Finding Relationships Between Chronic Illness and Weather Patterns with Flaredown App Data

Group members: Hyun Jo Park Parinitha Kompala Taylor Hudson zhaofeng liu

Abstract:

This data set is extrapolated from an app called Flaredown. Flaredown is an app that is designed for people with chronic illness in order to track their own symptoms in real-time with additional information on daily therapies and environmental factors called "trackables”. Users can "check-in" each day, recording the intensity of symptoms and illnesses, treatments used and dosage information, as well as any environmental circumstances. The dataset is unique because it encompasses not a specific disease but a multitude of illnesses. Our group would like to conduct both a relational and predictive analysis with user reporting of treatment and environmental factors and how this relates to symptom reporting.

Using this data from a symptom reporting app, Flaredown, we are trying to find if there is a correlation between weather conditions and an increased rating in the discomfort of symptoms? If so, can we use this weather data as a predictor for increased symptom rating?

Background:

It has long been accepted that symptom flare-up can be influenced by weather. Patients with chronic pain frequently feel that their discomfort is caused by the weather. The occurrence is so well-known that there is even a term created to describe symptoms associated with weather conditions; meteopathy[1]. Although this connection is recognized, development of models to predict symptoms based on weather are sparse, in part because it's difficult to obtain a large dataset of patients who report their pain symptoms often and in a range of weather situations. For many individuals with chronic conditions, a change in the severity of their symptoms tend to be aligned with the coming of winter season as well as other factors such as pollution[2]. Cold weather, rain, and a shift in barometric pressure might aggravate symptoms for some people, while others may feel better as the summer heat fades away[3]. Around three-quarters of people living with arthritis believe their pain is affected by the weather[4]. Despite extensive investigation into the existence and nature of the weather-pain link, there is a lack of effective assistance to healthcare providers regarding pain management projection.[5] A predictive model to anticipate flare-up based on weather could be a great asset to clinicians and hospitals in order to prepare for the influx of patients and/or gain better insight on the factors influencing patients complaints.

Smartphone usage has rapidly increased, which opens up new and exciting options for health research[6]. The Flaredown App data contains a wealth of information obtained from users who have chronic illnesses. There are thousands of user reports of symptom flareup, weather reports, food intake, and many other variables. Our primary steps to analyze trends in weather is to extrapolate symptoms reporting and user reporting of weather. These values can be categorized by dates and regions in order to accurately report historical weather patterns. Given these categories, the frequency of symptom reporting can then be tallied. Weather data will be merged with this symptom frequency information by using the app's internal weather check-in that the user indicates. Additionally, historical weather data from the national weather service, and can provide additional information on temperature, precipitation, atmospheric pressure and cloud coverage. This data set will be split into 80% “train” and 20% “test” for a predictive analysis. Poisson linear regression will identify trends in the frequency of symptoms of reporting as a response to temperature, cloud coverage, pressure and precipitation. After forming the best model to predict the frequency of symptom reporting, we will apply the model to our testing dataset to verify the results. The final model will serve as a tool for a future system that would help patients and health care providers better manage their health by providing pain projections.

Methods and Application:

Hypotheses: Based on a long standing assumption that weather conditions are linked to agitation of some conditions and/or injuries, we hypothesize that there will be a positive correlation between a higher pain scale(1-4) of symptom reporting and increased reporting of atmospheric pressure, increased humidity and days with precipitation.

H_a = There is an increased rating of the symptoms by users when there is increased atmospheric pressure, increased humidity and days with precipitation.

H_o = There is no change in the frequency of symptom reporting when there is increased atmospheric pressure, increased humidity and days with precipitation.

Data sources: (link)

The data is extracted from an app designed for users with chronic illnesses to track their own symptoms and other components that may influence their symptoms. The dataset has various variables such as the user ID, sex, age, country, check-in date, and trackable data such as trackable type, trackable dataset value(e.g. dosage of medication) , trackable name(user specified trackable item). The trackable type, category of user reporting (eg. symptoms, weather(weather is pulled automatically for the user's postal code), treatment (anything a patient uses to improve their symptoms), tag(a string representing an environmental factor), food, HBI (Harvey Bradshaw Index -- a standard metric for the level of activity of Crohn's Disease specifically).

Variables in the original dataset

- The variable **“trackable_type**” has 7 unique factors namely, "Condition", "Symptom", **"Weather",** "Treatment" , "Tag", "Food", "HBI" .
- And then will be looking into the variable **“trackable_name”,** this variable will include the information regarding the weather condition
- **“Country”,** this variable indicates the region**.**
- **“Check-in date”,** this is the date when the symptom was reported by the subject.

Variables used for analysis post-data wrangling

- Dependent variable(self-reported): Rating of symptom severity on a 1-4 scale, *ordinal*
- Predictor variables(automatically imported with app using the user postal code):
    - Atmospheric pressure, continuous
    - Weather icon, nominal
    - Humidity, nominal
    - Precipitation intensity, continuous
    - Temperature, continuous
- Potential influencers/effect modifiers or confounders
    - Treatment, nominal
    - Sex
    - Age

Orginal datset from flaredown app

user_id	age	sex	country	checkin_date	trackable_id	trackable_type	trackable_name	trackable_value
QEVuQwEASFRf7Xf1KLwTCxFMeTqirg==	33	female	US	2017-11-01	532	Condition	Idiopathic hypersomnia	4
QEVuQwEASFRf7Xf1KLwTCxFMeTqirg==	33	female	US	2017-11-01	380	Symptom	Lack of Motivation	4
QEVuQwEASFRf7Xf1KLwTCxFMeTqirg==	33	female	US	2017-11-01	136	Treatment	Adderall	2 x 20mg; 4 x 10mg
QEVuQwEASFRf7Xf1KLwTCxFMeTqirg==	33	female	US	2017-11-01	37338	Weather	temperature_min	30.0

Table 1: Each row contains a single tackable input from a user.There are 7,976,223 rows of user check-ins

Weather and symptoms reported by same user and same data

user_id	checkin_date	Symptom_reported	symptom_value	icon	temperature_min	temperature_max	precip_intensity	pressure	humidity
QEVuQwEAah0GoJ6IChS3X/WniyYKUA==	2017-02-15	exhaustion	2	snow	26	35	0.0014	1005	77
QEVuQwEAIscm8xlvubdDkqERi9WV0A==	2017-02-11	breathing	0	partly-cloudy-day	53	74	0.0012	1015	33
QEVuQwEAfEGUEstNreltgXJRvbgcfA==	2017-05-07	numbness in feet	1	partly-cloudy-day	41	68	0.0017	1009	61
QEVuQwEA3e3MDYba0HBuLjkew7YvHw==	2017-02-24	Migraine	0	partly-cloudy-night	51	74	0.0005	1011	77
QEVuQwEAxZP31SEiUvDf/+pSmJa6BA==	2017-04-21	Weight gain	4	partly-cloudy-day	44	59	0.0006	1032	73

Table 2: Each row contains a all weather data and symptoms reporting by a user per day.There are 180,658 rows of user check-ins

Univariate analyses:

To understand the data distributions and check if model assumptions are met, assess the degree of missing data:

Missing data:

Part of the data wrangling process will trim down much of the data and will only keep data with certain variables filled out, weather, symptoms, weather values, etc. Remaining missing values in the data will be handled

Data Distribution:

Symptom severity scale is ordinal.histogram, boxplot, qq plot of, dependent variable; frequency of symptom reporting

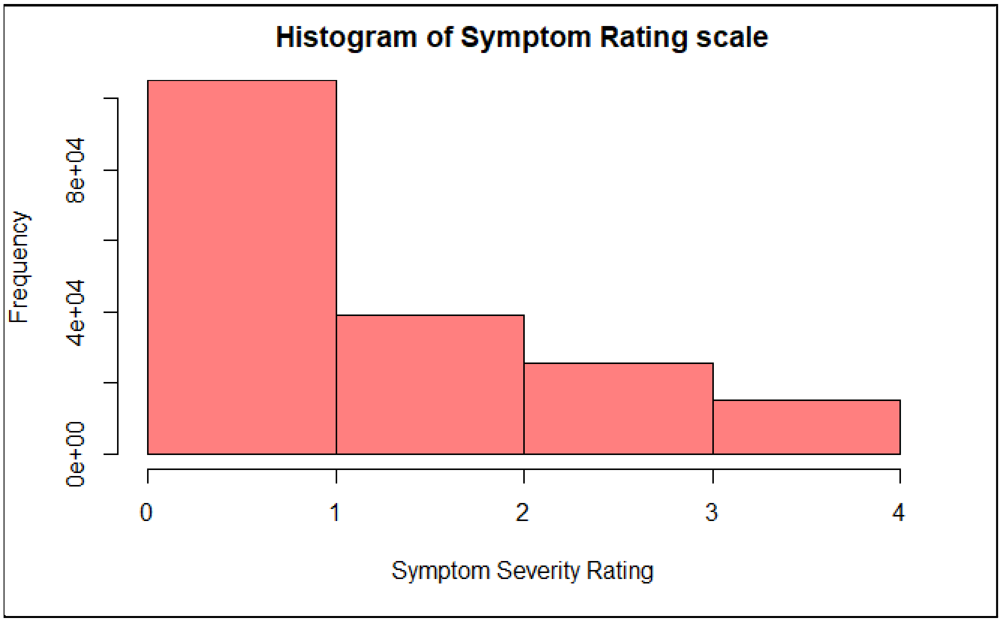


Figure 1: Histogram showing distribution of the symptom rating scale across all user

Table 1: Summary by each symptom rating scale

	0 (N=56500)	1 (N=48635)	2 (N=38880)	3 (N=25580)	4 (N=15155)
temperature_min					
Mean (SD)	48.1 (16.9)	47.7 (16.9)	47.7 (16.9)	48.0 (16.8)	49.0 (17.3)
Range	-44.0 - 93.0	-33.0 - 96.0	-44.0 - 94.0	-47.0 - 93.0	-42.0 - 89.0
temperature_max					
Mean (SD)	65.1 (18.7)	64.6 (18.7)	64.5 (18.7)	64.9 (18.6)	65.7 (19.0)
Range	-18.0 - 117.0	-15.0 - 115.0	-13.0 - 119.0	-18.0 - 118.0	-14.0 - 115.0
precip_intensity					
Mean (SD)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Range	0.0 - 1.1	0.0 - 0.8	0.0 - 1.1	0.0 - 0.7	0.0 - 0.4
pressure					
Mean (SD)	1016.6 (7.4)	1016.3 (7.7)	1016.3 (7.8)	1016.4 (7.7)	1016.5 (7.5)
Range	978.0 - 1048.0	953.0 - 1046.0	971.0 - 1049.0	938.0 - 1048.0	979.0 - 1051.0
humidity					
Mean (SD)	70.6 (15.4)	70.8 (15.6)	71.1 (15.3)	71.0 (15.2)	71.4 (14.7)
Range	5.0 - 100.0	6.0 - 100.0	6.0 - 100.0	8.0 - 100.0	7.0 - 100.0

Table 1: Continuous weather measurements man, standard deviation and range groups by symptom rating

Data-wrangling:

Group by users who have checked in multiple trackable types (ex. Symptoms or weather type) on the same day to assure the weather reporting corresponds with symptom reporting. Any data points from users that only input one type in one day will be removed. For example, one user only reported their symptoms on 1/18/17, since there is no other data type to be associated with the symptoms it would be removed.

Preliminary analyses:

To identify potential confounders, check other model assumptions (e.g., Are bivariate associations linear?), initial filtering steps for ML models:

For ordinal linear regression: check the quantity of non-zero and zero counts of symptoms reporting, check sample size

Primary analysis/analyses:

To define how the primary hypothesis will be tested (or how the primary goal will be assessed):

Ordinal regression

Include the primary test statistic(s), define threshold for statistical significance:

P-value for z-statistic and confidence interval derived from the ordinal regression model and possibly adjusted by sandwich standard error is over-dispersion observed

Define how model fit or performance will be assessed:

Over-dispersion

→ Check deviances are no too large with Chi-square with n-k-1 degrees of freedom

Clustering for repeated measure

→ Try seeing how much the model differs if repeated measures of the same users are deemed random effects

Confounders or effect modification:

→ Use other user reported data; food intake and treatment, to add to the model to see if coefficient estimates for weather change.

Include any secondary analyses and/or sensitivity analyses:

Use model to predict rates of symptom reporting in testing subset

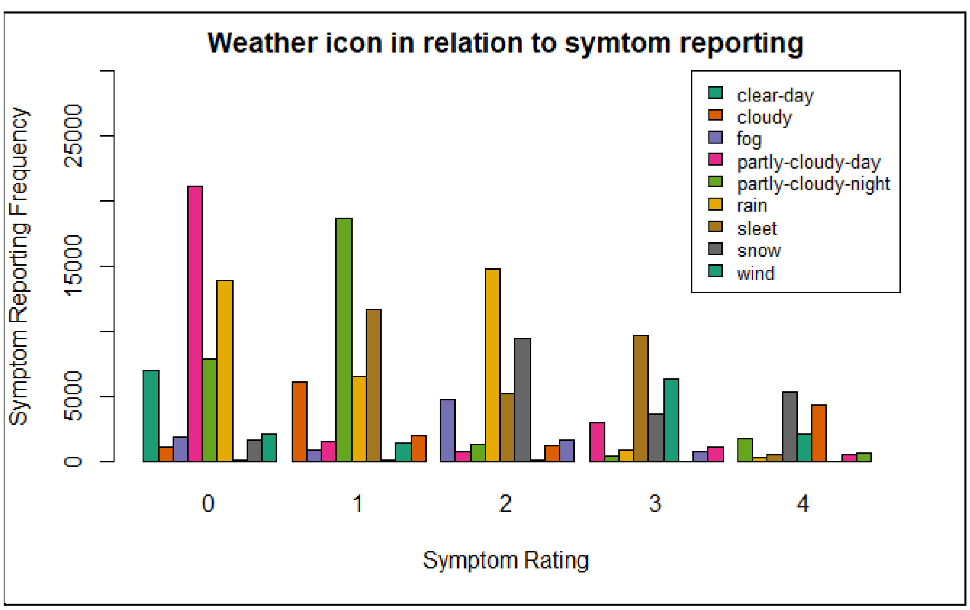


Figure 2: Frequency of Symptoms rating scale grouped by weather icon.

Repeat the process for any secondary analyses:

This data set will be split into 80% “train” and 20% “test” for a predictive analysis. Ordinal linear regression will identify trends in the frequency of symptoms of reporting as a response to temperature, cloud coverage, pressure and precipitation. After forming the best model to predict the frequency of symptom reporting, we will apply the model to our testing dataset to verify the results.

Address limitations:

How might measurement errors affect your findings?

All the data obtained were entered by the subject themselves so there might be some faulty data. This includes but is not limited to human error in inputting data and bias from false reporting.

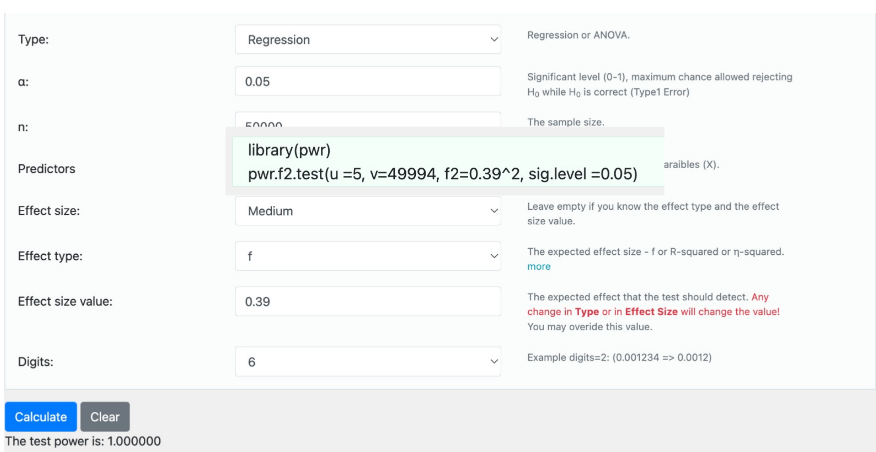
How generalizable are the findings?:

The finished model will be used as a tool for a future system that would deliver pain forecasts to help individuals and health care providers better manage their health, so this can be applicable to mostly all people of all age groups and from any environment.

Power calculation:

A power calculation was conducted for poisson regression which resulted in a test power estimate which was close to 1.0. This is likely due to the large sample size of our data (n = 50000).

```
library(pwr)
pwr.f2.test(u =5, v=49994, f2=0.39^2, sig.level =0.05)
```



MARKET RESEARCH:

Resource 1 (link)

- **The reason why you have chosen the particular resource.**

  The first resource that we choose for our market research is BCCResearch Library. The reason we chose this resource was that it was more user-friendly than the others. All the content on the website was categorized so that we were able to easily search for terms. Also, BCC Research provides objective, unbiased measurement, and assessment of market opportunities with detailed market research reports and we were able to get more insights.

- **What have you learned from the chosen report?**
  - In BCCResearch Library we looked into the “Technologies for Long-term Care and Home Healthcare: GlobalMarkets”.
  - Medical science and technology are continuously evolving for the betterment of mankind and the healthcare industry. Medicine and public health, along with increased consciousness about nutrition, environmental impact, and personal hygiene have led to a sharp increase in life expectancy, globally. Advancements and changing lifestyles encourage market growth.
  - Then we also looked into “Mobile Health (mHealth) Technologies and Global Markets”, and we obtained from that was that “the global mHealth market should reach $46.2 billion by 2021 from $13.2 billion in 2016 at a compound annual growth rate (CAGR) of 28.6%, from 2016 to 2021”. Healthcare providers will progressively use mHealth solutions that enable them to raise productivity and improve efficiencies as the industry evolves toward a patient-centric value-based healthcare system.

- **How is this information relevant for your current group’s project?**

  The reason why we looked into this is that our project is based on data from a mobile app and we are looking into the weather data there and checking if that has any influenza or the symptom reported by the patient all the data is self-reported and comes under “Home health care”. So based on this report we can come to the conclusion that how the world is changing from the traditional methods of healthcare to modern methods.

  Scope of the report- This new research on mobile health will give a quick overview of the industry's current state and recent advances. It explains how the environment is changing in terms of new difficulties and opportunities for app development, remote monitoring, and medical data networking. In terms of downloads and revenues, the research examines industry trends, key service providers, therapeutic markets, and the most popular mHealth applications.

Resource 2 (link)

- **The reason why you have chosen the particular resource.**

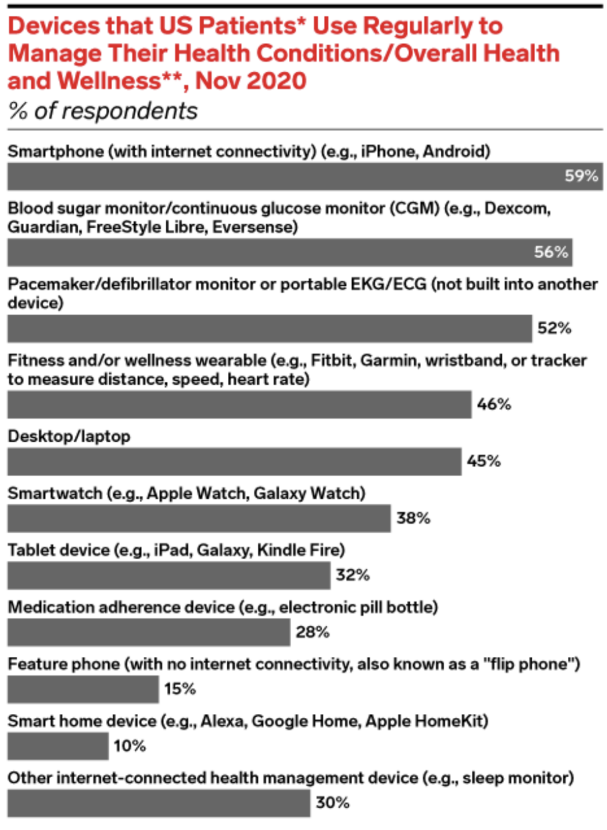
  We also looked at eMarketer to get data on digital health and specifically telehealth. The particular report we looked at focused on Mobile Health Apps for Disease Management: Patients Turn to Smartphones for Condition-Specific Help .This particular market research report has a Digital Atlas of internet and mobile phone users around the world, along with healthdata, which is a point of intersection of our research topic. This resource aggregates data from thousands of agencies and research firms which would be suitable to give us a comprehensive idea about the market. Also, eMarketer's reports provided a lot of visuals and charts to supplement the statistics.

- **What have you learned from the chosen report?**

  The pandemic has had a significant impact on the mobile health market — 37% of patients are now using a mobile app to manage their health, a significant increase since before COVID-19. Specifically, an interesting point to note from this resource is that chronic illness patients find condition-specific apps more helpful than general health and wellness apps that do not focus on a condition. However, these same patients find it difficult to find the right apps for this. Further, 68% of US employers are likely to invest more in digital health in the next five years, adding support to the profitability and opportunities in this market.

- **How is this information relevant for your current group's project?**

  This report highlights the importance of our research. Given the rising prominence of mHealth apps, the significance of our research can extend not only to the current literature but also to the improvement of patients' health outcomes. For instance, the report stated that 59% of chronic health patients who own a smartphone said they use it regularly to manage their health and wellness, which was the highest percentage among multipurpose devices. As briefly stated in the section above, given the large proportion of chronic illness patients who use mHealth to track their symptoms and health (81% across all age groups use at least one device), 76% believe these apps are beneficial for their health management. mHealth can also benefit providers/clinicians as well, allowing them to have more quality time with their patients.



This image shows us detailed stats on what device is used regularly by people in The US to manage their health conditions and wellness.

Resource 3 (link)

- **The reason why you have chosen the particular resource.**

Then another article in the BCCResearch Library also caught our attention. And the particular article we looked into called “Global Mobile Health And Fitness Sensors Market Size By Product, By Application, By Geographic Scope And Forecast”.We choose this particular resource on BCCResearch because it has great categories which help us to search and identify what the content we are looking into is fast and effective. Our research focuses on data gathered from mobile apps which the resources in gathering data is very specific. We can easily search on this BCCResearch Library to identify researches and articles which fitted our project the most. And the resources are authorized and the articles are mostly without subjective opinion and the writer only talks about what they got based on research so it is more unbiased.

- **What have you learned from the chosen report?**

First, the increasing amount of remote monitoring systems is attributed to the growth of the mobile health market which will therefore lead to high growth for the amount of health apps in the future.

Second, the report is all-inclusive and talks about how the mobile health market grows and is related to different factors, and taking into consideration the impact of various social, geographic, and economic factors associated with the growing mobile health market.

Third, this report goes into details to take the various participants into consideration, which takes into consideration that different groups of people are using the mobile app to report their life expectancy and so on.

- **How is this information relevant for your current group's project?**

The article we chose here actually fit our project very well. First of all, the range it focuses on is global which is what our scope is aimed at. Second, this talking about how geographic difference influences the mobile health and sensors market size. The market size has an association with home health care. We can come to a conclusion based on this association here. Furthermore, this research here takes the geography into consideration, and different geographic leads to different weather conditions. It can help us better understand the data we gathered based on each region with different weather features.

Reference

1. Hansel NN, McCormack MC, Kim V. The Effects of Air Pollution and Temperature on COPD. COPD. 2016;13(3):372-379. doi:10.3109/15412555.2015.1089846
2. Angela R-I, Johanna S, Peter S. Pain Among Children and Adolescents: Restrictions in Daily Living and Triggering Factors. Search.library.dartmouth.edu. Accessed January 25, 2022. (link)
3. Daher A, Matthes M, Keszei A. Characterization and Triggers of Dyspnea in Patients with Chronic Obstructive Pulmonary Disease or Chronic Heart Failure: Effects of Weather and Environment. Springer. (link). Accessed January 25, 2022.
4. Erin E. Krebs MD. Opioid vs nonopioid medications on pain-related function. JAMA. Published March 6, 2018. Accessed January 25, 2022. (link)
5. Perkin MR, Logan K, Bahnson HT, et al. Efficacy of the enquiring about tolerance (EAT) study among infants at high risk of developing food allergy. Journal of Allergy and Clinical Immunology. Published December 4, 2019. Accessed January 25, 2022. (link)
6. Tekur P, Nagarathna R, Chametcha S, Hankey A, Nagendra HR. A comprehensive yoga program improves pain, anxiety and depression in chronic low back pain patients more than exercise: An RCT. Complementary Therapies in Medicine.Published January 27, 2012. Accessed January 25, 2022. (link)
7. Alfred J S. Solar ultraviolet radiation exposure does not appear to... : AIDS. Solar ultraviolet radiation exposure does not appear to exacerbate HIV infection in homosexual men. Published November 15, 1997. Accessed January 25, 2022. (link)

8. Garcia-Vidal C, Labori M, Viasus D, et al. Rainfall is a risk factor for sporadic cases of legionella pneumophila pneumonia. PloS one. Published April 16, 2013. Accessed January 25, 2022. (link)
9. Benouaich V, Soler P, Gourraud PA, Lopez S, Rousseau H, Marcheix B. Impact of meteorological conditions on the occurrence of acute type A aortic dissections. OUP Academic. Published March 1, 2010. Accessed January 25, 2022. (link)
10. Tekur, P., et al. “A Comprehensive Yoga Programs Improves Pain, Anxiety and Depression in Chronic Low Back Pain Patients More than Exercise: An RCT.” *Complementary Therapies in Medicine*, vol. 20, no. 3, June 2012, pp. 107–18. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ctim.2011.12.009.