Development and Internal Validation of a Multivariable Prediction Model for Individual Episodic Migraine Attacks Based on Daily Trigger Exposures

Katherine K. Holsteen, PhD,

Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA

Michael Hittle, BS,

Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA

Meredith Barad, MD,

Department of Anesthesia, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

Lorene M. Nelson, PhD, MS

Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA

Abstract

Objective.——To develop and internally validate a multivariable predictive model for days with new-onset migraine headaches based on patient self-prediction and exposure to common trigger factors.

Address all correspondence to L. M. Nelson, Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA, lnelson@stanford.edu.

Background.——Accurate real-time forecasting of one's daily risk of migraine attack could help episodic migraine patients to target preventive medications for susceptible time periods and help decrease the burden of disease. Little is known about the predictive utility of common migraine trigger factors.

Methods.——We recruited adults with episodic migraine through online forums to participate in a 90-day prospective daily-diary cohort study conducted through a custom research application for iPhone. Every evening, participants answered questions about migraine occurrence and potential predictors including stress, sleep, caffeine and alcohol consumption, menstruation, and self-prediction. We developed and estimated multivariable multilevel logistic regression models for the risk of a new-onset migraine day vs a healthy day and internally validated the models using repeated cross-validation.

Results.——We had 178 participants complete the study and qualify for the primary analysis which included 1870 migraine events. We found that a decrease in caffeine consumption, higher self-predicted probability of headache, a higher level of stress, and times within 2 days of the onset of menstruation were positively associated with next-day migraine risk. The multivariable model predicted migraine risk only slightly better than chance (within-person C-statistic: 0.56, 95% CI: 0.54, 0.58).

Conclusions.——In this study, episodic migraine attacks were not predictable based on self-prediction or on self-reported exposure to common trigger factors. Improvements in accuracy and breadth of data collection are needed to build clinically useful migraine prediction models.

Keywords

migraine; headache; predict; forecast; trigger

INTRODUCTION

The unpredictable nature of migraine attacks creates significant challenges for individuals living with episodic migraine. In theory, accurate real-time forecasts of oncoming migraine attacks could help patients to target medication use and other preventive strategies for susceptible time periods and help decrease the disease burden.[1,2] Several different types of data may correlate with day-to-day migraine risk, and thus may help to predict migraine;[1,2] these include trigger factors,[3,4] premonitory symptoms,[5–7] self-prediction,[5] and physiological signals.[8,9] Though trigger factors have been formally studied for at least 60 years,[3] and certain transient exposures (eg, psychological stress, changes in sleep, and menstrual cycles) are commonly thought to bring about migraine attacks, as of yet there is limited work toward assessing their predictive utility.[1,10–12] Predictive modeling of migraines is promising not only for the longer term goal of applying real-time migraine prediction for individual patients, but also in the near term to set a baseline for predictive accuracy and to understand the relevance of different factors to this end.[13]

The current best effort toward migraine prediction based on self-reported data, by Houle et al., used a model based on stressful event frequency and current headache status, and estimated an out-of-sample concordance statistic (area under the receiver-operating characteristic curve) of 0.65.[11] This work laid important groundwork for future

migraine prediction studies and sparked several commentaries or re-analyses describing the importance of and room for improvement in headache forecasting.[1,12,14] It also brought to light key areas for methodological improvement in predicting migraine days. For one, as the authors acknowledged, it would be intuitively useful to focus on prediction of new attacks during non-headache times (rather than predicting next-day headache regardless of current headache status).[11] Second, migraine triggers may stack up in an additive way to increase migraine risk, therefore, including multiple different trigger factors in an additive model might boost predictive accuracy.[1,15] Third, with a multilevel data structure and an intended application of predicting day-to-day migraine risk within individual patients, it is most fitting to evaluate predictive performance using within-person metrics.[16,17]

In light of these opportunities for improvement, the objective of this study was to develop and internally validate a multivariable predictive model for days with new-onset episodic migraine attacks, based on self-reported exposure to specific trigger factors, headache self-prediction, and passively collected weather data.

METHODS

Subject Recruitment and Inclusion and Exclusion Criteria.—

Between October 2018 and March 2019, we used ResearchMatch, Facebook Ads, and other online forums to recruit patients with episodic migraine for a 90-day, iPhone app-based prospective daily-diary cohort study. We assessed study eligibility via an online survey (Supporting Section S1). Eligible participants were at least 18 years of age, living in the United States, and had an iPhone 5s or later. They had suffered from severe headaches for at least 1 year, had migraine headaches in the past 3 months as assessed by the 3-question ID Migraine diagnostic tool,[18] and had 2–10 headaches per month with fewer than 15 headache days per month. We administered the eligibility survey and managed study data using REDCap (Research Electronic Data Capture) electronic data capture tools hosted at our institution. [19,20]

Data Collection.—

The iPhone application for daily data collection was programmed using ResearchKit (Apple; Cupertino, CA, USA) and transmitted study data to a secure server managed by Stanford University. After completing the online eligibility survey, eligible persons were invited to download the application onto their personal iPhones. Within the app, they could complete informed consent, enroll in the study, and submit all study-related data. At baseline, the application presented one-time surveys on health history and headache history. For the next 90 days, the application administered nightly surveys about headache timing and symptoms, sleep timing, perceived stress, caffeine and alcohol consumption, medication changes, menstruation, premonitory symptoms, and headache self-prediction (survey questions in Supporting Section S2). We designed the daily surveys to measure factors that were (1) common potential trigger factors or premonitory symptoms in the literature; (2) affirmed by our migraine specialist co-author (MB); and (3) quantifiable within a simple daily survey on a small phone screen. Participants were asked to complete the survey every evening, close to their bedtime but before 3 am. Completed participants received $70 in Amazon

gift cards and a personal summary of their data. The final participant completed the 90-day follow-up in June 2019. To obtain daily weather data, we used the Google Maps Geocoding Application Programming Interface (API)[21] and the Dark Sky weather API[22] to gather daily summaries of temperature, pressure, wind, precipitation, and humidity for the participant's home ZIP code. This study was approved by the Stanford University Institutional Review Board. This is the primary analysis of these data and the analysis was preplanned.

Data Analysis.—

Outcome.——Each participant-reported headache was classified as a migraine if it met the following criteria, taken from a recent controlled trial of a monoclonal antibody for episodic migraine prevention:[23]

- At least 30 minutes in duration and meets criteria C *or* D for migraine without aura in the International Classification of Headache Disorders, 3rd edition (ICHD-3),[24]

or

- Participant reported taking an ergotamine or triptan for this headache.

These criteria were intentionally looser than the ICHD-3 migraine diagnostic criteria in order to more sensitively capture migraines and probable migraines among patients who may already have a working rescue medication regimen. Other recent trials have applied similar criteria.[25,26]

We defined the units of analysis as "social days" starting at 3:00 am one day and ending at 2:59 am the next day.[27] After identifying the timing of migraine headaches, we classified each participant-day as At-Risk (as of day start, 3:00 am, no migraine was present within the past 24 hours) or Recovery (as of 3:00 am, there had been a migraine present within the past 24 hours).[28] Recovery days were excluded from the analysis.

Missing Data.——Because the surveys generally did not allow skipping questions, surveys were either entirely present or entirely missing for each person-day. To minimize missing data, participants were allowed, but not encouraged, to complete surveys up to 24 hours late. Missing data were multiply imputed 5 times using the R package *Amelia*.[29] The imputation model included the outcome and all of the potentially relevant predictors, including leads and lags, from the surveys and weather data. Imputed values were used to predict migraine on subsequent days but outcomes were not predicted for days with missing surveys.

Sample Size.——At least 10–20 events per variable are necessary for predictive modeling with logistic regression.[30,31] With approximately 180 participants, and an average of 10 migraines per person, there are 1800 total events permitting 90–180 candidate predictors.

Candidate Predictors.——In order to predict the current day's migraine risk, day-level predictors were based on survey responses from the previous day(s), with the exception of menstruation and workday/off day status, which were unlikely to be affected by migraine occurrence. In the primary and penalized models, we included menstruation (lag 0), an

indicator for a workday or day off of work (lag 0 and lag 1), and self-prediction of headache probability (lag 1). For each of 4 categories of trigger factors – sleep, stress, caffeine/alcohol, and weather – we defined between 1 and 4 exposure variables (Table 1) based on the raw survey responses or raw passively collected data without reference to the headache outcomes. For each trigger category, we considered several candidate model specifications which varied the predictor construct, timing (lag) of the exposure measurements, and functional form of the predictor/outcome relationship (Table 2). We considered linear relationships and nonlinear relationships parametrized by natural cubic splines with 3 knots at the 10th, 50th, and 90th percentiles.[32] These predictor constructs, lags, and functional forms were predetermined without examination of the data. Each model also included person-level means of each day-level factor, to distinguish between within-person and between-person correlates of migraine risk.[33,34] To help explain between-person variation in risk, each model included 3 self-reported person-level factors collected at baseline with reference to the past 3 months: number of monthly headaches, days per month with headache, and whether the participants' headaches had changed in frequency or severity in the past 3 months.

In the Supporting Section S3, we describe the predictor susceptibility to measurement error and the timing of collection of each data element with respect to the intended time of risk estimation. We did not include the measured premonitory symptoms as predictors because we found that their day-to-day variation correlated strongly with concurrent migraine attacks but had essentially no correlation with future migraines. We did not include lagged migraine status because descriptive plots revealed a relatively constant hazard of migraine over the time since recovery from the last one (Supporting Section S4).

Model Development.——The primary prediction model used multilevel logistic regression with varying intercepts. As an initial step toward the future goal of accurate within-person migraine prediction, the 2-fold purpose of this model was to (1) optimize within-person predictive accuracy and (2) interpret the estimated coefficients with proper distinction between the within-person and between-person variation present in this multilevel data. We carried out variable selection within each of 4 trigger categories – sleep, stress, caffeine/alcohol, and weather – by using the Akaike information criterion (AIC) to select the best-fitting specification among those listed in Table 2. We combined all of the predictors from those 4 best-fitting models, along with the prespecified menstruation, workday, and self-prediction variables, to compose the primary model.

As an alternative method, we used grouped-lasso penalized logistic regression which carried out one-step variable selection and estimation on all of the predictors at once.[35] The purpose of this model was simply to optimize within-person predictive accuracy. We submitted 95 total candidate predictors including 3 lags of each raw exposure value (listed in Table 1) and 2 lags of their change scores, with natural cubic splines for each of the lag-1 predictors. The additional lags and spline terms were included to allow for the possibility of various timing of maximal correlation between exposures and outcome risk and for potential cumulative effects of exposures over multiple days. The grouped-lasso grouped the pairs of linear terms that parametrized each spline so that both terms were either included or excluded together for all of the lag-1 predictors. We selected hyperparameters λ (penalty strength, of 25 values

automatically generated by the fitting routine) and α (ratio of penalty on the group-lasso vs ridge, of values 0.25, 0.5, 0.75, and 1) to minimize the AIC. We used the R package *grpreg* to select the hyperparameters and estimate the grouped-lasso models.[36]

Model Evaluation.——We evaluated the out-of-sample predictive performance of these model development procedures using 10-fold cross-validation repeated 10 times, applied at the participant level to preserve the multilevel structure of the data. Cross-validation of the multilevel model used the average person-specific intercept to calculate the predicted probabilities of migraine for held-out participants. The statistic of most interest was the within-person C-statistic, defined as a meta-analytic summary of the person-specific proportions of discordant pairs of at-risk days that correctly had a higher predicted risk for the day with migraine.[37] This statistic is more relevant than the standard C-statistic for daily risk prediction within persons.[16] We also evaluated within-person calibration estimated via multilevel logistic regression[17,38] and the variance of the predicted probabilities. Statistical significance was determined by $P < .05$, 2-tailed testing.

Sensitivity Analyses.——We conducted sensitivity analyses to assess the impact of key methodological choices on the estimated coefficients and predictive performance of our primary and penalized models. We executed the same data preparation, model development, and internal validation procedure with variations in the input data. First, to assess the impact of potential misclassification of migraine headaches, we defined the outcome to include all reported headaches (migraine or not). Second, we included only migraine headaches as defined by the ICHD-3 criteria (1.1 migraine without aura or 1.5.1 probable migraine without aura). Third, to assess any impact of the exclusion of post-migraine recovery days from the analysis, we defined at-risk days as those migraine-free as of 3 am. Fourth, to examine whether predictive accuracy might be improved in a more homogeneous group of patients, we included only the female participants who reported having menstrual cycles. Finally, we examined the impact of excluding person-level means of the predictor variables from the set of candidate predictors. We used R version 3.6.0 for all analyses.[39]

RESULTS

In total, 290 participants were enrolled in the study over the course of the 6-month recruitment period (Fig. 1). Of these, 62 (21%) were withdrawn during the study (primarily for failing to complete 12 of the first 14 evening surveys on time) and another 50 were excluded from analysis due to chronic headache or a medication change within the first 30 days of the study. There were 178 participants (166 women and 12 men) available for the primary analysis (Table 3).

Participants were on average 37 years old, and 42% (74/176) had a graduate degree, including 53% (39/73) of those recruited from ResearchMatch. Participants reported an average Migraine Disability Assessment Test (MIDAS) score of 32.6 (indicating severe headache-related disability), an average of 7 headache days per month, and 16 years since migraine diagnosis. Forty-three percent of the sample (76/176) had never taken daily headache preventive medications.

Available Data.—

Out of 15,418 person-days of follow-up (an average of 86.6 days per participant), only 197 days (1.3%) had a missing survey and 1376 (9%) had their survey completed up to 24 hours late. The distributions of predictor variables over all person-days, within-person means, and within-person standard deviations are given in the Supporting Section S6.

Number of Outcomes.—

The surveys included reports of 3546 headaches, of which 2831 (80%) were classified as migraine. After combining migraines that were not separated by a complete migraine-free recovery day, and removing recovery days, we observed a total of 1870 individual migraine attacks on 10,696 at-risk days. Participants experienced an average of 10.5 migraines each, with a range of 2–22 (corresponding to a 1% to 48% daily risk of migraine, Fig. 2).

Primary Model.—

Of the candidate specifications listed in Table 2, the best-fitting category-specific models were: Stress: model 1S; Sleep: model 1L; Caffeine/Alcohol: model 3; and Weather: model 5L. Figure S4 shows the estimated odds ratios (ORs) from these category-specific models. The primary model combined the predictors from these best category-specific models for a total of 34 predictors (including an intercept) and an intraclass correlation of 0.025. Statistically significant predictors of higher migraine risk for a given person-day included caffeine consumption (ie, higher caffeine 2 days ago and lower caffeine 1 day ago), being within 2 days of the start of a menstrual period, self-prediction of higher headache probability on that day, and higher stress ratings 1 day ago (Fig. 3). Close similarity between the estimated ORs from the separate models (Fig. S4) and the combined primary model (Fig. 3) indicated that adjustment for the other categories did not meaningfully affect the coefficient estimates.

This model yielded a cross-validated within-person C-statistic of 0.56 (95% CI: 0.54, 0.58) (Table 4). The average calibration-in-the-large intercept was 0 (95% CI: −0.08, 0.07). The weak calibration slope of 0.73 (95% CI: 0.57, 0.90) was less than 1, suggesting an overfitted model. Figure 4 shows that the predicted probabilities clustered tightly around the overall observed migraine risk with a standard deviation of approximately 0.06. The full coefficient estimates are listed in Table S4.

Penalized Model.—

The grouped-lasso model hyperparameters were selected to be $\alpha = 1$ (corresponding to an exclusively group-lasso penalty, and no ridge penalty) and $\lambda = 0.003$ (a moderately strong penalty). Of the 95 candidate predictors, the procedure selected between 38 and 42 nonzero predictors on each imputation. The full results of the coefficient estimates are given in Table S5.

Overall the penalized and primary models had very similar findings in terms of strongest coefficient estimates and predictive performance. The day-level predictors with the largest absolute log-ORs were menstruation days 1 ± 2 (OR = 1.28), headache self-prediction (OR for 2-SD increase = 1.22), the lag-1 change in cups of caffeinated beverages (OR for 10th vs

50th percentile = 1.21), and the lag-1 stress rating (OR for 90th vs 50th percentile = 1.15). The penalized model yielded a within-person C-statistic of 0.56 (95% CI: 0.54, 0.58) (Table 4).

Sensitivity Analyses.—

The 5 sensitivity analyses included between 1266 and 2462 headache outcomes (overall event risks of 0.13–0.22) (Table 5). Results were similar across all sensitivity analyses in terms of the strongest predictors and the within-person concordance statistics.

DISCUSSION

In this study, we developed and internally validated a multivariable predictive model for days with new-onset migraine headaches among patients with episodic migraine. We hypothesized that predictive accuracy might be improved with the use of multiple different features commonly thought to trigger or predict migraines. The model performed slightly better than chance in terms of within-person discrimination, but the performance was inadequate for practical use.

This study contributes a relatively large, comprehensive, prospective analysis to the nascent literature on migraine prediction. Even with limited predictive ability, this study helps to (1) evaluate the predictive relevance of trigger factors that are commonly cited and/or found associated with migraine risk in previous small studies; (2) compare various constructs of such factors in terms of predictive ability; and (3) measure the room for improvement in scientific understanding of migraine predictors.[13] We found that a multivariable model with several lags of various constructs of potential predictors performed poorly. In the current model, many of the candidate predictors showed a weak and/or noisy correlation with migraine risk, which contributed to the poor discrimination. The day-level factors that most strongly associated with higher migraine risk in this sample were a reduction in caffeine consumption yesterday compared to the previous day, higher self-predicted headache probability, higher stress rating yesterday, and being within 2 days of the onset of menstruation (measured for the 124 female participants who reported having menstrual cycles). These findings corroborate previous literature describing similar prospective associations[11,28,40–42] but their effect sizes were too small to move the predicted risks very far from the overall observed migraine risk. Future studies may take into account the direct relationship between ORs and predictive discrimination to evaluate the utility of various potential predictors *a priori* based on hypothesized effect sizes.[43]

The within-person and overall discrimination of the current model were worse than the previously published overall discrimination for a model based only on yesterday's perceived stress and headache status (Houle et al[11]). Several methodological choices may help explain this difference. *First*, the current model only predicted new-onset migraines, and excluded days with continuing migraine. Predicting only new-onset migraines is more difficult statistically since it does not take advantage of the obvious first-order autocorrelation arising from multiday headache attacks.[44] We maintain that predicting only new-onset migraines is intuitively more relevant and should be the standard for future migraine prediction efforts. *Second,* we evaluated the current model using a within-person discrimination statistic. In

multilevel data, the standard overall C-statistic must be interpreted with care, because it combines discordant pairs of days within persons *and* discordant pairs of days across different persons, and weights these 2 categories based on the sample composition.[37] Greater variation in headache risk across persons in the sample can boost an overall C-statistic, if the model can estimate person-level average risk with accuracy, but this has no bearing on the ability of the model to estimate differences in headache risk from day to day for any person. The within-person statistic is more relevant for the goal of accurately predicting migraine risk for a particular person.[16] *Third*, Houle et al[11] utilized a much longer and more comprehensive stress measurement (the 58-item Daily Stress Inventory) than the current study. We used an abbreviated single-item Self-Reported Stress Score[45] to minimize participant burden of the full nightly survey. This coarser measurement may have led to noisier or less variable stress measurements and attenuated the statistical effect of stress on migraine risk.

This study had several strengths. *First*, we employed state-of-the-art mobile app methodology to collect daily survey data in an efficient and user-friendly way, which yielded 90% on-time completion of the daily surveys. *Second,* we recruited a relatively large sample size with a nearly 80% study completion rate, making this the largest study yet of migraine prediction in terms of the number of participants. *Finally,* we implemented careful statistical analysis – including the use of multiple different trigger factors, consideration of various constructs and lags, predicting new-onset migraines, evaluating predictive performance within persons, adjusting for person-level means in the models, and examining the effect of key choices in sensitivity analyses – in order to achieve a more accurate, if less optimistic, representation of the predictive utility of the available data.

This study had several limitations that must be taken into account. *First*, the participants were self-selected from a highly selective study population who saw the recruitment advertisements and owned iPhones. They may have been more enthusiastic or more educated than the general population of episodic migraine patients who would fit our eligibility criteria, and may have been more likely to complete the study than might be expected in another sample. Of note, if the participants in this sample were more or less susceptible to common triggers than the target population, then the main findings would be biased and would not properly generalize to a broader population of episodic migraine patients. *Second*, the mobile-app-based surveys were susceptible to measurement error due to the brief, simple questions in the end-of-day survey (as discussed in Supporting Section S3). Noisy data for the key predictors may have hindered the model performance. Future studies might seek to validate daily diary questions and quantify the effect of measurement error on predictive accuracy. *Third*, the population of episodic migraine patients is heterogeneous. The relationships between trigger factors and migraine risk may be more similar among a subset of patients with similar symptoms, perceived triggers, and comorbidities,[46,47] and thus prediction may be superior in such a subgroup. In this study, we did not pursue n-of-1 personalized migraine prediction because of the extended duration of data collection that would have been required to collect the standard 10–20 events per person per candidate feature. In theory, a predictive model that is fairly accurate at the population or subgroup level could form a starting point to update and personalize for individual patients in the future.[48,49]

In the context of increasing popularity of mobile apps for research data collection, this work illustrates that the use of a mobile app in itself does not guarantee strong data quality or predictive accuracy. Researchers considering the use of mobile apps for future studies should carefully evaluate the available options for data collection methodology and seek to optimize data quality, feasibility, and participant compliance for their specific objectives and study design. Ongoing expansion of resources for simpler and cheaper research app development, maintenance, and data storage may enable a growing number of studies to effectively employ mobile apps. Whether collecting data with cutting-edge technology or traditional methods, future studies must rely on established epidemiologic principles in order to yield information and impact.

There is a great deal of room for improvement in the development of accurate, real-time migraine prediction. To explain more of the variation in day-to-day migraine risk, future studies may need to explore additional common trigger factors or premonitory symptoms such as diet, hydration, mood, and perception of odors or light.[3] It may be helpful to measure relevant factors and generate predictions several times per day, to gain more accurate measurements and take advantage of intra-day effects. Additional passive data collection will help maximize accuracy and ease participant burden. For example, wearable devices could measure sleep timing and quality with actigraphy, and short-term stress response based on heart rate variability, breathing rate, and skin temperature. A parallel avenue for improvement could be to focus on prediction within on a subgroup of episodic migraine patients who appear to have similar underlying pathophysiology; for example, women with menstruation-related migraine, or patients with certain classes of comorbidities such as metabolic or psychiatric conditions.

CONCLUSION

A multivariable prediction model incorporating several lags and various constructs of commonly cited migraine triggers yielded predictive performance barely better than random in this sample of episodic migraine patients. Incorporating additional predictors, improving the accuracy and frequency of measurement, and studying a more homogeneous migraine population may enable future predictive models to perform better, enable targeted medication use, and reduce the unpredictability of episodic migraine attacks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

Study data were collected using the mHealth platform at Stanford University. The mHealth platform is developed and operated by Stanford Medicine Research IT team. The mHealth platform services at Stanford are subsidized by Stanford School of Medicine Research Office.

Study data were managed using REDCap electronic data capture tools hosted at Stanford University. The Stanford REDCap platform is developed and operated by Stanford Medicine Research IT team. The REDCap platform services at Stanford are subsidized by (1) Stanford School of Medicine Research Office; and (2) the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1 TR001085.

Abbreviations:

|	
AIC	Akaike Information Criterion, a measure of model fit which includes a penalty for the number of estimated parameters. Smaller AIC values indicate better fit, App smartphone application
API	Application Programming Interface, a set of functions allowing the engineering of software applications that access the features or data of another system or service
CI	confidence interval, C-statistic concordance statistic, a measure of discrimination in a predictive model, equal to the area under the receiver-operating characteristic curve
HIPAA	Health Insurance Portability and Accountability Act of 1996, which consists of data privacy and security provisions to protect medical information
ICHD-3	International Classification of Headache Disorders, 3rd edition
OR	odds ratio
MIDAS	Migraine Disability Assessment Test
PC	principal component
SD	standard deviation

REFERENCES

1. Lipton RB, Pavlovic JM, Buse DC. Why migraine forecasting matters. Headache 2017;57:1023–1025. [PubMed: 28699331]

2. Pavlovic JM, Buse DC, Sollars CM, Haut S, Lipton RB. Trigger factors and premonitory features of migraine attacks: Summary of studies. Headache 2014;54:1670–1679. [PubMed: 25399858]

3. Pellegrino ABW, Davis-Martin RE, Houle TT, Turner DP, Smitherman TA. Perceived triggers of primary headache disorders: A meta-analysis. Cephalalgia 2018;38:1188–1198. [PubMed: 28825314]

4. Peroutka SJ. What turns on a migraine? A systematic review of migraine precipitating factors. Curr Pain Headache Rep 2014;18:454. [PubMed: 25160711]

5. Giffin N, Ruggiero L, Lipton R, et al. Premonitory symptoms in migraine: An electronic diary study. Neurology 2003;60:935–940. [PubMed: 12654956]

6. Schoonman GG, Evers DJ, Terwindt GM, van Dijk JG, Ferrari MD. The prevalence of premonitory symptoms in migraine: A questionnaire study in 461 patients. Cephalalgia 2006;26:1209–1213. [PubMed: 16961788]

7. Kelman L The premonitory symptoms (prodrome): A tertiary care study of 893 migraineurs. Headache 2004;44:865–872. [PubMed: 15447695]

8. Siirtola P, Koskimäki H, Mönttinen H, Röning J. Using sleep time data from wearable sensors for early detection of migraine attacks. Sensors 2018;18.

9. Pagán J, De OMI, Gago A, et al. Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data. Sensors 2015;15: 15419–15442. [PubMed: 26134103]

10. Hoffmann J, Schirra T, Lo H, Neeb L, Reuter U, Martus P. The influence of weather on migraine – Are migraine attacks predictable? Ann Clin Transl Neurol 2015;2:22–28. [PubMed: 25642431]

11. Houle TT, Turner DP, Golding AN, et al. Forecasting individual headache attacks using perceived stress: Development of a multivariable prediction model for persons with episodic migraine. Headache 2017;57:1041–1050. [PubMed: 28699328]

12. Turner DP, Lebowitz AD, Chtay I, Houle TT. Forecasting migraine attacks and the utility of identifying triggers. Curr Pain Headache Rep 2018;22. [PubMed: 29541938]

13. Shmueli G To explain or to predict? Stat Sci 2010;25:289–310.

14. Turner DP, Lebowitz AD, Chtay I, Houle TT. Headache triggers as surprise. Headache 2019;59:495–508. [PubMed: 30919414]

15. Spierings ELH, Donoghue S, Mian A, Wöber C. Sufficiency and necessity in migraine: How do we figure out if triggers are absolute or partial and if partial, additive or potentiating? Curr Pain Headache Rep 2014;18:455. [PubMed: 25163437]

16. van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y. Assessing discriminative ability of risk models in clustered data. BMC Med Res Methodol 2014;14:5. [PubMed: 24423445]

17. Wynants L, Vergouwe Y, van Huffel S, Timmerman D, van Calster B. Does ignoring clustering in multicenter data influence the performance of prediction models? A simulation study. Stat Methods Med Res 2018;27:1723–1736. [PubMed: 27647815]

18. Cousins G, Hijazze S, van De Laar FA, Fahey T. Diagnostic accuracy of the ID migraine: A systematic review and meta-analysis. Headache 2011;51:1140–1148. [PubMed: 21649653]

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–381. [PubMed: 18929686]

20. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208. [PubMed: 31078660]

21. Google Inc. Google Maps Platform: Geocoding API 2019. Available at: https://developers.google.com/maps/documentation/geocoding/start. Accessed April 3, 2020.

22. The Dark Sky Company LLC. Dark Sky API 2019. Available at: https://darksky.net/dev. Accessed April 3, 2020.

23. Goadsby PJ, Reuter U, Hallström Y, et al. A controlled trial of erenumab for episodic migraine. N Engl J Med 2017;377:2123–2132. [PubMed: 29171821]

24. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. Cephalagia 2018;38:1–211.

25. Stauffer VL, Dodick DW, Zhang Q, Carter JN, Ailani J, Conley RR. Evaluation of galcanezumab for the prevention of episodic migraine: The EVOLVE-1 randomized clinical trial. JAMA Neurol 2018;75:1080–1088. [PubMed: 29813147]

26. Dodick DW, Silberstein SD, Bigal ME, et al. Effect of fremanezumab compared with placebo for prevention of episodic migraine a randomized clinical trial. JAMA 2018;319:1999–2008. [PubMed: 29800211]

27. Burg MM, Schwartz JE, Kronish IM, et al. Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional stress-exercise association at the group and person (n of 1) level. Ann Behav Med 2017;51:799–809. [PubMed: 28290065]

28. Turner DP, Smitherman TA, Penzien DB, Porter JAH, Martin VT, Houle TT. Nighttime snacking, stress, and migraine activity. J Clin Neurosci 2014;21:638–643. [PubMed: 24417796]

29. Honaker J, King G, Blackwell M. Amelia: A Program for Missing Data. R package version 1.7.5 2018.

30. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 2017;26:796–808. [PubMed: 25411322]

31. Wynants L, Bouwmeester W, Moons KGM, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. J Clin Epidemiol 2015;68:1406–1414. [PubMed: 25817942]

32. Harrell FE. General aspects of fitting regression models. In: Regression Modeling Strategies, 2nd ed. New York, NY: Springer; 2015:24–27.

33. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: A overview. Ann Intern Med 2001;135:112–123. [PubMed: 11453711]

34. Berlin JA, Kimmel SE, Ten Have TR, Sammel MD. An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. Biometrics 1999;55:470–476. [PubMed: 11318202]

35. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B Stat Methodol 2006;68:49–67.

36. Breheny P, Zeng Y. grpreg: Regularization Paths for Regression Models with Grouped Covariates. R package version 32–1. 2019.

37. van Oirbeek R, Lesaffre E. Assessing the predictive ability of a multilevel binary regression model. Comput Stat Data Anal 2012;56:1966–1980.

38. Bouwmeester W, Twisk JW, Kappen TH, van Klei WA, Moons KG, Vergouwe Y. Prediction models for clustered data: Comparison of a random intercept and standard regression model. BMC Med Res Methodol 2013;13:19. [PubMed: 23414436]

39. R Core Team. R: A Language and Environment for Statistical Computing 2019. Available at: https://www.r-project.org/.

40. Martin PR. Stress and primary headache: Review of the research and clinical management. Curr Pain Headache Rep 2016;20:45. [PubMed: 27215628]

41. Martin VT, Vij B. Diet and headache: Part 1. Headache Curr 2016;56:1543–1552.

42. Macgregor EA. Menstrual migraine: A clinical review. J Fam Plann Reprod Heal Care 2007;33:36–47.

43. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol 2012;12:82. [PubMed: 22716998]

44. Houle TT, Penzien DB, Rains JC. Time-series features of headache: Individual distributions, patterns, and predictability of pain. Headache 2005;45:445–458. [PubMed: 15953261]

45. Lipton RB, Buse DC, Hall CB, et al. Reduction in perceived stress as a migraine trigger: Testing the "let-down headache" hypothesis. Neurology 2014;82: 1395–1401. [PubMed: 24670889]

46. Schürks M, Buring JE, Kurth T. Migraine features, associated symptoms and triggers: A principal component analysis in the Women's Health Study. Int Headache Soc 2011;31:861–869.

47. Lipton RB, Serrano D, Pavlovic JM, et al. Improving the classification of migraine subtypes: An empirical approach based on factor mixture models in the American Migraine Prevalence and Prevention (AMPP) Study. Headache 2014;54:830–849. [PubMed: 24527745]

48. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76–86. [PubMed: 18083464]

49. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: A study on sample size and shrinkage. Stat Med 2004;23:2567–2586. [PubMed: 15287085]

50. Peris F, Donoghue S, Torres F, Mian A, Wöber C. Towards improved migraine management: Determining potential trigger factors in individual patients. Cephalalgia 2017;37:452–463. [PubMed: 27179352]

51. Park JW, Chu MK, Kim JM, Park SG, Cho SJ. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. PLoS ONE 2016;11:e0149577. [PubMed: 26901341]

52. Scheidt J, Koppe C, Rill S, Reinel D, Wogenstein F, Drescher J. Influence of temperature changes on migraine occurrence in Germany. Int J Biometeorol 2013;57:649–654. [PubMed: 22895651]

53. Goadsby PJ. Something expected, something unexpected. Neurology 2014;82:1388–1389. [PubMed: 24670890]

54. van Dongen HPA, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003;26:117–126. [PubMed: 12683469]

55. Hamilton NA, Affleck G, Tennen H, et al. Fibromyalgia: The role of sleep in affect and in negative event reactivity and recovery. Health Psychol 2008;27:490–497. [PubMed: 18643007]

56. Schilling OK, Diehl M. Reactivity to stressor pile-up in adulthood: Effects on daily negative and positive affect. Psychol Aging 2014;29:72–83. [PubMed: 24660797]
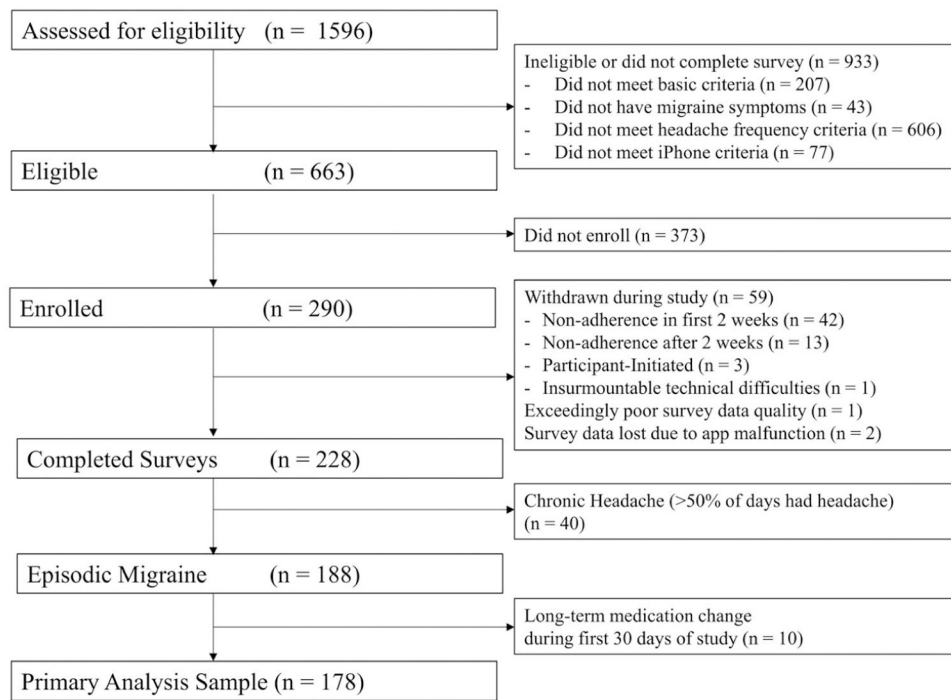
Fig. 1.—.
Participant flow diagram.
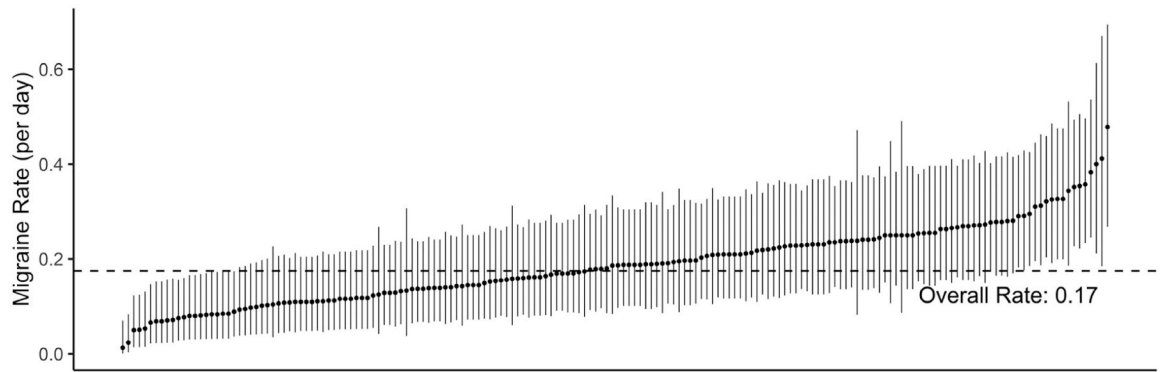
Fig. 2.—.

Migraine rates across participants. The raw observed migraine rate (number of new-onset migraine days, divided by number of at-risk days) for each participant is marked with a dot, and ranked in order from lowest to highest. These rates were calculated using only the raw analytic dataset before any regression modeling. The 95% binomial confidence interval for each participant's migraine rate is marked with a line.
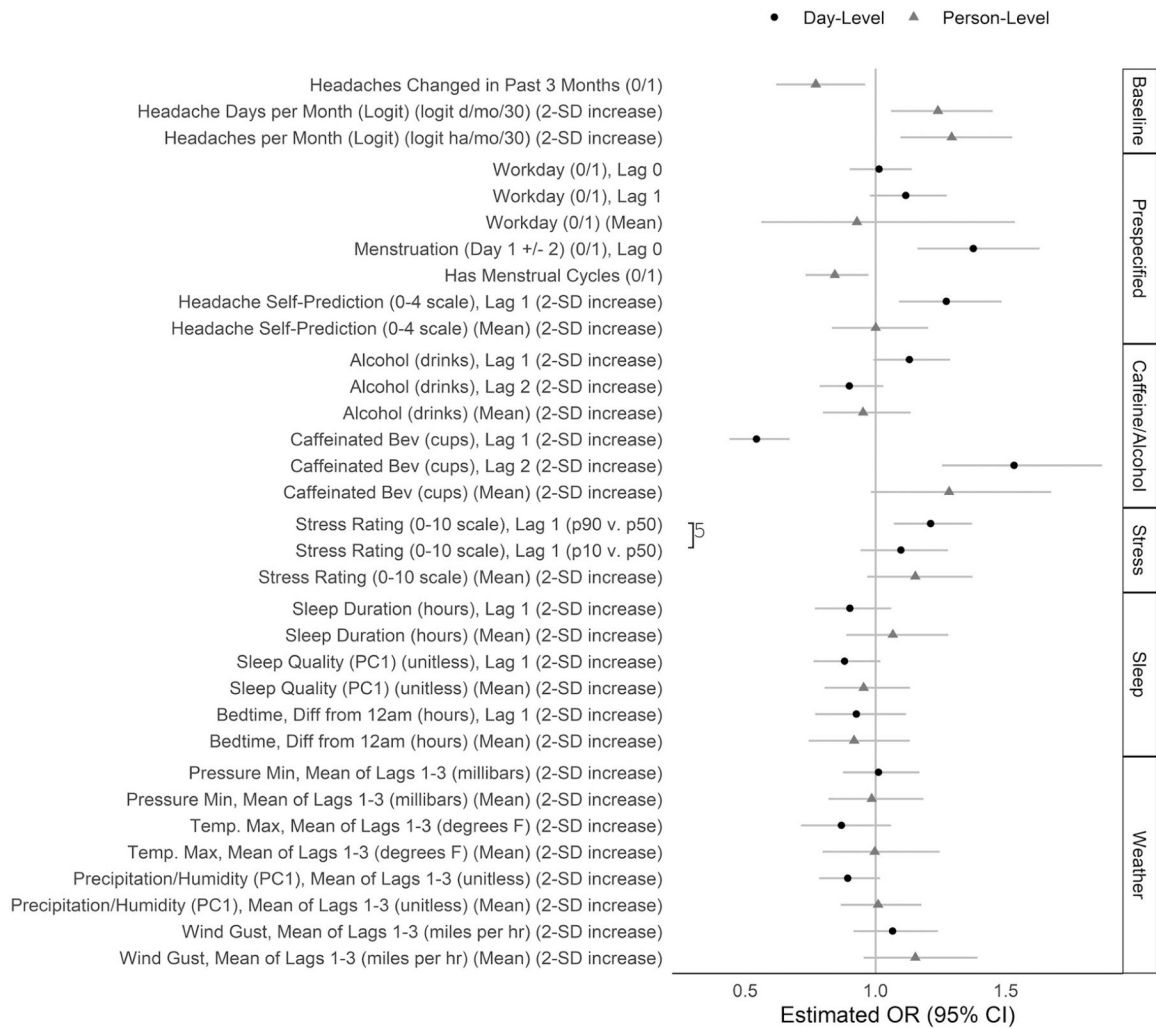
Fig. 3.—.

Coefficient estimates from the primary model. The coefficients and 95% Wald confidence intervals were estimated in the primary multilevel logistic regression model for migraine risk including all 34 predictors shown here plus an intercept. Predictors were measured at the day level (varying across persons and across days within persons) (black) or the person level (varying across persons, but constant across days within persons) (gray). Lag 0 denotes the day for which migraine risk is predicted; lag 1 denotes the day before, and lag 2 denotes 2 days prior. Predictors modeled by 3-knot natural cubic splines were represented in the model by 2 separate linear terms and are shown in this plot by 2 separate estimates to summarize the nonlinear relationship: odds ratios for the 90th vs 50th, and 50th vs 10th percentiles of the overall distribution of the predictor. On this plot, these pairs of terms are connected by a "]" with an integer denoting the number of imputations (of 5) in which the pair of linear terms was jointly statistically significant. The data for this plot are given in Table S4.
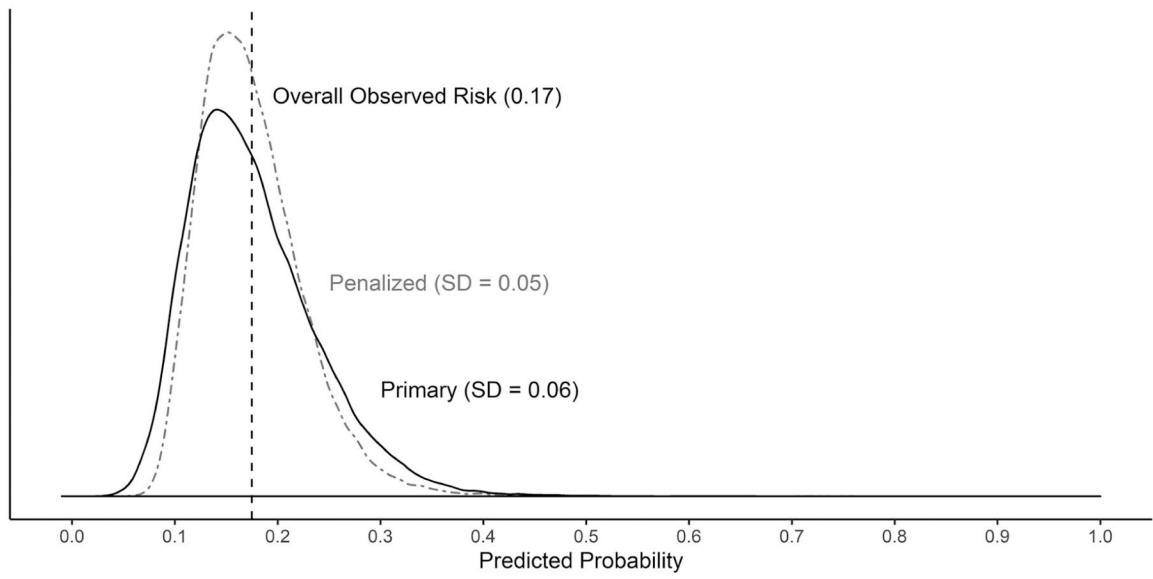
Fig. 4.—.

Distribution of predicted probabilities for primary and penalized models.

Table 1.—

Trigger Exposures in Each Category

Category	Exposure	Survey Questions Used to Construct Exposure
Sleep	Sleep quality (1 PC: Higher scores correspond to higher sleep latency and awake duration, and lower sleep quality)	16, 18, 20
	Bedtime (difference from midnight, in hours)	15
	Sleep duration (hours)	15, 16, 17, 19, 20
Stress	Overall stress level for day (0–10 scale)	23
Caffeine/Alcohol	Cups of caffeinated beverages (cups)	24
	Drinks of alcohol (drinks)	26, 28
Weather	Minimum barometric pressure (millibars)	N/A: passively collected
	Maximum temperature (degrees F)	
	Precipitation/humidity (1 PC: Higher scores correspond to higher humidity and higher probability and amount of precipitation)	
	Wind speed (mph)	
Prespecified	Workday indicator (vs weekend/holiday)	14
	Menstruation indicator (Day 1 ± 2)	32
	Headache self-prediction (0–4 scale, 0 meaning "very unlikely")	33

Day-level trigger exposures were constructed based on the raw survey responses (or, in the case of the weather category, raw passively collected data). The numbered survey questions are listed in Supporting Section S2.

Table 2.—

Candidate Model Specifications

ID	Description of Predictors	Rationale	Functional Form	Category-Specific Degrees of Freedom			
				Sleep	Stress	Caffeine/Alcohol	Weather
1L	Raw levels: Lag-1	Used raw exposure values from previous day. Simplest; commonly found in the literature.[11,50,51]	Linear	6	2	4	8
1S			Spline	9	3	5	12
2L	Change in level: Lag-2 to Lag-1	Used change in exposure level from one day to the next, to predict migraine on the third day. Changes in some factors may predict migraine as well or better than the raw values.[10,45,52,53]	Linear	6	2	4	8
2S			Spline	9	3	5	12
3	Raw levels: Lag-1 and Lag-2	Extension of model 1L; allowed various timing of maximal correlation between exposure and migraine risk.	Linear	9	3	6	8
4	Change in level: Lag-2 to Lag-1, and Lag-3 to Lag-2	Extension of model 2L; allowed various timing of maximal correlation between change in exposure and migraine risk.	Linear	9	3	6	12
5L	Cumulative: Mean of Lag 1, 2, and 3	Used mean exposure value over lags 1–3, as a low-dimensional summary of the cumulative exposure. Cumulative levels of stress and sleep deprivation have been correlated with other transient health outcomes,[54–56] but have not been well studied with respect to migraine risk.	Linear	6	2	4	8
5S			Spline	9	3	5	12
6	Change: Max absolute 1-day change over Lags 1–3	Summarized change in exposure over the past 3 days with the maximum absolute 1-day change between either lag 4 to lag 3, lag 3 to lag 2, or lag 2 to lag 1.[52]	Linear	6	2	4	8

The candidate model specifications shown above were based on the day-level trigger exposures defined in Table 1. These models were estimated with multilevel logistic regression and their AICs compared for each of the 4 categories. From each category, the predictors from the best-fitting model specification were included in the primary prediction model combining all categories.

Table 3.—

Characteristics of Analysis Sample, by Recruitment Source

Characteristics	Overall	Research Match	Facebook	Other	Standardized Mean Difference	N
N	178	73	71	34		
Age (mean [SD])	37.8 (11.8)	38.0 (11.5)	39.1 (12.0)	34.5 (11.6)	0.26	178
Sex = Female (%)	166 (93)	71 (97)	68 (96)	27 (79)	0.39	178
Body mass index (mean [SD])	27.2 (6.6)	26.7 (6.8)	28.6 (6.4)	25.0 (6.4)	0.37	176
Education level (%)					0.56	176
High school/GED or less	7 (4)	1 (1)	6 (9)	0 (0)		
Some college (1–3 years)	41 (23)	13 (18)	22 (31)	6 (18)		
Bachelor's degree	54 (31)	20 (27)	18 (26)	16 (48)		
Graduate degree	74 (42)	39 (53)	24 (34)	11 (33)		
Rural area of residence (%)	28 (16)	6 (8)	19 (27)	3 (9)	0.34	178
Region of residence (%)					0.99	178
Northeast – New England	6 (3)	1 (1)	4 (6)	1 (3)		
Northeast – Middle Atlantic	23 (13)	7 (10)	13 (18)	3 (9)		
Midwest – East North Central	47 (26)	21 (29)	18 (25)	8 (24)		
Midwest – West North Central	19 (11)	10 (14)	7 (10)	2 (6)		
South – South Atlantic	18 (10)	11 (15)	6 (8)	1 (3)		
South – East South Central	11 (6)	7 (10)	1 (1)	3 (9)		
South – West South Central	10 (6)	6 (8)	4 (6)	0 (0)		
West – Mountain	12 (7)	3 (4)	9 (13)	0 (0)		
West – Pacific	32 (18)	7 (10)	9 (13)	16 (47)		
Self-reported monthly headache days (mean [SD])	6.9 (3.1)	6.7 (3.3)	7.2 (3.0)	6.5 (3.1)	0.15	178
Years since initial migraine diagnosis (mean [SD])	16.6 (11.0)	16.5 (10.7)	18.1 (11.0)	13.2 (11.4)	0.29	170
Uses ergotamines or triptans for headache relief (%)	125 (71)	58 (79)	47 (67)	20 (61)	0.28	176
Daily headache preventive medication (%)					0.55	176
Current use	44 (25)	23 (32)	17 (24)	4 (12)		
Previous use only	56 (32)	19 (26)	30 (43)	7 (21)		
No current or previous use	76 (43)	31 (42)	23 (33)	22 (67)		
MIDAS Score (mean [SD])	32.6 (23.9)	29.3 (25.7)	36.4 (22.9)	31.5 (21.2)	0.20	176

Rural areas of residence were determined based on participant-reported mailing address ZIP codes defined as rural by the Federal Office of Rural Health Policy (https://www.hrsa.gov/rural-health/about-us/definition/datafiles.html, accessed 3 April 2020).

Regions of residence were determined based on the U.S. Census regions and divisions of states (https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf, accessed 3 April 2020). These divisions are Northeast – New England (CT, ME, MA, NH, RI, VT), Northeast – Middle Atlantic (NJ, NY, PA), Midwest – East North Central (IL, IN, MI, OH, WI), Midwest – West North Central (IA, KS, MN, MO, NE, ND, SD), South – South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV, DC), South – East South Central (AL, KY, MS, TN), South – West South Central (AR, LA, OK, TX), West – Mountain (AZ, CO, ID, MT, NV, NM, UT, WY), and West – Pacific (AK, CA, HI, OR, WA).

Table 4.—

Predictive Performance of Primary and Penalized Models

Dimension	Statistic	Primary Model		Penalized Model	
		Original (Within-Sample)	Cross-Validated (Out-of-Sample)	Original (Within-Sample)	Cross-Validated (Out-of-Sample)
C-statistic (Discrimination)	Within-person	0.58 (0.57, 0.60)	0.56 (0.54, 0.58)	0.58 (0.56, 0.60)	0.56 (0.54, 0.58)
	Overall	0.65 (0.64, 0.67)	0.58 (0.56, 0.6)	0.62 (0.6, 0.63)	0.59 (0.57, 0.60)
Calibration in the large	Mean intercept	0.01 (−0.04, 0.06)	0 (−0.08, 0.07)	−0.01 (−0.07, 0.06)	−0.01 (−0.08, 0.07)
	SD of intercepts	0	0.29	0.18	0.26
Weak calibration	Mean intercept	0.38 (0.20, 0.56)	−0.42 (−0.68, −0.16)	0.41 (0.15, 0.67)	−0.1 (−0.39, 0.19)
	SD of intercepts	0	0.14	0.03	0.12
	Mean slope	1.25 (1.13, 1.37)	0.73 (0.57, 0.9)	1.27 (1.1, 1.45)	0.94 (0.75, 1.13)
	SD of slopes	0	0.12	0.13	0.09
Predicted probabilities	Variance	0.004	0.004	0.002	0.002

The C-statistic was calculated within-person (by calculating the C-statistic for each person separately, and combining the statistics across persons with a random effects meta-analysis)and overall (comparing all pairs of days in the standard way).16,37 The calibration statistics were calculated based on multilevel logistic regression models for the observed outcome based on the linear predictor as an offset (calibration in the large) or as the univariable random effect (weak calibration) plus a random intercept. Standard deviations of intercepts and slopes from these models represent the standard deviation of the estimated random effects across persons. Wald 95% confidence intervals are shown for the key statistics.

Table 5.—

Summary of Sensitivity Analysis Results

	Main Analysis	1 ICHD-3 Migraines Only	2 All Headaches	3 No Recovery Days	4 Women with Menstrual Cycles	5 No Within-Person Means
Sample: Descriptive statistics						
Participants (N)	178	174	178	178	124	178
At-risk days (N)	10,696	11,280	9969	13,136	7554	10,696
Events (N)	1870	1411	2213	2462	1266	1870
Overall event risk	0.17	0.13	0.22	0.19	0.17	0.17
Predictive performance: Cross-validation statistics						
C-Statistic						
Within-person	0.56 (0.54, 0.58)	0.54 (0.51, 0.57)	0.57 (0.55, 0.58)	0.56 (0.54, 0.58)	0.56 (0.54, 0.58)	0.57 (0.55, 0.59)
Overall	0.58 (0.56, 0.6)	0.57 (0.54, 0.59)	0.58 (0.56, 0.59)	0.59 (0.57, 0.60)	0.56 (0.54, 0.58)	0.59 (0.57, 0.60)
Calibration in the large						
Mean intercept	0 (−0.08, 0.07)	−0.01 (−0.11, 0.1)	0 (−0.07, 0.07)	−0.01 (−0.08, 0.07)	0 (−0.1, 0.09)	0 (−0.08, 0.07)
SD of intercepts	0.29	0.49	0.24	0.32	0.29	0.25
Weak calibration						
Mean intercept	−0.42 (−0.68, −0.16)	−0.9 (−1.34, −0.47)	−0.32 (−0.53, −0.11)	−0.33 (−0.54, −0.12)	−0.7 (−1.03, −0.36)	−0.28 (−0.53, −0.04)
SD of intercepts	0.14	0.37	0.17	0.08	0.15	0.12
Mean slope	0.73 (0.57, 0.9)	0.55 (0.33, 0.76)	0.75 (0.58, 0.91)	0.78 (0.64, 0.93)	0.56 (0.35, 0.77)	0.82 (0.66, 0.98)
SD of slopes	0.12	0.10	0.17	0.15	0.06	0.10
Predicted probabilities						
Variance	0.004	0.003	0.005	0.004	0.003	0.003