

Short report

Seasonal variation in antidepressant prescriptions, environmental light and web queries for seasonal affective disorder

Thomas Lansdall-Welfare, Stafford Lightman and Nello Cristianini

Summary

The state of an individual's mental health depends on many factors. Determination of the importance of any particular factor within a population needs access to unbiased data. We used publicly available data-sets to investigate, at a population level, how surrogates of mental health covary with light exposure. We found strong seasonal patterns of antidepressant prescriptions, which show stronger correlations with day length than levels of solar energy. Levels of depression in a population can therefore be determined by proxy indicators such as web query logs.

Furthermore, these proxies for depression correlate with day length rather than solar energy.

Declaration of interest

None.

Keywords

Seasonal affective disorder; web queries; day length.

Copyright and usage

© The Royal College of Psychiatrists 2019.

Seasonal affective disorder (SAD) comprises episodes of depression that occur regularly during the winter months and remit in the following spring or summer. This condition is much more common in women, and has a prevalence of 3.44% in a Zurich cohort studied for over 20 years¹ and about 5% of the USA population, who remain symptomatic for approximately 40% of the year.² It has been suggested that shorter photoperiods in winter trigger SAD in susceptible individuals, perhaps through differences in retinal melatonin-activated retino-hypothalamic pathway activity.³ This has provoked numerous studies on the effectiveness of lighting interventions to improve mood. Surprisingly, there have been no studies comparing different levels of illuminance, and a recent Cochrane database review concluded there was only low-quality evidence that light therapy might improve mood in SAD.⁴ We have used an alternative approach: the use of publicly available databases to obtain information on the pattern of antidepressant prescriptions, web queries for SAD and details of day length and solar energy. The aims of this study were to investigate whether web queries could provide proxy indicators for population changes in mental health over time, and to assess whether these changes correlate with circannual variations in sunlight.

the community in the UK. Prescriptions written in England but dispensed outside England are included. The data includes prescriptions written by general practitioners and other non-medical prescribers (such as nurses and pharmacists) who are attached to general practitioner practices.

Web queries for SAD

Trends of web queries for SAD between August 2010 and November 2017 were obtained from Google Trends for the UK, and are publicly available (<https://trends.google.com/>). The data is an 88-point time series that represents an unbiased random sample of Google search data that has been anonymised, categorised by topic and aggregated by location (UK). Google provides data based either on terms, which show matches for all terms related to the query, or topics, which include all terms related to the same query. Which queries form a topic is determined by Google and cannot be controlled by us, but it is explained generally as 'queries related to the same concept, e.g. equivalent queries in different languages'. We chose to search for topic timelines for SAD, which was broken down by location. Additionally, monthly web query data for SAD for each of the 50 states in the USA (excluding Alaska) were obtained from the same source, averaged by month between August 2010 and November 2017, forming a 12-point time series for each state.

Method

Data sources

Practice-level prescribing data

Data on prescriptions in England covering the period between August 2010 and November 2018 were obtained from the National Health Service Digital website, where it is publicly available (<https://digital.nhs.uk/data-and-information/publications/statistical/practice-level-prescribing-data>). The data forms an 88-point monthly time series for all medicines, dressings and appliances that are prescribed by all practices in England and dispensed in the community each month. For each practice in England, the following information is presented at presentation level for each medicine, dressing and appliance (by presentation name): the total number of items prescribed and dispensed, the total net ingredient cost, the total actual cost and the total quantity. The data covers National Health Service prescriptions written in England and dispensed in

Solar energy data

Data on the average solar energy in each of the states in the USA (excluding Alaska) per month of the year were obtained from the National Renewable Energy Laboratory and are publicly available (<https://www.nrel.gov/gis/assets/docs/solarsummaries/solarsummaries.xlsx>), forming a 12-point time series for each state. The solar energy data relates to the direct normal irradiance, also known as beam radiation, and provides monthly average insolation values averaged over surface cells of 0.1 degrees in both latitude and longitude, relating to an area approximately 10 km in size. The data were created using the SUNY Satellite Solar Radiation model,⁵ where the data are averaged from hourly model output over 11 years between 1998 and 2009. The model uses hourly radiance images from geostationary weather satellites, daily snow cover data and monthly

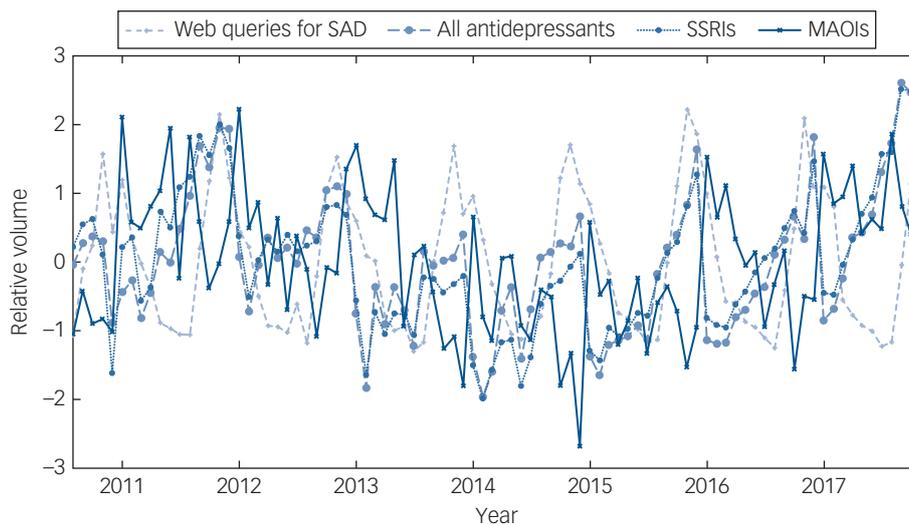


Fig. 1 Time series between prescriptions for antidepressants and its relevant subcategories in the UK between August 2010 and November 2017, and web queries for seasonal affective disorder in the UK during the same period. Time series are detrended and standardised.

MAOIs, monoamine-oxidase inhibitors; SAD, seasonal affective disorder; SSRIs, selective serotonin reuptake inhibitors.

averages of atmospheric water vapor, trace gases and the level of aerosols in the atmosphere to calculate the hourly total insolation (sun and sky) falling on a horizontal surface. The direct beam radiation is then calculated using the atmospheric water vapor, trace gases and aerosols, which are derived from a variety of sources. Where possible, existing ground measurement stations are used to validate the data.⁶

Day-length data

Day-length data for the population centre of each state in the USA (excluding Alaska, where no solar energy data was available) were computed for 2017 from the United States Naval Observatory website (http://aa.usno.navy.mil/data/docs/Dur_OneYear.php).⁷ Day lengths were then averaged over each day of the month to give a repeating 12-point time series representing the average day length in the given location for which we wished to compare data types (web queries, solar energy, day length, etc.). Geographical coordinates for the centres of population mass were obtained from the latest census (2010) from the United States Census Bureau and are publicly available (http://www2.census.gov/geo/docs/reference/cenpop2010/CenPop2010_Mean_ST.txt). The concept of the centre of population is analogous to a centre of mass for which each weight is represented by the location of one person.

Time series pre-processing

For each of the data sources, we obtained monthly time series for the variables of interest to the study. To obtain meaningful comparisons between these disparate data-sets, we needed to perform some pre-processing of the data, as described here.

The time series for each prescription in the practice-level prescribing data (<https://www.hpft.nhs.uk/information-and-resources/pharmacy-and-medicines-optimisation/medicines-formulary/bnf-chapter-4-central-nervous-system/>) was first normalised by the total prescription volume within the same month, to account for differences between months such as number of available working days or number of patients for whom a prescription was issued. The normalised prescription time series for a specific medicine, dressing or appliance can be considered an estimate of the probability that,

given a prescription is written, what is the probability that the prescription is for that specific item (i.e. a conditional probability).

The monthly time series from each data source were further pre-processed to remove their underlying trend and ensure fair correlations when analysing the relationship between different data sources (without this detrending step, any two increasing functions will have a high correlation, regardless of their seasonal form). This was performed by fitting a regression line, using ordinary least-squares to each time series, before subtracting the regression line to remove any (linear) trend.

Finally, time series were standardised so that each had a zero mean and unit variance, allowing the time series to all fall within the same range of values for comparison and visualisation. It should be noted that this step has no effect on the correlation between different series, and is performed mainly for convenience and visualisation purposes.

Hypothesis testing and explaining variation

Establishing web queries as a proxy for antidepressant prescriptions

Focusing on prescriptions for antidepressant medications, we wished to test whether the number of web queries for SAD in a given month and location can be used as a proxy for antidepressant prescriptions in the same month and location, using the Pearson's correlation coefficient (ρ) as our test statistic. This can be restated in terms of an alternative hypothesis where the number of web queries for SAD are related to antidepressant prescriptions in the

Table 1 Pearson's correlation coefficient (ρ) between time series of prescriptions for antidepressants and its relevant subcategories in the UK between August 2010 and November 2017, and web queries for seasonal affective disorder in the UK during the same period

	Web queries	Antidepressants	SSRIs	MAOIs
Web queries	–	0.3229	0.2625	–0.1538
Antidepressants	0.3229	–	0.9475	0.0041
SSRIs	0.2625	0.9475	–	0.1424
MAOIs	–0.1538	0.0041	0.1424	–

Coefficients in bold are significant with a P -value <0.05.
SSRIs, selective serotonin reuptake inhibitors; MAOIs, monoamine-oxidase inhibitors.

Table 2 Pearson's correlation coefficient (ρ) between the time series for web queries for seasonal affective disorder in the USA, solar energy, day length in hours and the deviation from a 12 h day for each of the states (excluding Alaska)

	Web queries	Solar energy	Day length (h)	Day length (deviation)
Web queries	–	–0.1965	–0.4646	–0.1657
Solar energy	–0.1965	–	0.6108	–0.0378
Day length (h)	–0.4646	0.6108	–	0.1776
Day length (deviation)	–0.1657	–0.0378	0.1776	–

Coefficients in bold are significant with a P -value <0.05 .

same month and location ($H_a: \rho \neq 0$), and a null hypothesis where they are unrelated ($H_0: \rho = 0$). Setting our significance level at $\alpha = 0.05$, we computed the Pearson's correlation coefficient between the monthly time series of web queries for SAD in the UK, and the time series for all antidepressant prescriptions as a group (<https://www.hpft.nhs.uk/information-and-resources/pharmacy-and-medicines-optimisation/medicines-formulary/bnf-chapter-4-central-nervous-system/>), monoamine-oxidase inhibitors and selective serotonin reuptake inhibitors (SSRIs) over a period of 88 months between August 2010 and November 2017 (Fig. 1). Tricyclic antidepressants showed a similar seasonal variation to SSRIs, but were not shown in the data to avoid an excessively busy figure.

Explaining variation in web queries with solar energy and day length

After establishing whether certain web queries can be used as a proxy for antidepressant prescriptions, we further wanted to assess if we can explain the variation in the number of web queries for SAD better using the levels of solar energy, measured using direct normal irradiance, the average day length or the deviation from a 12 h day in the same month and location. Because longitudinal data on solar energy for the UK was not available, and to more easily differentiate between variation related to solar energy and day length (as locations with longer days will also receive more solar energy, on average), we focused on comparing web queries from each of the states in the USA where solar energy data was available (all states except Alaska), enabling us to include locations with a much wider range of day lengths and levels of solar energy than using one location (England). Using the 12-point time series described above, we computed the average Pearson's correlation across all states for web queries, solar energy, day length hours and deviation from a 12 h day.

Ethics

All data was obtained from publicly available websites and no data could be related to any individual person. This work was ratified by Research and Enterprise Development at the University of Bristol and did not need further ethics permission.

Results

We can see (Table 1) that prescriptions of antidepressants are highly seasonal, generally climbing from around May each year, with a peak in December, before reaching a nadir in February (a time series of searches for 'depression' rather than SAD showed a different and much less pronounced seasonal variation). Furthermore, web queries for SAD are significantly correlated with the number of prescriptions for antidepressants, with correlations between 26% (SSRIs) and 32% (antidepressants in general) with web queries, suggesting that web queries may represent a possible proxy indicator. Monoamine-oxidase inhibitors are not significantly correlated with web queries. Looking within the USA, a

direct comparison of the time series for web queries, solar energy received, day length in hours and deviation from a 12 h day shows a significant correlation for all tests (Table 2), with the day length hours explaining the most variance (ρ^2) in web queries (21.6%), followed by solar energy (3.86%) and finally by deviation from a 12 h day (2.75%).

Discussion

Web queries have been used to follow many diseases as disparate as HIV, coronary artery disease and illness related to climate change. The best known example of this approach, however, is the association between web queries and influenza epidemics.^{8,9} We have been able to use publicly available data-sets to quantify circannual variation both in the prescription of antidepressant drugs and in web query logs for SAD. This demonstrated clear correlations between web queries for SAD and prescriptions for antidepressants. In addition, analysis of environmental light data shows how changes in day length most closely correlate with both SAD and antidepressant prescriptions. This suggests that research on the use of light therapy for SAD should concentrate on studies increasing the duration of daylight. This data also fits with a recent study showing that the major association between chronotype and geographical location is determined by the timing of sunset.¹⁰ We do stress that our method can only detect correlations, which do not imply causality, but can be used to rule out possible explanations that would be incompatible with the correlations we observe and also enable planning of appropriate clinical studies.

These results show that web queries can provide proxy indicators for population changes in mental health over time, and that these changes correlate with day length. This methodology also holds potential for future modelling of additional factors (such as chronotype/genotype) that could be tested for any effect on reducing the variance of our data.

Thomas Lansdall-Welfare, PhD, Research Associate, Intelligent Systems Laboratory, University of Bristol, UK; **Stafford Lightman**, FMedSci FRS, Professor of Medicine, Henry Wellcome Laboratories for Integrative Neuroscience and Endocrinology, University of Bristol, UK; **Nello Cristianini**, PhD, Professor of Artificial Intelligence, Intelligent Systems Laboratory, University of Bristol, UK

Correspondence: Nello Cristianini, Intelligent Systems Laboratory, Merchant Venturers Building, University of Bristol, Bristol BS8 1UB, UK. Email: nello.cristianini@bris.ac.uk

First received 9 Oct 2018, final revision 20 Dec 2018, accepted 20 Jan 2019

References

- Wirz-Justice A, Ajdacic V, Rössler W, Steinhausen HC, Angst J. Prevalence of seasonal depression in a prospective cohort study. *Eur Arch Psychiatry Clin Neurosci* 2018, in press.
- Kurlansik SL, Ibay AD. Seasonal affective disorder. *Am Fam Physician* 2012; **86** (11): 1037–41.

- 3 Roeklein KA, Wong PM, Miller MA, Donofry SD, Kamarck ML, Brainard GC. Melanopsin, photosensitive ganglion cells, and seasonal affective disorder. *Neurosci Biobehav Rev* 2013; **37**(3): 229–39.
- 4 Pachito DV, Eckeli AL, Desouky AS, Corbett MA, Partonen T, Rajaratnam SM, et al. Workplace lighting for improving alertness and mood in daytime workers. *Cochrane Database Syst Rev* 2018; **3**: CD012243.
- 5 Perez R, Ineichen P, Moore K, Kmiecik M, Chain C, George R, et al. A new operational model for satellite-derived irradiances: description and validation. *Solar Energy* 2002; **73**(5): 307–17.
- 6 Perez R, Moore K, Wilcox S, Renne D, Zelenka A. Forecasting solar radiation - preliminary evaluation of an approach based upon the national forecast database. *Solar Energy* 2007; **81**(6): 809–12.
- 7 Urban S, Seidelmann PK. *Explanatory Supplement to the Astronomical Almanac* (3rd edn). University Science Books, 2012.
- 8 Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. *PNAS* 2015; **112**(47): 14473–8.
- 9 Samaras L, García-Barriocanal E, Sicilia MA. Syndromic surveillance models using web data: the case of influenza in Greece and Italy using google trends. *JMIR Public Health Surveill* 2017; **3**(4): e90.
- 10 Porcheret K, Wald L, Fritschi L, Gerkema M, Gordijn M, Merrow M, et al. Chronotype and environmental light exposure in a student population. *Chronobiol Int* 2018; **35**: 1365–74.