Practice of Epidemiology

Methods for Estimating Remission Rates From Cross-Sectional Survey Data: Application and Validation Using Data From a National Migraine Study

Jason Roy* and Walter F. Stewart

* Correspondence to Dr. Jason Roy, Center for Clinical Epidemiology and Biostatistics, School of Medicine, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104 (email: jaroy@upenn.edu).

Initially submitted July 7, 2010; accepted for publication December 2, 2010.

Knowledge about remission rates can affect treatment decisions and facilitate etiologic discoveries. However, little is known about remission of many chronic episodic disorders, including migraine. This is partly due to the fact that medical records do not fully capture the history of these conditions, since patients might stop seeking care once they no longer have symptoms. For these disorders, remission rates would typically be obtained from prospective observational studies. Prospective studies of remission for chronic episodic conditions are rarely conducted, however, and suffer from many analytical challenges, such as outcome-dependent dropout. Here the authors propose an alternative approach that is appropriate for use with cross-sectional survey data in which reported age of onset was recorded. The authors estimated migraine remission rates using data from a 2004 national survey. They took a Bayesian approach and modeled sex- and age-specific remission rates as a function of incidence and prevalence. The authors found that remission rates were an increasing function of age and were similar for men and women. Follow-up survey data from migraine cases (2005) were used to validate the methods. The remission curves estimated from the validation data were very similar to the ones from the cross-sectional data.

Bayesian inference; Bayes theorem; cross-sectional studies; epidemiologic methods; migraine disorders; remission induction; spline

Abbreviation: AMPP, American Migraine Prevalence and Prevention.

Migraine is a common, chronic neurologic disorder, characterized by recurrent symptoms such as headaches with moderate or severe pain (1). While prevalence and incidence rates for migraine have been reported (1, 2), less is known about remission rates. In this article, we focus on migraine remission rates, but first we consider general challenges in estimating remission rates for chronic episodic conditions.

Estimates of remission are largely confined to studies of treatment for chronic progressive disorders (e.g., ulcerative colitis, Crohn’s disease, and different cancers, among others), where the interest is in “curing” the disease (3–5). Notably, observational studies of the natural history of most chronic diseases do not examine remission outside of a treatment framework, largely because the overwhelming majority of patients with such conditions progress to a more serious stage of disease. The natural history of progression, not remission, is the dominant interest. In contrast, there are many chronic episodic symptomatic conditions that emerge between the ages of 10 and 55 years that are very common and are often associated with substantial disability and health-care costs. Understanding remission and recurrence rates for these conditions is crucial. Typically, studies of remission involve 2–3 surveys separated by long time intervals, where the follow-up rate decreases with each subsequent survey and continued participation can be correlated with persistence of symptoms (6–10). Remission data could be potentially useful to etiologic research and to clinical care. Unlike the case with chronic progressive disorders, information about remission and recurrence is crucial for addressing chronic episodic conditions. Here we consider challenges in estimating remission for such conditions and offer a potentially useful solution.
Migraine, depression, asthma, allergic rhinitis, rhinosinusitis, epilepsy, and urinary incontinence are among the many chronic episodic disorders characterized by variation in age of onset, the amount of time that people experience exacerbating episodes of symptoms, and the frequency of exacerbating episodes (11, 12). An understanding of remission for these conditions and who is likely to enter remission could be important to research on heterogeneity of causes (e.g., genetic, psychosocial, environmental), to patient understanding of what to expect, and to treatment decision-making, yet remarkably little is known about remission for these conditions. While it is widely recognized that persons with these conditions often cease having exacerbating or symptomatic episodes, relatively few epidemiologic researchers have actually estimated remission rates. To some degree, considerable resources are required to document that a chronic episodic condition has remitted and to estimate remission rates—resources that are not usually available.

For serious chronic progressive disorders (e.g., cancer, ulcerative colitis), patients routinely seek medical care, and outcomes are monitored and documented. In contrast, the experience of persons with most chronic episodic disorders is only partially captured by health-care data. If remission occurs, patients typically stop seeking care rather than report to their physician that they are no longer having symptoms. As a consequence, medical records cannot be used to reliably estimate incidence or remission of such conditions. Instead, prospective follow-up studies are required to estimate remission rates for these disorders, where data need to be obtained on a periodic schedule to assess symptom status and to document when a person last experienced an episode. Analysis of such data is then required to formally define the temporal criteria (i.e., the duration of time without an exacerbation of symptoms) for remission, as distinct from recurrence of the same condition. Such studies are uncommon. Moreover, prospective studies of these types of conditions are fraught with logistical and methodological challenges. Continued follow-up can often be associated with whether or not a participant continues to have the condition of interest. Persons whose condition remits may be more likely to drop out. Frequent contact over many years is required to document persistence or remission. Given these challenges, we considered how cross-sectional survey data could be used as an efficient alternative to prospective studies to estimate remission rates.

This work builds on prior studies, where cross-sectional data were used to estimate age- and sex-specific incidence rates for migraine (2, 13, 14). We use a Bayesian approach and model remission indirectly as a function of incidence and prevalence. We apply this new method to national survey data collected on persons with migraine headache in the American Migraine Prevalence and Prevention (AMPP) Study. To validate the approach, we use data from the second phase of the AMPP Study. Phase 2 includes follow-up data from persons who had a migraine in phase 1. We use phase 2 data to estimate remission rates using the standard, longitudinal approach. We then compare the migraine remission rates estimated from the new and standard approaches.

MATERIALS AND METHODS

Migraine study (cross-sectional)

Phase 1 of the AMPP Study was conducted in 2004. The AMPP Study was modeled on the methods of American Migraine Studies 1 and 2, which are described in detail elsewhere (1, 15, 16). In the AMPP Study, 120,000 US households were selected via stratified random sampling from the 600,000-household nationwide panel maintained by National Family Opinion, Inc. (Greenwich, Connecticut). Sampling blocks of 5,000 households each comprise the panel, which are constructed to be representative of the US population. Census and demographic details (e.g., age, sex, educational level) on household members are routinely updated by National Family Opinion.

The AMPP questionnaire was sent to the designated head of the household, who was instructed to identify all household members suffering from at least occasional self-defined severe headache. Each household member with severe headaches was asked to complete a symptom screening questionnaire, consisting of 21 questions, assessing all headache features important to the detection of migraine. The screening questionnaire has been validated (using the second edition of the International Classification of Headache Disorders (17)) as a reference in a population sample of migraine sufferers and controls with other types of headache (18). Surveys were returned by 162,562 persons from 77,879 households. Response rates were similar by sex, geographic region, population density, and household income. Survey participants were between 12 and 100 years of age; 55% were female.

Diagnosis of migraine, limited to active cases (i.e., at least 1 reported attack in the previous year), was based on established criteria from the International Classification of Headache Disorders, second edition (17), and was assigned if a respondent reported having had at least 1 severe headache in the previous 12 months but fewer than 15 severe headaches in the prior month, with unilateral or pulsatile pain and either nausea, vomiting, phonophobia with photophobia, or visual or sensory auras before the headache. If these criteria were not met, respondents with severe headache were classified as having “other severe headaches.” A total of 4,386 (5%) males and 14,604 (14%) females met the criteria for migraine.

In addition to migraine status, age, and sex, the other variable we were interested in was age at onset. As part of the symptom screening questionnaire, subjects with severe headaches were asked about the age at which they had first had severe headaches (not the age of diagnosis). Thus, for current migraine cases, we observed reported onset age.

Variables and definitions

In this paper, we are interested in estimating age- and sex-specific remission probabilities. Here we define remission as having had no attacks in the past year. While other definitions of remission are possible, 1-year symptom inactivity rates are what we can identify from the data without making assumptions about disease recurrence probabilities.
Let the indicator variable $Y_j$ denote that the subject was an active migraine case (defined above) at age $j$. Denote the subject’s current age by $A$. In the AMPP survey, age was reported as an integer. The hypothetical complete outcome data for a subject who is currently age $j$ consist of $\{Y_1, \ldots, Y_j\}$. Although they are not fully observed, we next use the complete data to define the parameters of interest.

The prevalence ($\text{Prev}$) of migraine at age $j$ for subjects who are currently age $j$ is $\text{Prev}_j = P(Y_j = 1 | A = j)$. That is, it is the marginal probability of having the outcome at age $j$. The incidence ($\text{Inc}$) at age $j$ is the probability of having the outcome at age $j$, given that the outcome did not occur in prior years; that is, $\text{Inc}_j = P(Y_j = 1 | Y_{j-1} = 0, A = j)$. Finally, we define remission as the situation where a subject who had migraine headaches in the preceding year did not have any in the current year. Specifically, the remission probability ($\text{Rem}$) at age $j$ is $\text{Rem}_j = P(Y_j = 0 | Y_{j-1} = 1, A = j)$.

Our primary interest is in the remission probability. However, we cannot obtain this directly using data from the AMPP Study, as we do not know the value of $Y_{j-1}$ for subjects with $A = j$ and $Y_j = 0$ (since onset age was not obtained for controls). The remission probability can be obtained indirectly, as a function of incidence and prevalence. Specifically, since

$$P(Y_j = 1 | A = j) = P(Y_j = 1 | Y_{j-1} = 0, A = j)P(Y_{j-1} = 0) + P(Y_j = 1 | Y_{j-1} = 1, A = j)P(Y_{j-1} = 1),$$

the remission rate at age $j$ as a function of prevalence at age $j$ and $(j - 1)$ and incidence at age $j$ is as follows:

$$\text{Rem}_j = 1 - \frac{\text{Prev}_j - \text{Inc}_j \times (1 - \text{Prev}_{j-1})}{\text{Prev}_{j-1}}. \quad (1)$$

### Remission rate estimation from cross-sectional data

In order to estimate the remission probabilities, we estimate the prevalence and incidence rates and then obtain the remission rate using equation 1.

Prevalence can be estimated directly. For example, $\text{Prev}_j$ could be estimated as the sample proportion of age $j$ subjects who are active cases. However, because we expect prevalence rates to be similar at similar ages, we can gain efficiency in estimation by assuming that the rates are a smooth function of age. We assume that the number of migraine cases for subjects currently age $j$ follows a binomial distribution with the probability $\text{Prev}_j$. We then fit a logistic regression model, relating the prevalence probability to (centered) age using a penalized spline (19). Specifically,

$$\text{logit}(\text{Prev}_j) = \beta_0 + \beta_1 \text{age}_j + \sum_{i=1}^L b_i (\text{age}_j - k_i)_+, \quad (2)$$

where $\beta_0$ and $\beta_1$ are regression coefficients, the $b_i$’s are random effects distributed as $N(0, \sigma^2)$, $k_1, \ldots, k_L$ are fixed knots, and $(\text{age}_j - k_i)_+ = I(\text{age}_j > k_i)(\text{age}_j - k_i)$, where $I()$ is the indicator function. In the migraine data analysis, we used 3 knots, located at tertiles of the (centered) age distribution. Prior distributions for $\beta_0$ and $\beta_1$ were specified as normal with mean 0 and standard deviation 20. This prior distribution was chosen because it is proper but fairly flat (noninformative). The standard deviation of the random effects, $\sigma$, was assumed to follow a uniform (0, 10) distribution. The upper limit of 10 is much larger than we would expect to see for a model of this type.

Incidence rates were modeled using the approach of Roy and Stewart (13). Briefly, the number of migraine cases in the at-risk set of subjects was assumed to follow a binomial distribution. The logit of the corresponding probability was assumed to be a smooth bivariate function of age and lag, where lag is the difference between current age and reported onset age. We specified a low-rank thin-plate spline as the basis function, which tends to work well for Bayesian models (20). Forty knots were selected using the space-filling function cover.design in the R fields package (21). The incidence rate corresponds to predicted values at each age, where lag is set to 0. The motivation for this is that if onset age was equal to current age, there would be no recall bias and incidence could be accurately estimated. Simulation studies demonstrated good performance of the methods in a variety of situations involving recall bias (13).

The incidence and prevalence models were fitted simultaneously using WinBUGS (22). The WinBUGS code is provided in the Web Appendix, which is posted on the Journal’s Web site (http://aje.oxfordjournals.org/). Draws of $\text{Rem}_j$ were obtained using equation 1, by plugging in the corresponding draws of $\text{Prev}_j$ and $\text{Inc}_j$. Using this approach, it is possible to obtain a draw of the remission probability that is not between 0 and 1. In our analysis, we rejected draws of this type. Separate models were fitted for males and females.

A total of 10,000 draws from 3 parallel chains (with different starting values) were discarded as part of the burn-in period. We obtained 10,000 additional draws and used them for inference. Convergence was checked using trace plots and the Gelman-Rubin statistic (23). Trace plots for the parameters from 3 different chains are given in Web Figure 1 and Web Figure 2. In general, the models appeared to have good mixing properties. Posterior median values and 95% credible intervals for remission were obtained for each sex-age combination. A 95% credible interval implies that the posterior probability that the parameter is between the 2 limits is 0.95 (23).

### Validation: migraine follow-up study

A follow-up study of the original AMPP Study took place in 2005 (the year following the phase 1 study) (24). A modified version of the AMPP questionnaire was sent to a random sample of adult (ages ≥18 years) respondents who reported active (i.e., past 12 months), severe headache in the screening survey. Of the 24,000 household members who were sent the questionnaire, 16,573 returned completed questionnaires (69.0% response rate). By following up subjects who were migraine cases in 2004, we had the ability to...
Remission rate estimation from validation data

The demographic characteristics of the 2005 respondents were similar to those of the target sample. There were no significant differences with regard to region of the country, state, race, sex of the head of the household, annual household income, or household size. Responders did not differ from nonresponders in terms of age and sex (24). Response rates did not seem to vary based on headache symptom frequency or severity.

Remission rate estimation from validation data

For each of the participants in the follow-up study, we observed whether the person had been an active case in the previous year and his or her current migraine status (if the participant had had an attack in the past year). Data consisted of all subjects who had had a migraine in the original 2004 study and returned a complete questionnaire in 2005. Migraine status in 2005, equal to 1 if the participant had had at least 1 attack in the past year and 0 otherwise, was the outcome variable. We assumed that the outcome followed a binomial distribution, with the remission probability modeled using logistic regression. We assumed that remission rates were a smooth function of age. Specifically,

\[
\text{logit}(\text{Rem}_j) = \alpha_0 + \alpha_1 \text{age}_j + \sum_{i=1}^{l} a_i (\text{age}_j - k_i)^+,\]

where \(\alpha_0\) and \(\alpha_1\) are regression coefficients and the \(a_i\)'s are normally distributed random effects. In the migraine data analysis, we used 3 knots, located at the tertiles of the (centered) age distribution. The same prior distributions were used for this model as were used in the prevalence model described above. The remission models were fitted separately for males and females. Again, after convergence, 10,000 draws were obtained from the posterior distribution using WinBUGS.

RESULTS

Estimated remission rates and 95% credible intervals for the cross-sectional approach and validation data are presented in Table 1. The results are presented separately for males and females and at ages 20–60 years, in increments of 10. In all cases, the 95% credible intervals for the cross-sectional and validation data overlap. The estimated remission rate curves and 95% credible intervals for the cross-sectional data are displayed in Figure 1, along with the estimated curves from the validation data (dashed curve). The most notable difference between the estimated curves occurs for persons under age 40 years. The validation curve tends to be flat from age 20 years to age 40 years, whereas the curve from the cross-sectional approach tends to continue decreasing as age decreases (particularly for females). In general, however, the curves appear to be very similar. Thus, at least for this example, the cross-sectional modeling approach proposed in this paper appears to be very effective relative to the gold standard.

The estimated remission rate for women gradually increases in a monotonic fashion with age, from a low of 1.7 cases/100 at age 20 years to a high of 7.1 cases/100 at age 60 years. The pattern differs for men, where remission declines from age 20 years (4.2 cases/100) to age 30 years (2.5 cases/100) and then increases monotonically at approximately the same rate as that observed in women, peaking at more than 8.3/100 at age 60 years. The remission rate for females and males indicates that prognosis for persons with active migraine improves with age.

DISCUSSION

Cross-sectional data were previously used to estimate the age-specific incidence of migraine, indicating that the cumulative lifetime risk of migraine is very high (i.e., 43% among females and 18% among males) (2). Given the 1-year period prevalence of active migraine (i.e., 17% among females and 6% among males) in the general population and the estimates of cumulative incidence, we speculate that the duration of illness, on average, is likely to be short for a majority of persons who have ever had migraine and for a substantial minority of active prevalent cases (2). The latter is relevant to prognosis, to the understanding...
of health-care needs, and to investigations of etiologic heterogeneity.

To better understand the dynamics of change for migraine in the population, we have proposed a new statistical approach for estimating remission rates from cross-sectional survey data that may be more generally applicable to other chronic episodic conditions like depression, back pain, and asthma, among others. The method requires that both current disease status and reported age of onset (for current cases) are known. The remission rates are estimated by fitting models for prevalence and incidence using existing methods and then obtaining the remission rate indirectly. We applied the method to data from a national migraine study and found that remission rates increase gradually with age. After age 40 years, the remission rates were very similar for males and females. A follow-up study of the original headache study was used for validation. We found that estimated remission rates from the cross-sectional and validation data were very similar, particularly for persons above age 40 years. There were differences between the curves between ages 20 years and 40 years, but the 95% credible intervals had substantial overlap, so we cannot conclude that the curves differed in that region.

For a variety of life-threatening conditions, estimates of remission rates are fundamentally important to deciding whether a treatment is effective. In contrast, natural remission is not usually a consideration in clinical care for common chronic episodic conditions, largely because little is known about remission rates for such conditions or about predictors of remission. However, access to information on remission is likely to be useful for both providers and patients in guiding health-care management and treatment decisions. Patients who are likely to have a shorter duration of disease may be reassured that their condition will not be lifelong. Patients with an expected longer course of disease may engage their health-care provider and consider approaches to more effective management and education.

While we did not specifically examine predictors of migraine remission, the estimates from this study reveal patterns that are informative. The prognosis for migraine is better for males than for females. The age-specific prevalence of migraine is, on average, 3 times higher in females than in males, although this ratio varies substantially from less than 2.0 before puberty to a high of 3.4. In contrast, remission estimates for men and women after age 30 years indicate that active cases will remit at a much faster rate among males than among females, accounting for the preponderance of migraine among females even after age 50 years; the prevalence ratio is approximately 2.6 after age 60 years (25). However, there could be confounding variables that explain some of the differences between men and women. We do not address that issue here.

There have been relatively few evaluations of migraine remission in population-based studies. In a workplace cohort, migraine changed to a nonmigraine headache or remitted 10 years later in 63% of active migraine patients aged 35–50 years at baseline, yielding an average remission rate of 6.3 cases/100 person-years. Lyngberg et al. (6) studied remission in a population sample of migraine patients aged 25–64 years at baseline. A total of 42% remitted 13 years later. While the numbers of persons with migraine at baseline were relatively small in these 2 studies, the remission rate estimates are consistent with our findings (Table 1).

### Table 1. Estimated Migraine Remission Rates Based on Cross-Sectional and Follow-up Data From the American Migraine Prevalence and Prevention Study, by Sex, 2004–2005

<table>
<thead>
<tr>
<th>Age, years</th>
<th>Women</th>
<th>Validation Data</th>
<th>Men</th>
<th>Validation Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-Sectional Approach</td>
<td></td>
<td></td>
<td>Cross-Sectional Approach</td>
</tr>
<tr>
<td></td>
<td>Estimate$^b$</td>
<td>95% CI</td>
<td>Estimate</td>
<td>95% CI</td>
</tr>
<tr>
<td>20</td>
<td>1.7</td>
<td>0.3, 4.8</td>
<td>4.5</td>
<td>2.9, 7.4</td>
</tr>
<tr>
<td>30</td>
<td>3.2</td>
<td>1.6, 6.3</td>
<td>4.3</td>
<td>3.5, 5.3</td>
</tr>
<tr>
<td>40</td>
<td>4.7</td>
<td>2.5, 8.6</td>
<td>4.6</td>
<td>3.6, 5.4</td>
</tr>
<tr>
<td>50</td>
<td>5.6</td>
<td>3.1, 10.2</td>
<td>5.9</td>
<td>5.0, 6.8</td>
</tr>
<tr>
<td>60</td>
<td>7.1</td>
<td>3.9, 12.9</td>
<td>7.6</td>
<td>6.2, 9.3</td>
</tr>
</tbody>
</table>

**Abbreviation:** CI, credible interval.

$^a$ The remission rate is the number of remitted cases per 100.

$^b$ Posterior median value from a Bayesian remission model.
et al. (8)) and challenges regarding the possibility of differential remission rates by follow-up status.

The relatively high annual remission rates for migraine may help to explain, in part, the apparent patterns of under-diagnosis and undertreatment. In our previous work, we reported that migraine is underdiagnosed (1). The relatively low rate of diagnosis may not be due to access to medical care or other health-care-related factors and may be explained by a short duration of time with migraine. The probability of migraine diagnosis and use of migraine-specific prescription medications is associated with age and, most likely, with duration of illness (1). In a free-standing population sample of persons with active migraine, many may have had recent onset and a short duration of illness and may later remit, eliminating the need for health care or prescription medication—a process that is impossible to assess cross-sectionally.

Etiologic research, including exogenous factors, familial aggregation, and genomic investigations, will be strongly influenced by how migraine sufferers are sampled from the general population, depending on the extent of correlation with age of onset, duration of illness, or age of remission. Typically, a random sample of the population is selected with a specific focus on active cases. Patients who experience remission are not usually considered. Moreover, active cases within a study are likely to be highly heterogeneous with regard to the above covariate, and studies are likely to vary substantially depending on the age distribution of cases. Future investigations of etiologic factors involved in onset, remission, persistence, and progression should take duration of illness into account.

The method described herein can be applied to existing cross-sectional data to estimate remission rates. At a minimum, the data set would need to document the age of onset among persons who currently have the condition. A limitation of our own data is that the inclusion criteria were confined to active migraine sufferers. We recommend that in the future, data be collected from all persons who have had the condition of interest, regardless of current symptom status. In addition, for those who are inactive, we recommend that researchers obtain information about the date on which symptoms last occurred. Access to such data offers an opportunity to empirically determine the duration of time that qualifies as true remission.

Two limitations of our study warrant attention in future studies. First, we arbitrarily defined remission as being migraine-free for 12 months. A more formal definition of remission will require either cross-sectional data with data on inactive cases and time since remission or longitudinal analysis of remission and the probability of recurrence. Our estimates of remission rates would be lower if recurrence of headache is common after a single year without headaches. On the other hand, if recurrence is unlikely after 1 year, then our remission rates are likely to be unbiased. Second, epidemiologic studies of remission are highly informative to clinical care providers, especially if definitive predictors of remission can be identified. Future cross-sectional surveys could offer valuable insights if data were collected on remitted cases, age of onset, and factors possibly related to remission.

ACKNOWLEDGMENTS

Author affiliations: Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania (Jason Roy); and Geisinger Center for Health Research, Geisinger Health System, Danville, Pennsylvania (Walter F. Stewart).

The American Migraine Prevalence and Prevention Study is funded through a research grant to the National Headache Foundation from Ortho-McNeil Neurologics, Inc. (Titusville, New Jersey).

REFERENCES