

GEMS Project: A Platform to Investigate Multiple Sclerosis Risk

Zongqi Xia, MD, PhD^{1,2,3}, Charles C. White, PhD^{1,3}, Emily K. Owen, RN¹, Alina Von Korff, BS¹, Sarah R. Clarkson, BS¹, Cristin A. McCabe, BS³, Maria Cimpean, BA¹, Phoebe A. Winn, BA¹, Ashley Hoelsing, MS¹, Sonya U. Steele, MS⁴, Irene C. M. Cortese, MD⁴, Tanuja Chitnis, MD^{1,2}, Howard L. Weiner, MD^{1,2}, Daniel S. Reich, MD, PhD⁴, Lori B. Chibnik, PhD, MPH^{1,2,3} and Philip L. De Jager, MD, PhD^{1,2,3}

1. Program in Translational Neuropsychiatric Genomics and Partners Multiple Sclerosis Center, Ann Romney Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital, Boston MA.
2. Harvard Medical School, Boston MA.
3. Program in Medical and Population Genetics, Broad Institute, Cambridge MA.
4. Division of Neuroimmunology and Neurovirology, National Institute for Neurologic Diseases and Stroke, Bethesda MD.

Count:

Title: 9 words, 64 characters (with spaces)

Abstract: 97 words (100 maximum)

Manuscript: 5,133 words

References: 43

Tables: 4

Figures: 3 (3 color figures)

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version record](#). Please cite this article as [doi:10.1002/ana.24560](https://doi.org/10.1002/ana.24560).

Supplementary Tables: 3

Key Words: multiple sclerosis, family member, genetics, environment, risk, prospective study

Running Title: Genes and Environment in Multiple Sclerosis Study

Corresponding Author:

Philip De Jager, MD, PhD

Department of Neurology, Brigham and Women's Hospital

77 Avenue Louis Pasteur, NRB 168

Boston, MA 02115

Phone: 617 525-4529

Fax: 617 525-5333

Email: pdejager@rics.bwh.harvard.edu

Accepted Article

Abstract

The Genes and Environment in Multiple Sclerosis (GEMS) project establishes a platform to investigate the events leading to MS in at-risk individuals. It has recruited 2,632 first-degree relatives from across the USA. Using an integrated genetic and environmental risk score, we identified subjects with twice the MS risk when compared to the average family member, and we report an initial incidence rate in these subjects that is 30 times greater than that of sporadic MS. We discuss the feasibility of large-scale studies of asymptomatic at-risk subjects that leverage modern tools of subject recruitment to execute collaborative projects.

Accepted

Introduction

The underlying inflammatory demyelinating disease process in multiple sclerosis (MS) likely predates its earliest clinical manifestation¹. This concept is supported by the observation that many individuals with incidental MS-like brain lesions subsequently fulfill criteria for diagnosis of MS² and that some individuals have incidental, asymptomatic MS-like lesions on autopsy³. Early detection of MS is important, as longitudinal neuroimaging studies have demonstrated accelerated brain atrophy after the first episode of neurological symptoms⁴, and as clinical trials have shown that early treatment with disease-modifying drugs delays the accumulation of disability^{5,6} and possibly reduces mortality⁷. Thus, developing the capacity to detect the earliest stages of the disease process and to identify affected individuals months or years before symptom onset is clinically meaningful.

Primary prevention strategies have not yet been tested in MS in part due to the low incidence of MS in the general population. There are certain populations that are at higher risk of developing MS, but the incidence rate and risk factors that operate in such targeted populations of high-risk individuals for the disease have not been well characterized to date. In particular, first-degree family members of MS patients are 20-40 times more likely to develop MS than the general population¹. Consistent with the notion that much of the disease is asymptomatic, clinically silent MS-like brain lesions are seen in 4-10% of MS family members⁸⁻¹¹. However, screening all MS family members with serial neuroimaging is not practical since the absolute risk of the disease remains modest.

Over the past decade, researchers have validated many genetic factors and environmental exposures that increase MS susceptibility in the general population. Large-scale genome-wide association studies resulting from international collaborative efforts confirmed that genetic variations within the major histocompatibility complex (MHC) exert the greatest individual effect on MS susceptibility but also identified more than 110 additional common genetic variants of more modest effect size outside of the MHC¹²⁻¹⁸. In parallel, epidemiological studies have firmly established the contribution of several environmental factors to MS risk, such as infectious mononucleosis, smoking, adolescent obesity, and Vitamin D deficiency¹⁹⁻²⁷. Taken together, these studies have laid the groundwork for new opportunities to combine risk factors and generate an individualized risk estimate for MS²⁸⁻³⁰. In the future, such a tool may be deployed to identify high-risk individuals (such as family members) prior to symptom onset, and such subjects would be excellent candidates for clinical trials of primary prevention. Our Genes and Environment in Multiple Sclerosis (GEMS) project shows that studies of presymptomatic individuals at risk of MS are feasible.

To tackle the challenge of early detection of MS in this high-risk population of family members, we initiated the GEMS project, a prospective natural history study that will map the sequence of events in the transition from health to MS. Here, we report the design of and initial findings from the GEMS cohort. We also highlight strategies that we have found to be effective in subject recruitment of this non-patient population (i.e., first-degree family members). Further, we introduce and report the efficacy of an integrated

Genetic and Environmental Risk Score (GERS) that provides a single aggregate estimate of MS risk for individual family members. Finally, we evaluate, in this unique cohort of high-risk subjects, the role of known environmental risk factors that were identified in the general population and provide the first estimate of the incidence of MS among first-degree family members. We conclude with a discussion of the strengths of the GEMS study as a platform for investigating risk and prevention of MS as well as our plans to overcome some of its limitations.

Recruiting First-Degree Relatives

Recruiting a subject population (such as family members) that is not typically seen in a clinical setting is challenging. For cost-efficient, rapid and large-scale subject recruitment of a primarily non-patient population, the GEMS study (**Figure 1A**) deployed an innovative recruitment approach that strategically leveraged the effective outreach efforts of patient advocacy groups such as the National Multiple Sclerosis Society, social media tools such as Facebook (<https://www.facebook.com/pages/Genes-and-Environment-in-Multiple-Sclerosis-GEMS-Research-Study/200900853288242>), and person-to-person electronic communication to reach eligible first-degree relatives (**Figure 1B**). In its first four years, the study has recruited 2,632 subjects from every state across the United States (**Figure 1C**).

The inclusion criteria for the family member portion of the GEMS study are: (1) being 18 to 50 years of age at enrollment, (2) having at least one first-degree relative

with a diagnosis of MS (i.e., parent, full-sibling, or child) (**Figure 1A**). We did not include subjects above 50 years of age because the mean age of MS diagnosis is in the early thirties¹ and the age-related appearance of nonspecific white matter lesions in older individuals could confound neuroimaging outcomes. Subjects with an existing diagnosis of MS who have a family member with MS are included to provide a crucial comparison group.

The GEMS project drew its inspiration from the diabetes autoimmunity study in the young (DAISY), a prospective study of high-risk siblings of type 1 diabetes patients that has made important contributions to our understanding of the onset of type 1 diabetes³¹. While the DAISY study and local studies of MS family members^{11,30} have established the feasibility of regional studies of high-risk individuals, the GEMS study has showcased the feasibility of a nationwide approach by leveraging social networks of patient advocacy organizations and the GEMS study Facebook site. These electronic interconnections are crucial for a nationwide strategy that rapidly recruits a large population of subjects who are at risk for a neurologic disease but do not interact regularly with a neurologist.

Building a Longitudinal Data Repository of Individuals at Risk for Multiple Sclerosis

Upon enrollment, each GEMS subject completes a detailed web-based questionnaire that captures demographic information, medical history, family history and

environmental exposures. Each subject returns by mail a saliva sample for extraction of DNA (OG-500 DNA Genotek, Ontario, Canada). The majority of the GEMS subjects have given consent to provide blood and stool samples, undergo neuroimaging, and participate in additional studies. Every three years, subjects are asked to complete questionnaire updates with the goal to follow each subject for 20 years (**Figure 1A**).

Genotyping to Determine the Multiple Sclerosis Genetic Burden

For each GEMS subject, targeted genotyping of validated and replicated single nucleotide polymorphisms (SNPs) that are significantly associated with MS susceptibility^{16,17} (**Supplementary Table 1**) is performed at the Broad Institute on the MassArray iPLEX platform (Sequenom, San Diego, CA) and processed for standard quality control (minor allele frequency > 0.01, genotype call rate > 0.89, batch effect). This list of 64 SNPs, including five within the major histocompatibility complex (MHC) that contains the human leukocyte antigen (HLA) genes, was the most up-to-date as of 2011 when targeted genotyping of GEMS subjects first began. Independent susceptibility SNPs were determined using a linkage disequilibrium (LD) threshold of $r^2 < 0.5$ between any pair of SNPs: if two SNPs had an $r^2 > 0.5$, the one with the lower p-value from previous literature was retained. Genotyping was done over the course of the reported study in four batches, and two SNPs were removed because they did not pass the quality control criteria. Because initial targeted genotyping was performed only on the known genetic variants associated with MS susceptibility (as of 2012), there was insufficient information to determine the genetic ancestry for each subject. However,

since the study focuses on the vast majority of subjects who have self-reported European ancestry (see **Table 1**), the lack of a precise measure of genetic ancestry will not significantly affect results. This would be a much bigger problem for admixed populations or more diverse populations. Ancestry-informative markers will be genotyped and incorporated into future estimation of an individual's risk, as such an approach would increase generalizability of the study results.

Demographics of the GEMS Cohort

As of September 2015, the GEMS study has enrolled 2643 subjects, including 134 with a self-reported MS diagnosis confirmed by a review of medical records. Thirteen subjects withdrew from the study after their initial enrollment. The rate of subject attrition due to loss from follow-up is still being assessed. To maintain engagement, we send an annual newsletter containing study update to subjects and provide regular updates through social media. In addition, each subject provides a secondary contact (e.g., relative, friend) in case we can no longer establish communication with the subject.

In our initial review of the family history data in our study, the majority of GEMS subjects have a single first-degree family member with MS. However, 6% have two, and an important minority (0.6%) has three or more first-degree relatives with MS. The observation that rare families have 3 (n=9) and 4 (n=1) first-degree relatives with MS suggests that further detailed study of these families (e.g., whole genome sequencing,

analysis of gene-environment interaction) may lead to the identification of novel genetic or environmental factors of large effect.

In 2014, we performed a cross-sectional analysis of the first 1,696 GEMS subjects with genotype data and completed questionnaires containing exposure history (see **Table 1** for demographics), including 1,583 asymptomatic subjects and 113 subjects with an existing diagnosis of MS at the time of enrollment (**Table 2**). All subjects, both MS and asymptomatic, have at least one first-degree relative with MS. The proportion of GEMS subjects that have a diagnosis of MS at enrollment (6.7%) is greater than the prevalence rate of MS among first-degree family members in prior reports (2-4%)¹. This likely reflects the fact that first-degree relatives who already have an MS diagnosis are more inclined to participate in our study than the average family member.

Overall, we observed an excess of women in the study (79% of the cohort), consistent with gender differences in human study participation that may be heightened in our study by public understanding of the increased risk of MS for women. Because of this high frequency of women in the study, the larger proportion of women among GEMS subjects who have MS (85%) is not significantly different from the proportion seen in asymptomatic subjects (79%).

Assessing the Role of Selected Environmental Multiple Sclerosis Risk Factors Among First-Degree Relatives

In a cross-sectional analysis of the first 1,696 GEMS subjects, we compared the prevalence of exposure to environmental risk factors between asymptomatic first-degree relatives and those with MS diagnosis at enrollment. We used χ^2 tests for categorical variables and independent sample t-tests for continuous variables. Covariate-adjusted analyses were performed using logistic regression. These analyses have already returned intriguing observations. We established the role of smoking in MS susceptibility in first-degree family members ($p=0.01$) (**Table 2**), but the current analysis is underpowered to confirm the role of body mass index at age 18 (**Table 2** and **Table 3**). Interestingly, a history of infectious mononucleosis was not associated with MS in our sample because of its high prevalence in asymptomatic family members (25%), which is similar to that of the family members with MS (27%) and that of published reports of sporadic MS patients from the general population (23-28%)^{32,33}. This prevalence is higher than that seen in healthy subjects from the Boston-based PhenoGenetic Project (18%) and in reports of the general population (10-15%)^{32,33}. The higher prevalence of mononucleosis in asymptomatic GEMS subjects is not attributable to known MS susceptibility variants: an individual's aggregate burden of MS risk alleles (the weighted MS genetic risk score, GRS) is not associated with a history of mononucleosis in GEMS subjects with MS ($p=0.35$) or in those without MS ($p=0.10$). Overall, these findings suggest that the high rate of infectious mononucleosis in asymptomatic family members may reflect a shared environmental history (e.g., co-infection) and/or a shared genetic component not captured by the known MS variants. This result needs validation but raises the possibility that the association of infectious

mononucleosis with MS susceptibility may be an epiphenomenon and not a causal element in the cascade of events leading to MS.

Determining the Genetic and Environmental Architecture of Multiple Sclerosis Risk Among First-Degree Relatives

For each subject, we calculated a weighted environmental risk score (ERS) and a weighted GRS. The ERS contains three validated non-genetic risk factors that are obtained from questionnaire data. Specifically, we counted the presence or absence of these factors, each weighted by the natural log of the published odds or risk ratio: sex (odds ratio, OR=3.54 for female versus male subjects)³⁴, infectious mononucleosis (OR=2.3 for a history of infectious mononucleosis versus none)³⁵, and smoking status (OR=1.4 for current smoker versus past or never smoker)³⁶⁻³⁸. While it is genetically determined, we include sex with the environmental factors given its vast effects on human biology during the life course. While we collected information on other MS risk factors, they were not included in our pre-planned analysis either because they were less robustly validated at the start of the study in 2011 (e.g., body mass index at age 18) or because they are less precisely ascertained from questionnaire data (e.g. history of Vitamin D intake and sunlight exposure). However, these and other factors will be considered in future versions of the individualized risk scores.

The GRS contains 64 SNPs that are significantly associated with the risk of developing MS based on published genome-wide association studies of MS

susceptibility^{16,17}. The weights of SNPs are based on published odds ratio from the replication phase of the genetic studies. Each SNP was coded additively by the established risk allele and weighted by the natural log of the odds ratio for MS susceptibility. A complete list of SNPs and their weights are included in **Supplementary Table 1**. As expected, the five SNPs tagging HLA alleles have a large weight in the final GRS given their reported effect sizes. Occasionally, a SNP would fail in only one or two of the genotyping batches (**Supplementary Table 1**). In these cases, since none of the failed SNPs were in the HLA region, we substituted the mean score for the SNP from the other genotyping batches of GEMS subjects.

The weighted Genetic and Environment Risk Score (GERS) that integrates genetic burden and environmental exposures is created for each subject (i) as shown in the equation below:

$$GERS_i = \left[\sum_{j=1}^{64} SNP_j \times w_j \right] + I_i[Female] \times \ln(3.54) + I_i[Mono] \times \ln(2.30) + I_i[CurrSmk] \times \ln(1.40)$$

where w_j is the natural log of the odds ratio for SNP_j , and SNP_j is coded as 0, 1 or 2 copies of the reported risk allele, and each of the environmental risk factors is incorporated separately.

Deriving an Individualized Multiple Sclerosis Risk Profile

Following our pre-planned study strategy, we calculated the ERS, GRS, and GERS in subjects from the GEMS cohort with completed genotype and questionnaire data, including subjects with diagnosis of MS at the time of enrollment. We additionally calculated the GRS in MS patients from the Partners MS Center and healthy control subjects from the PhenoGenetic project, but ERS and GERS for these subjects could not be calculated due to incomplete information on smoking status and infectious mononucleosis.

Given the small number of environmental risk factors included in the ERS and the currently modest number of MS subjects in the GEMS cohort, we see a trend toward a greater ERS in GEMS MS subjects relative to the asymptomatic subjects, but this difference is not significant ($p=0.096$ after adjusting for age; **Figure 2A; Table 4**).

As anticipated, asymptomatic GEMS subjects have a greater burden of MS risk alleles than random individuals from the general population. Using a GRS derived from the 63 (out of the 64) MS SNPs that are available in *both* GEMS subjects and the healthy subjects of the PhenoGenetic project, we found a greater mean GRS among asymptomatic GEMS subjects [mean GRS \pm SD: 9.28 ± 0.88] when compared to healthy control subjects [mean GRS \pm SD: 8.89 ± 0.83] ($p=1.8\times 10^{-13}$) (**Supplementary Table 2³⁹**). However, the mean GRS of asymptomatic GEMS subjects remains smaller than that of MS cases from the Partners MS Center, a large MS clinic in Boston ($p=3.0\times 10^{-6}$) (**Supplementary Table 2^{40,41}**) and that of the GEMS MS subjects ($p=1.5\times 10^{-5}$) (**Figure 2B; Table 4**), consistent with the expectation that the majority of asymptomatic family members will not develop MS. MS subjects from the GEMS cohort are all cases from

multiplex families (i.e., they have a diagnosis of MS in addition to having at least one first-degree relative with MS), whereas MS subjects from the MS clinic are predominantly sporadic cases. Consistent with a prior study of multiplex MS cases³⁰, we observed that MS GEMS subjects have a greater mean GRS ($p=0.01$) when compared to the clinic-based population (**Supplementary Table 2**). This GRS difference appears to be primarily driven by a greater burden of HLA alleles in the MS subjects from the GEMS cohort ($p=0.006$) rather than differences in the 59 non-HLA alleles ($p=0.57$).

When we evaluate the GERS, the summary estimate of all risk factors, we note that MS subjects from the GEMS cohort have a higher mean GERS than asymptomatic subjects ($p=4.8 \times 10^{-6}$) (**Figure 2C; Table 4**). In **Figure 2D**, we illustrate the distribution of risk among all GEMS subjects along both the GRS and ERS dimensions. In this smoothed two-dimensional histogram, each family member is found somewhere along the surface of risk.

Evaluating the Utility of the Genetic and Environmental Risk Score (GERS) for Risk Stratification

The GEMS study is designed to capture a subject population that would be similar to the one that may seek medical attention to evaluate their risk of MS given their family history. Three years after its launch, we can already use a cross-sectional approach to assess the utility of the GERS in stratifying the risk of MS in a high-risk population that all share a family history of MS. Using the subset of GEMS subjects with

MS at study enrollment for comparison, we evaluated whether the GERS can identify a stratum of family members who have the highest risk of developing MS.

The GERS was analyzed as both a continuous and a categorical variable. First, we partitioned the continuous measure into seven groups of risk determined by the distribution in the asymptomatic subjects (control group). Using the mean and standard deviation (SD) of these control subjects, we defined the seven subject groups as 0.25, 0.75 and 1.25 SD from the mean, with the extreme groups either less than or greater than 1.25 SDs from the mean of the controls. Our detailed methods were previously reported⁴². For the analyses of the categorical GERS (or GRS), we used the median group (i.e., group 4) as the reference since it represents the average risk in the study population of first-degree family members. This approach avoids exaggerating the difference in risk that comes from comparing the extreme subsets (e.g., group 1 versus group 7). To calculate the p-value for a linear trend, we treated the groups (1 to 7) as continuous. All GEMS subjects were divided into seven strata using their individual GERS (**Supplementary Table 3**), and we plot the proportion of subjects with an MS diagnosis that is present in each stratum (**Figure 3**). Then, in relation to the median stratum of risk (the reference group 4), we calculated the odds ratio for developing MS in each group.

Overall, in the GEMS cohort of first-degree family members, an increasing GERS is associated with a greater likelihood of a diagnosis of MS ($p\text{-trend}=1.4\times 10^{-5}$), with the proportion of MS subjects rising from 4% in the first stratum to 20% in the seventh stratum. For subjects belonging to the two highest strata, there is a significantly

increased odds of developing MS when compared to subjects in the median (4th) stratum: OR [95% CI] of 2.12 [1.70 – 4.18] for the 6th stratum and 2.63 [1.35 – 5.12] for the 7th stratum (**Supplementary Table 3, Figure 3**). Thus, having a GERS in either of these two top strata places a MS family member in the highest risk category.

The method by which we generate an aggregate measure of risk for MS is based on robust predictive tools⁴³. An important feature of these algorithms is the weighing of each risk factor's contribution to account for the large difference in effect size between genetic and environmental risk factors. The weight of each risk factor is derived from published replication studies^{16,17}, minimizing the over-estimation of effect size that is common in discovery studies. However, these replication studies did not use MS family members, and we therefore made the practical assumption that effect sizes will be similar in family members and the general population. Over time, we will be able to establish these weights in our own cohort of subjects.

Estimating an Initial Incidence Rate of Multiple Sclerosis Among First-Degree Relatives

While the prevalence of MS in first-degree family members has been documented in different populations⁴⁴⁻⁴⁶, the incidence of MS in family members has not been estimated. This information is critical to the design of prospective studies of individuals at risk of MS, and the GEMS study provides the opportunity to assess the incidence rate of MS among first-degree family members.

A subject who is asymptomatic at enrollment to the GEMS study but is subsequently diagnosed with MS after enrollment is defined as an MS converter. Because enrollment occurs on a rolling basis, the follow-up duration for subjects is variable (**Table 1**). We calculated the incidence of MS among the asymptomatic subjects within GEMS cohort by assessing the number of MS converters over person-years of follow-up.

Leveraging responses to a follow-up questionnaire deployed to all GEMS subjects in 2014, we identified four subjects who were diagnosed with MS by their local neurologists after enrolling in the GEMS study out of the initial 1,583 GEMS subjects who were asymptomatic at study enrollment. In all four cases, an MS specialist was able to document a true conversion event that meets a diagnosis of MS by McDonald criteria⁴⁷. Thus, given 3,258 person-years of observation among the 1,583 asymptomatic GEMS subjects considered in our analyses to date, we estimate an incidence rate of 123 cases per 100,000 first-degree family members annually, which is over 30 times greater than the reported incidence of sporadic MS in the United States⁴⁸ or worldwide⁴⁹. With only four converters, we cannot yet provide meaningful statistics in regards to risk factors. Notably, one of these four converters has the highest GERS among all of the 1,583 asymptomatic GEMS subjects.

Given this incidence rate and the calculated odds ratio of MS for different strata among first-degree family members, we can now design a properly powered prospective study of MS onset. Without stratification, we would require 10,000 first-degree family members to capture 62 clinical conversions to an MS diagnosis over 5

years. Limiting an analysis to the two highest GERS strata could double this number. Focusing on a younger population (e.g., <30 years of age) may further increase the incidence in the cohort. Finally, additional subjects are likely to develop asymptomatic lesions that are detectable by neuroimaging over this period of time, permitting well-powered studies of intermediate phenotypes. We estimate that such a study would have >99% power to detect a novel MS risk factor with an effect size similar to that of infectious mononucleosis.

Lessons in designing a Platform to Investigate Risk Factors and Prevention in Multiple Sclerosis

Here, we report the successful launch of a large prospective natural history study of first-degree family members at risk for MS. With its recruitment target of 5000 subjects, GEMS is well powered to investigate the sequence of events leading to MS, and its nationwide scope will allow the evaluation of environmental risk factors. However, it is clear that a larger study would accelerate progress of discovery, improve estimate of the incidence rate, and provide critical information regarding the generalizability of our results if conducted internationally. A large-scale study would also enable us to support clinical trials for primary prevention strategies more effectively. Engagement of the MS research community that includes collaborative efforts in bringing together expertise and resources is critical to the impact of such a larger study both to test the most compelling candidate early interventions and to analyze the vast amount of phenotypic and molecular data that will be generated. Within ethical

guidelines, rapid data sharing prior to publication is becoming standard and should soon show its utility in accelerating the generation of insights into human disease.

As with any prospective study, a major challenge going forward will be subject retention. The GEMS subjects are motivated to participate in this study due to their familial connection, but ongoing engagement through social media and potentially a virtual environment where subjects that choose to do so can interact with one another will be crucial, especially in a study focusing on younger individuals. Efforts to collect blood and stool samples as well as neuroimaging data are ongoing and collection of novel phenotypes using biometric devices and other self-reported instruments are about to begin. These efforts will also benefit from active subject engagement. It is clear that genetics, even when coupled with a rich environmental risk factor history, will probably not be sufficient to support clinical decision-making. Thus, additional information such as neuroimaging and blood measures as well as new forms of outcome measures from self report and wearable devices will contribute to the development of clinical algorithms for quantifying and mitigating the risks of MS. Further, because certain preventive strategies will only be effective at specific stages of the prodromal phase of MS, a clear delineation of each subject's state at a given time will be essential for future study designs.

The GEMS project is therefore the first iteration of a research platform and collaborative resource with which to investigate the onset of MS in high-risk individuals in the future. We have gained important experience, insights and results in this first phase of the study, which is guiding further improvement in the study operations and

prioritization of pilot study designs in the GEMS subject population. This grand rounds forum is a wonderful opportunity for us to share this experience and provide important details of this new resource for the MS community. We encourage investigators to contact us with ideas for collaboration. Overall, we seek to engage colleagues and family members to advance the understanding of MS susceptibility as rapidly as possible and to establish a platform that effectively carry out primary prevention trials. Realizing these goals is essential to allow us to bring individualized prevention to MS.

Limitations and Future Directions

In the design of the GEMS study, we operated under a number of different constraints that lead to limitations. First, issues of privacy limited our ability to systematically contact additional family members and the probands. Thus, some GEMS subjects may not have a first-degree relative with a validated diagnosis of MS, a possibility that approximates the clinical situation in which individuals will seek medical attention due to a concern for MS risk based on family history. However, this number is likely to be very small given that the distribution of genetic risk factors in the asymptomatic GEMS subjects (**Figure 2B**) is very different from that of the general population. There are a number of different reasons that make it difficult to obtain documentation of the family history from every subject, such as the affected family member being deceased, unwilling to participate, or unwilling to communicate with the subject. Having such a requirement would not only limit subject recruitment, sample size, and the utility of the resource, but may also raise ethical considerations with regard

to confidentiality. (Currently, we ask each subject whether “there is at least one first-degree relative with MS”, their total family size, the number of first-degree relatives with known MS diagnosis and their relationships to the subject.) In the future, we plan to expand the study to allow review of medical records of the probands and assess their MS diagnosis. We also plan to factor genetic ancestry information into future version of the risk estimate. Inclusion of as many MS probands as possible would be useful to understand the proportion of GEMS subjects that do not truly have a family member with MS. As noted above, our current genetic data suggests that this number is likely to be small. As we move toward testing primary prevention strategies, understanding (1) the proportion of GEMS subjects that do not have a family member with a validated diagnosis of MS will be important and (2) whether the efforts to obtain such information have a meaningful impact on the study given the availability of genetic and other information.

A second limitation is that the calculation of MS incidence is preliminary, based as it is on the small number of converters identified during a relatively short subject follow-up period. Further, the small number of converters prevents us from adequately adjusting for age or sex.

Third, this comparison between MS and asymptomatic first-degree family members is a cross-sectional analysis using baseline questionnaire data and a moderate number of MS cases. Nonetheless, the number of MS subjects is sufficient to yield statistically significant results in our pre-planned analysis, and the longitudinal

component of the study is ongoing and will more accurately evaluate risk factors as the GEMS study progresses and more MS converters emerge.

Fourth, the GEMS cohort of first-degree family members is not population-based and may thus be subject to selection bias. Overall, a traditional population-based study of high-risk individuals is not feasible in a country such as the United States given the low prevalence of the disease. We appreciate that our sampling method contains some biases inherent in other similar studies. For example, the GEMS cohort has more females than the overall population of first-degree relatives. However, the equal distribution of males to females between MS and asymptomatic GEMS subjects helps control for influence of selection biases.

Finally, in the first phase of the GEMS study, we did not include subjects below 18 years of age. In the future, we plan to include children, as many events critical to MS risk appear to be active in adolescence.

Although the current version of the risk score is not yet clinically deployable, it enables the design of an adequately powered, prospective study of the higher-risk subset of MS family members. This risk algorithm can be updated as additional risk factors are identified and can include interaction terms should well-validated evidence of interaction between two risk factors emerge. Our long-term goal is to leverage the GEMS platform for conducting investigations that map the sequence of events leading from health to disease and for trials of primary prevention of MS in high-risk individuals. Such an ambitious program cannot be accomplished in isolation. It will require contributions from many different groups of investigators that bring different expertise.

Further, the accumulated data may not be sequestered in a closed environment, as limited data access would hamper the role of GEMS as a vehicle to realize individualized disease prevention. Using appropriate safeguards for subject confidentiality, the data will be made available to the community of MS researchers, ensuring that all reasonable ideas can be explored. Such an open concept for human investigations is rapidly evolving in translational research, and the GEMS study represents an exciting opportunity to test this approach in the context of MS.

Accepted Article

Author Contributions: ZX, LBC and PLD contribute to the concept and study design. ZX, CCW, EKO, AVK, SRC, CAM, MC, PAW, AH, SUS, ICMC, TC, HLW, DSR, LBC, PLD contribute to the data acquisition and analysis. ZX, CCW, LBC and PLD contribute to the drafting of the manuscript and figures.

Conflict of Interest Disclosures: None

Acknowledgments: We thank all of study participants for being part of the GEMS study. We thank Professor Alberto Ascherio and Dr. Cassandra Munger (Harvard School of Public Health) for their invaluable advice on study design. Dr. Xia was a recipient of the Clinician Scientist Development Award from the National Multiple Sclerosis Society and the American Academy of Neurology and is supported by NIH K08-NS079493. Dr. De Jager was a Harry Weaver Neuroscience Scholar of the National Multiple Sclerosis Society. The GEMS study is supported by grant RG-5003-A-2 from the National Multiple Sclerosis Society and NIH K08-NS079493, and the intramural Research Program of NINDS.

References

1. Compston A, Coles A. Multiple sclerosis. *Lancet* 2008;372(9648):1502–1517.
2. Okuda DT, Mowry EM, Beheshtian A, et al. Incidental MRI anomalies suggestive of multiple sclerosis: the radiologically isolated syndrome. *Neurology* 2009;72(9):800–805.
3. Engell T. A clinical pathoanatomical study of clinically silent multiple sclerosis. *Acta Neurologica Scandinavica* 1989;79(5):428–430.
4. Dalton CM, Chard DT, Davies GR, et al. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. *Brain* 2004;127(Pt 5):1101–1107.
5. Comi G, Martinelli V, Rodegher M, et al. Effect of glatiramer acetate on conversion to clinically definite multiple sclerosis in patients with clinically isolated syndrome (PreCISe study): a randomised, double-blind, placebo-controlled trial. *Lancet* 2009;374(9700):1503–1511.
6. Kappos L, Freedman MS, Polman CH, et al. Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. *Lancet Neurology* 2009;8(11):987–997.
7. Goodin DS, Reder AT, Ebers GC, et al. Survival in MS: a randomized cohort study 21 years after the start of the pivotal IFN β -1b trial. *Neurology*

- 2012;78(17):1315–1322.
8. Lynch SG, Rose JW, Smoker W, Petajan JH. MRI in familial multiple sclerosis. *Neurology* 1990;40(6):900–903.
 9. Fulton JC, Grossman RI, Mannon LJ, et al. Familial multiple sclerosis: volumetric assessment in clinically symptomatic and asymptomatic individuals. *Multiple Sclerosis* 1999;5(2):74–77.
 10. De Stefano N, Cocco E, Lai M, et al. Imaging brain damage in first-degree relatives of sporadic and familial multiple sclerosis. *Annals of Neurology* 2006;59(4):634–639.
 11. Gabelic T, Ramasamy DP, Weinstock-Guttman B, et al. Prevalence of radiologically isolated syndrome and white matter signal abnormalities in healthy relatives of patients with multiple sclerosis. *American Journal of Neuroradiology* 2014;35:106–112.
 12. Consortium IMSG, Hafler DA, Compston A, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *New England Journal of Medicine* 2007;357(9):851–862.
 13. Bahlo M, Booth DR, Broadley SA, et al. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nature Genetics* 2009;41(7):824–828.
 14. De Jager PL, Jia X, Wang J, et al. Meta-analysis of genome scans and replication

- identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nature Genetics* 2009;41(7):776–782.
5. Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Human Molecular Genetics* 2009;18(4):767–778.
 6. Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFN β -1b and a CCR1-Antagonist, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of Neurology* 2011;70(6):897–912.
 7. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011;476(7359):214–219.
 8. Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics* 2013;45(11):1353–1360.
 9. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Annals of Neurology* 2007;61(6):504–513.
 0. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Annals of Neurology* 2007;61(4):288–299.

1. Ascherio A, Munger KL, Lünemann JD. The initiation and prevention of multiple sclerosis. *Nature Reviews Neurology* 2012;8(11):602–612.
2. Ebers GC. Environmental factors and multiple sclerosis. *Lancet Neurology* 2008;7(3):268–277.
3. Hedstrom AK, Lima Bomfim I, Hillert J, et al. Obesity interacts with infectious mononucleosis in risk of multiple sclerosis. *Eur J Neurol* 2015;22(3):578–e38.
4. Hedström AK, Olsson T, Alfredsson L. High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. *Multiple Sclerosis* 2012;18(9):1334–1336.
5. Hedström AK, Lima Bomfim I, Barcellos L, et al. Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. *Neurology* 2014;82(10):865–872.
6. Langer-Gould A, Brara SM, Beaber BE, Koebnick C. Childhood obesity and risk of pediatric multiple sclerosis and clinically isolated syndrome. *Neurology* 2013;80(6):548–552.
7. Munger KL, Bentzen J, Laursen B, et al. Childhood body mass index and multiple sclerosis risk: a long-term cohort study. *Multiple Sclerosis Journal* 2013;19(10):1323–1329.
8. De Jager PL, Simon KC, Munger KL, et al. Integrating risk factors: HLA-DRB1*1501 and Epstein-Barr virus in multiple sclerosis. *Neurology* 2008;70(13 Pt

- 2):1113–1118.
9. De Jager PL, Chibnik LB, Cui J, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurology* 2009;8(12):1111–1119.
 10. Gourraud P-A, McElroy JP, Caillier SJ, et al. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. *Annals of Neurology* 2011;69(1):65–74.
 1. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 1996;39(7):807–812.
 2. Hernan MA, Zhang SM, Lipworth L, et al. Multiple sclerosis and age at infection with common viruses. *Epidemiology* 2001;12(3):301–306.
 3. Simon K, Schmidt H, Loud S, Ascherio A. Risk factors for multiple sclerosis, neuromyelitis optica and transverse myelitis. *Multiple Sclerosis Journal* 2014;
 4. Wallin MT, Culpepper WJ, Coffman P, et al. The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service. *Brain* 2012;135(6):1778–1785.
 5. Thacker EL, Mirzaei F, Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. *Annals of Neurology* 2006;59(3):499–503.
 6. Hernán MA, Jick SS, Logroscino G, et al. Cigarette smoking and the progression

- of multiple sclerosis. *Brain* 2005;128(Pt 6):1461–1465.
7. Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology* 2009;73(9):696–701.
 8. Hedström AK, Sundqvist E, Bäärnhielm M, et al. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. *Brain* 2011;134(Pt 3):653–664.
 9. Raj T, Rothamel K, Mostafavi S, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 2014;344(6183):519–523.
 0. Xia Z, Secor E, Chibnik LB, et al. Modeling disease severity in multiple sclerosis using electronic health records. *PLoS ONE* 2013;8(11):e78927–e78927.
 1. Xia Z, Chibnik LB, Glanz BI, et al. A putative Alzheimer's disease risk allele in PCK1 influences brain atrophy in multiple sclerosis. *PLoS ONE* 2010;5(11):e14169.
 2. Karlson EW, Chibnik LB, Kraft P, et al. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. *Annals of the Rheumatic Diseases* 2010;69(6):1077–1085.
 3. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97(18):1837–1847.

4. Ebers GC, Sadovnick AD, Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* 1995;377(6545):150–151.
5. Friczka-Nagy Z, Bencsik K, Rajda C, et al. Epidemiology of familial multiple sclerosis in Hungary. *Multiple Sclerosis* 2007;13(2):260–261.
6. Hader WJ, Yee IM. The prevalence of familial multiple sclerosis in saskatoon, Saskatchewan. *Mult Scler Int* 2014;2014(2):545080–7.
7. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 2011;69(2):292–302.
8. Baum HM, Rothschild BB. The incidence and prevalence of reported multiple sclerosis. *Annals of Neurology* 1981;10(5):420–428.
9. Alonso A, Hernán MA. Temporal trends in the incidence of multiple sclerosis: a systematic review. *Neurology* 2008;71(2):129–135.

Figure Legends

Figure 1. The Genes and Environment in Multiple Sclerosis (GEMS) Study: illustrating the nationwide scope of this new resource of subjects and samples.

(A) The diagram summarizes the overall design of the GEMS study. *MS*, multiple sclerosis. (B) The pie chart displays the relative contribution of different subject recruitment strategies and highlights the importance of electronic communications in subject recruitment. *NMSS*, National Multiple Sclerosis Society. (C) The maps show the location of enrolled subjects as of August 2015. Maps were generated with permission using Map Data ©2015 Google, INEGI, via Mapalists™. Subjects have been recruited from each of the 50 states in the United States.

Figure 2. Density distribution of risk scores for the GEMS subjects (first-degree family members with or without multiple sclerosis at the time of study enrollment).

(A) Environmental Risk Score (ERS), (B) Genetic Risk Score (GRS), and (C) the combined Genetic and Environmental Risk Score (GERS). The environmental risk score contains three factors. The genetic risk score contains 64 validated genetic variants associated with MS susceptibility. Please refer to the section entitled

“Determining the Genetic and Environmental Architecture of MS Risk Among First-

Degree Relatives” for details. (D) Two dimensional histogram combining the data from A and B to present the distribution of risk among all GEMS subjects along both the genetic and environmental dimensions simultaneously. Each subject is located somewhere along this surface of risk profile.

Figure 3. Quantile plot presenting the risk of multiple sclerosis in each of the seven strata of GEMS subjects defined by their risk scores. Groups are stratified using the GRS (**A**) and the GERS (**B**). Please refer to the section entitled “*Evaluating the Utility of the Genetic and Environmental Risk Score (GERS) for Risk Stratification*” for details. The bars show the relative frequency distribution of the asymptomatic subjects in the strata. Group 1 is the lowest risk category and Group 7 the highest. The odds ratio (red triangle) for MS susceptibility in each risk group is superimposed in red with the 95% confidence interval (CI) for that estimate (red line). The p-value is calculated as a linear trend across the groups. Odds ratio are determined as the odds of MS in each group divided by the odds of MS in the reference group (Group 4, the median group).

Table 1. Demographic information of the initial GEMS cohort of first-degree family members (n=1696).

Parameter	Percentage of the Study Population
Female	79.2%
Race (Self-reported)	
European descent, non-Hispanic	95.5%
African American	1.7%
Multi-racial	1.8%
Others	1.0%
Born in continental US	96.8%
Existing MS diagnosis at enrollment	6.7%
Smoking status	
Smoking, current	8.1%
Smoking, ever	26.9%
History of infectious mononucleosis	25.0%
	Mean (SD)
Follow-up Duration (years) *	2.1 ± 0.77

* As of June 2014

Table 2. Comparison of the demographic parameters between asymptomatic first-degree family members and those first-degree family members with an existing diagnosis of multiple sclerosis at GEMS study enrollment.

Demographic Parameter	MS n=113	Asymptomatic n=1583	<i>p</i>-value	<i>p</i>-value <i>Adjusted</i> *
Age at enrollment: year, Mean (SD) ¹	39.2 (7.5)	33.8 (8.5)	<0.0001	N/A
European descent, non-Hispanic: n (%)	103 (95%)	1458 (96%)	0.86	N/A
Female: n (%)	96 (85%) ²	1248 (79%)	0.12	0.14
Smoking, current: n (%)	16 (14%)	122 (8%)	0.015	0.01
IM: n (%)	31 (27%)	394 (25%)	0.55	0.39
% first-degree relatives with MS: Mean (SD)	23% (12%)	26% (12%)	0.25	N/A
BMI at age 18: Mean (SD)	23.1 (5.3)	22.7 (4.2)	0.39	N/A

Abbreviation: *IM*, history of infectious mononucleosis; % *first-degree relatives with MS*: percentage of a subject's total number of first-degree relatives with MS; *BMI*, body mass index; *N/A*: not applicable.

* Adjusted for age since MS subjects in the GEMS study have a higher mean age at enrollment than asymptomatic subjects.

Note 1: Age at GEMS study enrollment is not the same as the age at MS diagnosis.

Note 2: Among the GEMS subjects with MS, the female to male ratio is 5.6 to 1.

Accepted Article

Table 3. Detailed comparison of the body mass index (BMI) at 18 years of age (based on recall) between asymptomatic first-degree family members and those first-degree family members with an existing diagnosis of MS at GEMS study enrollment.

BMI at age 18	MS	Asymptomatic	Odds Ratio (95% CI)	Odds Ratio (95%CI) <i>Adjusted *</i>
Categories: N (%)	n=113	n=1583		
< 18.5	9 (8%)	167 (11%)	0.67 (0.32 – 1.42)	0.62 (0.29 – 1.32)
18.5 to < 21	34 (31%)	413 (27%)	1.0	1.0
21 to < 23	27 (25%)	356 (24%)	0.92 (0.55 – 1.54)	0.99 (0.59 – 1.67)
23 to < 25	15 (14%)	260 (17%)	0.72 (0.39 – 1.33)	0.83 (0.44 – 1.55)
25 to < 27	10 (5%)	141 (9%)	0.88 (0.43 – 1.82)	1.04 (0.50 – 2.17)
27 to < 30	5 (5%)	92 (6%)	0.68 (0.26 – 1.77)	0.89 (0.34 – 2.37)
≥ 30	9 (8%)	85 (6%)	1.32 (0.62 – 2.83)	1.74 (0.80 – 3.81)

* Adjusted for age

Accepted Article

Table 4. Comparison of the risk scores between asymptomatic first-degree family members and those first-degree family members with an existing diagnosis of multiple sclerosis at GEMS study enrollment.

Risk Score	MS n=113 Mean (SD)	Asymptomatic n=1583 Mean (SD)	<i>p</i> -value	<i>p</i> -value <i>Adjusted</i> *
ERS	1.3 (0.6)	1.2 (0.6)	0.12	0.096
GRS	10.2 (0.8)	9.9 (0.9)	8.3×10^{-6}	1.5×10^{-5}
GERS	11.6 (1.1)	11.1 (1.1)	3.8×10^{-6}	4.8×10^{-6}

Abbreviation: *ERS*, environmental risk score; *GRS*, genetic risk score; *GERS*, genetic and environmental risk score

* Adjusted for age

Supplementary Table 1. List of single nucleotide polymorphisms and their weights used for calculating the genetic risk score.

SNP	CHR	Gene	Risk Allele	Alt. Allele	Risk Allele Frequency	Weight	% Total GRS	REF
rs10201872	2	<i>SP140</i>	T	C	0.187	0.1391	0.0150	16
rs10466829	12	<i>CLECL1</i>	A	G	0.450	0.0982	0.0106	16
rs1077667	19	<i>TNFSF14</i>	C	T	0.832	0.1519	0.0164	16
rs11129295	3	<i>EOMES</i>	T	C	0.392	0.1078	0.0117	16
rs11154801	6	<i>AHI1</i>	A	C	0.368	0.1197	0.0129	16
rs1132200	3	<i>TMEM39A</i>	C	T	0.872	0.1330	0.0144	17
rs11581062	1	<i>SLC30A7</i>	G	A	0.317	0.1075	0.0116	16
rs12048904	1	<i>EXTL2</i>	T	C	0.354	0.1056	0.0114	16
rs12122721	1	<i>KIF21B</i>	G	A	0.737	0.1081	0.0117	17
rs12212193	6	<i>BACH2</i>	G	A	0.461	0.0889	0.0096	16
rs12368653	12	<i>AGAP2</i>	A	G	0.526	0.0970	0.0105	16
rs12466022	2	NA	C	A	0.739	0.1210	0.0131	16
rs1250550	10	<i>ZMIZ1</i>	A	C	0.329	0.1143	0.0124	16
rs12722489	10	<i>IL2RA</i>	C	T	0.878	0.2184	0.0236	17
rs13192841	6	NA	A	G	0.298	0.1217	0.0132	16
rs13333054	16	<i>IRF8</i>	T	C	0.233	0.1019	0.0110	16
rs1393122	5	<i>Intergenic</i>	A	G	0.857	0.0619	0.0067	17
rs17066096	6	<i>IL22RA2</i>	G	A	0.258	0.1355	0.0146	16
rs170934	3	<i>EOMES</i>	T	C	0.501	0.1360	0.0147	17

rs17174870	2	<i>MERTK</i>	C	T	0.777	0.1037	0.0112	16
rs1738074	6	<i>TAGAP</i>	C	T	0.705	0.1137	0.0123	16
rs17445836	16	<i>IRF8</i>	G	A	0.814	0.1120	0.0121	17
rs17824933	11	<i>CD6</i>	G	C	0.242	0.1314	0.0142	17
rs1790100	12	<i>MPHOSPH9</i>	G	T	0.251	0.1034	0.0112	17
rs1800693	12	<i>TNFRSF1A</i>	C	T	0.428	0.1269	0.0137	17
rs180515	17	<i>RPS6KB1</i>	G	A	0.376	0.1005	0.0109	16
rs2019960	8	<i>PVT1</i>	C	T	0.217	0.1027	0.0111	16
rs2119704	14	<i>GPR65</i>	C	A	0.947	0.2095	0.0226	16
rs2293152	17	<i>STAT3</i>	C	G	0.652	0.1693	0.0183	17
rs2300603	14	<i>BATF</i>	T	C	0.776	0.0933	0.0101	16
rs2300747	1	<i>CD58</i>	A	G	0.903	0.2102	0.0227	17
rs2303759	19	<i>DKKL1</i>	G	T	0.279	0.1186	0.0128	16
rs2395175	6	<i>HLA-DRB1*04:01</i>	G	A	0.909	0.4415	0.0477	PC
rs2744148	16	<i>SOX8</i>	G	A	0.203	0.1107	0.0120	16
rs2760524	1	<i>RGS1</i>	G	A	0.856	0.1237	0.0134	17
rs2762932	20	<i>CYP24A1</i>	C	T	0.153	0.0589	0.0064	16
rs2844821	6	<i>HLA-A*02:01</i>	T	C	0.882	0.3577	0.0387	PC
rs3129889	6	<i>HLA-DRB1*1501</i>	G	A	0.224	1.0639	0.1150	17
rs354033	7	<i>ZNF767</i>	G	A	0.752	0.1080	0.0117	16
rs3817964	6	<i>HLA-DRB1*04:04</i>	A	T	0.051	0.2445	0.0264	PC
rs4285028	3	<i>SLC15A2</i>	A	C	0.740	0.0952	0.0103	16
rs4308217	3	<i>CD86</i>	C	A	0.687	0.0745	0.0080	16
rs4410871	8	<i>MYC</i>	C	T	0.741	0.1018	0.0110	16
rs4613763	5	<i>PTGER4</i>	C	T	0.139	0.1964	0.0212	16
rs4648356	1	<i>MMEL1</i>	C	A	0.700	0.1475	0.0159	16
rs4680534	3	<i>IL12A</i>	C	T	0.362	0.1037	0.0112	17
rs4902647	14	<i>ZFP36L1</i>	C	T	0.540	0.1031	0.0111	16

rs498422	6	HLA-DRB1*14:01	T	G	0.958	0.1847	0.0200	PC
rs6074022	20	CD40	C	T	0.275	0.1329	0.0144	17
rs650258	11	CD5	C	T	0.671	0.1228	0.0133	16
rs6718520	2	Intergenic	A	G	0.459	0.0852	0.0092	17
rs6897932	5	IL7R	C	T	0.770	0.1119	0.0121	17
rs703842	12	METTL1	A	G	0.714	0.1154	0.0125	17
rs7089861	10	IL2RA	C	G	0.882	0.1806	0.0195	17
rs7090512	10	IL2RA	C	T	0.360	0.0843	0.0091	16
rs7200786	16	CLEC16A	A	G	0.487	0.1654	0.0179	16
rs7238078	18	MALT1	T	G	0.789	0.1085	0.0117	16
rs744166	17	STAT3	G	A	0.441	0.1147	0.0124	17
rs763361	18	CD226	T	C	0.494	0.0439	0.0047	17
rs771767	3	NFKBIZ	A	G	0.255	0.0963	0.0104	16
rs802734	6	PTPRK	A	G	0.734	0.1126	0.0122	16
rs8112449	19	TYK2,CDC37	G	A	0.687	0.0894	0.0097	16, PC
rs874628	19	MPV17L2	A	G	0.739	0.1343	0.0145	16
rs9846534	3	CBLB	T	C	0.825	0.0765	0.0083	17

Abbreviation: *SNP*, single nucleotide polymorphism; *CHR*, chromosome; *Risk Allele*, reported risk allele; *Alt. Allele*, alternative allele; *Risk Allele Frequency*, risk allele frequency from all batches of genotyping of GEMS subjects (not including Partners Multiple Sclerosis Center or healthy subjects); *Weight*, weight used in genetic risk score calculation (natural log of the published odds ratio); *% total GRS*, percentage of total genetic risk score; *REF*, reference (see main text); *PC*, personal communication from International Multiple Sclerosis Genetics Consortium.

Note: The list is based on significant and validated association with multiple sclerosis susceptibility from published meta-analysis of genome-wide association studies^{16,17} and is up-to-date as of 2011 when targeted genotyping of GEMS subjects first began. Of note, rs9282641 failed in more than two genotyping batches and rs12708716 has high LD with rs7200786.

Accepted Article

Supplementary Table 2. Comparing the genetic risk score of GEMS subjects (first-degree family members with and without MS diagnosis) with MS patients from the Partners Multiple Sclerosis Center and healthy subjects from the PhenoGenetic Project.

	N	GRS64, Mean (SD)	GRS63, Mean (SD)
GEMS, asymptomatic	1583	9.89 (0.91)	9.28 (0.88)
GEMS, MS	113	10.28 (0.84)	9.62 (0.83)
MS Center, MS ^A	1114	10.05 (0.91)	9.49 (0.89) ^C
Healthy Control ^B	344	N/A	8.89 (0.83) ^D

Abbreviation: *GRS64*, genetic risk score containing 64 MS susceptibility risk variants; *GRS63*, same as *GRS64* except for the absence of rs2844821 (see Note B).

Note A: MS patients from the Partners Multiple Sclerosis Center (Boston) are sporadic cases whereas MS subjects from the GEMS study are multiplex cases (based on study inclusion criterion of having at least one first-degree relative with MS). Demographics of the MS subjects from a clinic-based prospective natural history study at the Partners Multiple Sclerosis Center have been previously reported (75% European descent, median age 40 years, 73% female)^{40,41}. For MS subjects from the Partners MS Center, genotyping data are obtained from previously published genome-wide association studies^{16,17}. Environmental factors are not available from all subjects to calculate GERS.

Note B: Healthy subjects from the PhenoGenetic Project (Brigham and Women's Hospital, Boston) are drawn from the general population of the Greater Boston metropolitan area and are self-reported to be free of chronic infectious, inflammatory, and metabolic diseases. Only subjects of non-Hispanic European ancestry were used for comparison in this report. Genotypes were extracted from genome-wide data generated as part of the ImmVar project³⁹. A previous report describes these subjects in detail (median age 26 years, 58% female)³⁹. Because one SNP (rs2844821) from GRS64 is missing from the existing genotype data for healthy subjects, GRS63 is used for comparison involving healthy subjects.

Note C: Comparing to MS patients from the Partners Multiple Sclerosis Center, asymptomatic GEMS subjects have a smaller mean GRS64 ($p=3.0 \times 10^{-6}$), whereas MS subjects from the GEMS study have a higher mean GRS64 ($p=0.01$).

Note D: Comparing to healthy subjects from the PhenoGenetic Project, both asymptomatic GEMS subjects ($p=1.8 \times 10^{-13}$) and MS subjects from the GEMS study ($p=6.3 \times 10^{-15}$) have a higher mean GRS63.

Supplementary Table 3. Stratifying first-degree family members of the GEMS cohort according to their GRS or GERS and the odds ratio of MS in each stratum when compared to the median group (group 4).

Group *	GRS			GERS		
	MS n (%)	Asymptomatic n (%)	OR	MS n (%)	Asymptomatic n (%)	OR
1	2 (2%)	157 (10%)	0.21 (0.05-0.93)	5 (4%)	171 (11%)	0.57 (0.21-1.59)
2	10 (9%)	212 (13%)	0.79 (0.35-1.76)	8 (7%)	176 (11%)	0.89 (0.37-2.12)
3	19 (16%)	306 (19%)	1.04 (0.53-2.04)	16 (14%)	295 (19%)	1.06 (0.20-2.16)
4	17 (15%)	284 (18%)	1.00	16 (14%)	313 (20%)	1.00
5	26 (23%)	262 (17%)	1.66 (0.88-3.13)	25 (22%)	272 (17%)	1.80 (0.94-3.44)
6	21 (19%)	197 (12%)	1.78 (0.92-3.46)	20 (18%)	185 (12%)	2.12 (1.07-4.18)
7	18 (16%)	165 (10%)	1.82 (0.91-3.63)	23 (20%)	171 (11%)	2.63 (1.35-5.12)
7 vs. 1			8.56 (1.96-37.48)			4.60 (1.71-12.37)
p for trend			5.6×10^{-5}			1.4×10^{-5}

Abbreviation: *GRS*, genetic risk score; *GERS*, genetic and environmental risk score.

* *GRS* and *GERS* are partitioned into seven groups of risk determined by the distribution of the controls (i.e., asymptomatic subjects at enrollment).

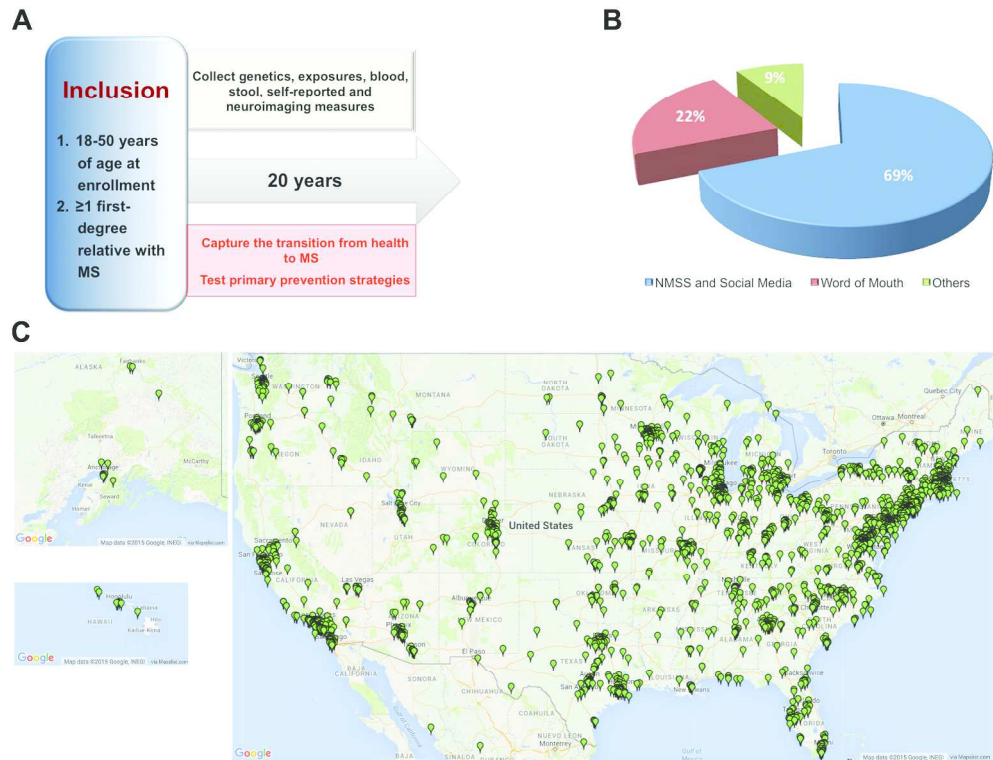


Figure 1. The Genes and Environment in Multiple Sclerosis (GEMS) Study: illustrating the nationwide scope of this new resource of subjects and samples. (A) The diagram summarizes the overall design of the GEMS study. MS, multiple sclerosis. (B) The pie chart displays the relative contribution of different subject recruitment strategies and highlights the importance of electronic communications in subject recruitment. NMSS, National Multiple Sclerosis Society. (C) The maps show the location of enrolled subjects as of August 2015. Maps were generated with permission using Map Data @2015 Google, INEGI, via MapalistTM. Subjects have been recruited from each of the 50 states in the United States. 603x461mm (300 x 300 DPI)

Acc

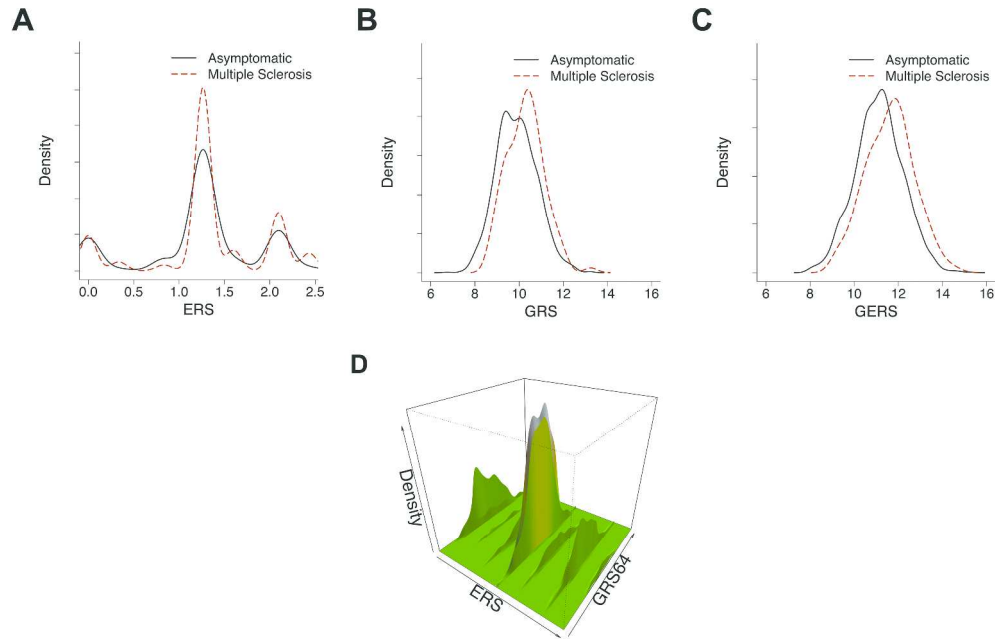


Figure 2. Density distribution of risk scores for the GEMS subjects (first-degree family members with or without multiple sclerosis at the time of study enrollment). (A) Environmental Risk Score (ERS), (B) Genetic Risk Score (GRS), and (C) the combined Genetic and Environmental Risk Score (GERS). The environmental risk score contains three factors. The genetic risk score contains 64 validated genetic variants associated with MS susceptibility. Please refer to the section entitled “Determining the Genetic and Environmental Architecture of MS Risk Among First-Degree Relatives” for details. (D) Two dimensional histogram combining the data from A and B to present the distribution of risk among all GEMS subjects along both the genetic and environmental dimensions simultaneously. Each subject is located somewhere along this surface of risk profile.

554x355mm (300 x 300 DPI)

Acce]e]

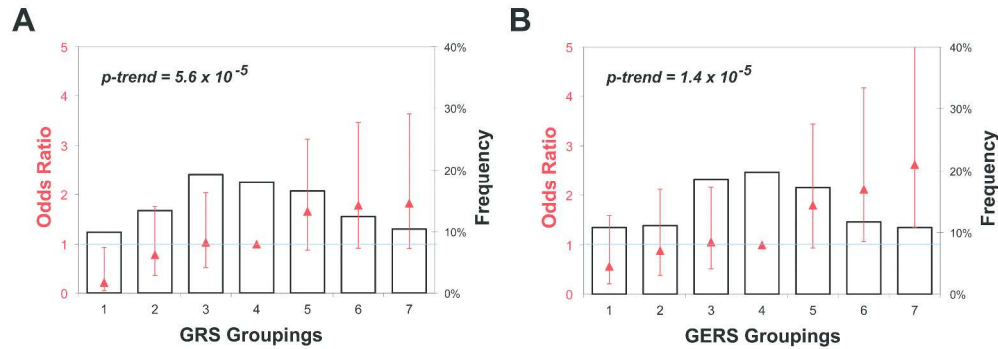


Figure 3. Quantile plot presenting the risk of multiple sclerosis in each of the seven strata of GEMS subjects defined by their risk scores. Groups are stratified using the GRS (A) and the GERS (B). Please refer to the section entitled "Evaluating the Utility of the Genetic and Environmental Risk Score (GERS) for Risk Stratification" for details. The bars show the relative frequency distribution of the asymptomatic subjects in the strata. Group 1 is the lowest risk category and Group 7 the highest. The odds ratio (red triangle) for MS susceptibility in each risk group is superimposed in red with the 95% confidence interval (CI) for that estimate (red line). The p-value is calculated as a linear trend across the groups. Odds ratio are determined as the odds of MS in each group divided by the odds of MS in the reference group (Group 4, the median group).

443x161mm (300 x 300 DPI)

Accepted