Lay versus expert interviewers for the diagnosis of migraine in a large sample of elderly people

C Tzourio, B Gagnière, M El Amrani, M-G Bousser, A Alépérovitch

J Neurol Neurosurg Psychiatry 2003;**74**:238–241

See end of article for authors' affiliations

Correspondence to:
Dr Christophe Tzourio,
INSERM U360, Hôpital
Salpêtrière, 75651 Paris
cedex 13, France;
tzourio@chups.jussieu.fr

Received 13 August 2002
Accepted
29 October 2002

Objective: To determine the agreement between lay interviewers and experts in the diagnosis of migraine by questionnaire.

Subjects: A population based sample of 1188 individuals aged 64 to 73 years.

Methods: Participants who declared that they had recurrent headaches ($n = 238$) answered a structured questionnaire by lay interviewers with special training in migraine. A migraine expert subsequently interviewed all the headache sufferers using the same questionnaire. Migraine was defined according to the International Headache Society criteria.

Results: In comparison with the expert, the diagnosis derived by the lay interviewers had high values for specificity (97%) and positive predictive value (86%), and a low sensitivity (50%) and negative predictive value (57%). Agreement between the expert and the lay interviewers was low, with a κ value of 0.36 (95% confidence interval 0.26 to 0.47). The most serious discrepancies concerned the duration of attacks, the worsening of headaches by physical activity, the presence of nausea or vomiting, and the unilaterality of headaches. As a result, the lifetime prevalence of migraine headaches was greatly underestimated by lay interviewers (6.5%) in comparison with the expert (11.1%).

Conclusions: A low level of agreement between lay interviewers and a headache expert in the diagnosis of migraine headaches by structured questionnaire may result in a substantial underestimation of migraine prevalence.

Although the diagnosis of migraine often requires some level of expertise, in large studies it is impractical for all the participants to be interviewed by an expert in headaches. Thus in most epidemiological studies the diagnosis of migraine is made by self administered questionnaire or after an interview with lay interviewers.¹ In 1988 the International Headache Society (IHS) set up an algorithm for the diagnosis of migraine² which can easily be incorporated into a structured questionnaire and administered by lay interviewers. There have been few studies involving substantial numbers of cases in which migraine diagnosis based on the IHS criteria obtained by lay interviewers has been compared with that made by a migraine expert.^{3,4}

Our aim in this study was to compare the diagnosis of migraine obtained by trained lay interviewers and by a headache specialist using the same IHS criteria-based structured questionnaire in a large population of elderly people.

METHODS

Study design

The EVA (epidemiology of vascular aging) study is a population based longitudinal study of cognitive and vascular aging which has been described in detail elsewhere.⁵ Briefly, men and women born between 1922 and 1932 were recruited from the electoral rolls of the city of Nantes, France. A standardised questionnaire was administered to obtain information about demographic background, occupation, medical history, drug use, and personal habits. Cognitive function was assessed by the mini-mental state examination (MMSE).⁶ Written informed consent was obtained from all participants, and the study was approved by the ethics committee of the Hôpital de Kremlin-Bicêtre.

Migraine diagnosis

The migraine study was undertaken during the second follow up visit, which was four years after the baseline visit and

involved 1188 participants. Migraine assessment involved the use of a structured questionnaire reproducing all the items of the IHS criteria.² This questionnaire had been tested previously in another study and showed high reproducibility between experts.⁷ The 1188 participants were first asked about recurrent attacks of headache during their lifetime as part of the general questionnaire during a face to face interview with two lay interviewers (fig 1). Those who did not complain of having had recurrent attacks of headache at any time ($n = 946$) were considered to be headache-free and the remainder of the questionnaire on headaches was ignored. The 238 participants who answered yes to the screening question were asked specific questions about their headaches. Instructions were given to the interviewers to obtain a lifetime history of headaches by asking participants questions about headaches in their young adulthood. It was decided not to include questions on aura, which requires an experienced interviewer. From the answers to this questionnaire it was therefore possible to reach a diagnosis of migraine by applying the IHS algorithm (lay interviewer diagnosis).

Participants who answered yes to the screening question were also asked to participate in a telephone interview with a headache specialist. The interview was undertaken by a neurologist specialising in headaches, using the same structured questionnaire as the lay interviewer. The expert was blinded to the results of the questionnaire by the lay interviewer, as well as to all clinical and laboratory data concerning the participants. With this questionnaire, a diagnosis of migraine based on the IHS algorithm was then derived (expert diagnosis). This expert based diagnosis was considered to be the reference diagnosis. Participants with migraine were also asked if their attacks were persistent—defined as at least one attack during the past year. The one year prevalence of migraine was derived from this question.

Among the 238 subjects screened positive for recurrent headaches, four had died by the time of the interview and one had hearing difficulties and could not participate in a

Figure 1 Flow chart of the study of migraine headaches in the epidemiology of vascular aging study.

Table 1 Agreement between lay interviewers and an expert in the diagnosis of migraine by questionnaire

Lay interviewers	Expert		
	No migraine	Migraine	Total
No migraine	85	64	149
Migraine	11	65	76
Total	96	129	225

Values are numbers of subjects.
Sensitivity 50.4%, specificity 96.6%, positive predictive value 85.5%, negative predictive value 57.0%.

telephone interview. The final sample thus consisted of 1179 participants (99% of the initial sample) among whom 233 were screened positive for recurrent headaches (fig 1).

Statistical methods

Sensitivity, specificity, positive and negative predictive values, κ index, and the McNemar test were used to compare the lay interviewers with the expert for the diagnosis of migraine. The accuracy of the lay interviewers in identifying migraine as positive and negative predictive values (per cent). Guidelines suggest that values of κ above 75% indicate excellent agreement, values between 75% and 40% indicate good to fair agreement, and those below 40% are considered to show poor agreement.⁸ The McNemar test compares the difference between false positive and false negative diagnoses and addresses the question of whether the two types of inconsistency are equally common. A 5% level of significance and 95% confidence intervals were used. All analyses were done using the SAS[®] statistical package (version 8.00).

RESULTS

The sample included 1179 individuals (689 women, 58.4%), mean age 69 years (range 63 to 75). Cognitive status was quite high, with a median MMSE of 28 (interquartile range 27 to 29) and no participant had severe cognitive impairment or dementia.

Among the 233 participants screened positive for recurrent headaches, lay interviewers were unable to collect precise information on some aspects of the headaches in eight, who were therefore considered as missing (fig 1). Migraine headaches were diagnosed in 76 participants (64 women) by the lay interviewers' questionnaire. The lifetime prevalence (95% confidence interval (CI)) of migraine based on the lay interviewers' questionnaire was therefore 6.5% (5.1% to 7.9%). The expert was able to interview all 233 participants with recurrent headaches, and diagnosed migraine in 131 of them (114 women, 17 men). The expert also determined that 18 of the participants with migraine but without aura also had attacks of migraine with aura. The lifetime prevalence of migraine headaches diagnosed by the neurologist was 11.1% (9.3% to 12.9%). The one year prevalence of migraine was 3.9% (2.7% to 5.0%) according to the lay interviewers' questionnaire and 6.8% (5.3% to 8.2%) according to the expert.

The level of agreement between the neurologist and the lay interviewers was low (table 1), with a κ value of 0.36 (95% CI,

Table 2 Comparison of the proportion of positive answers for each IHS criterion among headache sufferers interviewed by both lay interviewers and an expert

IHS criteria	Lay	Expert	Agreement*	False positive†	False negative‡	p Value§
At least five attacks during life	98.0	98.5	96.5	1.5	2.0	0.71
Attacks lasting 4 to 72 hours	54.0	77.3	64.7	6.1	29.3	<0.0001
At least two of the headache characteristics below	80.3	77.5	73.5	14.7	11.8	0.41
Unilateral	67.3	42.1	65.8	29.7	4.5	<0.0001
Pulsating	36.3	41.7	69.1	12.8	18.1	0.17
Moderate to severe	89.2	89.7	87.7	5.9	6.4	0.84
Aggravated by physical activity	41.6	67.2	61.8	6.3	31.9	<0.0001
At least one associated symptom	68.4	71.8	79.4	8.6	12.0	0.29
Nausea and/or vomiting	40.2	56.9	74.7	4.3	21.1	<0.0001
Photophobia and phonophobia	59.8	65.6	75.1	9.6	15.3	0.096

Values are per cent.

*Agreement rate including positive and negative responses.

†Proportion of subjects with the IHS criterion according to the lay interviewers and without it according to the expert.

‡Proportion of subjects with the IHS criterion according to the expert and without it according to the lay interviewers

§McNemar test.

IHS, International Headache Society.

Table 3 Agreement between lay interviewers and an expert on migrainous disorders fulfilling all IHS criteria except one (code 1.7 of the IHS classification)

Lay interviewers	Expert		Total
	No migraine	Migraine	
No migraine	43	44	87
Migraine	24	120	144
Total	67	164	231

Values are numbers of subjects.

Sensitivity 73.2%, specificity 64.2%, positive predictive value 83.3%, negative predictive value 49.4%.

IHS, International Headache Society.

0.26 to 0.47). Compared with the expert diagnosis, the diagnosis of migraine based on the lay interviewers' questionnaire was characterised by a sharp contrast between a high specificity and positive predictive value (96.6% and 85.5%, respectively) and a low sensitivity and negative predictive value (50.4% and 57.0%, respectively). Misclassification by the lay interviewers occurred in 49.6% of migraine sufferers (64/129) erroneously classified as non-sufferers, and in 11.5% of non-sufferers (11/96) erroneously classified as sufferers (McNemar test, $p < 0.0001$).

To determine whether this lack of sensitivity of the lay interviewer assessment could be related to the characteristics of the participants, we compared the participants incorrectly diagnosed as non-migrainous by the lay interviewers (false negative, $n = 64$) with those correctly diagnosed as migrainous (true positive, $n = 65$). There was no difference between these two groups with regard to characteristics that might have had some influence on the recall of headache characteristics (age, sex, occupation, body mass index, alcohol and tobacco consumption, cognition, education, severity of headaches, and persistence of attacks at the time of interview (data not shown)). Participants' characteristics did not explain why those with migraine according to the expert were incorrectly classified as non-migrainous by the lay interviewers.

We then compared the answers for each IHS criterion in all patients who had been interviewed by both the lay interviewer and the headache specialist ($n = 225$) (table 2). Compared with the headache specialist, lay interviewers underestimated the duration of attacks, the worsening of headaches by physical activity, and the presence of nausea or vomiting (false negative between 20% and 30%). Conversely, unilateral headache was more frequent on the lay interviewers' assessment than on the expert's, resulting in a 30% false positive rate.

To evaluate whether the overall agreement could be improved by a simple modification of the algorithm, we used the code 1.7 of the IHS classification—that is, headache attacks fulfilling all IHS criteria but one, and not fulfilling any criterion for tension type headache. With this definition, the prevalence of migraine was 12.6% (10.8% to 14.5%) according to the lay interviewers and 14.2% (12.2% to 16.2%) according to the expert. The sensitivity improved to 73.2%, but the specificity and positive predictive value deteriorated (specificity = 64.2%; positive predictive value = 83.3%). Overall agreement did not improve ($\kappa = 0.34$ (0.22 to 0.47)) (table 3).

DISCUSSION

In this population based study we found that the agreement between a headache specialist and trained lay interviewers for migraine diagnosis on the basis of the IHS criteria was poor ($\kappa = 0.36$). Misclassification of migraine cases by the lay interviewers was characterised by a low sensitivity (47.4%) and a high specificity (87.5%) and positive predictive value (85.5%).

The lay interviewers had specific training for interviewing participants with headaches and they used a structured questionnaire with detailed questions. The expert's diagnosis was derived from the application of the IHS criteria algorithm to the answers obtained from the same questionnaire. The difference between the lay interviewers and the expert was therefore limited by design to the skill in collecting the information, and was not directly dependent on expert opinion. In spite of this, and of the training of the lay interviewers, κ was less than 0.50 to 0.60, which is usually considered to represent the minimum acceptable agreement.⁹ We could not identify any characteristics of the participants—such as age, sex, cognitive status, depressive symptoms, or education—that could explain the poor level of agreement observed. The use of code 1.7 of the IHS criteria (migrainous disorders fulfilling all but one of the IHS criteria, and not fulfilling any tension type headache criteria) improved the sensitivity but led to a deterioration in positive predictive value and specificity, while the κ value remained almost unchanged. Other simple modifications of the IHS algorithm did not lead to substantial improvement in agreement (data not shown).

These results of low sensitivity and relatively high specificity are reflected in those obtained in two other studies done to validate migraine diagnosis by questionnaire, either self administered (sensitivity 51%; specificity 92%),¹⁰ or administered by lay interviewers (sensitivity of 42%; specificity 91%).⁴

Analysis of the agreement for each IHS criterion showed that the greatest misclassification occurred for four items: the duration of attacks, aggravation of headaches by physical activity, the presence of nausea or vomiting, and unilateral headache. Granella *et al* noted that the worst agreement between four clinicians was obtained for the duration of the attacks.¹¹

Our study had limitations and strengths. It was based on a sample of elderly participants recruited from the general population. Any observations concerning the validity of the diagnosis are therefore applicable only to participants in that age group and should not be directly extended to younger populations. Because of their relatively old age the participants might have had memory impairment, leading to an underestimation of migraine frequency. However, they were closely monitored for cognitive function, as this was an important outcome variable in the study. In fact, the overall cognitive level was relatively high and there were no cases of dementia among the study population at the time of interview.⁵

Another potential limitation of the study was the lack of any physical examination of the patients with headaches.² Participants in the EVA study were, however, followed up over a long period and most secondary causes of headache could therefore be excluded. Further, 70% of the participants (828/1179) had cerebral magnetic resonance imaging as part of the EVA study,¹² and no haematoma or brain tumour was diagnosed in patients with recurrent headaches.

An important strength of the study was the large sample and the high participation rate. There were very few missing data, and all the participants with recurrent headaches agreed to take part in the expert interview. One expert undertook all the interviews, and it was deemed impracticable to test his reproducibility in the setting of the study. However, in a previous study the questionnaire proved to be highly reproducible between experts in the same team ($\kappa = 0.83$).⁷

The lifetime prevalence of migraine was 6.5% according to the lay interviewers' questionnaire results—much less than the 11.1% obtained with the expert interview. Similarly, the one year prevalence was much lower in the lay interviewers' assessment (3.9%) than in the expert's assessment (6.8%). Thus, if the diagnosis of migraine had been based only on lay interviewers' questionnaire, there would have been an important underestimation of the prevalence of migraine in this

population. Moreover, about half the participants with migraine would have been misclassified as non-migrainous, which would have been a major limitation of any study of the correlates of and risk factors for migraine in this population.

Overall, our results suggest that in elderly people the best study design to obtain a reliable estimate of the prevalence of migraine and to avoid misclassification would be a two step one. The first step could be a screening procedure with simple questions on recurrent headaches. This could be done by lay interviewers or possibly by a self administered questionnaire. The second step would be an interview of the participants screened as positive by an expert in migraine blinded to all other characteristics of the participants. This strategy seems to be a good compromise between a high quality diagnosis and expert availability. If this optimal design could not be used, and if interviews have to be performed by non-specialists, then the training of lay interviewers should be concentrated on the items which were shown to be highly discordant in our study. Further, the agreement between lay interviewers and headache specialists should be monitored closely, including during the collection of the data, in order to avoid significant misclassification.

ACKNOWLEDGEMENTS

The EVA study was carried out under an agreement between INSERM (Institut National de la Santé et de la Recherche Médicale), the Merck, Sharp and Dhome-Chibret Laboratories (West Point, Pennsylvania), and the EISAI Company.

.....

Authors' affiliations

C Tzourio, B Gagnière, A Alperovitch, INSERM U 360, Hôpital Salpêtrière, Paris, France

M El Amrani, M-G Bousser, Department of Neurology, Hôpital Lariboisière, Paris

Competing interests: none declared

REFERENCES

- 1 **Breslau N**, Rasmussen BK. The impact of migraine – epidemiology, risk factors, and co-morbidities. *Neurology* 2001;**56**:S4–12.
- 2 **Headache Classification Committee**. International Headache Society. Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. *Cephalalgia* 1988;**8**(suppl 7):1–96.
- 3 **Rasmussen BK**, Jensen R, Olesen J. Questionnaire versus clinical interview in the diagnosis of headache. *Headache* 1991;**31**:290–5.
- 4 **Michel P**, Dartigues JF, Henry P, *et al*. Validity of the International Headache Society criteria for migraine. GRIM. Groupe de Recherche Interdisciplinaire sur la Migraine. *Neuroepidemiology* 1993;**12**:51–7.
- 5 **Tzourio C**, Dufouil C, Ducimetiere P, *et al*. Cognitive decline in individuals with high blood pressure: a longitudinal study in the elderly. *Neurology* 1999;**53**:1948–52.
- 6 **Folstein MF**, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;**12**:189–98.
- 7 **Tzourio C**, Tehindrazanarivelo A, Iglesias S, *et al*. Case-control study of migraine and risk of ischaemic stroke in young women. *BMJ* 1995;**310**:830–3.
- 8 **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- 9 **Brennan P**, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;**304**:1491–4.
- 10 **Rasmussen BK**, Jensen R, Schroll M, *et al*. Epidemiology of headache in a general population – a prevalence study. *J Clin Epidemiol* 1991;**44**:1147–57.
- 11 **Granello F**, D'Alessandro R, Manzoni GC, *et al*. International Headache Society classification: interobserver reliability in the diagnosis of primary headaches. *Cephalalgia* 1994;**14**:16–20.
- 12 **Dufouil C**, deKersaintGilly A, Besancon V, *et al*. Longitudinal study of blood pressure and white matter hyperintensities – the EVA MRI cohort. *Neurology* 2001;**56**:921–6.