Video Article

A Method of Trigonometric Modelling of Seasonal Variation Demonstrated with Multiple Sclerosis Relapse Data

Tim Spelman[1,2], Orla Gray[3], Robyn Lucas[4], Helmut Butzkueven[1,2]

[1]Department of Neurology, Royal Melbourne Hospital

[2]Department of Medicine (RMH), The University of Melbourne

[3]Department of Neurology, Ulster Hospital

[4]National Centre for Epidemiology and Population Health, Australian National University

Correspondence to: Tim Spelman at tim@burnet.edu.au

URL: http://www.jove.com/video/53169
DOI: doi:10.3791/53169

Keywords: Medicine, Issue 106, Multiple Sclerosis, relapse, residual plots, trigonometric regression, sine regression, seasonality, latitude

Date Published: 12/9/2015

Citation: Spelman, T., Gray, O., Lucas, R., Butzkueven, H. A Method of Trigonometric Modelling of Seasonal Variation Demonstrated with Multiple Sclerosis Relapse Data. *J. Vis. Exp.* (106), e53169, doi:10.3791/53169 (2015).

Abstract

This report describes a novel Stata-based application of trigonometric regression modelling to 55 years of multiple sclerosis relapse data from 46 clinical centers across 20 countries located in both hemispheres. Central to the success of this method was the strategic use of plot analysis to guide and corroborate the statistical regression modelling. Initial plot analysis was necessary for establishing realistic hypotheses regarding the presence and structural form of seasonal and latitudinal influences on relapse probability and then testing the performance of the resultant models. Trigonometric regression was then necessary to quantify these relationships, adjust for important confounders and provide a measure of certainty as to how plausible these associations were. Synchronization of graphing techniques with regression modelling permitted a systematic refinement of models until best-fit convergence was achieved, enabling novel inferences to be made regarding the independent influence of both season and latitude in predicting relapse onset timing in MS. These methods have the potential for application across other complex disease and epidemiological phenomena suspected or known to vary systematically with season and/or geographic location.

Video Link

The video component of this article can be found at http://www.jove.com/video/53169/

Introduction

The most common form of multiple sclerosis (MS) is Relapsing Remitting Multiple Sclerosis (RRMS). RRMS is characterized by episodic deteriorations in neurological function, followed by partial or complete recovery. Globally, the incidence and prevalence of MS increase with increasing distance away from the equator in both hemispheres.[1-3] Whether the frequency of relapse events that occur specifically in RRMS also vary with latitude, and whether there is any underlying seasonal variation in any such association, remains unclear. To date studies exploring seasonality in relapse timing have been limited to single clinical centers, limiting any inferences regarding seasonal trends in relapse timing to solitary geographical locations and thus unable to explore broader latitudinal influences. [4-14] These studies have been further limited by small sample sizes and sparse relapse data. A 2000 meta-analysis of ten studies from clinical centres in Europe, the United States and Canada, where each study included a minimum of thirty cases reporting the season-of-onset of relapses, described a clear seasonal trend in the timing of relapse onset, with relapses peaking in spring and with a winter trough[4]. Similar cyclical annual trends in onset have been observed in subsequent, albeit smaller, studies in both Japan[15] and Spain[16]. However, a comparable United States study failed to corroborate this pattern[17]. To date, these studies and observations have been limited to the northern hemisphere. The MSBase study group recently analyzed a large global dataset of MS relapses across both northern and southern hemispheres to explore seasonal trends in the timing of relapse onset in addition to latitudinal influences on the relationship between peak relapse probability and seasonal ultraviolet radiation (UVR) trough[18]. Central to these methods was the application of trigonometric regression to visualize and evaluate trends in the timing of relapse onset and UVR distributions.

The overall goal of this study was to test the hypothesis that temporal variation in the timing of relapse onset in MS varied predictably with season in both the northern and southern hemispheres and this seasonality was influenced by latitude. The rationale for the use of trigonometric modelling to investigate these questions was its flexibility for characterizing two- or three-dimensional phenomena that are known or suspected to describe discrete, predictable and consistent shapes or patterns, such as the annual cycle of peaks and troughs commonly observed in biological or epidemiological phenomena possessing seasonality.[19-22] A disadvantage of conventional time-series analyses, including Fourier analysis, is the presumption that time series are often characterized by stochastic processes.[21,23,24] By contrast, incorporating trigonometric functions into a regression type model has the advantage of both facilitating exploration of regular and systematic structures in periodic data whilst exploiting the regression model structure to explore other correlates or adjust for confounders of seasonality.

Copyright © 2015 Journal of Visualized Experiments

Trigonometric regression has previously been used widely in the medical epidemiological literature to explore temporality in topics as diverse infectious disease outbreak detection, the role of circadian rhythms in everything from autonomic nervous system dysfunction to preterm placental abruption through to seasonal correlates of congenital malformations and the timing of presentations of accident and emergency.[25-32] Such modelling typically demands larger sample sizes than more conventional time-series analyses and as such this is the first time it has been applied to a global dataset of MS relapse onset. Trigonometric regression as described here is suitable tool for investigators exploring any phenomena which is known to or suspected of cycling systematically over time. Not only can such modelling help characterise and visualize these patterns, it further permits the user to explore potential drivers and correlates of these trends.

Regarding the specific example of MS relapse onset presented here, the use of scatter and residual plots to visualize and assess how closely a hypothesized trigonometric model form fits the data constitutes the critical step in determining: 1) whether the observed data provide sufficient evidence to support a hypothesis of seasonality or other temporal trends in the timing of relapse onset; and 2) whether the frequency and arrangement of sine and cosine functions which define a particular trigonometric model is adequate to permit use of this model for subsequent inference and prediction. Regression modelling also permits control for important confounders of any observed seasonal or latitudinal effect such as patient-level propensities for relapse, particularly factors which in themselves are time-varying such as the duration of pre-relapse exposure to disease-modifying drug (DMD) treatment. Isolating independent geographic and temporal predictors and correlates of relapse onset timing in MS has the potential to guide biological investigation into the mechanisms of relapse events which in turn may inform the development of future treatment interventions aimed at preventing or delaying disease exacerbation.

The MSBase Registry

MS patients contributing relapse data to this analysis were sourced from the international MSBase registry. Established in 2004, the registry longitudinally collates demographic, disease activity, clinical examination and investigation characteristics and metrics from consenting patients attending MS clinic using an internet-based, physician-owned and operated system.[33] Member centers follow a common protocol that defines the minimum dataset required to be uploaded at agreed regular intervals to ensure outcome data such as relapse events are consistently and prospectively compiled. The date of relapse onset is included as a mandatory minimum dataset variable. In addition relevant clinical data associated with these relapse events is commonly collected including corticosteroid treatment and functional system affected. The use of the common iMed data entry system further ensures a unified approach across centers to data collection and reporting. This project holds Human Research Ethics Committee approval or exemption at each contributing center. Informed consent according to local laws from all patients included in the analysis is mandatory.

Inclusion criteria

A total of 9811 patients contributing 32,762 relapse events were included in the analysis. Clinical MS centres with a minimum of 20 registered patients consented, uploaded and tracked in the registry as of the 1st December 2013 (date of data compilation) were eligible for inclusion in the analysis. To ensure all relapse events included in the analysis were prospectively observed, only relapse onsets dated subsequent to the first recorded patient disability assessment (using the Kurtzke Expanded Disability Status Score (EDSS)) were included in the analysis. All patients contributing relapse data to the analysis satisfied formal diagnostic criteria for MS.[34,35]

Outcome measures

This study considered two primary outcomes: 1) whether there was temporal variation in the probability of relapse onset at the level of the geographic location, the hemisphere and/or globally; and 2) whether there was a relationship between latitude and the lag, in months, between the timing of seasonal UVR trough and the subsequent peak relapse probability date. The MSBase Study group hypothesized that as absolute vitamin D levels are lower in regions further away from the equator and location-specific seasonal population level vitamin D nadirs are likely reached earlier following the winter solstice in such distal locations, then the effect of low vitamin D levels on increased MS relapse probability would similarly describe such temporal and latitudinal patterns.

Relapse definition and dates

A relapse was defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 hours, in the absence of concurrent illness or fever, and occurring at least 30 days after a previous attack. This definition has previously been applied in an MSBase relapse phenotype analysis. [36] The follow-up period for each eligible patient across which relapse events could be observed was defined as the period spanning the date of first EDSS assessment through to the date of the most recent EDSS assessment recorded in the registry prior to the data of data extract and compilation. In instances where the exact day of relapse onset was unavailable or unable to be determined for a particular month, clinics used either the 1st or 15th day of the month as default dates. Of the 32,762 relapses analysed in this report, 7913 (24.2%) and 4594 (14.0%) were recorded on the 1st and 15th day of the month respectively, significantly higher than the proportions recorded on any other day of the month which ranged from 0.8% through 5.6%. To correct for this, relapses recorded on either the 1st of 15th day of the month were randomized to a day within a 15 day interval either side of both these default dates. The internal validity of this approach was confirmed via sensitivity analyses which demonstrated that the modelled estimate of peak relapse date under default date randomization was not significantly different from a model using either the original reported dates or excluding default dates entirely.

Protocol

NOTE: Each step described corresponds to a section of Stata code with the same number in the code file provided. Stata command names have been italicised in the following protocol.

1. Prepare and Plot the Observed Relapse Onset Data

1. Open a do file by clicking on the "New Do-file Editor" button and use the *generate* command to calculate the number of relapse onsets dated to each of the twelve calendar months for each of the three geographic levels to be modelled: location, hemisphere and global. Action command by clicking the "Execute (do)" do-file action button in the do-file.
2. Use the *swilk orsktest* command to test the underlying distribution of relapse counts for normality using a Shapiro-Wilk or modified Jarque-Bera test for location aggregate relapse data or disaggregate patient level relapse data respectively. Select code and click "Execute (do)".
   NOTE: In the presence of significant skew, apply a natural log transformation and subsequently test the log-transformed relapse count variable for approximate normality by reapplying the Shapiro-Wilk or modified Jarque-Bera test as appropriate.[38,39]
3. Use the *generate* command to create a new variable "north_month" for southern hemisphere calendar months offset by +6 to allow plotting of both northern and southern hemisphere relapses by season along the same horizontal axis. Select code and click "Execute (do)".
   1. Graph a scatterplot of observed monthly relapse onsets with relapse frequency on the y-axis and calendar month on the x-axis for each hemisphere using the *twoway scatter* command. Repeat for each location. Observe pattern of peaks and troughs in relapse onset over the calendar year by viewing each plot in the graph viewer the automatically opens to screen.

4. Use the *radar* command to draw radar plots of the distribution of relapse frequency by calendar month with each radar axis capturing a single month ordered in a clockwise manner. Select code and click "Execute (do)".
   1. Repeat for all sites. Observe pattern of peaks and troughs in relapse onset over the calendar year by viewing each plot in the graph viewer the automatically opens to screen.

5. Run the *seast* command to apply an Edward's test of seasonality across the observed relapse data.[40-42] Repeat for all geographic levels.

2. Model Building and Selection

1. Use the *generate* command to specify the annual cycle sine and cosine trigonometric functions to be used in the regression. Select code and click "Execute (do)".
2. Use the *regress* command specify the form of the base model with relapse count as the dependent outcome variable and the sine and cosine terms calculated in step 2.1 as the primary explanatory variables.
   1. Add location-specific UVR[37] to the base model as an additional adjusting covariate and use the analytic weight *aweight* option to weight the model for the number of patients contributed by each location. Select code and click "Execute (do)".
      NOTE: Record the model coefficient of determination (R^2) and the residual error in the results window that automatically opens to screen. Ultraviolet radiation: Daily average erythemally-weighted ambient UVR for each month from 1979 to 2004 inclusive was sourced from the National Aeronautics and Space Administration Earth Probe Total Ozone Mapping Spectrometer for all individual locations included in the analysis. [37]

3. Store the model predicted monthly log(relapse) using the *predict* command. Convert log relapses back to integer relapse counts by exponentiating the log(relapse) term using the *generate* command. Select code and click "Execute (do)". Repeat for all sites.
4. Overlay the exponentiated predicted monthly relapse estimates from 2.3 over the observed monthly relapse data using the *twoway scatter* command. Select code and click "Execute (do)".
   1. Repeat for all sites. View each plot in the graph viewer.

5. Use the *regress* command to expand the model specified in 2.2 by adding an additional harmonic sine/cosine pair. Select code and click "Execute (do)".
   NOTE: Record the residual error and the coefficient of determination. Save and transform model estimates as per 2.3 and plot model estimates over observed data as per 2.4. Repeat for all sites.
6. Use the *regress* command to further expand the model specified in 2.2 by adding two additional harmonic sine/cosine pairs. Select code and click "Execute (do)".
   NOTE: Record the residuals and the coefficient of determination. Compare this model directly with the base model using a likelihood ratio test. Use the estat ic post-estimation command to generate Akaike and Bayesian Information Criteria. Save and transform model estimates as per 2.3 and plot model estimates over observed data as per 2.4. Repeat for all geographic levels.

3. Estimating Peak Relapse Probability

1. Use the non-linear combination of estimators function (*nlcom*) to calculate the point estimate and 95% confidence interval for the phase-shift, using the best-fitting model identified from steps 2.1 through 2.6. Select code and click "Execute (do)".
   1. Convert these point estimates and associated confidence intervals to numbers representing calendar dates of peak relapse frequency (T_{max}) and trough relapse frequency (T_{min}) , where 1=1st January and 365=31st December and T_{max} = phase-shift + (365/4) and T_{min} = phase-shift + ((365/4)*3). Repeat for all geographic levels. Match T_{max} and T_{min} to a calendar date via the Excel look-up file.

2. Use the *generate* command to calculate peak-to-trough difference (T_{max} minus T_{min}) for each location, standardized for every 100 patients per site. Use a Wilcoxon rank-sum test to compare standardized peak-to-trough difference by latitude range. Select code and click "Execute (do)".

4. Modelling Ultraviolet Radiation Data

1. Run *use* command to load the UVR data. Calculate median monthly UVR for each location using the *egen* command. Select code and click "Execute (do)".

2. Graph a scatterplot of monthly UVR (y-axis) by calendar month (x-axis) for each location using the *twoway scatter* function. View each plot in the graph viewer the automatically opens to screen.
3. Repeat step 1.2 for the UVR data and use the *regress* command to specify a base model of location-level annual UVR trend where monthly UVR is specified as the dependent outcome variables and the sine and cosine trigonometric functions specified in step 2.1 are incorporated into the model as the explanatory variables.
4. Repeat steps 2.4 through 2.6 for the UVR model and limited to the location-specific models only. This involves rerunning the *twoway scatter* command to overlay predicted estimates on observed data and using the *regress* command to run the expanded harmonic model alternatives.
5. Using the best-fitting model of location-specific monthly UVR identified in steps 4.2 through 4.4 use the *generate* command to calculate the phase-shift point estimate and associated 95% confidence interval for UVR by again applying the double-angle formulae specified in step 3.1. Calculate T_{min} (date of trough UVR) for each location using the formula described in step 3.1. Select code and click "Execute (do)".

5. Modelling UVR-trough-to-relapse-peak Lag

1. Append the model-estimated dates of seasonal UVR trough from step 4.5 and relapse peak dates from step 3.1 for each location using the *merge* command. Use the *generate* command to calculate the time lapsed in months between UVR trough date and subsequent relapse peak date. Select code and click "Execute (do)".
2. Use the *sktest* command to test the UVR-trough-to-relapse-peak lag variable for significant departures from normality using a Shapiro-Wilk test Select code and click "Execute (do)".
3. Append location-level latitude data to dataset using the *merge* command. Convert relative latitude to absolute latitude using the *abs(x)* function. Select code and click "Execute (do)".
4. Using the *regress* command, test the linearity of the relationship between lag and absolute latitude by running both linear and quadratic regressions and comparing residuals. Select code and click "Execute (do)".
5. Using *regress*, specify a linear mean regression model with UVR-trough-to-relapse-peak lag as the dependent outcome variable and absolute latitude in units of 10 degrees as the predictor variable. Weight the model for the number of patients contributed by each location using the *aweights* regress option. Select code and click "Execute (do)".
6. Use the *twoway scatter command to plot* absolute latitude on the y-axis against UVR-trough-to-relapse-lag in months on the x-axis. Overlay a line of best fit using *lfit* graph option. Visualise the relative patient weights of each location using the *aweight* analytical weights option. Select code and click "Execute (do)".

6. Sensitivity Analyses of Patient-level Relapse Propensities

1. Use the *mepossion* command to specify a mixed-effects Poisson regression where monthly relapse count is the dependent outcome variable, the sine and cosine trigonometric functions specified in step 2.1 are again incorporated into the model as the fixed variables, baseline EDSS, age at MS onset and prior exposure to MS specific disease-modifying treatment are included as potential confounders and unique patient identifier is specified as a random effect. Select code and click "Execute (do)".
2. Repeat steps 2.4 through 2.6 to identify the best-fitting Poisson model. This involves rerunning the *twoway scatter* command to overlay predicted estimates on observed data and using the *regress* command to run the expanded harmonic model alternatives.
3. Use the non-linear combination of estimators function (*nlcom*) to calculate the point estimate and 95% confidence interval for the phase-shift and calculate the date of peak relapse frequency. Compare results with the primary analysis.
4. Use the *generate* command to recalculate UVR-trough-to-relapse-peak lag in months for each location as described in step 5.1, using the patient-level Poisson model estimates of peak relapse date derived in step 6.3. Select code and click "Execute (do)".
5. Use the *regress* command to remodel absolute latitude as a predictor of lag as described in step 5.5 and compare results with the primary analysis. Select code and click "Execute (do)".

Representative Results

The application of trigonometric regression to 32,762 relapse events sourced from 46 clinical centers across 20 countries was the basis for providing a defensible statistical argument for the observation that the timing of relapse onset in MS is cyclic and seasonal across both hemispheres and that the duration between seasonal UVR trough and subsequent relapse peak correlates with latitude. Critical to this was the reliance upon plot analysis to guide the necessarily iterative process of model development, evaluation and refinement.

Analysis of relapse frequency by calendar month scatter plots of the observed data suggested an annual cycle with a spring peak and autumn trough across all geographical levels. Radar plots of the global relapse data confirmed northern hemispheric relapses peaked in May **(Figure 1A)**. This spring peak persisted when southern hemisphere relapse onset data was combined with the northern data **(Figure 1B),** with southern locations demonstrating a November peak. Autumnal troughs were also recorded in both hemispheres with the lowest frequency of northern and southern hemispheric relapses observed in November and May respectively. An Edward's test further confirmed that relapse onset demonstrated significant departures from a uniform, non-seasonal distribution. Taken together, these results suggested that the periodic temporal variation observed in MS relapse onset at all three levels of geography best described a single annual cycle consisting of a single peak and a single trough separated by a regular six month interval. Thus a trigonometric regression model specified with a single pair of sine and a cosine functions was selected as the base case model across both hemispheres (**Figure 2**). When compared to competing trigonometric model solutions expanded to include two or three period harmonics, the base model across northern hemispheric locations minimised the residual square error and returned a superior fit of the observed data ($p<0.0001$, adjusted $R^2 = 0.263$) when compared to either a model incorporating an additional harmonic ($p=0.0001$, adjusted $R^2 = 0.198$) or an additional two harmonics ($p=0.0014$, adjusted $R^2 = 0.181$). Similarly the same base model out-performed the extended-harmonic alternatives when applied to the southern hemisphere with the base model ($p<0.0001$, adjusted $R^2 = 0.241$) again minimizing the residual differences between the observed and estimated data relative to the model incorporating two additional harmonics ($p<0.0001$, adjusted $R^2 = 0.167$); the one-additional harmonic model described a similar fit relative to the base ($p<0.0001$, adjusted R^2

= 0.243). Importantly for the modelling of location-specific latitude as a predictor of UVR-trough-to-relapse-peak lag, the base model again out-performed either of the extended-harmonic models at the level of the individual geographical locations.

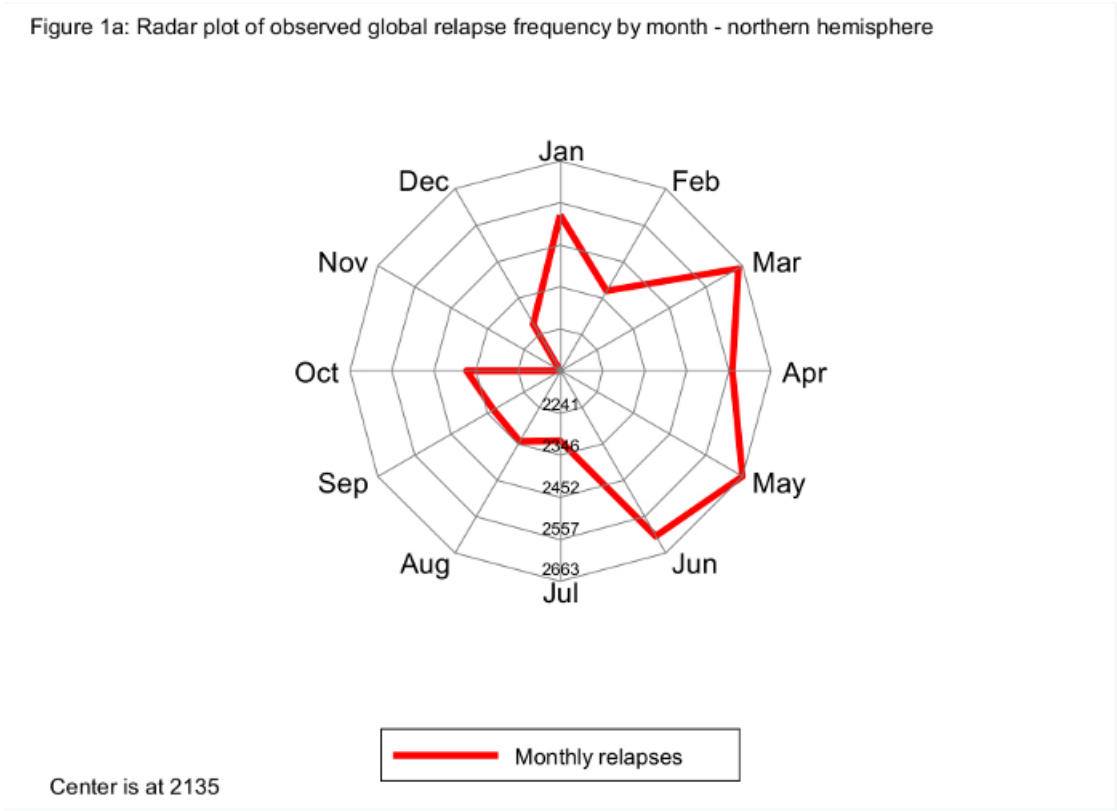
Using the base model specified on a single sine/cosine pair, the phase-shift across all relapses globally was estimated at -24.8 (95% CI -45.8, -3.9), translating into an estimated northern hemispheric peak relapse onset date of the 7th March (95% CI: 10th February, 28th March) and a southern hemispheric peak date of the 5th September (95% CI: 10th August, 26th September). There was no difference in the phase-shift estimate by hemisphere (test of interaction: p=0.254). Mean (SD) standardized peak-to-trough relapse difference was 7.6 (6.6) relapses per 100 patients. Although centers located at an absolute latitude of 40 degrees or more recorded a larger peak-to-trough difference (mean 8.6, SD 7.6) relative to sites located within an absolute latitude range of 20 through 39 degrees (mean 5.7, SD 3.3), this difference was not statistically significant (p=0.135).

Scatterplot analysis of UVR by calendar month suggested that the base model defined on a single sine/cosine harmonic pair as described above was similarly appropriate for UVR seasonality, at all geographical levels. As illustration, **Figure 3** depicts the regression modelled monthly UVR estimates overlaid on the observed UVR data for four selected individual locations, two from each hemisphere. What can be appreciated from these plots is just how closely the modelled estimates, based on an annual cycle single peak and trough sine regression, confers to the observed data. The base UVR model again outperformed either of the extended-harmonic models in terms of minimizing residuals and a superior coefficient of determination.

Overlaying the cyclical UVR sinusoid curve over the equivalent curve for relapse onset suggested that trough UVR consistently preceded peak relapse onset probability. Furthermore this lag appeared to shrink the further north or south a particular location was sited away from the equator. Applying a linear regression of the mean, every 10 degrees of latitude away from the equator in either hemisphere was associated with a statistically significant decrease in this lag of 28.5 days in the UVR-trough-to-relapse-peak lag (95% CI: 3.29, 53.7; p=0.028). As **Figure 4** demonstrates, as absolute latitude increased away from the equator in both hemispheres, the sooner relapses peaked following the winter UVR trough. There was no difference in this association by hemisphere (test of interaction p=0.811).

The patient-level mixed effects Poisson extension of the primary trigonometric sine regression returned very similar results with a peak relapse date estimated at just two days later than that estimated by the primary base model (9th March compared with the 7th March for northern hemisphere locations, 7th September versus 5th September for southern locations). Similarly the UVR-trough-to-relapse peak lag was comparable under either the primary or sensitivity models, with the patient-level Poisson extension demonstrating a mean only 4.1 days different in lag (mean lag =24.8 days, 95% CI 2.0, 49.2) relative to the primary location-level model. Again, there was no difference in this association by hemisphere (test of interaction, p=0.671).

Copyright © 2015 Journal of Visualized Experiments



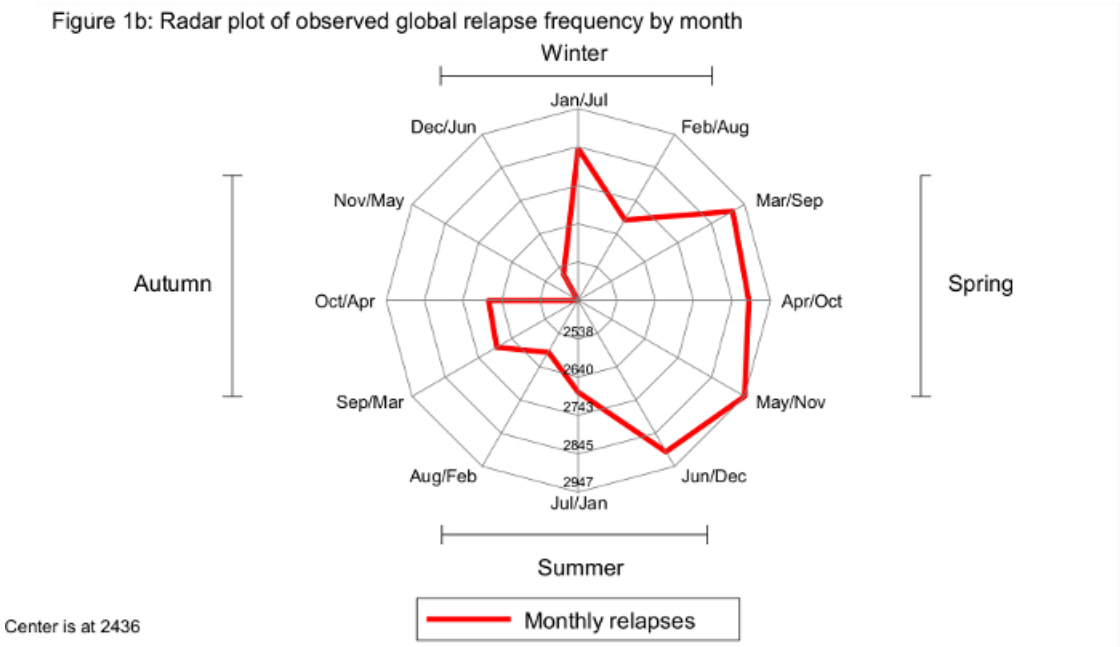


Figure 1. Radar plots of observed global relapse frequency by month. (A) northern hemisphere, **(B)** combined northern and southern hemispheres Please click here to view a larger version of this figure.

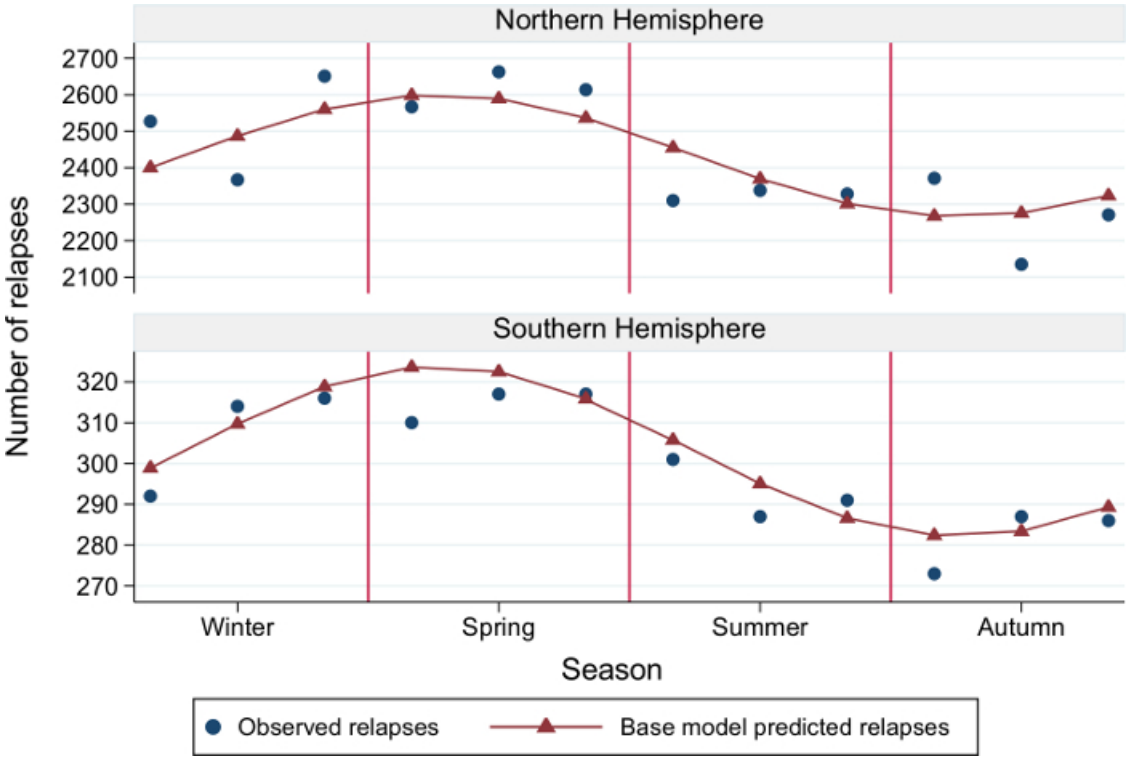


Figure 2. Base model predicted vs observed relapses. Plots comparing observed monthly relapses by hemisphere with predicted relapses using the base-case trigonometric model describing a single annual cycle of one peak and one trough separated by six months. Please click here to view a larger version of this figure.

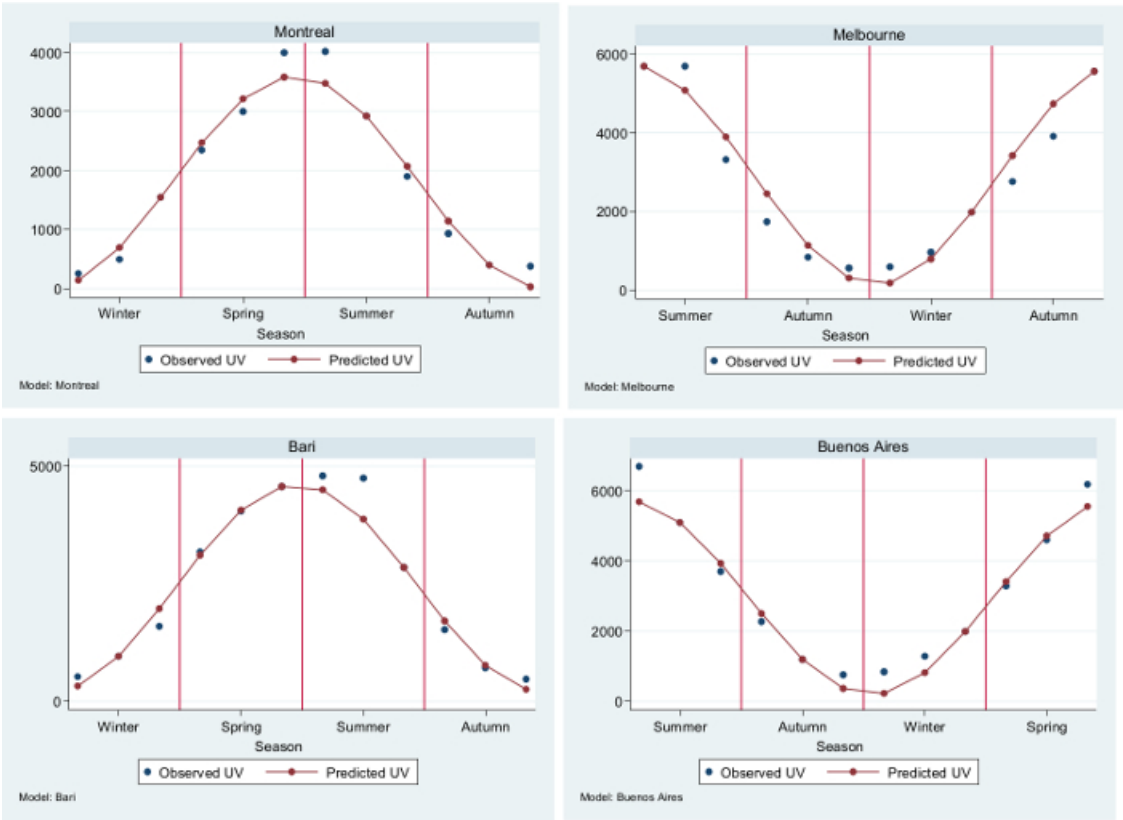


Figure 3. Base model predicted vs observed relapses. Plots comparing observed median monthly UVR with base model predicted UVR for Montreal, Canada; Melbourne, Australia; Bari, Italy & Buenos Aires, Argentina. Please click here to view a larger version of this figure.

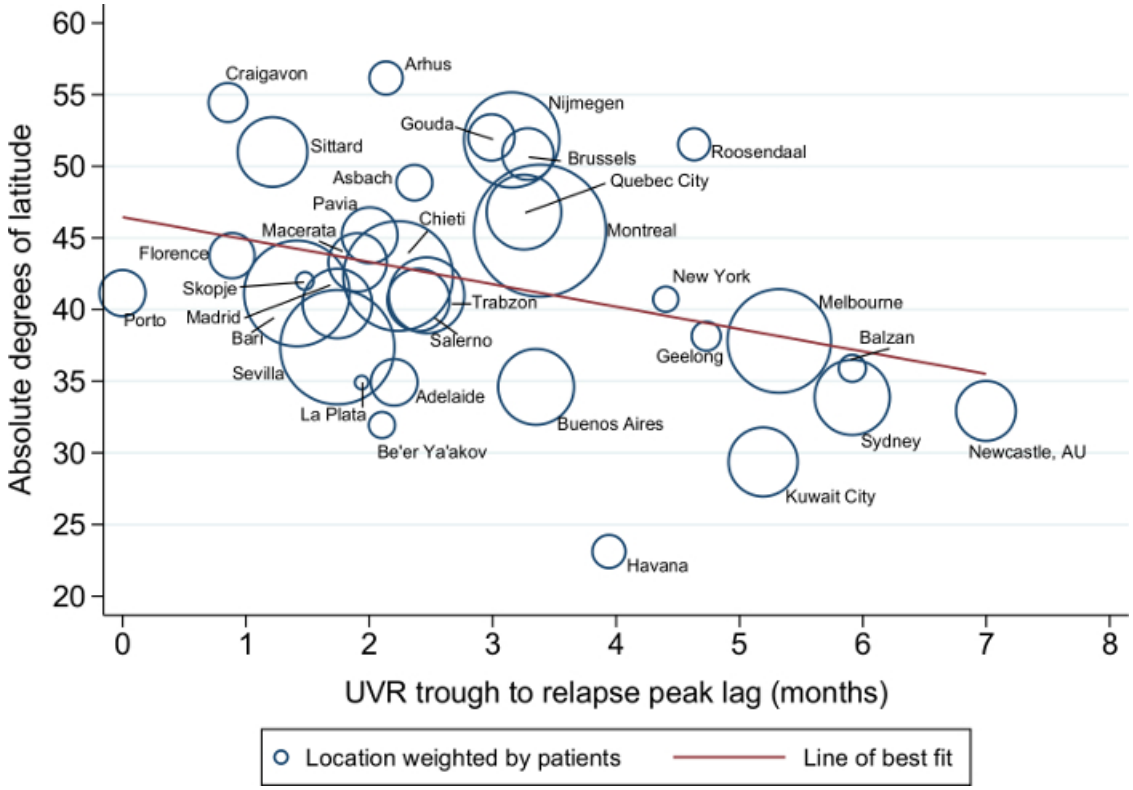


Figure 4. Weighted line of best fit between absolute latitude and UVR-trough-to-relapse-peak lag. Please click here to view a larger version of this figure.

Discussion

The protocol described herein details a systematic regression based technique, guided by visual plot analysis, of global MS relapse onset data. It takes as a starting point a relatively simple descriptive analysis of relapse data from 20 countries across both hemispheres, allowing the user to explore theories regarding the temporality of relapse onset timing in MS and testing these theories formally through the use of trigonometric models. Through a step-wise process of first plotting global relapse onset data and then systematically graphing and assessing candidate geometric fits of the observed data, a latitude-dependent correlation between seasonal trough UVR and subsequent peak relapse onset probability was observed, a correlation hitherto unprecedented in MS epidemiology. Furthermore by combining trend visualization with formal statistical modelling, this analysis also confirmed prior meta-analysis suggesting seasonality was a factor in relapse onset timing in the northern hemisphere and, also for the first time, extended this observation to the southern hemisphere.

Trigonometric regression modelling is a flexible tool for formally exploring cyclical, time- or season-dependent periodic phenomena, permitting statistical characterisation of trend data that conforms to geometric shapes such as the annual cyclical sinusoid curve observed in both the relapse onset timing and UVR data explored in this report. However given the range of shapes and structures that complicated, multi-factorial epidemiological trend phenomena such as relapse onset timing may potentially assume, visualization of both the original data and the differences between such observed data and those predicted by a particular model (*i.e.*, the residuals) are critical for both the hypothesis-generating (relapse onset timing varies seasonally across the year) and hypothesis-testing phases of this study (this seasonality is predictable and best described using a sine regression). The result is a suite of novel, empirically-grounded inferences regarding the potential global influence of season and latitude in disease exacerbation patterns in MS.

The critical step within the protocol was perhaps the simplest to execute, the visualization of the observed relapse onset data using simple descriptive scatterplots. Given the multitude and diversity of possible temporal structures periodic data may take, simple graphs of observed data provide both the an empirical basis for forming a hypothesis around relapse onset patterns as well as the starting point for building models that best capture these trends and which can subsequently be used for statistical inference and prediction. A key modification engineered into the protocol was the systematic comparison of the base model against alternative models incorporating additional trigonometric harmonic functions. "Best" fit is a relative state and only by testing the performance of the base model against plausible alternatives was best fit in this case able to be determined. The other key step was replicating each of the relapse probability and UVR models at all three distinct levels of geography - global, hemisphere and location. Not only did this provide internal validation of the primary results (the higher powered trends observed at the global and hemispheric levels were replicated at the location level) it also permitted troubleshooting of the code used to run the plots and models. Unexpected results or implausible model fits, not always evident at the level of the global or hemispheric analysis, derived at the level of the location were used as a red flag for quality checking the code used over all levels of geography. This provided confidence that the seasonal cycles and latitude patterns observed globally were not an artefact of data aggregation or miscoding. A further advantage of this protocol is that not only can it capture and describe seasonality and the influence of hemisphere and latitude with an appropriate robustness, it also adjusts these associations for potential confounding from patient-level propensities for relapse including differing levels of disability and varying disease-modifying drug exposure prior to relapse. This allows us to better isolate season and latitude as independent predictors of relapse probability

resulting in estimates of effect that better approximate the truth. This is particularly important given the potential clinical consequences of this research.

The observation of a predictable, latitude-dependent lag between winter trough UVR levels and subsequent relapse peak frequency may in part relate to an influence of changing vitamin D status at a given geographical location, each with its own unique UVR profile. Several vitamin D-mediated immunomodulatory correlates with MS relapse onset probability have previously been observed including shifting T helper lymphocytes away from a pro-inflammatory Th1 profile to the less inflammatory Th2[43-46] and inhibition of dendritic cells and IgM/IgG antibody production[43,47-50]. Coupling this to the observation of a potential role for both season and latitude in the kinetics of relapse timing, this suggests a role in clinical practice for latitude-specific, location-appropriate vitamin D supplementation for reducing the probability of future relapse. Of course despite this suggestion, the MSBase study did not collect longitudinal data on patient-level vitamin D status nor formal UVR skin exposure quantification and thus this theorized inverse correlation between vitamin D status and subsequent relapse probability remains exactly that, a hypothesis only. Formal, appropriately powered randomized clinical trials are required to establish causality. Two such trials of vitamin D monotherapy, the Australian/New Zealand PREVANZ trial (registration ACTRN12612001160820) and the French "D-lay MS" study (registration PHRC-N/2012/ET), are currently in progress.

Perhaps most notably, this study is illustrative of the possible synergy available to epidemiologists from the combination of formal statistical modelling and diagnostics with data visualization techniques. The significance of this technique relative to other forms of time-series analyses lies with its rejection of the assumption of conventional time series analyses that any underlying temporality is predominantly a random process. By comparison trigonometric regression explicitly seeks out structures in the temporal variation of cyclic, periodic phenomena such as MS relapse. As such trigonometric models are exquisitely reliant upon systematic visualization of both observed data and modelled estimates to guide and corroborate the model building and evaluation process, every step of the way. Neither the visualization or the modelling would have been sufficient in isolation - plot analysis was necessary for establishing realistic hypotheses regarding the presence and structural form of seasonal and latitudinal influences of relapse probability and then testing the performance of the resultant models whilst trigonometric regression was necessary for both quantifying these relationships, adjusting for important confounders, and providing a measure of certainty as to how plausible these associations are.

The technique described herein is a powerful method for isolating the role or influence of seasonality or latitude on complex, multifactorial events such as the timing of MS relapse. As such it has future potential for wide application for studying other clinical or biological phenomena which are known or suspected of varying systematically with season and/or latitude. This technique would be particularly relevant for prediction in disease epidemiology, both in terms of communicable and non-communicable disease where the timing of key events such as an infection or disease progression are complex and often driven by a multitude of both environmental factors (season, temperature, latitude) and patient-level characteristics (age, comorbidities, exposure to treatment). Such a tool may assist in risk-stratification of patients more likely to experience an adverse health event and thus guide earlier interventions.

Disclosures

Tim Spelman received honoraria for consultancy and funding for travel from Biogen Idec Inc; Orla Gray received travel support from Biogen Idec, Merck Serono and Novartis; compensation for serving on scientific advisory boards from Biogen Idec, Genzyme, Novartis and Merck Serono; Robyn Lucas did not disclose any competing interests and Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.

Acknowledgements

The authors would like to thank Ivan Hanigan for his support in extracting and interpreting the ultraviolet radiation satellite data. The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. RL is supported by a NHMRC Career Development Award [ID 1004898].

References

1. Simpson Jr, S., Blizzard, L., Otahal, P., Van der Mei, I., & Taylor, B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J Neurol Neurosurg Psychiatry.* **82** (10): 1132-41 (2011).
2. Risco, J., *et al*. Latitudinal prevalence gradient of multiple sclerosis in Latin America. *Mult Scler.* **17** (9), 1055-1059 (2011).
3. Hollingworth, S., Walker, K., Page, A., & Eadie, M. Pharmacoepidemiology and the Australian regional prevalence of multiple sclerosis. *Mult Scler.* **19**(13), 1712-1716 (2013).
4. Jin, Y., de Pedro-Cuesta, J., Soderstrom, M., Stawiarz, L., & Link, H. Seasonal patterns in optic neuritis and multiple sclerosis: a meta-analysis. *J Neurol Sci.* **181**(1), 56-64 (2000).
5. Bamford, C.R., Sibley, W.A., & Thies, C. Seasonal variation of multiple sclerosis exacerbations in Arizona. *Neurol.* **33**(6), 697-701 (1983).
6. Bisgard, C. Seasonal variation in disseminated sclerosis (Danish). *Ugeskrift for Laeger.* **152**(16), 1160-1161 (1990).
7. Callaghan, T. S. Multiple sclerosis and sinusitis. *Lancet.* **328**(8499), 160-161 (1986).
8. Gay, D., Dick, G., & Upton, G. Multiple sclerosis associated with sinusitis: a case-controlled study in general practice. *Lancet.* **327**(8485), 815-819 (1986).
9. Goodkin, D.E., & Hertsgaard, D. Seasonal variation of multiple sclerosis exacerbations in North Dakota. *Arch Neurol.* **46**(9), 1015-1018 (1989).

Copyright © 2015 Journal of Visualized Experiments

10. Hopkins, C.E., & Swank, R.L. Multiple sclerosis and the local weather. *Arch Neurol.* **74**(2), 203-207 (1955).
11. Reilly, M.A.R., & O'Reilly, P.M.R. Temporal influences on relapses of multiple sclerosis. *Eur Neurol.* **31**(6), 391-395 (1991).
12. Schapira, K. The seasonal incidence of onset and exacerbations in multiple sclerosis. *J Neurol Neurosurg Psychiat.* **22**(4), 285 (1959).
13. Sibley, W.A., & Foley, J.M. Seasonal variation in multiple sclerosis and retrobulbar neuritis in Northeastern Ohio. *Trans Am Neurol Assoc.* **90**, 295-297 (1965).
14. Sosa, E. M., Betancor, L. P., Rosas, C., & Navarro, M. C. Multiple sclerosis in the province of Las Palmas (Spanish). *Archivos de Neurobiologia.* **46**(3), 161-166 (1982).
15. Ogawa, G., Mochizuki, H., Kanzaki, M., Kaida, K., Motoyoshi, K., & Kamakura, K. Seasonal variation of multiple sclerosis exacerbations in Japan. *Neurol Sci.* **24**(6), 417-419 (2004).
16. Abella-Corral, J., Prieto, J. M., Dapena-Bolaño, D., Iglesias-Gòmez, S., Noya-Garcìa, M., & Lema, M. Seasonal variations in the outbreaks in patients with multiple sclerosis. *Rev Neurol.* **40**(7), 394-396 (2004).
17. Koziol, J. A., & Feng, A. C. Sesonal variations in exacerbations and MRI parameters in relapsing-remitting multiple sclerosis. *Neuroepidemiology.* **23**(5), 217-223 (2004).
18. Spelman, T., *et al*. Seasonal variation of relapse rate in multiple sclerosis is latitude dependent. *Ann Neurol.* **76**(6), 880-890 (2014).
19. Gallier, J. H. *Curves and surfaces in geometric modeling: theory and algorithms.* Morgan Kaufmann. (2000).
20. Agoston, K. *Computer Graphics and Geometric Modelling: Implementation & Algorithms.* Springer Science & Business Media. ISBN 978-1-84628-108-2 (2005).
21. Cox, N. J. Speaking Stata: in praise of trigonometric predictors. *Stata Journal.* **6**(4), 561-579 (2006).
22. Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., & Armstrong, B. Time series regression studies in environmental epidemiology. *Int J Epidemiol.* (2013).
23. Bracewell, R.N. *The Fourier Transform and Its Applications.* New York: McGraw-Hill. (2000).
24. Korner, T.W. *Fourier Analysis.* Cambridge: Cambridge University Press. (1998).
25. Rigdon, S.E., *et al*. Detection of Outbreak Signals Using R. *Online J Public Health Inform.* **6**(1), (2014).
26. Ziemssen, T., Reimann, M., Gasch, J., & Rüdiger, H. Trigonometric regressive spectral analysis: an innovative tool for evaluating the autonomic nervous system. *J Neural Transm.* **120**(1), 27-33 (2013).
27. Luque-Fernandez, M. A., *et al*. Absence of circadian rhythms of preterm premature rupture of membranes and preterm placental abruption. *Ann Epidemiol.* **24**(12), 882-887 (2014).
28. Luteijn, J. M., *et al*. Seasonality of congenital anomalies in Europe. *Birth Defects Res A Clin Mol Teratol.* **100**(4), 260-269 (2014).
29. Giardini, V., Russo, F. M., Ornaghi, S., Todyrenchuk, L., & Vergani, P. Seasonal impact in the frequency of isolated spina bifida. *Prenat Diagn.* **33**(10), 1007-1009 (2013).
30. Eghtesady, P., Brar, A., & Hall, M. Seasonality of hypoplastic left heart syndrome in the United States: A 10-year time-series analysis. *J Thorac Cardiovasc Surg.* **141**(2), 432-438 (2011).
31. Abiona, T. O., Adebowale, S. A., & Fagbamigbe, A. F. Time Series Analysis of Admission in the Accident and Emergency Unit of University College Hospital, Ibadan, Southwestern Nigeria. *Am. J. Comput. Appl. Math.* **2**(1), 1-9 (2012).
32. Cantwell, K., Dietze, P., Morgans, A. E., & Smith, K. Ambulance demand: random events or predicable patterns? *Emerg Med J.*, **30**(11), 883-887 (2012).
33. Butzkueven, H., *et al*. MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis. *Mult Scler.* **12**(6), 769-774 (2006).
34. Poser, C.M., *et al*. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol.* **13**(3), 227-231 (1983).
35. McDonald, W.I., *et al*. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol.* **50**(1), 121-127 (2001).
36. Kalincik, T., *et al*. Risk of relapse phenotype recurrence in multiple sclerosis. *Mult Scler.* (2014).
37. Total Ozone Mapping Spectrometer on board the Earth Probe spacecraft. Accessed 1 October 2013. Available from: http://iridl.ldeo.columbia.edu/SOURCES/.NASA/.GSFC/.TOMS (2013).
38. D'Agostino, R. B., Belanger A. J., & D'Agostino Jr, R. B. A suggestion for using powerful and informative tests of normality. *Am Stat.* **44**(4), 316-321 (1990).
39. Gould, W. W., & Rogers, W. H. Summary of tests for normality. *Stata Technical Bulletin.* 3:20-23, (1991).
40. Stolwijk, A. M., Straatman, H., & Zielhuis, G. A. Studying seasonality by using sine and cosine functions in regression analysis. *J Epidemiol Community Health.* **53**(4), 235-238 (1999).
41. Brookhart, M. A., & Rothman, K. J. Simple estimators of the intensity of seasonal occurrence. *BMC Med Res Methodol.* **8** (1), 67 (2008).
42. Fernández-Durán, J. J., & Gregorio-Domìnguez, M. M. Testing for seasonality using circular distributions based on non-negative trigonometric sums as alternative hypotheses. *Stat Methods Med Res.* **23**(3), 279-292 (2011).
43. Lemire, J.M., Archer, D.C., Beck, L., & Spiegelberg H.L. Immunosuppressive actions of 1,25-dihydroxyvitamin D3: preferential inhibition of Th1 functions. *J Nutr.* **125**(6 Suppl), 1704S-1708S (1995).
44. Tsoukas, C.D., *et al*. Inhibition of interleukin-1 production by 1,25-dihydroxyvitamin D3. *J Clin Endocrinol Metab.* **69**(1), 127-33 (1989).
45. Lemire, J.M. Immunomodulatory actions of 1,25-dihydroxyvitamin D3. *J Steroid Biochem Mol Biol.* **53**(1-6), 599-602 (1995).
46. van Etten, E., & Mathieu, C. Immunoregulation by 1,25-dihydroxyvitamin D3: basic concepts. *J Steroid Biochem Mol Biol.* **97**(1-2), 93-101 (2005).
47. Tsoukas, C. D., Provvedini, D. M., & Manolagas, S. C. 1,25-dihydroxyvitamin D3: a novel immunoregulatory hormone. *Science.* **224**(4656), 1438-1440 (1984).
48. Smolders, J., Menheere, P., Kessels, A., Damoiseaux, J., & Hupperts, R. Association of vitamin D metabolite levels with relapse rate and disability in multiple sclerosis. *Mult Scler.* **14**(9), 1220-1224 (2008).
49. Provvedini, D.M., & Manolagas, S.C. 1 Alpha,25-dihydroxyvitamin D3 receptor distribution and effects in subpopulations of normal human T lymphocytes. *J Clin Endocrinol Metab.* **68**(4), 774-779 (1989).
50. Provvedini, D.M., Tsoukas, C.D., Deftos, L J., & Manolagas, S.C.1 alpha,25-Dihydroxyvitamin D3-binding macromolecules in human B lymphocytes: effects on immunoglobulin production. *J Immunol.* **136**(8), 2734-2740 (1986).

Copyright © 2015 Journal of Visualized Experiments