

Chapter 6

Epidemiological Methods for Headache Studies



Lars Jacob Stovner and Timothy J. Steiner

6.1 Introduction

The importance of epidemiological studies for headache science is now widely recognized. When headache was ranked the second most disabling disorder worldwide in the Global Burden of Disease 2016 study ([1]; also, Chaps. 5, 9 and 14), this was in large part the result of decades of headache epidemiological studies in many countries [2].

While much money and time had been committed to these studies, their methods and quality were variable [2–7]. A review of all previous studies found they had been performed, and reported, in quite different ways [2, 7]. This made it very difficult to interpret and summarize their results and, particularly, to compare the findings of studies performed in different settings and countries or at different times. Further, while quality varied, there were some studies that appeared well-enough performed, but reported in too little detail to make comparisons possible.

Headache epidemiological studies inform needs assessment, underpin service policy and gain acceptance of headache disorders as a public-health priority [6]. In addition, they can increase our understanding of causes and risk factors, thereby

L. J. Stovner (✉)

Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway

Norwegian Advisory Unit on Headaches, Department of Neurology and Clinical Neurophysiology, St Olavs University Hospital, Trondheim, Norway

e-mail: lars.stovner@ntnu.no

T. J. Steiner

Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway

Division of Brain Sciences, Imperial College London, London, UK

e-mail: t.steiner@imperial.ac.uk

improving opportunities for treatment and prevention (see Chap. 10). Quality therefore matters in these studies. The need for better and standardized methodology, supported by guidelines for the design and performance of these studies, had been evident for some years [3, 4] when, in 2014, such guidelines were published [5]. The initiative came from *Lifting The Burden* (LTB), a UK-registered charitable non-governmental organization, which directs the Global Campaign against Headache in official relations with the World Health Organization (WHO). It was supported financially by LTB, the International Headache Society (IHS), the World Headache Alliance and the Norwegian University of Science and Technology in Trondheim, Norway.

Quality improvement was the main aim, not only to improve reliability of headache epidemiological studies but also so that resources directed towards them would be better spent [5]. Further, while the guidelines would primarily be useful in the planning and performance of new studies, they could also, it was believed, be applied to evaluation of the quality of studies previously published [5]. This chapter is based on the parts of these guidelines that deal with how to measure prevalence. The parts on how to measure burden and cost are the basis of Chap. 7.

An expert consensus group was convened, with two considerations in mind: inclusion of experience and competence in headache epidemiology and epidemiology in general, and international and cross-cultural relevance. To the latter end, members were drawn from all six WHO world regions. After a first draft had been circulated, and feedback received from all members, the group convened in Trondheim in September 2011 for a 3-day meeting. A preliminary manuscript then posted on the IHS website invited worldwide open consultation, which, along with feedback from staff at WHO, informed the final, published version [5].

The recommendations here are drawn from these guidelines. They relate specifically to studies of headache, or address issues of particular relevance to headache, with more general supporting discussion when needed. It is strongly advised that headache epidemiological studies be planned and performed in collaboration with a local epidemiologist, familiar with the population of interest and culture(s). While the focus of the recommendations is on adult studies, many apply equally to studies of children and/or adolescents and others can be applied with adaptation. There are some considerations specific to studies of the elderly, and attention is drawn to these. With regard specifically to the *reporting* of studies, the STROBE statements (<http://www.strobe-statement.org/>) [8, 9] should be consulted.

The recommendations apply most readily to primary headache disorders, principally migraine and tension-type headache (TTH) (see Chap. 2). They are not intended to be exclusive to these: the principles are relevant also to epidemiological studies of secondary headaches, provided that adequate definitions can be applied in questionnaires or other survey instruments. However, burdens arising from secondary headaches are, in general, more correctly attributed to the underlying disorders. Medication-overuse headache (MOH), a secondary headache by definition, is something of an exception. The recommendations expressly encompass MOH firstly because it arises only from a pre-existing, usually primary headache disorder and, secondly, because it unquestionably contributes to public ill health. They also embrace the broader group of headaches occurring on 15 or more days per month,

again because these unquestionably contribute to public ill health. It is acknowledged later that these may be poorly characterized and, within a survey conducted by enquiries at single points in time, impossible to diagnose more specifically than as frequent headache.

6.2 Headache Epidemiology

Epidemiology is aptly described as “the study of distribution and determinants of disease frequency in human populations” [10]. Epidemiological studies are often classified as *descriptive* (setting out the distribution of disease among different groups) or *analytical* (elucidating determinants: i.e., risk factors or causes).

The usual focus of descriptive epidemiology in headache is *prevalence*, an estimate of how common a disease is, and expressed as a proportion (number of cases in a population divided by the number of individuals in that population). Also important is *incidence*, a measure of the risk of developing a condition within a specified period of time, expressed as a rate. *Remission* is the probability of a case becoming a non-case, through natural history or intervention, again within a specified period of time, and expressed as a rate. *Duration* of disease is the period between onset and remission. *Mortality* is important in epidemiology in general, but it has little relevance to primary headache disorders.

Incidence, duration, remission and prevalence are related not only conceptually but also mathematically, the last being the steady-state consequence of the first two, often summarized by the formula $P = ID$, where P = (point) prevalence, I = incidence per year and D = duration in years.

Headache terminology is unfortunately inconsistent, owing to the fact that headache is a chronic disorder with episodic manifestations. An “active headache disorder” is essentially characterized by the occurrence, at least once, of symptoms within the previous year [11]. Prevalence studies that adopt this definition of a case (i.e., a person who reports at least one headache episode during this time) necessarily use a timeframe of 1 year and usually report the findings as 1-year prevalence. Strictly speaking, however, such estimates describe the number of *current* cases (i.e., point prevalence). A different enquiry, defining a case only when symptoms are actually present (“headache now”), also estimates point prevalence, but of headache attacks, or headache as a symptom, not of a headache disorder. The terms “incidence” and “remission” can, similarly, be applied either to headache attacks or to headache disorders, although in practice they have generally referred to the latter, with specified time periods (e.g., incidence rate = number with first onset of headache per 100 person-years; remission rate = % of cases of headache disorder who then have no further attacks during 1 year of follow-up). These terms must be used carefully to avoid confusion.

These recommendations are of equal relevance to descriptive and analytical studies. But which of these is the *purpose* of the study must always be clear at the outset: it affects, fundamentally, the design of the study, the choice of study population, the size of the study sample and the information to be collected.

6.2.1 *Ethical Issues*

Ethical issues in epidemiological studies arise in their planning, conduct and reporting. Many are general. Ethical principles firmly and universally established in medical practice also apply to research in the form of headache epidemiological studies. These principles include respect for autonomy, non-maleficence, beneficence and justice [12], the last with particular reference to resource allocation in a context of limited resources (distributive justice, or equity). The ethics of the medical profession include veracity (truth-telling), fidelity (the keeping of promises) and confidentiality, and are based on the needs of patients (including potential and future patients), the responsibilities of doctors, the good of society as a whole and deserts. These have been discussed in detail elsewhere [13, 14].

Before they commence, epidemiological studies require approval by an appropriate ethics review board (ERB), usually local but, where one does not exist, from another legitimate, authoritative and competent source, such as WHO's Research Ethics Review Committee. Data protection also requires due consideration [14], and may, according to the laws of the country in which the study is conducted, require additional approvals.

Consent by participants in surveys is often implicit, and ongoing: respondents renew consent each time they provide an answer to an enquiry. Formal written consent at the start of the enquiry provides only evidence of consent at that moment, and may serve little purpose for that reason (although an ERB may require it). Consent must be informed, which requires that the purpose and nature of the survey are explained to each participant's satisfaction. Consent must also be voluntary, and participants allowed to withdraw from the study whenever they may wish.

Inducements to take part in a study that carries no risk of harm to the participant do not directly raise concerns. Both parties benefit: the researcher from the subject's participation and the subject from the inducement. However, they may raise concerns indirectly, since inducements are not of equal value to all potential subjects. Monetary inducements are more attractive to poorer people, and this does not respect the principle of equity. For the same reason, monetary inducements (and probably inducements of any sort) are likely to increase participation bias (an important quality issue discussed later).

When the inducement is the offer of needed healthcare, either free or to which the research subject would not otherwise have ready access, he or she may have little choice but to participate. Still the participant benefits. In the absence of risk of harm, even this may be considered acceptable, although a counterargument is that collecting data from needy people while offering them nothing in return is objectionable. Justification is forthcoming when the purpose of the survey is needs assessment—to inform the development of health services, which, later, will be provided for the benefit of the population being surveyed. But there is another important question, which goes beyond the ethics of reward: to what extent is there a duty of care upon researchers when untreated and possibly serious illness is discovered by a survey? This question is sometimes raised—especially in low-income

countries. These recommendations cannot give a general answer: this must be a matter for local ERBs. Two points can be noted: first, surveys are commonly made by lay interviewers, who do not themselves diagnose and have no skills to recognize illness, let alone do anything to alleviate it; second, research that may benefit a society cannot be made too onerous, or it will never be done.

Data protection, and consent relating to the use of personal data, generally requires that participants are explicitly informed of each of the following:

- Where, in what form, how and by whom data relating to them will be held
- Who will have access to them
- The purpose(s) for which they will be used, with guarantees that they will be used for no other purpose (this implies that, if data are to be stored long-term for other purposes not yet foreseen, at least a general explanation of this intention should be given)
- How they will be destroyed once the purposes are achieved.

To reduce the possibility of misuse of personal data, the duration of their storage should be as short as possible. On the other hand, it is desirable, and regulators often require it, that original research data are stored for several years, for documentation and to enable detection of fraud in science. Clearly, personal data should be handled and stored safely, respecting the privacy of participants and in accordance with the laws and regulations for data storage in the country.

Resources are limited. Studies that waste them (whether financial resources or the willingness of subjects to take part) are unethical because of the opportunity cost: other studies will not be possible as a result. Under-resourced studies that cannot achieve their purpose are likely to be unethical because the resources are probably wasted. Unscientific studies certainly waste resources, and are unethical for this reason alone. Worse, they may generate misleading results.

Adherence to these recommendations should ensure appropriate allocation of resources, and their effective use, in headache epidemiological research. However, the need for *efficiency* calls for careful deliberation of whether a particular new headache epidemiological study is required at all, and of the need for high diagnostic precision, with large sample size and resource-demanding methods of data collection. In some circumstances, a new stand-alone study can be adequately replaced, with conservation of resources, by joining a larger health survey (see later).

6.2.2 *Methodological Issues*

The methodological issues arising with regard to the design of a headache epidemiological study are largely general, and are dealt with in general textbooks on epidemiology. Here, as noted earlier, we discuss those relating specifically or addressing issues of particular relevance to studies of headache. More general supporting discussion is added only when needed.

6.2.2.1 The Study Design

This should match the purpose of the study and take due account of available resources and the general conditions in the area(s) where the study will be performed. It must be described in sufficient detail that the study can be replicated.

Most studies on headache epidemiology are descriptive, with a cross-sectional design in which prevalence and burden are assessed at the same point in time. Studies with more analytical aims, to define causes of or risk factors for headache, usually have case–control or cohort designs. Case–control studies typically compare cases (who have the disease) with controls (who are similar persons without the disease) for prior exposure to one or more putative risk factors. In cohort studies, a group (cohort) of disease-free individuals are followed and assessed periodically to determine whether they have developed the disease of interest. Within the cohort, participants are categorized as exposed or not to a suspected causal factor; incidence rates are compared after a specified time in the exposed and unexposed categories.

Methodologies differ mostly in how study samples are selected; the principles for collecting data, engaging with participants and diagnosing headaches are similar. Therefore, while these recommendations mostly concern cross-sectional studies, they will be useful in all study designs.

6.2.2.2 The Population of Interest

This is sometimes referred to as the *sampling frame*, and is the population that it is wished to study. It includes every person so defined, and is always defined geographically as well, often, as by one or more additional characteristics. It should match the purpose of the study, and the selection should be explicitly justified. It must also be described in sufficient detail that the study can be compared with others.

In headache research, the population of interest is usually, but not always, the population of a whole country, or of a region larger or smaller than a country. However, depending on the aim of the study, the population of interest may be restricted to a specific age group (e.g., adults of working age, adolescents, school or pre-school children), to members of groups defined by ethnicity, culture or language, to workers in certain trades or professions or university students, to people with another particular disease, etc. These recommendations remain relevant to studies of these more selected groups.

Headache *patient* populations are rarely legitimate populations of interest, not just because they are highly selected but also because the criteria by which they are selected (often self-selected) are generally indeterminable. A study of such populations tells little about, and cannot be extrapolated to, either the general population or any more broadly defined population. An arguable exception occurs in the case of severe headache disorders (discussed later, under *Special issues*).

It is an advantage when the characteristics of the population of interest are known (distributions of age, gender, educational levels and socio-economic status,

proportions living in rural and urban areas, etc.), since this allows evaluation of representativeness of the participating sample, and statistical adjustments with regard to these features when necessary.

6.2.2.3 Bias

Bias refers to errors that are systematic rather than random. In epidemiological studies, biases are of two main types: selection (or participation) bias and information (or measurement) bias.

Selection bias is commonly the consequence of an imperfect sampling procedure and/or a low participation proportion, resulting in a sample unrepresentative of the population of interest. With regard to gender and age composition, statistical adjustments for imbalance can be made if these properties are known for the population of interest. *Participation bias* (“interest bias”) can result from the higher tendency of people with headache to participate in headache studies, because they have more personal interest in them.

Information bias, or a systematic error in measuring disease or exposure, may occur when the manner in which information is gathered varies systematically. For example, two interviewers, one in a city and the other in a rural area, may conduct the interview differently, with differences in findings that are erroneously attributed to area of habitation. Similarly, one interviewer who completes all interviews in one location first, then all those in another, may introduce systematic differences over time in the manner of interviewing (due to a learning curve, for example), again with spuriously different results from the two locations.

Biases are almost inevitable, but should be minimized. The likely causes of bias should be identified at outset, and due steps taken to manage them. Bias may not necessarily be of much consequence, depending on its magnitude and on whether it is differential (affecting some participants more than others) or non-differential (affecting all participants equally). The potential for bias, and the need to avoid or at least minimize it, drive many aspects of the planning, execution and analysis of a study.

6.2.2.4 Sample Selection

To include every member of the population of interest in an epidemiological survey is feasible only in a minority of studies. Instead, it is necessary to choose a smaller group of people (*the sample*) to whom access is possible. It is essential that this sample is *representative* of the population of interest, in order to be able to generalize the results from the sample to the whole population of interest. “Representative” means similar to the population of interest in all properties of relevance to (i.e., likely to influence) the objects of measurement (here, headache prevalence and/or burden). There is an assumption here that knowledge exists of what these properties are, which may not be entirely true. In the context of headache, representativeness

clearly encompasses age and gender, which are known to affect headache prevalence. Also of relevance, probably, are socio-economic status, employment, area of habitation (rural or urban) and ethnicity, and possibly, in some settings, native language and/or tribal group. Methods that ensure or at least optimize representativeness with regard to a range of identified variables such as these are more likely to achieve the same with other, unrecognized variables.

Sampling introduces multiple opportunities for selection bias, which should be recognized and controlled. Sampling aims, first and foremost, to produce a group within the population of interest who are both accessible for survey and representative of it. Sampling *identifies* individuals to be surveyed, but usually does not, of itself, provide a *means of access* to these individuals, while the means of access is an important consideration when determining the sampling method.

Sampling uses either *probability* or *non-probability* methods. With the former, each member of the population of interest has an initial probability (which is larger than zero) of being selected, and this probability can be accurately determined; this is not the case with the latter. Non-probability methods include convenience sampling (selecting those who are at hand), judgemental sampling (selecting those judged to be best suited for the purpose), quota sampling (selecting quotas fulfilling particular traits) and snowball sampling (letting participants recruit future participants). With these methods, the degree to which the sample differs from the population of interest is uncontrolled, and selection bias is impossible to assess.

Probability sampling methods are therefore much to be preferred. With simple random sampling, all individuals within the population of interest have equal probability of being selected. Nonetheless, the method is vulnerable to sampling error: *by chance*, important characteristics of the sample such as gender or age distributions may not well reflect those of the whole population. Stratified sampling reduces this chance by dividing the population into sub-populations (strata), different with regard, for example, to age, gender and/or habitation, and randomly drawing the sample from within each of these strata, in parts in proportion to their size.

Both random sampling and stratified sampling are relatively easy when an overview of the population of interest exists—usually in the form of a register of all members of it. Selection can then be made directly from the register (usually by computer). A map showing all households in an area to be sampled can also serve the purpose of a register. Sampling by telephone is an established method [15]. Where telephones (landline or mobile) are widespread, but no population overview exists in the form of a complete telephone directory, random digit-dialling (area codes, followed randomly by as many digits as are typical for phone numbers in the areas) is an effective method of obtaining a random sample [15], but it risks bias because telephones are not evenly distributed among different age, gender and socio-economic groups.

Cluster sampling is an alternative in areas with no pre-existing overview of the population, but often preferable anyway because it is logistically efficient, reducing travel costs and time. Usually it involves selecting participants only from a limited number of defined geographical areas (e.g., blocks, streets or parts of villages, or

perhaps schools) that are themselves chosen randomly. Areas can also be stratified according to socio-economic status, urban/rural location, etc. In *multistage cluster sampling*, smaller areas are selected randomly within larger areas, and this is repeated in many stages until the requisite number of small, surveyable areas are identified, spread around the region or country. In these ultimate units, all individuals, or a random selection of individuals or households, are contacted.

In many studies, rather than contacting individuals, it will be easier to contact *households* (people living together and sharing the same kitchen). Generally, in headache studies, only one of a household should be selected, because members of families are similar genetically, share their environment and have common lifestyles, effectively reducing variance when two or more are included. In order to avoid bias with regard to who will be home and answer the door, the interviewer should list all members of the household, then select the participant randomly from all those who are eligible, returning by appointment for the interview when that person is not present.

The study protocol should set out the method for *replacement*, when it is impossible to contact a selected person: for example, by pre-selecting more individuals than needed, or by extending the sampling process by visiting more randomly selected households than initially specified.

The *sample size* must be sufficient to achieve the study purpose(s), but not so large as to waste resources. In determining sample size, expected prevalence and desired precision of the estimates are the only factors to consider (not, as often believed, the size of the population of interest). Desired precision is not free from issues of resources and ethics: a larger sample than needed is wasteful, whereas a study with too small a sample for its purpose is futile, and also wasteful. Several statistical programs include sample-size calculations, and there is also a simple table in the published guidelines [5].

If the requirement is to estimate, with acceptable precision, the prevalence of different headache disorders, or of subtypes of a disorder (migraine in general and migraine with aura, for example), sample size is calculated with regard to the disorder or subtype assumed to be least prevalent. The same is true when estimates within subgroups or comparisons between them (e.g., males versus females, rural versus urban) are part of the purpose: the sample size must be calculated to include sufficient of the population in the smallest subgroup. To avoid inflating the overall sample unnecessarily, it is possible to “oversample” people in that particular subgroup (i.e., select more than the proportion in the population). Such people then have a higher (but known) chance of being selected than those in other subgroups, and correction is necessary when calculating overall prevalence.

A larger sample may also be needed to estimate burden than to estimate prevalence, because burden is not distributed equally among cases: most of it is accounted for by a minority of those with the disorder. For example, 3–4% of the population have most of the burden of migraine [16]; among all people with migraine, TTH or MOH, the relatively few with MOH have the highest individual burden [17].

Cluster sampling is assumed to reduce natural variance, and therefore requires larger sample sizes to obtain the same precision of estimates [18].

The sampling method should be explicitly justified, and it should be described in sufficient detail that the study can be replicated. Biases that might have arisen from the sampling procedure should be identified and discussed.

6.2.2.5 Accessing and Engaging Participants

The ways in which members of the sample are contacted (*access*), and their willing commitment to the enquiry procured (*engagement*), are important determinants of how carefully and completely they will respond and, therefore, of data quality.

Means of access clearly depends on what means of communication exist (telephone registries, email or home address lists, up-to-date maps of residential areas) and on infrastructure (e.g., means and ease of travel). Access methods compatible with *probability sampling* include visiting households and calling by telephone (landline or mobile), usually without prior warning (cold-calling) in either case, and letters mailed or emailed to participants selected from registers that provide addresses. In some settings, participants may be invited to come to the interviewer's office.

Cold-calling at households tends to yield a higher participation proportion (among people who are at home) than telephone interviews, which are easier for the interviewee to terminate. Both methods may give rise to selection bias, because certain types of person are more likely to stay at home, open the door or answer the phone. To increase participation proportion, it is often necessary to make more than one attempt to contact a person who does not answer first time. The study protocol should define how many attempts are made (commonly three), and when, before a person or a household is deemed impossible to contact. Certain types of household are more likely to be empty, or their phones unanswered, at particular times of the day: for example, working households will be selectively uncontactable during normal working hours, while older people may not open doors to strangers in the evening. The protocol should take these factors into consideration when stipulating schedules for repeated contacts.

Accessing participants by mail is cheap, and by email even cheaper, but these methods have two major disadvantages. They presume the use of self-administered questionnaires, which can give rise to misunderstandings and therefore data of low quality. They inevitably lead to selection bias because certain types of person are less inclined to reply.

Inviting prospective participants to the office of the interviewer allows face-to-face interview, and examination if necessary, but it is time-consuming for participants, likely therefore to discourage participation and promote selection bias with regard to who has the time to attend.

Methods involving *non-probability sampling* include stopping prospective participants in the street and using lists of telephone numbers or email addresses that happen to be available rather than complete registries. These are convenience samples, biased by whatever factors cause people to be on the street, or in a selective list, and rarely useful in headache epidemiology.

Engagement, the procurement of prospective participants' willing cooperation in a study, directly affects participation proportion and, therefore, participation bias. How it is done depends upon the means of access, and also is determined by certain characteristics of the prospective participants, such as literacy, language, cognitive ability and culture.

Face-to-face interviews are the most direct method of engagement, and the only method useful in populations with poor literacy. Their disadvantage is that they are time-consuming and therefore expensive. *Telephone interviews* are almost as direct but, unlike face-to-face interviews, do not allow physical examination.

Both face-to-face and telephone interviews allow clarification of questions. While generally thought to lead to more accurate answers, clarification may instead give rise to information bias, not only because the information given to those who ask is different from that given to those who do not, but also because different interviewers may give different clarifications. Therefore, if clarification is permitted, there should be clear, pre-specified limits to the extent of it, aiming to ensure that all interviewers do it in the same way. In a computer-assisted telephone interview (CATI), interviewers follow a script driven by a computer program. Face-to-face interviews can be computer-assisted in the same way. These methods permit only pre-scripted clarifications.

Self-administered questionnaires are a relatively cheap method, but require a high degree of literacy and some familiarity with answering questionnaires. The method provides no encouragement to respond and no opportunity for clarifications, generally resulting in low participation proportions and incomplete questionnaires.

Engagement in groups (e.g., by a teacher posing questions collectively to the pupils in the classroom) can be a cost-effective way of performing a study. Engagement through a third person may be the only way to gain contact with or information from some participants, for example small children through their parents. In some cultures, people can be engaged only through village elders or heads of households. With these methods, the risk of misunderstandings must be carefully evaluated. Additionally, lower sensitivity and specificity for detecting headache must be expected from such remote engagements.

Careful selection and adequate training of interviewers are of paramount importance, whether interviews are conducted face-to-face or by telephone. As to their selection, interviewing is a skill in itself: it should not be assumed that health personnel such as nurses, medical students or, especially, doctors are the best qualified to do it. Unless the interviewer is a headache specialist (see below), diagnoses should not be made during the interview but later, by applying an algorithm to the questionnaire responses [19]. It is therefore doubtful, in most surveys, whether clinical skills are important: it may be better to engage professionally trained interviewers who understand interview methodology and who follow the questionnaire and operations manual.

Adequate training of interviewers embraces a clear understanding of the nature and purpose of the survey, and a recognition of which questions are of particular importance or may need clarification. When there are multiple interviewers, training

should be identical for each of them to ensure that data are collected in the same way by all, without introduction of information bias. While multiple interviewers almost inevitably introduce some degree of variability, more interviewers reduce the duration of the study, which, if long, can also be a source of variation.

When the interviewer is a headache specialist (a physician skilled in headache diagnosis and familiar with the culture and language of the respondent), not only can the diagnosis be made during the interview but, further, multiple diagnoses, where appropriate, can be made in the same participant. If examination and supplementary investigations are made, secondary headaches can be diagnosed. No validation of the diagnostic method is needed, since it can be assumed that optimal diagnostic methods are employed. Headache specialists, of course, may not explicitly use ICHD diagnostic criteria [11]: they have at their disposal and are likely to apply, as in the clinic, a broader, experienced-based and more inclusive set of criteria for diagnosis, which ICHD-based questionnaires can never match. High sensitivity can be expected for detecting relatively minor headache complaints and rare headaches.

In low-income countries, headache specialists are simply not available.

6.2.2.6 Participation and Non-participation

The *participation proportion* is an important result of a study; in particular it is important for evaluating bias. But it has no universal definition [9]. It is generally understood as the proportion of those *selected, contacted and eligible* who actually participate in the study meaningfully (i.e., provide answers to most questions). Calculation of participation proportion therefore excludes those in the preselected sample who a) were not contactable (because they had died, or moved away since the study was planned, or because no-one answered the phone or opened the door), or b) were contacted but found to be ineligible (because it was not possible to seek their consent, for whatever reason, or they did not fulfil the eligibility criteria [wrong age or gender, for example]).

In almost all cases, *non-participation* (among those contacted and eligible) results either from outright refusal to participate or, among those consenting to take part, from inability to cooperate usefully, perhaps answering only a few questions or providing conflicting responses.

Only a very high participation proportion, rarely achieved, guarantees *representativeness* (assuming correct sampling). Participation proportions $>80\%$ are considered excellent, and $\geq 70\%$ acceptable, but even these do not secure representativeness. On the other hand, a low participation proportion does not necessarily mean the included sample is unrepresentative: this depends on the factors responsible for non-participation [9]. But usually, because probability of participation tends to vary between different subgroups of a sample (e.g., it may be particularly low among young males), a low participation proportion influences the overall results (i.e., it is a source of bias). In other words, both how many and who actually participate in the study are crucial to representativeness.

Since non-participation cannot be altogether avoided, all reasons for it (declined to participate, too sick to be interviewed, did not return the questionnaire, completed the questionnaire inadequately or inappropriately, etc.) should be listed, and, if possible, coupled with at least a demographic description of each non-participant. They can then be taken into account in evaluations of sample representativeness, and the likelihood, magnitude and influence of participation bias.

When participation proportion is low, representativeness may be estimated through a study of non-participants, although this is possible only when the initial sampling was from a register including contact details. Enquiry is usually by telephone, calling a random sample of non-participants, and it must be limited to a few key questions (e.g., age, gender, the screening question(s) for headache, perhaps one on headache frequency and a very few on diagnosis). This minimal dataset, at least showing whether non-participants are similar to or very different from participants in the main study, is highly valuable in assessing various types of selection bias. An additional question, asking why they did not participate initially, will inform discussion of selection bias. It may also help in the better design of future studies.

6.2.2.7 Method of Enquiry: The Study Instrument

A *structured questionnaire* (prescribed questions, with predefined response options) is usual both to identify (and diagnose) headache caseness and for enquiry into headache burden. Some questions may call for open answers (e.g., number of days with headache, names of medicines) but, generally, open answers are difficult to interpret and categorize; for diagnosis, they do not permit algorithmic determination (see below).

Structured questionnaires should, depending on the study purpose(s), include identifier(s), demographics (age, gender, education, employment, personal and/or household income, habitation [urban, rural], ethnicity [when relevant], etc.), screening and sieve questions, diagnostic questions, enquiry into symptom burden (headache frequency, duration, intensity, etc.) and one or more other elements of burden (disability, time loss, family burden, cost, etc.) ([19]; also, Chaps. 4 and 7).

Quality of the questionnaire is fundamental to the quality of the entire study: nothing can compensate for failures in data collection. Time and resources are well spent in developing a good questionnaire, and much attention should be paid to layout (clarity and ease of use reduce error rates), intelligibility, acceptability and content (how many diagnoses are necessary, and what aspects of burden are to be measured?).

Questionnaires should be *parsimonious*, and not include any questions that do not contribute to the study's purpose(s): irrelevant questions not only irritate participants but also create unnecessary workload. Further, questionnaires should avoid enquiries that are irrelevant to particular participants, by directing them past sections that are not applicable. Ideally, of course, all questions pertinent to the survey are answered clearly by every participant, but certain questions (screening question,

diagnostic questions and, perhaps, key questions on burden) *must* be answered if the respondent is not to be classified as a non-participant.

A *pre-pilot study* in a small convenience sample drawn from patients in the clinic may be most useful to test whether questions are culturally inoffensive and that questionnaire length (time to complete) is acceptable.

Meticulous translation is then essential, following a rigorous translation protocol [20]. A larger *pilot study* in a convenience sample drawn from the population of interest should be conducted after translation to ensure that questions are understood correctly and discover those that may cause problems and require clarification. The pilot studies may uncover need for amendment(s) to the questionnaires or retranslation (and, potentially, retesting).

If the questionnaire is newly developed, validation in the language and specific setting of the study is highly desirable, at least of the diagnostic part (see below).

6.2.2.8 Case Definition and Diagnosis

In all headache epidemiological studies, it is of fundamental importance to define who is a case and who is not. How caseness is defined can greatly influence the results, and this concerns the definition of headache, and of its types or subtypes. “Migraine” may include all its types and subtypes (ICHD-3 codes 1.1–1.6) [11] in some studies, but only migraine with or without aura (ICHD-3 codes 1.1 and 1.2) [11] in others. “Migraine with aura” may or may not include the subtypes aura without headache and aura with non-migrainous headache. “Migraine” and “tension-type headache” may include both episodic and chronic types or be restricted to the former, with chronic types subsumed within the general category of headache occurring on 15 or more days per month.

In all studies, the timeframe should be defined. Caseness in headache epidemiological studies usually implies an active headache disorder, defined in ICHD-II as headache during the last year [21]; hence *1-year prevalence* is most used, and it allows the most comparisons with previous studies. Shorter timeframes (6-month and 3-month) are also quite common, but adopt different definitions of caseness.

Recall errors probably generate some information bias, not least because they are greater in the elderly. Obviously, these are greater over longer periods: very short and recent timeframes, such as *1-day prevalence* (headache today, or yesterday) avoid recall problems almost completely. Estimates of 1-day prevalence do not describe the proportion of the population with an active headache disorder, but they yield very accurate information on burden in the population ([22]; also, Chap. 7).

Headache yesterday is probably most correctly estimated when respondents are contacted directly by face-to-face interview or telephone, and not given the opportunity to choose when to answer. If the question is posed by letter, for example, it is conceivable (indeed, not unlikely) that people with headache on the day of receiving it will postpone answering until the first headache-free day, resulting in a spuriously high reported prevalence of headache yesterday.

Life-time prevalence (“Have you ever had headache?”) has mostly been used for the rarer headache disorders such as cluster headache because it increases the likelihood of finding cases. It is also relevant in genetic epidemiological studies, which must eliminate those who have ever had the disorder from control groups. Here, recall error may be problematic, especially in the elderly; young people may have better recall, but there is no “long ago”.

Lifetime, 1-year and 1-day prevalences can be ascertained in a single study with appropriate design and enquiry.

To detect cases, most studies use a two-stage procedure. First, a *screening question* asks participants whether they have headache or not (within the designated timeframe); second, those answering affirmatively are posed the diagnostic questions. In some studies, screening employs a relatively simple self-administered questionnaire, whereas, in the second stage, screen-positive participants are subjected to more thorough face-to-face or telephone interviews. Two-stage procedures are time-saving and avoid irrelevant enquiry, but are at risk from false-negative screening. Hence, sensitivity may be lower, particularly for minor headache complaints, than is achieved by subjecting all participants to full personal interview.

The screening question(s) that define headache caseness are of crucial importance, since the study results depend to a large degree on their exact phrasing. A neutral question (such as “Have you had headache during the last year?”) will include almost all cases, even those for whom headache is only a minor nuisance, yielding higher prevalence estimates than a non-neutral question specifying degree, frequency, intensity or circumstances of its occurrence [23] (“Have you *suffered from* headache?”, “Have you had *frequent* headache?”, “Have you had *bad* headache?”, “Have you had headache *not due to* hangover, head injury, flu or common cold?”). Screening questions should always be reported *verbatim*.

After a neutral question, additional questions can then *sieve* participants: identify, and perhaps exclude, those with very low frequency (e.g., <1/month), intensity or functional impairment, who are of less interest from medical or public-health perspectives.

Most headache epidemiological studies wish to distinguish between different *headache types*. For this purpose, diagnostic questions built into the structured questionnaire should, ideally, apply ICHD-3 criteria [11]. However, these criteria were primarily designed for clinical use and not for epidemiological enquiry, for which they are not particularly well-suited: their strict application requires that all participants are personally interviewed by a competent clinician and, in many cases, examined, which is rarely feasible. Additionally, they are expressed in technical language, which is particularly relevant in distinguishing between migraine and TTH, the two most common disorders. Questions about headache duration [2] require patients to consider untreated attacks, which they may never have, or last had long ago. There are no easy lay explanations of photo- and phonophobia. In epidemiological studies, there must be modifications of these criteria, which, therefore, should be tested in a validation study (see below). All modifications should be reported (not engulfed in the unhelpful expression, “modified ICHD criteria”).

Migraine aura is very difficult to diagnose by questionnaire. For this reason, it should not dictate the broader diagnosis of migraine.

The group of disorders characterized by headache on 15 or more days per month present a particular problem, since precise diagnosis, from enquiry at a single point in time, is generally difficult and often impossible (or, at least, highly unreliable). The important subgroup with medication overuse can be identified, but labelled only as “probable MOH”, since causation cannot be established.

Diagnoses are best made by *algorithm*, applied after interview to the recorded responses [19]. This separates the (non-expert) interviewer from the diagnostic process, and, when there are multiple interviewers, ensures uniformity in the process. The algorithmic flow is important: probable MOH should be diagnosed before the primary headaches, and migraine before TTH. Probable migraine and probable TTH follow (see below).

An alternative approach to diagnosing headache types is a *recognition-based method*, presenting descriptions (case vignettes) or pictorial representations of different headaches to participants [24]. Recognition-based diagnosis is appropriate and convenient—and may give the most accurate results—in certain settings and populations (especially young children).

6.2.2.9 “Definite” and “Probable” Diagnoses

ICHD-3 allows the diagnosis of *probable* headache types when all but one of the diagnostic criteria for a disorder are met, provided that not all are met for another disorder [11]. For example, a headache cannot be diagnosed as probable migraine if it meets the criteria for TTH. In epidemiological studies, common experience is that about half of diagnoses are probable according to these rules [25, 26], while validation studies in subgroups of the same populations find, according to expert diagnoses, that fewer than 10% are probable. This may be a consequence of applying modified ICHD-3 criteria in the main study, but not in the validation; however, empirically, duration of (supposedly untreated) attacks appears to be unreliably reported.

“Probable migraine” and “probable TTH” are not separate diagnostic entities, and it is unhelpful to report them as though they were. Cases of “probable migraine”, not meeting the criteria for TTH, are more probably migraine than anything else, and vice versa. The guidelines recommend that “definite migraine” and “probable migraine” be reported separately, then combined (“all migraine”) for further analysis; and, similarly, “definite TTH” and “probable TTH” (“all TTH”) [5]. When a diagnostic algorithm is applied, diagnoses must be made in the order dictated by ICHD-3 [11]: migraine, TTH, probable migraine, probable TTH.

6.2.2.10 How Many Diagnoses in Each Individual?

Participants in epidemiological surveys may have more than one headache type. This poses problems, since it may be difficult for them to remember which features belong to which headache type, and attribute them accordingly; only a headache

specialist as interviewer can sort this out correctly. A solution, valid for public-health purposes, is to ask the participant to identify the *most bothersome headache* (the one, in his or her mind, that interferes most with life), and focus solely on this when responding. This approach, adopted in the Headache-Attributed Restriction, Disability, Social Handicap and Impaired Participation (HARDSHIP) questionnaire ([19]; also, Chap. 7), will tend to underestimate less troubling headache types, therefore systematically neglecting TTH among participants with both TTH and migraine. Arguably, the observed prevalence of TTH in studies where this is done should be inflated by the same proportion of those with migraine. This is rarely done: conservatism is preferred.

Multiple diagnoses require more questions, discouraging participation and leading to a lower participation proportion; studies should carefully assess their value. When the purpose is to assess population headache burden, it is neither realistic nor necessary to assess the burden attributable to each headache type in each person. For public-health purposes, it is more important to avoid double counting, which this might encourage.

6.2.2.11 Validation of the Questionnaire

Diagnostic accuracy is usually important. *Diagnostic validation* is performed preferably in a randomly selected sub-sample of participants in the main study; stratified sampling, according to caseness and diagnosis (no headache, migraine, TTH or probable MOH), can then ensure adequate numbers of each. Otherwise, a separate sample is drawn from the population of interest, but stratification is then not possible. Headache patients are *not* an acceptable substitute: their headache disorders are not representative of those of the population of interest, and, often with more knowledge of headache and experience of anamnesis, their performance is conditioned when answering questionnaires.

Validation requires that participants are re-interviewed by a headache expert (a physician skilled in headache diagnosis and familiar with the culture and language of the respondent), applying clinical skills to diagnose (the “gold standard”) while being ignorant of the questionnaire diagnoses. To minimize discrepancies due to change in the headache itself, expert interviews should follow soon after the original diagnoses (no more than 1 month). They are preferably conducted face-to-face, but telephone interviews are an acceptable resource-conserving alternative when respondents are widely spread geographically.

Questionnaire-derived and expert diagnoses are then compared, which allows calculation of the instrument’s sensitivity, specificity, positive and negative predictive values and kappa value for each diagnosis [27]. The precision (confidence interval) of the estimate for each diagnosis is dependent on the number with that diagnosis in the validation sample. Validations performed in the last decade have typically included 180–500 participants overall, drawn randomly as a sub-sample from the main study, giving relatively narrow confidence intervals for migraine and TTH [25–29] but not always for probable MOH [27]. Validation of the screening question (“Have you had headache ...?”) is achieved in those who answered “No”.

The validation study can inform and optimize the diagnostic algorithm: small amendments can be made (e.g., excluding from the diagnostic criteria for migraine the minimum duration of 4 h, which has been found empirically to cause problems), and the comparisons remade between questionnaire-derived and expert diagnoses, which may or may not lead to improvement. There is usually a trade-off between sensitivity and specificity: one increases at the expense of the other. Genetic studies require caseness certainty, and therefore high specificity; for most other studies, the algorithm that gives the highest overall diagnostic accuracy (sum of sensitivity and specificity) is preferable.

Validation is clearly not possible in countries with no headache experts (no gold standard available); yet, in such countries, studies of headache burden may be of particular importance. In these circumstances, it is advisable to employ a diagnostic questionnaire used and validated already, in multiple languages and cultures.

6.2.2.12 Data-Entry

Data should be recorded in the most practical way, which is usually on paper, but sometimes directly into computer. Highly portable tablet computers promise direct computer-entry as the way of the future, with the very great advantage of obviating data transcription.

When data must be transferred from source documents to computer, rigorous quality controls are essential, ideally employing full double data-entry (two people independently transcribe all data) to produce two full datasets. These are electronically compared to detect errors (inconsistencies), which are resolved by reference to the source documents. At minimum, the proportion of errors should be estimated in a subset of the data (e.g., 10% of items), entered twice. If this proportion (including its confidence interval) is judged likely to have more than negligible influence on the results to be reported, full double data-entry is required.

6.2.3 Special Issues

6.2.3.1 Studies of Particular Age Groups

In countries with obligatory schooling, representative samples of *children and adolescents* of school age can be obtained by selecting all, or a random sample, of the pupils of a representative sample of schools. This is a cluster-sampling method, and selected schools should reflect the diversities of the country or region with regard to socio-economic status and area of habitation (rural or urban). Interviews may be undertaken by specially trained interviewers, or by a teacher or school nurse. For small children, some information may have to be given by parents or teachers.

Somewhat different diagnostic criteria apply in these age groups, although still essentially based on ICHD-3 ([30]; also, Chap. 3). Diagnostic instruments require

modification accordingly, and, of course, these and other measuring instruments must reflect the language and comprehension skills of the age range. Validation of the diagnostic method is as important as in studies on adults.

Studies in *elderly populations* (usually defined as >65 years of age) can expect lower prevalences of primary headache disorders and higher prevalences of secondary headaches, more comorbidities, reduced mental and physical capacities (which may affect engagement and ease of data collection), and lower workforce participation (which affects the impact of headache and the way burden is estimated). For these reasons, larger samples may be needed, and there may be more focus on diagnosis of secondary headaches—at least to avoid their misdiagnosis as primary headache disorders when these are the object of interest. Comorbidities, as potential confounders in estimations of burden, may also require greater consideration, and more importance attached to adjustments for comorbidities in analysis.

6.2.3.2 Studies of Rare Headaches

Prevalence of rare headaches is difficult to assess. Very large samples are needed and, because they are not easily identified through self-administered questionnaires, headache experts must usually make the diagnoses.

However, only rare headaches that are severe, for example, cluster headache and trigeminal neuralgia, are of any public-health interest. Conditions such as these demand medical attention, which, in societies with good access to medical care and well-kept electronic (i.e., searchable) records, makes it possible to detect potential cases by screening large clinic populations such as general practitioners' lists. Diagnoses, in far smaller numbers of candidates, can then be ascertained by expert interview [31].

Most published prevalence estimates of rare headaches are of lifetime prevalence, because cases are more numerous; but current prevalence (1-year or 1-day) is likely to be of greater interest.

6.2.3.3 Stand-Alone Studies Versus Large Health Surveys

A relatively long questionnaire, such as HARSHIP ([19]; also, Chap. 7), is most suited to a study wholly dedicated to headache (*stand-alone study*).

The large burden arising from headache disorders should equally motivate analytical epidemiological studies to determine risk factors and causes of headache (see Chap. 10), some of which may be preventable. These purposes may not be achievable within the same studies that measure prevalence and burden: they require, in addition, collection and registration of relevant exposure data, and follow-up studies, possible only in *large health surveys* that register multiple disorders and potential risk factors.

However, large health surveys cannot feasibly include a full headache questionnaire; instead, the necessary minimum is a valid screening question for active

headache disorder, some questions allowing determination of severity (frequency, intensity and duration), and ideally a question set for the diagnosis of migraine and TTH (as mutually exclusive diagnoses). Validation of these questions remains as important as in stand-alone studies.

6.3 Concluding Remarks

This chapter recounts, and slightly updates, consensus methodological guidelines for headache epidemiological studies first published in 2014 [5]. It makes recommendations covering ethical aspects, key epidemiological concepts and the methodological issues around study design, sample selection, accessing and engaging participants, study instruments, diagnosis of headache types and sources of error.

Its purpose is not to inform those who are planning studies (who should refer to the original) so much as to assist critical evaluation of published reports of headache prevalence and burden, given that these are of very variable quality.

References

1. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulkader RS, Abdulle AM, Abebo TA, Abera SF, Aboyans V, Abu-Raddad LJ, Ackerman IN, Adamu AA, Adetokunboh O, Afarideh M, Afshin A, Agarwal SK, Aggarwal R, Agrawal A, Agrawal S, Ahmadi H, Ahmed MB, Aichour MTE, Aichour AN, Aichour I, Aiyar S, Akinyemi RO, Akseer N, Al Lami FH, Alahdab F, Al-Aly Z, Alam K, Alam N, Alam T, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1211–59.
2. Stovner LJ, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, Steiner TJ, Zwart JA. The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia*. 2007;27:193–210.
3. Stovner LJ, Scher AI. Epidemiology of headache. In: Olesen J, Goadsby PJ, Ramadan NM, Tfelt-Hansen P, Welch KMA, editors. *The headaches*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 17–26.
4. Steiner TJ, Stovner LJ, Al Jumah M, Birbeck GL, Gururaj G, Jensen R, Katsarava Z, Queiroz LP, Scher AI, Tekle-Haimanot R, Wang S-J, Martelletti P, Dua T, Chatterji S. Improving quality in population surveys of headache prevalence, burden and cost: key methodological considerations. *J Headache Pain*. 2013;14:87.
5. Stovner LJ, Al Jumah M, Birbeck GL, Gururaj G, Jensen R, Katsarava Z, Queiroz LP, Scher AI, Tekle-Haimanot R, Wang SJ, Steiner TJ. The methodology of population surveys of headache prevalence, burden and cost: principles and recommendations from the Global Campaign against Headache. *J Headache Pain*. 2014;15:5.
6. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, et al. Years lived with disability

- (YLD) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2163–96.
7. Stovner LJ, André C. Prevalence of headache in Europe: a review for the Eurolight project. *J Headache Pain*. 2010;11:289–99.
 8. Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JPA, Kirsch-Volders M, Matullo G, Phillips DH, Schoket B, Stromberg U, Vermeulen R, Wild C, Porta M, Vineis P. Strengthening the Reporting of OBservational studies in Epidemiology—Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. *PLoS Med*. 2011;8:e1001117.
 9. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007;18:805–35.
 10. Hennekens CH, Buring JE. *Epidemiology in medicine*. Philadelphia: Lippincott, Williams & Wilkins; 1987.
 11. Headache Classification Committee of the International Headache Society. The international classification of headache disorders. 3rd edition. *Cephalalgia*. 2018;38:1–211.
 12. Beauchamp TL, Childress JF. *Principles of biomedical ethics*. 6th ed. New York: Oxford University Press; 2008.
 13. Steiner TJ. Ethical issues in headache research and management. Report and recommendations. *Cephalalgia*. 1998;18:505–29.
 14. Steiner TJ. Ethical issues arising from commercial sponsorship and from relationships with the pharmaceutical industry—report and recommendations of the Ethics Subcommittee of the International Headache Society. *Cephalalgia*. 2008;28(Suppl 3):1–25.
 15. Olson SH, Kelsey JL, Pearson TA, Levin B. Evaluation of random digit dialing as a method of control selection in case-control studies. *Am J Epidemiol*. 1992;135:210–22.
 16. Linde M, Dahlöf C. Attitudes and burden of disease among self-considered migraineurs—a nation-wide population-based survey in Sweden. *Cephalalgia*. 2004;24:455–65.
 17. Linde M, Gustavsson A, Stovner LJ, Steiner TJ, Barré J, Katsarava Z, Lainez JM, Lampl C, Lantéri-Minet M, Rastenyte D, Ruiz de la Torre E, Tassorelli C, André C. The cost of headache disorders in Europe: the Eurolight project. *Eur J Neurol*. 2012;19:703–11.
 18. Schaeffer RL, Ott RL, Mendenhall W, Gerow KG. *Elementary survey sampling*. 7th ed. Belmont, CA: Duxbury Press; 2011.
 19. Steiner TJ, Gururaj G, André C, Katsarava Z, Ayzenberg I, Yu SY, Al Jumah M, Tekle-Haimanot R, Birbeck GL, Herekar A, Linde M, Mbewe E, Manandhar K, Risal A, Jensen R, Queiroz LP, Scher AI, Wang SJ, Stovner LJ. Diagnosis, prevalence estimation and burden measurement in population surveys of headache: presenting the HARSHIP questionnaire. *J Headache Pain*. 2014;15:3.
 20. Peters M, Bertelote JM, Houchin C, Kandoura T, Steiner TJ. Translation protocols. *J Headache Pain*. 2007;8(suppl 1):S40–7.
 21. Classification Subcommittee of the International Headache Society. The international classification of headache disorders. 2nd edition. *Cephalalgia*. 2004;24(Suppl 1):1–160.
 22. Yu S, He M, Liu R, Feng J, Qiao X, Yang X, Cao X, Zhao G, Fang Y, Steiner TJ. Headache yesterday in China: a new approach to estimating the burden of headache, applied in a general-population survey in China. *Cephalalgia*. 2013;33:1211–7.
 23. Stovner LJ, Zwart JA, Hagen K, Terwindt G, Pascual J. Epidemiology of headache in Europe. *Eur J Neurol*. 2006;13:333–45.
 24. Zwart JA, Dyb G, Stovner LJ, Sand T, Holmen TL. The validity of ‘recognition-based’ headache diagnoses in adolescents. Data from the Nord-Trøndelag health study 1995–97, head-HUNT-youth. *Cephalalgia*. 2003;23:223–9.
 25. Yu S-Y, Cao X-T, Zhao G, Yang X-S, Qiao X-Y, Fang Y-N, Feng J-C, Liu R-Z, Steiner TJ. The burden of headache in China: validation of diagnostic questionnaire for a population-based survey. *J Headache Pain*. 2011;12:141–6.
 26. Ayzenberg I, Katsarava Z, Mathalikov R, Chernysh M, Osipova V, Tabeeva G, Steiner TJ. The burden of headache in Russia: validation of the diagnostic questionnaire in a population-based sample. *Eur J Neurol*. 2011;18:454–9.

27. Hagen K, Zwart JA, Aamodt AH, Nilsen KB, Brathen G, Helde G, Stjern M, Tronvik EA, Stovner LJ. The validity of questionnaire-based diagnoses: the third Nord-Trøndelag health study 2006–2008. *J Headache Pain*. 2010;11:67–73.
28. Lipton RB, Dodick D, Sadosky R, Kolodner K, Endicott J, Hettiarachchi J, Harrison W. A self-administered screener for migraine in primary care: the ID migraine validation study. *Neurology*. 2003;61:375–82.
29. Kukava M, Dzagnidze A, Mirvelashvili E, Djibuti M, Fritsche G, Jensen R, Stovner LJ, Steiner TJ, Katsarava Z. Validation of a Georgian language headache questionnaire in a population-based sample. *J Headache Pain*. 2007;8:321–4.
30. Wöber C, Wöber-Bingöl C, Uluduz D, Aslan TS, Uygunoglu U, Tüfekçi A, Alp SI, Duman T, Sürgün F, Emir GK, Demir CF, Balgetir F, Özdemir YB, Auer T, Siva A, Steiner TJ. Undifferentiated headache: broadening the approach to headache in children and adolescents, with supporting evidence from a nationwide school-based cross-sectional survey in Turkey. *J Headache Pain*. 2018;19:18.
31. Torelli P, Beghi E, Manzoni GC. Cluster headache prevalence in the Italian general population. *Neurology*. 2005;64:469–74.