Multiple sclerosis (MS) is an inflammatory disorder that causes demyelination and axonal injury within the central nervous system. The underlying etiology remains elusive but both environmental and genetic factors are believed to play a role\textsuperscript{1,2} culminating in the over-activation of various immune subsets that accumulate in the central nervous system to produce injury. Familial inheritance, cigarette smoking, vitamin D deficiency, and ultraviolet (UV) light exposure may all contribute to the risk of MS\textsuperscript{1,3}. While previous studies have attempted to quantify the risk of these factors, largely as a single entity and at a regional scale, no combined quantification has been made of these purported environmental effects in the context of each other, nor has the risk been investigated on a global scale.

In this study, we have collated prevalence data on MS from 54 published studies to obtain a global perspective, and we have captured geographic and demographic information, and population health indicators related to these locations. We then employed various mathematical analyses and numerical methods to examine the purported environmental effects and provide a relative weighting of their risk and relative

contribution to MS in the context of each other, and on a worldwide scale.

We have found a very significant negative correlation between MS prevalence and available ultraviolet radiation, thereby providing quantification of the suspected inverse association of MS with sunlight. Importantly, the lack of available UV radiation outweighs other risk factors by at least 20 fold.

**METHODS AND MATERIALS**

**MS Study Ascertainment**

A search was conducted of the MEDLINE, EMBASE and Cochrane databases for all articles published from November 1998 until October 2008 using PubMed, Google Scholar and the Cochrane website as search engines. No language restrictions were imposed. The following was used as a baseline search schema in PubMed: (multiple sclerosis or MS or optic neuritis) and (incidence or prevalence). Based on the title of the publication and its abstract, the full articles were either downloaded or requested from our medical library. To locate unpublished material and to decrease publication bias, references from original and review articles were manually searched. Some symposia proceedings from large neurology association meetings were also searched.

**Inclusion Criteria**

Abstracts and full articles were retrieved and screened for the following:

- Established diagnosis of MS using the available clinical criteria of the time (eg Poser, McDonald).
- Statement of disease prevalence.
- Reporting of the study population denominator, and sex ratio (or total male and female numbers).

**Exclusion Criteria**

Studies that did not state their method of case ascertainment were excluded.

Fifty-four MS prevalence studies met the above criteria. These are listed in the supplemental references.

**Data Extraction**

A standardized data collection form was used to increase uniformity and reduce bias in reporting. Extracted data included:
- first author, year of publication, journal, sex ratio (female (F), Male (M)), population size, an ordinal definition of the study area size (1-city, 2-region/county, 3-province, 4-country), and the population denominator for the study catchment area (Supplemental Table). The “Methods” designation in the

![Figure 1](https://example.com/figure1.png)

**Figure 1:** MS prevalence correlates with available UV radiation in both hemispheres. a) MS prevalences (red diamond) are plotted vs. latitude for all of the 54 prevalence studies meeting our inclusion criteria (see Methods). The most northern studies (Scandinavian prevalences) are highlighted as blue circles. The available UV radiation (as calculated from NASA TOMS data) has been singly regressed (scaled and shifted) for all included studies, and its inverse (open squares) follows the same trend as the MS prevalence. b) The available UV radiation for each prevalence study is plotted vs. each study’s prevalence; again, the blue circles represent the most northern (Scandinavian) studies, suggesting an outlier group. The regression line is plotted in black. c) The prevalences of all 54 studies are displayed and ordered by latitude, demonstrating that the outlier group has a physical relationship.
Supplemental Table denotes the extent of data sources used to derive the prevalence (1-multiple sources, 2-one source only, 3-questionnaire) and is meant to determine if potential underreporting of cases is related to the number of data sources employed.

The life expectancy of the country (from the World Health Organization (WHO) website – www.who.int) was included both because it may relate the incidence to the prevalence and because socioeconomic factors may alter overall disease reporting. Gross domestic product (GDP) per capita (available from the WHO website) was also included in the analysis in order to account for socioeconomic factors.

For each study, population smoking statistics were abstracted from the World Health Organization website. If smoking statistics were not available from the WHO, a literature search using MEDLINE or EMBASE was used to identify missing population smoking rates.

Year of prevalence date was included in the regression analysis because an overall increase in prevalence has been observed over the last many decades. The sex ratio was also included because an overall increase in female to male ratio has been observed over the past century.

The catchment population size of each study was included in the analysis to account for bias in the ascertainment of cases. We hypothesized that studies on smaller populations may be subject to counting errors and that studies done on larger populations could be subject to under-ascertainment from incomplete availability of records. This population size was included both as the denominator used for the prevalence and as an ordinal representing city, region/county, province, and country. The second representation was included in case institutional differences exist from institutional versus ministerial/governmental sources.

Latitude was included to compare to the significance of the available UV radiation. Longitude was included in the analysis because although a north-south gradient has been observed in many studies over the last several decades, previous analyses for an east-west gradient in MS prevalence are sparse and contradictory. The longitude was represented two ways: as a continuous gradient from east to west, and as a reflected gradient with the zero point set at ten degree increments with Greenwich as the reference point. This incremental system was analyzed to search for a similar gradient as observed with the apparent reflected north-south gradient observed at the equator. The latitude and longitude of the study population was derived from Google Earth site data. The approximate geometric mean of the study population (i.e. the location of most of the study population) was estimated to define the latitude and longitude of the study.

**Determination of available UV radiation**

Total Ozone Mapping Spectrometer (TOMS) is a satellite-mounted optical sensor used to measure the albedo (reflected power) of the earth's atmosphere at six narrow spectral bands, including several ultraviolet B (UVB) bands. Erythemal exposure represents the potential for biological damage due to solar UV radiation. Total Ozone Mapping Spectrometer erythemal exposure is calculated by NASA using UV irradiance

![Figure 2: Single regression analysis of environmental and potential confounding variables. Although some factors are statistically significant, the inverse of available ultra violet (UV radiation) is found to be the most statistically significant variable, outweighing the next significant factor, latitude, by a factor of 20.](https://www.cambridge.org/core/).
reaching the surface of earth at noon weighted by the susceptibility of Caucasian skin to sunburn (erythema) and preferentially weights those frequencies that cause peripheral conversion of vitamin D in the skin.\textsuperscript{11} The TOMS data is available through NASA from the years 1997 to 2002 with a resolution of approximately 85km x 100km.

A computer program (available on request) was developed from standard software provided by the TOMS project to use the daily data for calculating a linearly interpolated measurement of daily noontime erythemal UV for each study location based on latitude and longitude. The average daily UV data over the period from Jan 1, 1997 to Dec 31, 2002 was then calculated.
Very little between-year variation of daily erythemal noontime UV measurements was noted.

Several equations are available to estimate the total sunlight hours for a given latitude and day; one was selected that permitted the specification of day, latitude, and the minimum angle of the sun with respect to the horizon. Given that the amount of available ultraviolet radiation decreases sharply to much less than 10% of the amount available from directly overhead as the zenith angle of the sun increases beyond 60° (the angle to the horizon is less than 30°), 30° was selected as a minimum angle to the horizon for these calculations.

The yearly ultraviolet exposure generally follows a semi-circular or triangular distribution. Therefore, a relative estimate of the total UV availability can be made by multiplying the average daily maximum by the average available sunlight hours throughout the year. This total available UV radiation was calculated for each study location (latitude and longitude were determined using Google Earth). Lack of availability can be estimated by either the negative or the inverse of the available UV radiation.

The total sunlight hours for a horizon angle of greater than 30° was multiplied by the average daily erythemal noontime UV radiation to derive the henceforth named available UV radiation. For plotting the confidence intervals, the mean, minimum and maximum of the interval was normalized to the mean value of the variable of interest to permit comparison between variables of different scales.

To explore extra geographical factors under multiple regression analysis, a two dimensional model of distance to a source was developed. This model repeatedly selects a random latitude and longitude (run 50 times) and iteratively calculates the distance to the random location from each of the 54 study locations. This distance variable is added to the available UVR model to create a double regression model of available UVR and distance to the random location. The algorithm then iteratively minimizes the p value of the double regression model by selecting adjacent locations and comparing p values using a recalculated distance. This algorithm was designed to walk towards a globally minimal p value using an iteratively updated distance in the double regression model.

RESULTS

Global MS Prevalence Trend

A total of 85 candidate studies were found for further examination using the search criteria above (see Methods). Fifty-four studies met our inclusion and exclusion criteria; only four studies were available from the southern hemisphere, reflecting the global geopolitical distribution. The distribution of MS prevalence vs. latitude for the 54 studies is shown in Figure 1A.

Quantification of Risk from Ultraviolet Radiation

Observation of the prevalence trend in Figure 1A is suggestive of an external, environmental contribution to disease prevalence given the mirrored reflection at the equator. Ultraviolet radiation, contained in sunlight, follows a very similar pattern and also demonstrates a change in trend at approximately 23° latitude (north and south). The available UV radiation is an environmental correlate that is a function of both the average maximum daily UV radiation and the number of daylight hours that sunlight is incident upon a location. Multiplication of both of these factors daily and then summing the daily values over the year, provides an estimate of the yearly available UV radiation for a given location.

For visualization, the available UV radiation was singly regressed (ultimately scaled, flipped and shifted) to the prevalence data and plotted with the results of the 54 MS prevalence studies above (Figure 1A). The correlation of MS prevalence to inverse available UV radiation provides a good fit (Figure 1A) and suggests that the latter factor is a dominant correlate of disease prevalence which outweighs the contribution of all other factors examined.

We next used single regression analysis on several environmental and (potentially) confounding factors to search for modulators of MS prevalence. We found that available UV radiation produced the most significant correlation with prevalence (Figure 2), and this out ranks latitude by a factor of 20, both in terms of contribution to the fit of the model and statistical significance. All other factors are small contributors compared with the available UV radiation.

Plotting the available UV radiation against prevalence (Figure 1B), a separable group of studies is apparent from the rest of the trend (represented by the closed red circles). A standard, unbiased estimator of clusters, the k-means algorithm (Matlab), was used to confirm that this is a significantly separable group in the two dimensional space of available UV radiation vs. prevalence. This group of studies represents all of the Scandinavian studies and a study of the Faroe Islands (eight studies total) in the 54 studies used. They are also the most northern of all the prevalence studies, they are geographically proximate, and are therefore of interest. These studies were removed for a separate single regression analysis of available UV radiation and the correlation was increased to an R² of 0.54 (p value of 7.2x10⁻⁶) for the remaining 46 studies. We plotted the prevalence of all studies ranked by latitude and the cluster of studies are the northernmost of all the studies (as expected).

For the single regression model of MS prevalence and available UV radiation, the final residuals with their confidence intervals were analyzed. Three of the 54 studies had confidence intervals outside of the expected range: Northern Ireland and Linguaglossa, Sicily had prevalences higher than expected and Belgrade had a prevalence lower than expected. Therefore, the prevalence for approximately 95% of the MS prevalence studies within the past ten years could be predicted reliably by this single regression model. Such a model that factors available UV radiation could be useful to health ministries both for estimating their MS patient population and for allocating appropriate resources for a given region.

Multiple Regression Analysis

To explore possible combined effects of the singly-regressed variables, multiple regression analysis was performed using the stepwise regression toolkit in Matlab. As expected, the most significant variable, available UV radiation, was included at a p value of 1.0x10⁻⁶. The only other two variables added into the model were latitude and longitude (to a p value of 9.2x10⁻⁶) – all other variables were non-significant after this initial three-variable model was developed. This suggested that further geographical correlation was inherent in the dataset and that it
involved both latitude and longitude components. For 50 separate analyses, a negative correlation was found to minimize the p value (and maximize the R^2 value) of the double regression model (available UV radiation and the extra geographical factor) with the location 54.8°N, 8.1°E (a location centered over Northern Ireland, which is geographically reasonably close to the latitudes of the Faroe Islands and of Scandinavia). This extra geographical factor confirms again that the prevalence from the Scandinavian studies significantly influences any simple modeling of the available UV radiation to MS prevalence. This minimal p value (2.0x10^-8) in the double regression model (including both the available UV radiation and the extra geographical factor) is less than the model with available UV radiation, latitude and longitude (9.2x10^-8, as above).

Further regression analysis in addition to the available UV radiation and the distance to 54.8°N, 8.1°E included two marginally (negatively) significant variables: methodology and smoking. A p value of 5.4x10^-9 (R^2=0.55) was achieved with the inclusion of all four of these variables – a very strong negative correlation was noted with the available UV radiation. Additional negative correlation was achieved with the distance from northern Europe, smoking, and the accuracy of the methods.

**DISCUSSION**

It was observed many decades ago that MS prevalence follows a north-south, equatorifugal (reflected at the equator) gradient, modeled using strictly latitude as the modifying environmental parameter. In order to characterize this distribution, 54 prevalence studies from the previous ten years were analyzed. This global approach permits quantification of the correlation between environmental factors, such as the available UV radiation, and worldwide MS prevalence trends, and yields a highly correlated result between these two factors. This relationship between available UV radiation and MS prevalence is more correlated than with latitude and MS prevalence given the shape of the curve – there is an abrupt change in the curve of both the MS prevalence and the regressed available UV radiation at approximately 23° latitude, coinciding with the Tropics of Cancer and Capricorn which are the most northern and southern extents respectively that the sun may appear directly overhead (zenith angle of zero) during a year. Therefore, this novel approach has added further evidence towards a link between MS prevalence and available UV radiation.

The inverse association of available UV radiation and MS prevalence implicates vitamin D, since ultraviolet B radiation (280 to 315 nm) converts 7-dehydrocholesterol to vitamin D3 in the epidermal and dermal layers and is the primary source of vitamin D3 in humans16. Vitamin D deficiency has previously17 and recently been suggested as a potential contributing factor in the pathogenesis of MS18-20. Due to the changing angle of declination of the sun, vitamin D insufficiency is common in the winter months in latitudes north of 42°N latitude21. Therefore, vitamin D is of interest as the biological correlate of available UV radiation.

Several comments are appropriate regarding other observations made with the regression analysis. The clustering of Scandinavian studies that measured a lower prevalence than the available UV radiation may be from a plateau effect above the Arctic Circle. An alternative explanation could be that the model of available sunlight hours is less accurate above certain latitudes22. A third explanation may be genetic/environmental interaction23-24. It has been hypothesized that the distribution of MS is related to a genetic trait originating from the Scandinavian countries with a declining genetic influence25. This cluster could also reflect chance alone.

Several sources have observed an increase in MS prevalence over the past few decades including one of our own studies7, but was non-significant in the current analysis (Figure 2). Possibly, MS prevalence has not increased significantly over the past decade given the generally increased availability of diagnostic tools and neurologists. The fact that our prevalence study noted a plateauing of incidence since the late 1990s may be relevant in this regard. The sex ratio is a non-significant contributor (shown to have increased over the past several decades in our and other studies) but this variation is likely too small to be significantly measurable over the single decade of this analysis.

In our single regression analyses, we found that population smoking rate is a minor contributor to risk (Figure 2). This supports the contention that while some meta-analyses of case control studies at the clinical level have implicated smoking26, the combined increased risk is low (odds ratio of 1.25) compared to available UV radiation which accounts for a 20 fold increase of risk from the equatorial regions to 60°N latitude (Figure 1A). These meta-analysis data are from clinical studies. As far as we could find, no correlation with MS prevalence and smoking has been analyzed at the population level as we have done. We have shown that either there is no effect when other factors are accounted for, or if there is an effect it is small. This is in keeping with the meta-analysis of the clinical data that shows only a minor increased risk of 25% with the small number of available studies. This does not rule out the possibility that in some individuals, there may be a biological effect.

Others have studied correlations with multiple purported causative agents23-27-30. However, the available UV radiation appears to be by far the single most important correlate of MS prevalence globally, suggesting that this factor may be one of the most important etiologic determinants of MS. Nonetheless, although this model estimates a general population trend with reasonable accuracy, it does not explain why certain individuals are more susceptible than others. Other factors such as genetic susceptibility on a population level may play a role. Additionally, mitigating factors such as skin pigmentation and use of sunscreen may alter serum vitamin D levels, although population studies have shown that vitamin D levels are correlated to both latitude and season through peripheral conversion21. Cultural modulations of serum vitamin D are also possible; countries with significant fortification of food or with diets containing higher vitamin D content may also change average population-based serum vitamin D concentrations.

**CONCLUSION**

Quantitative analyses of global epidemiologic and prevalence data collated from 54 studies emphasize that lack of available ambient UV radiation is the most significant environmental factor affecting the prevalence of MS. A low level of incident UV radiation outweighs other suspected environmental factors.
by at least 20 fold as a contributor to MS risk. Available UV radiation and vitamin D as key determinants of reducing the risks of developing MS.

REFERENCES


SUPPLEMENTAL REFERENCES


