Published in final edited form as:

Headache. 2013 April ; 53(4): 636–643. doi:10.1111/head.12075.

Natural experimentation is a challenging method for identifying headache triggers

Timothy T. Houle, PhD* and Dana P. Turner, MSPH

Department of Anesthesiology, Wake Forest School of Medicine, Winston-Salem, NC 27157

Abstract

Objective—In this study we set out to determine whether individual headache sufferers can learn about the potency of their headache triggers (causes) using only natural experimentation.

Background—Headache patients naturally use the covariation of the presence-absence of triggers with headache attacks to assess the potency of triggers. The validity of this natural experimentation has never been investigated. A companion study has proposed three assumptions that are important for assigning causal status to triggers. This manuscript examines one of these assumptions, constancy in trigger presentation, using real-world conditions.

Methods—The similarity of day-to-day weather conditions over four years, as well as the similarity of ovarian hormones and perceived stress over a median of 89 days in nine regularly cycling headache sufferers were examined using several available time-series. An arbitrary threshold of 90% similarity using Gower's index identified similar days for comparison.

Results—The day-to-day variability in just these three headache triggers is substantial enough that finding two naturally similar days for which to contrast the effect of a fourth trigger (e.g., drinking wine versus not drinking wine) will only infrequently occur. Fluctuations in weather patterns resulted in a median of 2.3 days each year that were similar (range: 0 to 27.4). Considering fluctuations in stress patterns and ovarian hormones, only 1.5 days/month (95%CI: 1.2 to 2.9) and 2.0 days/month (95%CI: 1.9 to 2.2), respectively, met our threshold for similarity..

Conclusion—Although assessing the personal causes of headache is an age-old endeavor, the great many candidate triggers exhibit variability that may prevent sound conclusions without assistance from formal experimentation or statistical balancing.

Keywords

headache triggers; causality; research design

Introduction

Human beings have been suffering from headaches for as long as there have been human beings.^{1,2} This unfortunate history combined with our natural proclivities for covariation

*Correspondence to: Timothy Houle, thoule@wakehealth.edu Department of Anesthesiology, Wake Forest Medical School, Winston-Salem, NC.

Conflict of Interest: None

assessment,³ of deducing that A causes B because A precedes or co-occurs with B, makes it likely that searching for the causes of one's headache through examination of naturally occurring events may be one of the primeval natural experiments. Indeed there is evidence that earlier humans had devoted considerable thought to what caused their headaches. In 300 BC, Hippocrates produced a list of factors, now called “triggers,” that he inducted as causes of headache.⁴ His trigger list would be eerily familiar to modern headache sufferers who might also believe that weather, menstrual cycle, alcohol, thirst, and hunting or other forms of exertion increase their chances of experiencing a headache. The insightful Hippocrates was hardly the first thinker to attend to the covariation of perceptible events in an attempt to gain control over the dreaded headache; the ancient Babylonians, Chinese, and doubtless others made similar efforts.² In the present day, despite valiant medical efforts to rid us of headaches, we still suffer from them, so our individual quests to identify the causes of our headaches continue.

The approach to searching for headache triggers no doubt varies from person to person, but must be conducted using natural experiments that include a range of methods from quasi-experimental procedures (e.g., the use of a diary for prospective evaluation of triggers assigned to days in a non-randomized time-series) to pre-experimental cognitive efforts (e.g., retrospective, “what-if” reasoning). No matter the approach, these investigations are widely attempted by highly motivated individuals who are acting as scientists seeking to utilize the information about their headache triggers to reduce the chances they will experience a headache (See: ⁵). When an experienced individual-scientist is queried about personal triggers, individuals typically list from 4 to 9 different trigger factors. Ninety-five percent of headache sufferers endorse the existence of at least one personal headache trigger from a provided list.⁶

The quest to identify personal headache triggers appears nearly universal and is certainly a normal part of suffering from headaches. Yet, the task of evaluating personal headache triggers is extremely difficult. The endeavor likely consists of headache sufferers attempting to make sense of their current headache by searching for extraordinary factors that had co-occurred with the attack. For example, uncharacteristically drinking a glass of wine and then comparing the effects to what might have happened had a counterfactual reality actually transpired (e.g., “what if I had not drank the wine?”) (See: ⁷). This assessment might naturally draw attention to a recent day where wine was not similarly consumed and a headache had not occurred. If in possession of perfect memory, the headache sufferer could formally estimate the association strength by comparing all previous wine drinking occasions to that of non-wine drinking occasions. This effort would also require a computer-like processor to generate the full 2×2 interaction of exposures (all wine versus no-wine occasions) in relation to their impact on headache states (headache versus no-headache). Because humans possess neither this type of memory nor analytic processor, we probably overly rely on the co-occurrence of rare events⁸ that have transpired in only very recent memory to formulate the association (i.e., “this is the first time I have drank wine in a while, and I now have a headache”). These limitations likely result in headache sufferers utilizing only limited present-absent parings of their candidate triggers in an attempt to isolate the factors that cause their headaches. Thus, although the quest to discover headache triggers may be ubiquitous, it may not be based on optimal information processing.

The scope of this study was to examine the value of personal assessments of headache triggers based on real world conditions. We have examined the natural variability in a number of popularly reported headache triggers to explore the likelihood that a sufferer can isolate the causal impact of a candidate trigger in the presence of other potential influences. In essence, we are evaluating the natural conditions encountered by an individual who is attempting to identify the influence of one particular trigger using natural experimentation.

Methods

Three assumptions that are important to accurately estimate the causal influence of a headache trigger using only limited experiences are proposed in the companion manuscript.⁹ In this formulation, an individual acting as a scientist could be naturally exposed to, or expose themselves to, a headache trigger on one occasion, and then to its opposite on a later (or if using memory, earlier), comparable day. To estimate the causal strength of the headache trigger, several things must be true. First, the pattern of possible confounding influences must be similar on both days (constancy in trigger presentation). Comparing the effect of wine to not drinking wine is not very informative if a second headache trigger is present on one day and not the other. Second, the effect of the headache trigger must have been the same on both days (constancy in trigger effect). The same trigger cannot be experienced as present and absent on the same occasion, so this assumption holds that the effect of the trigger should be the same if it were experienced on the first or second occasion. Third, the headache sufferer must be comparable on both occasions (constancy in sufferer). This precludes the two occasions being on adjacent days if the first trigger occasion induced a headache that was still present on the second day (as the second trigger exposure then cannot initiate a new attack, only make an existing attack worse) or if the two occasions were too far apart such that the headache sufferer changed during this time.

These assumptions seem reasonable, but to what extent can they be satisfied in reality? This study examines only the constancy in trigger presentation assumption by examining the natural variability in several popular headache triggers over time (stress, ovarian hormones, weather). These triggers were chosen because they exhibit some degree of natural variability and they are present to some extent for all female headache sufferers. By examining the naturally occurring variability in these triggers, we can estimate the frequency of similar days that could be used in natural experimentation (i.e., how likely are we to find two days that are similar in their weather conditions so that we can see if drinking wine impacts headache on one but not the other?). In this way, the validity of the trigger presentation assumption can be directly assessed. The analyses are conducted on time-series data that were collected from several studies on popular headache triggers. Each dataset is described below:

Ovarian hormone data and perceived stress data

These data are being collected through an effort designed to examine the predictive utility of physiological arousal, ovarian hormones, and stress on headache activity. The effort has been approved by the Institutional Review Board of Wake Forest School of Medicine and is funded by the NIH (NIH/NINDS 1R01NS06525701). At the time of analysis, the

convenience sample consisted of $N = 9$ women who contributed a median of 89 days of stress/hormone measurements (range: 59 to 90) that were collected between July 2009 and July 2011. All participants had regular menstrual cycles and were diagnosed with migraine either with or without aura.

The women completed a daily diary consisting of several instruments, but for the purposes of the analysis contained the Daily Stress Inventory (DSI).¹⁰ The DSI is a well-validated 58-item Likert-type scale that asks a participant to identify stressful events that they have experienced in the last 24 hours. For each item that is endorsed, a stressfulness impact rating is made ranging from 1 (“occurred but was not stressful”) to 7 (“caused me to panic”). Three indices are available from the scale: number of endorsed events (FREQ), sum of ratings (SUM), and average intensity rating for the endorsed items (AIR). All three indices reflect a somewhat different aspect of the daily stress experience so all three were used in the analyses.

In addition, women collected their first morning urine void and stored the sample in their freezer for later assay. The urine was assayed for E1G and PdG using a professionally prepared assay (Immunometrics, UK). The samples were prepared in duplicate and assayed using the standard protocols (E1G EIA Kit, Cat IM113; PDG EIA Cat IM114) of the kit by the same lab technician.

Weather data

Various aspects of weather (e.g., barometer, temperature, etc.) have been linked to headache in numerous surveys and daily diary studies.¹¹ To assess the natural variability in weather conditions, hourly weather data for the period 1/1/2009 to 1/1/2012 were downloaded from the website of National Climatic Data Center (<http://www.ncdc.noaa.gov>). The data were from the Smith Reynolds weather station (319539/93807). The data were examined for the following hourly weather variables: maximum wind speed, minimum visibility (miles), median temperature, median dew point, median station pressure, maximum temperature, minimum temperature, and 24 hour precipitation (inches).

Estimating Similarity in patterns of the triggers

To calculate the similarity, or dissimilarity, between two selected days on a number of headache triggers, we needed a statistical index. Gower's distance (difference) is a metric used in many branches of science to describe distances between entities in multidimensional space.¹² Many such distance measures exist and could have been used, but Gower's is flexible in allowing different data types and distributions that might be encountered in headache trigger research (e.g., nominal, ordinal, interval, or ratio data can be used). In our approach, the index calculates a score that represents the average difference between two days on a group of measurements (e.g., weather) as referenced to the maximal variability in the measurements. In this way, the score ranges from 0 (the two days are identical on all of the measurements) to 1.0 (the difference in the two days represents the maximal variability observed on all of the measurements). We selected $d = 0.10$ to serve as a threshold for similarity based on several considerations. First, headache trigger effects are likely to be subtle, so accurate estimation of these impacts would require a high degree of similarity

between small numbers of occasions and preclude days with even modest differences between them. Second, evaluation of the various series revealed that $d < 0.10$ provided an acceptable similarity in the various measurements that was intuitively valid. Figures 1c, 2c, and 3c display examples of how similar the various measurements are given the distance between them. For example in Figure 1c, even with a $d \sim 0.10$ the maximum temperature for the example day was different by 10 degrees Fahrenheit, with the other measurements being quite similar. This difference occurred because the metric represents the average similarity between the measurements, so variability in the distances are observed across measurements. This fact underscores that differences larger than $d > 0.10$ this are likely to represent substantial confounds for an individual acting as a scientist conducting his or her natural experiment.

Statistical simulation of expected similarity

The choice of similarity threshold and the number of considered triggers substantially impacts the uncovered estimates. The choice to use $d = 0.10$ in the analysis is intended to be viewed simply as an illustration of the difficulties encountered by an individual acting as a scientist in finding similar days to experiment. To illustrate the impact of this arbitrary choice of threshold on the expected number of similar days, we conducted an overly-simplistic mathematical simulation. For the simulation we estimated the probability that a similar day could be found using cutoffs of Gower's distance of $d = 0.10, 0.15$, and 0.20 in a variable number of series ranging from 2 to 20 (this simulates finding a similar day considering between 2 and 20 other potential triggers). Each series was generated using a normal random number generator with $\sim N(0, 1)$ that produced uncorrelated series (i.e., these data do not resemble time-series data in their autocorrelation structure). One thousand replicates were generated for each scenario that each contained $N = 100$ theoretical days. The probability was estimated as the sum of the number of days that could be found below the threshold for each simulated day, divided by the total number of following days (i.e., only unique days are counted).

Statistical Analysis

To determine how often subsequent days were similar to the current day in terms of the pattern of headache triggers, we simply counted the number of days *after* each day in our data that had a distance (using Proc DISTANCE) less than or equal to 0.10 (the calculation would have been the same moving forward or backward in time). These days were then modeled while considering how many days occurred after this day in the set using an intercept-only model. Point estimates and 95% confidence intervals were generated for the stress and hormone data as per 30-day time frames (i.e., a 30-day month). The model was conducted using generalized estimating equations (Proc GENMOD) with a negative binomial distribution for count data using days as a repeated measures nested within individuals and an unstructured covariance matrix. The distributions of similar days across subjects were plotted using a histogram. All analyses were conducted using SAS 9.2 (SAS, Inc., Cary, NC).

Results

It proves difficult to find occasions with similar patterns of headache triggers to the current day in a reasonable time period. This is partly the case because of the sheer number of headache triggers and partly because of the natural variability of these triggers in a typical headache sufferer. Figure 1A displays the similarity of weather over time to a selected day that is representative of median weather conditions. Using a cut-off of $\leq 10\%$ difference for similarity to the present day in seven weather conditions, there was only a median of 2.3 comparable days each year (range: 0 to 27.4) in the years 2009 to 2012 in Winston-Salem, NC (Figure 2B). If weather were the only headache trigger, there would be only 1 out of every 159 days that could be used for a valid potential comparison with the current day.

Weather is not the only headache trigger. It has been stated that “everything” is a headache trigger,¹³ and survey studies support this diversity with factors such as stress, ovarian hormones (menstrual cycle), mood, sleep, diet, physical activity and many others being reported to be perceived triggers by some individuals (See: ¹⁴). Demanding similarity in estrogen and progesterone levels produces only 2.0 days/month (95%CI: 1.9 to 2.2) of similar days (Figure 3a). Concurrent stress, or daily hassles, is an oft-reported headache trigger, with a different pattern of daily variability stemming from three different measures. Alternatively, the alleviation of stress (i.e., “let down”) is often reported among sufferers as a potent trigger. Using these aspects of stress, similar stressful days can only be found about 1.5 days/month (95%CI: 1.2 to 2.9), see Figure 4a.

If finding comparable days in one trigger is difficult, finding similar days using multiple triggers becomes increasingly difficult. If there were no covariation in 12 different triggers considered together, using a statistical simulation we would expect a similar day occurring only once every 684 days (~ 2 years; see Figure 4). Because in the real world, these factors slightly predict themselves (i.e., autocorrelation) and each other (i.e., for example, stress ratings are loosely, but significantly, associated with temperature), our median patient could probably find a comparable day to use for quasi-experimentation slightly more often. However, in the real world these rates are attenuated to even a greater extent because many headache sufferers experience multiple attacks each month that last for more than one day per attack. For these carry-over days, natural experimentation is even more tenuous because an attack is already in progress (see ⁹). If the criteria for similarity were decreased, a similar day could be found once every 26 days (15% dissimilarity) or even once every 4 days (20% dissimilarity), but these days would have substantially different patterns of background triggers.

Discussion

This study revealed that it is difficult to find days with similar patterns of headache triggers when relying on natural variability. The consequences of this realization are that a core assumption regarding causal inference is likely violated during natural experimentation. If an individual were to examine the role of wine in causing their headaches using the memory of several recent wine drinking occasions contrasted with recent non-wine drinking occasions, it is very likely that the considered days were appreciably different on a host of

other factors (e.g., stress levels, hormones, weather). These imbalances in the considered days make it difficult to assign causal status to drinking wine because the effect could also have been due to some other trigger that varied by occasion. The implications of this finding call into question the value of an individual person's efforts to uncover the causes of their headache using only natural experimentation. Further, because most of our knowledge about headache triggers is derived from surveys of individuals, this finding calls into question what we really know about the potency of most headache triggers.

This study was certainly not large in terms of sample size, or diverse in terms of considered triggers or environments. None of the considered triggers were under direct behavioral control in that they were free to vary; food triggers or other triggers that are directly manipulated through choice may exhibit entirely different patterns of variability. This work is simply a thought experiment examining how one important assumption to assign a causal relationship to a variable is expected to be performed when examined using real-world data. In this regard, a larger study with increased diversity is very likely to lead to similar conclusions and similarly highlight the vast complexity of using simple covariation assessments to assign causal-type status to any two variables amongst a sea of candidates.

These similarity rates should not be surprising when considering the low expected degree of chance covariation with increasing numbers of loosely related time-series. However, the low rates of comparable days make the ancient personal quest of searching for headache triggers using natural experimentation of questionable worth. Without the benefit of formal statistical procedures or the experimental method to aggregate the expectation of trigger potency (see: ⁹), it is likely that only extremely potent headache triggers will be detectable given the natural variation in triggers. The quest to identify personal headache triggers is a difficult natural experiment.

Acknowledgement

The data collection and analysis was funded by NIH/NINDS R01NS065257.

References

1. Alvarez WC. Was there sick headache in 3000 BC? *Gastroenterology*. 1945; 5:524. [PubMed: 21012748]
2. Silberstein, SD., Lipton, RB., Goadsby, PJ. Historical Introduction. Martin Dunitz Ltd; London: 2002. p. 1-8.
3. Cheng PW, Novick LR. Covariation in natural causal induction. *Psychol Rev*. 1992; 99:365. [PubMed: 1594730]
4. Hippocrates., Adams, F. The genuine works of Hippocrates. W. Wood and company; New York: 1886. p. 2
5. Martin PR. How do trigger factors acquire the capacity to precipitate headaches? *Behav Res Ther*. 2001; 39:545. [PubMed: 11341251]
6. Kelman L. The triggers or precipitants of the acute migraine attack. *Cephalalgia*. 2007; 27:394. [PubMed: 17403039]
7. Roese NJ. Counterfactual thinking. *Psychol Bull*. 1997; 121:133. [PubMed: 9000895]
8. McKenzie CR, Mikkelsen LA. A Bayesian view of covariation assessment. *Cogn Psychol*. 2007; 54:33. [PubMed: 16764849]

9. Turner DP, Smitherman TA, Martin VT, Penzien DB, Houle TT. Causality and Headache Triggers. Headache. in submission.
10. Brantley PJ, Waggoner CD, Jones GN, Rappaport NB. A Daily Stress Inventory: Development, reliability, and validity. J Behav Med. 1987; 10:61–74. [PubMed: 3586002]
11. Prince PB, Rapoport AM, Sheftell FD, Tepper SJ, Bigal ME. The effect of weather on headache. Headache. 2004; 44:596. [PubMed: 15186304]
12. Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971; 857
13. Blau JN, Thavapalan M. Preventing migraine: a study of precipitating factors. Headache. 1988; 28:481. [PubMed: 3243710]
14. Martin PR, MacLeod C. Behavioral management of headache triggers: Avoidance of triggers is an inadequate strategy. Clin Psychol Rev. 2009; 29:483. [PubMed: 19556046]

Figure 1.

1A. Using Gower's distance, a measure of similarity of multiple measurements that ranges from 0 (identical on all dimensions) to 1.0 (maximal variability in relation to each measurement's range), the similarity of subsequent days following the median weather day are plotted. Five days (black dots) could be found in the 3 years from that exhibited a difference of $d = 0.10$, making them comparably similar to the selected day. 1B. A histogram of the number of similar ($d = 0.10$) days that could be found for each calendar day from Nov 2009 to Jan 2011. 1C. An example of the similarity of two days with a difference of $d = 0.097$ on the considered weather domains.

Figure 2.
2A. Using Gower's distance, the similarity of subsequent days following the median ovarian hormone day are plotted. Two days (black dots) could be found in this 3 month period that exhibited a difference of $d = 0.10$, making them comparably similar to the selected day. 2B. A histogram of the number of similar ($d = 0.10$) days that could be found for each day over a 90-day observation period. 2C. An example of the similarity of two days with a difference of $d = 0.102$ on progesterone (PdG) and estrogen (E1G).

Figure 3.

3A. Using Gower's distance, the similarity of subsequent days following the median stress day are plotted. Four days (black dots) could be found in the 3 month period that exhibited a difference of $d = 0.10$, making them comparably similar to the selected day. 3B. A histogram of the number of similar ($d = 0.10$) days that could be found for each day over a 90-day observation period. 3C. An example of the similarity of two days with a difference of $d = 0.099$ on the sum of the number of daily stress ratings (SUM), the number of stressful events (FREQ), and the average intensity rating of the events (AIR).

Probability of observing a Gower's distance for a given threshold with increasing number of variables

Figure 4.

Increasing the threshold for similarity produces a large increase in the probability that a similar day can be found. For example, considering 12 predictors (like in the present study) results in a similar day being found ($1 / \text{probability}$) once in approximately 684 days, 26 days, or 4 days for a threshold of 0.10, 0.15, and 0.20, respectively.