▪ Headache Medicine 2026, 17(3):226-231 ▪ p-ISSN 2178-7468, e-ISSN 2763-6178

Headache Medicine

▪ DOI: 10.48208/HeadacheMed.2026.29

Review

How many patients do we need? Sample size calculation in headache research

Marcos Antonio Inacio de Oliveira Filho[1], Marcelo Moraes Valença[2]

[1]BAKIS, Sharjah, United Arab Emirates
[2]Federal University of Pernambuco, Recife, Pernambuco, Brazil

Marcos Antonio Inacio de Oliveira Filho
maiof@hotmail.com

Edited by:
Juliana Ramos de Andrade

Keywords:
Sample size
Statistical power
Migraine
Headache
Cluster headache
Clinical trials
Research methodology

Abstract

An adequately planned sample size is a fundamental component of clinical research. Studies that enroll too few participants may fail to detect clinically important effects, whereas unnecessarily large studies consume resources and may expose more participants than needed. This narrative review provides a practical introduction to sample size calculation for clinicians and investigators conducting headache research. We review the null and alternative hypotheses, type I and type II errors, statistical power, target difference (effect size), dropout, pilot studies, clustered designs, and repeated measurements, using examples from migraine, cluster headache, acute-treatment trials, preventive trials, observational studies, and interventional headache research. Particular attention is given to clinically meaningful differences in monthly migraine or headache days, responder outcomes, 2-hour pain-freedom endpoints, variability and missing data in headache diaries, and recruitment constraints in uncommon headache disorders. We emphasize that the calculation should be aligned with the prespecified primary outcome and planned analysis, that every assumption should be justified and reported transparently, and that feasibility should not be used retrospectively to legitimize an underpowered efficacy study. Early collaboration with a biostatistician remains strongly recommended.

Received: July 7, 2026
Revised: July 19, 2026
Accepted: September 21, 2026

© Copyright 2026

Introduction

Sample size calculation estimates, before data collection, how many participants or observations are needed for a study to answer its primary research question with an acceptable probability of detecting a clinically relevant effect. Although the arithmetic is often delegated to a statistician or software package, the assumptions are scientific decisions. Investigators must define the primary outcome, the target difference, the expected variability or event rate, the significance level, the desired power, the allocation ratio, and any allowance for missing outcome data or attrition. Current CONSORT guidance emphasizes transparent reporting of these assumptions so that the calculation can, in principle, be reproduced (1) .

This issue is particularly relevant to headache medicine. Headache outcomes are often recurrent, patient-reported, and variable over time. Preventive migraine trials may use change in monthly migraine days (MMDs), monthly headache days (MHDs), or responder rates; acute migraine trials commonly use binary endpoints such as pain freedom at 2 hours; cluster headache studies may analyze attack frequency; procedural studies may use pain intensity or headache-day outcomes; and observational studies may estimate prevalence, associations, diagnostic accuracy, or prognosis (2–5). A sample size calculation that is appropriate for one of these questions may be inappropriate for another.

The aim of this narrative review is to explain the central concepts of sample size planning in non-technical language and to show how they apply specifically to headache research. The goal is not to replace a biostatistician, but to help clinical investigators formulate better questions, select defensible assumptions, recognize common errors, and report calculations transparently.

Fundamental concepts

Null and alternative hypotheses

A sample size calculation should be linked to a clearly defined primary outcome and analysis. The null hypothesis (H0) generally states that there is no difference between groups, whereas the alternative hypothesis (H1) states that a difference exists. The outcome scale and planned statistical model determine what information is required for the calculation. For example, a preventive migraine trial based on mean change in MMDs requires assumptions about the target mean difference and variability, whereas an acute-treatment trial based on 2-hour pain freedom requires expected response proportions in the treatment and control groups (6).

Type I error and significance level

Type I error (α) is the probability, under the null hypothesis, of incorrectly concluding that a difference exists. A two-sided α of 0.05 is common in clinical research. This should be interpreted as a 5% long-run false-positive rate for the specified test when H0 is true, not as a guarantee that exactly one of every 20 studies will be falsely positive. Lowering α, while holding other assumptions constant, increases the required sample size (6,7).

Type II error and statistical power

Type II error (β) is the probability of failing to detect the prespecified target difference when that difference truly exists. Statistical power is 1 − β. Values of 80% or 90% are commonly used. Higher power requires a larger sample when the significance level, target difference, and variability are unchanged. Prospective power is useful for design; post hoc power calculated from the observed effect in a completed study is generally uninformative because it is mathematically tied to the observed P value and effect estimate (6–8).

Target difference and clinical importance

The target difference is the magnitude of difference that the study is designed to detect reliably. It should be plausible and, whenever possible, clinically important. The DELTA2 guidance emphasizes that investigators should use prior evidence, stakeholder judgment, and sensitivity analyses rather than select an effect size merely because it yields a convenient sample size (7). In headache research, this distinction is crucial: a statistically detectable difference in MMDs is not automatically a difference that matters to patients.

For continuous outcomes, the target difference can be expressed in natural units, such as a between-group difference of 1.5 MMDs, together with an expected standard deviation. Standardized measures such as Cohen's d may be useful, but conventional labels of 0.2, 0.5, and 0.8 should not be treated as universal thresholds of clinical importance (9). For binary outcomes, the calculation requires anticipated proportions in each group, such as the proportion achieving pain freedom at 2 hours or a ≥50% reduction in MMDs.

Why headache research requires special attention

The International Headache Society (IHS) has developed disorder- and treatment-specific recommendations for clinical trials, including acute migraine, preventive episodic migraine, preventive chronic migraine, neuromodulation, and cluster headache (2–5). These guidelines underscore a central methodological point: the sample size calculation must reflect the outcome that the trial is actually designed to test.

Headache frequency is intrinsically variable. A participant reporting 12 migraine days during one month may report a different number in the next month even without an effective intervention. Longitudinal trials therefore need to consider baseline frequency, repeated observations, within-participant correlation, and the

statistical model planned for the primary analysis. A simple independent-means calculation can be conservative or misleading when the final analysis uses repeated measures, covariate adjustment, count models, or mixed-effects models. In more complex settings, simulation-based sample size estimation may be preferable.

Headache diaries create another distinctive problem. Participant dropout and missing diary data are not synonymous. A participant may remain enrolled while providing too few diary entries for an efficacy period. The protocol should therefore define evaluability and missing-data rules in advance and distinguish inflation for expected attrition from assumptions about incomplete outcome ascertainment. CONSORT 2025 specifically emphasizes transparent handling and reporting of missing data (1).

Table 1. Common headache outcomes and their sample size inputs

Clinical question	Typical primary outcome	Outcome structure	Key inputs for sample size
Migraine prevention	Change in monthly migraine days	Continuous / longitudinal	Target mean difference, SD, baseline-adjustment or within-person correlation, α, power
Migraine prevention	≥ 50% responder rate	Binary	Expected response proportions in treatment and control groups, α, power
Acute migraine treatment	Pain freedom at 2 hours	Binary	Expected proportions in each group, allocation ratio, α, power
Acute migraine treatment	Freedom from most bothersome symptom	Binary	Expected proportions in each group, α, power
Cluster headache prevention	Weekly attack frequency	Count / longitudinal	Expected rates or rate ratio, dispersion, repeated-measure structure, α, power
Interventional headache study	Change in NRS/VAS or headache days	Continuous / longitudinal	Target difference, SD, within-person correlation when repeated
Prevalence study	Migraine or headache prevalence	Proportion	Expected prevalence, desired margin of error, confidence level, design effect
Diagnostic study	Sensitivity and/or specificity	Proportion	Expected sensitivity/specificity, desired precision, disease prevalence or case-control sampling plan

Worked examples in headache research

Example 1: a preventive migraine trial with a continuous outcome

Suppose investigators plan a two-arm, 1:1 randomized superiority trial and select change in MMDs as the primary outcome. They judge a between-group difference of 1.5 MMDs to be clinically relevant and expect a common standard deviation of 4.0 days. Using a two-sided α of 0.05 and 80% power, a simple independent-means calculation requires approximately 112 participants per group. If 15% attrition is anticipated, the enrollment target should be inflated by dividing by 0.85, yielding approximately 132 participants per group (264 total). This example is deliberately simple: if the primary analysis uses baseline adjustment or repeated measurements, the sample size calculation should reflect that analysis and its correlation assumptions.

Example 2: an acute migraine trial with a binary outcome

Consider a trial in which 2-hour pain freedom is the primary endpoint. If investigators anticipate pain freedom in 30% of participants receiving active treatment and 18% receiving placebo, the calculation should compare two proportions. The assumed control response and target absolute difference of 12 percentage points must be justified from prior trials or other credible evidence. Because the required sample size is sensitive to both proportions, sensitivity analyses using alternative plausible placebo and treatment response rates are useful before the final enrollment target is chosen (2,7).

Example 3: small interventional studies

Peripheral nerve blocks, trigger-point procedures, neuromodulation, and other interventional headache studies are often performed in single centers with limited recruitment. A small feasible sample does not automatically make a study a valid efficacy trial. Investigators should first decide whether the objective is efficacy, estimation, safety, or feasibility. If a definitive efficacy sample is unattainable, the study should be framed honestly as exploratory or pilot research and should emphasize estimation with confidence intervals rather than overinterpretation of nonsignificant P values.

Additional considerations in practice

Adjustment for anticipated dropout

If n participants must complete the primary outcome assessment and the expected dropout proportion is d, the enrollment target is n/(1 − d), not n + nd. Thus, if 500 evaluable participants are required and 10% attrition is expected, 500/0.90 = 555.6, so approximately 556 should be enrolled. The manuscript should state whether the reported target already includes this inflation (1,6).

Pilot and feasibility studies

Pilot studies should primarily address feasibility: recruitment, retention, acceptability, intervention delivery, diary completion, variance estimation, or unforeseen procedural problems. They are usually too small to provide a precise estimate of treatment efficacy. Viechtbauer and colleagues showed that, under a simple problem-detection framework, 29 participants provide approximately a 95% probability of observing at least one event when the event occurs in 10% of participants, whereas 59 are required when the event probability is 5% (10–12). These numbers should not be misused as universal pilot-study sample sizes; the appropriate number depends on the pilot objective.

Clustered, crossover, and repeated-measures designs

When participants are grouped within clinics or centers, outcomes may be correlated within clusters. Cluster-randomized studies therefore require inflation for the design effect, commonly related to cluster size and the intracluster correlation coefficient. Crossover and within-person designs require assumptions about within-participant correlation, period effects, and potential carryover. Repeated-measures headache trials should similarly align the sample size method with the planned longitudinal analysis. When analytic formulas are inadequate, simulation is often the most transparent approach (1,6).

Rare headache disorders and recruitment constraints

Cluster headache and other uncommon primary headache disorders illustrate the tension between statistical requirements and recruitment feasibility. The statistically desirable sample may exceed the number of eligible patients available at a single center. Feasibility, however, should not be used retrospectively to justify an underpowered efficacy claim. Multicenter recruitment, appropriate crossover or within-person designs, repeated-measures methods, or explicitly labeled pilot/feasibility studies may be more defensible. Guidance for cluster headache trials stresses standardized outcomes and prespecified treatment effects so that studies can be interpreted and compared across settings (5).

Observational and diagnostic headache studies

Not all headache research is randomized. Cross-sectional prevalence studies require assumptions about the expected prevalence and desired precision of the estimate. Case-control and cohort studies require assumptions about exposure frequency, event rates, effect measures, follow-up, and confounding structure. Diagnostic studies may be designed around desired precision for sensitivity and specificity. Studies of biomarkers, CGRP, neuroimaging, quantitative sensory testing, or anatomical measurements may require continuous-outcome, correlation, regression, or multivariable prediction approaches. The sample size method should therefore follow the primary scientific question rather than a generic formula.

A particular caution applies to multivariable analyses. Choosing a sample size only from an arbitrary rule such as a fixed number of participants per predictor can be inadequate. The expected outcome frequency, number of candidate parameters, anticipated model performance, shrinkage, and validation strategy should be considered, preferably with contemporary design guidance and statistical collaboration.

How should sample size be reported?

The current CONSORT 2025 statement requires authors of randomized trials to explain how sample size was determined and to report the assumptions supporting the calculation (1). For a conventional two-arm superiority trial, readers should be able to identify the primary outcome and time point, target difference, expected variability or control event rate, α and whether it is one- or two-sided, power, allocation ratio, resulting sample size, any inflation for attrition or other adjustments, and the formula, method, or software when relevant. DELTA2 further recommends explaining why the target difference is important or realistic and exploring sensitivity to uncertain assumptions (7).

For headache trials, authors should additionally make clear how diary completeness, baseline headache frequency, repeated measurements, and responder definitions were incorporated when relevant. The calculation described in the Methods should correspond to the actual prespecified primary analysis; a calculation for one endpoint cannot validate a different primary analysis performed later.

Practical recommendations for headache investigators

1. Define one primary outcome and its assessment time point before calculating sample size.
2. Choose a target difference that is clinically important and realistic; justify it from prior headache trials, patient-relevant evidence, or explicit expert/stakeholder judgment.
3. Match the calculation to the planned statistical analysis, including repeated measures, baseline adjustment, clustering, crossover structure, or count outcomes when applicable.
4. Distinguish participant attrition from incomplete headache-diary data and prespecify how each will be handled.
5. Use sensitivity analyses when key assumptions such as placebo response, standard deviation, intracluster correlation, or dropout are uncertain.
6. Do not use feasibility alone to justify an underpowered efficacy study; redesign, collaborate across centers, or label the work appropriately as pilot/exploratory.
7. Report every assumption transparently and follow CONSORT 2025 and relevant IHS trial-design guidance.
8. Involve a biostatistician during protocol development rather than after data collection.

Conclusion

Sample size calculation is not a bureaucratic requirement but part of the scientific question itself. In headache research, the challenge is amplified by recurrent symptoms, diary-based outcomes, within-person variability, diverse endpoint types, and recruitment constraints in uncommon disorders. A defensible calculation begins with a clinically meaningful primary outcome, uses assumptions that are both transparent and plausible, reflects the planned analysis, and anticipates missing data and attrition. Adequate planning protects participants, improves interpretability, and makes negative as well as positive findings more informative.

References

1. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025;389:e081123. Doi:10.1136/bmj-2024-081123.
2. Diener H-C, Tassorelli C, Dodick DW, Silberstein SD, Lipton RB, Ashina M, et al. Guidelines of the International Headache Society for controlled trials of acute treatment of migraine attacks in adults: Fourth edition. Cephalalgia 2019;39:687–710. Doi:10.1177/0333102419828967.
3. Diener H-C, Tassorelli C, Dodick DW, Silberstein SD, Lipton RB, Ashina M, et al. Guidelines of the International Headache Society for controlled trials of preventive treatment of migraine attacks in episodic migraine in adults. Cephalalgia 2020;40:1026–44. Doi:10.1177/0333102420941839.
4. Tassorelli C, Diener H-C, Dodick DW, Silberstein SD, Lipton RB, Ashina M, et al. Guidelines of the International Headache Society for controlled trials of preventive treatment of chronic migraine in adults. Cephalalgia 2018;38:815–32. Doi:10.1177/0333102418758283.
5. Medrea I, Tepper SJ, Wang D, Mathew PG, Burish M. Commentary on 2022 guidelines on clinical trial design in cluster headache and further suggestions. J Headache Pain 2024;25:32. Doi:10.1186/s10194-024-01732-3.
6. Ranganathan P, Deo V, Pramesh CS. Sample size calculation in clinical research. Perspect Clin Res 2024;15:155–9. Doi:10.4103/picr.picr_100_24.
7. Cook JA, Julious SA, Sones W, Hampson L V, Hewitt C, Berlin JA, et al. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. BMJ 2018:k3750. Doi:10.1136/bmj.k3750.
8. Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. J Bone Joint Surg 2001;83:397–402. Doi:10.1302/0301-620X.83B3.10582.
9. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
10. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015;68:1375–9. Doi:10.1016/j.jclinepi.2015.04.014.
11. Miot HA. Tamanho da amostra em estudos clínicos e experimentais. J Vasc Bras 2011;10:275–8. Doi:10.1590/S1677-54492011000400001.
12. Althubaiti A. Sample size determination: A practical guide for health researchers. J Gen Fam Med 2023;24:72–8. Doi:10.1002/jgf2.600.

Marcos Antonio Inacio de Oliveira Filho
https://orcid.org/0000-0003-4836-2302
Marcelo Moraes Valença
https://orcid.org/0000-0003-0678-3782

Author contributions: MAIOF: Conducted the literature search, drafted the manuscript; MMV: Conception, critically revised the content.
Conflict of interest: The authors declare no conflicts of interest.
Funding: This work received no specific funding.