longitudinal 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: ($R \approx 1$)—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: Longitudinal models: Proportional hazard) and that ($R \approx 1$). 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.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

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

REFERENCES

- Compston A, Confavreux C, Lassman H, *et al*. *McAlpine's Multiple Sclerosis*. Elsevier: Churchill Livingstone, 2006:287–346.
- 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.
- Goodin DS, Khankhanian P, Gourraud PA, *et al*. The nature of genetic and environmental susceptibility to multiple sclerosis. *PLoS One* 2021;16:e0246157.
- 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.
- Willer CJ, Dymont 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.
- International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 2019;365:eaav7188.
- 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.
- Sadovnick AD, Ebers GC, Dymont DA, *et al*. The Canadian collaborative study group. Evidence for genetic basis of multiple sclerosis. *Lancet* 1996;347:1728–30.
- Willer CJ, Dymont DA, Sadovnick AD, *et al*. Timing of birth and risk of multiple sclerosis: population based study. *BMJ* 2005;330:120.
- Staples J, Ponsobly 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.
- Pantavou KG, Bagos PG. Season of birth and multiple sclerosis: a systematic review and multivariate meta-analysis. *J Neurol* 2020;267:2815–22.
- 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.
- 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.
- 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.
- Munger KL, Levin LI, Hollis BW, *et al*. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA* 2006;296:2832–8.
- 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.
- Sadovnick AD, Dymont DA, Ebers GC, *et al*. Evidence for genetic basis of multiple sclerosis. *The Lancet* 1996;347:1728–30.
- 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.
- 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.
- Ebers GC, Sadovnick AD, Dymont DA, *et al*. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. *Lancet* 2004;363:1773–4.
- Sadovnick AD, Yee IML, Ebers GC, *et al*. Multiple sclerosis and birth order: a longitudinal cohort study. *Lancet Neurol* 2005;4:611–7.
- Dymont DA, Yee IML, Ebers GC, *et al*. Multiple sclerosis in step siblings: recurrence risk and ascertainment. *J Neurol Neurosurg Psychiatry* 2006;77:258–9.
- 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.
- Canadian Census. 2010. Available: https://www150.statcan.gc.ca/n1/en/pub/89-503-x/2010001/article/11475-eng.pdf?st=WV19_Ggm [Accessed 15 May 2023].
- Witte JS, Carlin JB, Hopper JL. Likelihood-based approach to estimating twin concordance for dichotomous traits. *Genet Epidemiol* 1999;16:290–304.
- Fisher LD, Belle G. *Biostatistics: A Methodology for the Health Sciences*. New York: John Wiley & Sons, 1993:369–73.
- Robertson NP, O'Riordan JI, Chataway J, *et al*. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. *Lancet* 1997;349:1587–90.
- Green B. *Until the End of Time*. New York, USA: Alfred A Knopf, Penguin Random House, 2020.
- Layzer D. *Why We Are Free: Consciousness, Free Will and Creativity in a Unified Scientific Worldview*. Information Publisher, 2021.