Published in final edited form as:

Headache. 2013 June ; 53(6): 901–907. doi:10.1111/head.12127.

Rethinking Headache Chronification

Dana P. Turner, MSPH¹, Todd A. Smitherman, PhD², Donald B. Penzien, PhD³, Richard B. Lipton, MD⁴, and Timothy T. Houle, PhD¹

¹Department of Anesthesiology, Wake Forest University Medical School, Winston-Salem, NC, USA

²Department of Psychology, University of Mississippi, Oxford, MS USA

³Head Pain Center, University of Mississippi Medical Center, Jackson, MS, USA

⁴Departments of Neurology, Epidemiology and Population Health and the Montefiore Headache Center, Albert Einstein College of Medicine, Bronx, NY, USA

Abstract

The objective of this series is to examine several threats to the interpretation of headache chronification studies that arise from methodological issues. The study of headache chronification has extensively used longitudinal designs with two or more measurement occasions. Unfortunately, application of these designs when combined with the common practice of extreme score selection as well as the extant challenges in measuring headache frequency rates (eg, unreliability, regression to the mean), induces substantive threats to accurate interpretation of findings. Partitioning the amount of observed variance in rates of chronification and remission attributable to regression artifacts is a critical yet previously overlooked step to learning more about headache as a potentially progressive disease. In this series on rethinking headache chronification, we provide an overview of methodological issues in this area (this paper), highlight the influence of rounding error on estimates of headache frequency (second paper), examine the influence of random error and regression artifacts on estimates of chronification and remission (third paper), and consider future directions for this line of research (fourth paper).

Keywords

Headache chronification; chronic migraine; methodology; statistics

Corresponding author: Timothy T. Houle, Ph.D. Department of Anesthesiology, Wake Forest University Medical School. Medical Center Boulevard, Winston-Salem, NC 27157. thoule@wfubmc.edu.

Conflicts of Interest: Dana P. Turner: Ms. Turner receives research support from Merck.

Todd A. Smitherman: Dr. Smitherman receives research support Merck.

Donald B. Penzien: Dr. Penzien receives research support from Merck.

Richard B. Lipton: Dr. Lipton serves/has served on scientific advisory boards for and received funding for travel from Allergan, Inc., Bayer Schering Pharma, Endo Pharmaceuticals, GlaxoSmithKline, Kowa Pharmaceuticals America, Inc., Merck Serono, Neuralieve Inc., and Ortho-McNeil-Janssen Pharmaceuticals, Inc.; serves as Associate Editor of Cephalalgia and on the editorial boards of Neurology® and Headache; receives royalties from publishing Headache in Clinical Practice (Isis Medical Media, 2002), Headache in Primary Care (Isis Medical Media, 1999), Wolff's Headache (Oxford University Press, 2001, 2008), Managing Migraine: A Physician's Guide (BC Decker, 2008), and Managing Migraine: A Patient's Guide (BC Decker, 2008); has received speaker honoraria from the National Headache Foundation, the University of Oklahoma, the American Academy of Neurology, the Annenberg Foundation, Merck Serono, GlaxoSmithKline, and Coherex Medical.

Timothy T. Houle: Dr. Houle receives research support from GlaxoSmithKline and Merck.

Introduction

The two most common primary headache disorders, tension type headache and migraine, are characterized by episodic attacks of headache that sometimes increase in frequency over time.¹ Headache ‘progression’ describes the clinical phenomenon of increasing attack frequency and is usually applied when episodic tension type headache (ETTH) or episodic migraine (EM) increases in frequency over time. ‘Chronification’ or ‘transformation’ refers to a specific form of progression in which headache frequency increases to 15 or more headache days per month for at least 3 months.

Clinical observation and epidemiologic studies suggest that episodic tension-type headache (ETTH) can transform into chronic tension type headache (CTTH) and episodic migraine (EM) can transform into chronic migraine (CM). Population studies have consistently demonstrated that 3% - 5% of adults suffer from chronic headache at any given point in time.^{2,3} Only a minority of persons with EM develop CM. Because CM is far more disabling for individuals and costly for society than EM, predicting and preventing chronification has emerged as a public health priority.^{4,5,6}

With the publication of a landmark analysis in 2000, Wang et al.⁷ initiated an effort to identify rates of headache chronification and the factors associated with this phenomenon. Since that time, numerous well-conducted studies have indexed the degree of chronification in the population^{2,8,9} and identified risk factors that predict chronification empirically.^{10,11} Collectively, these studies indicate that approximately 2.5% of individuals with episodic headache are at risk for transforming to a chronic headache phenotype.⁹ Others have identified two broad classes of risk factors, those that are non-modifiable (eg, demographic variables, head trauma) and those that are potentially amenable to modification (eg, high attack frequency, medication overuse, obesity, snoring, stressful life events).^{2,10,11}

Research on rates and risk factors for migraine chronification has employed variations of several well-established study designs (eg., case control studies, cohort studies, etc.). Despite the importance of this work on headache chronification, and the rich findings that have resulted, several methodological issues threaten the validity of these studies and the resulting conclusions. Calculation of chronification rates and associated predictors likely are complicated by substantive issues related to the research designs and statistical analyses of these studies. In what follows, these issues will be systematically introduced to consider how they may impact existing conclusions on headache chronification.

Longitudinal Designs with Two Measurement Occasions

Longitudinal studies of headache prognosis assess headache frequency at two or more points in time (ie, baseline and one or more follow-up times). Variations of this longitudinal design have been the dominant design used to study headache chronification; their characteristics are displayed in Table 1. Scher et al.,² Katsarava et al.,⁸ and Bigal et al.⁹ each used two assessments that were interceded by a period of 1 year. Lu et al.¹² and Wang et al.⁷ assessed headache frequency in their samples at intervals separated by 2 years, though Wang et al.⁷ actually aggregated two collection periods into one follow-up assessment. Hagen et al.¹³

utilized a lengthy pre-post interval with 11 years between the two measurement occasions. Appropriate to these research designs, no experimental manipulations existed in these studies; the frequency of headache attacks as reported in surveys was simply compared at the two measurement occasions. As there are no experimentally manipulated groups in these designs, groups of participants must be created/selected from within the overall sample. In studies of headache chronification, these groups are constructed based on differences in baseline headache frequency (eg, episodic [< 15 days/month] versus chronic [≥ 15 days/month] patients, or some further delineation within these frequency groups). This practice is useful for calculating changes in frequency or comparing risk factors across groups and corresponds with internationally accepted diagnostic guidelines. Unfortunately, the threshold of 15 or more headache days per month is at best a fallible indicator of headache biology and creates methodological problems due to extreme score selection, as discussed in the following section.

Extreme Group/Score Selection

The issue of extreme group/score selection is a salient threat to the interpretation of longitudinal research results. Typically, extreme score selection involves the biased enrollment of patients whose illness is either particularly severe or particularly mild at the time of enrollment. In clinic- or health plan-based samples, patients may seek care when their illness is at its worst and then regress towards their own or the population mean even in the absence of effective intervention. In an observational study, this return over time to a more typical level illness severity (headache frequency) may be falsely attributed to an intervention. In the absence of an external comparison group (ie, a control group from a randomized controlled trial), whether any observed change is simply a function of extreme score selection cannot be ascertained. The literature is replete with reports of severely ill patients appearing to benefit more from an intervention than less severely ill patients (see: ¹⁴), when these patients certainly exhibited extreme scores at baseline and simply may have benefitted from a “natural” return to their normal state.

A similar problem occurs when groups of participants are internally stratified based on baseline disease severity. In migraine research, this process may contribute to the high rates of CM remission reported in observational studies. Individuals whose usual attack frequency is approximately 12 days per month, on occasion might experience a brief period of headache exacerbation during which they may have report 15 or more headache days per month and thus be misclassified as CM. Upon returning to their usual level of attack frequency of 12 headache days, it would falsely appear such patients had experienced a remission to EM from CM. In an observational study, this phenomenon can inflate estimates of remission rate. Likewise, in an uncontrolled treatment study, it can lead to overestimation of treatment effects. In a randomized controlled trial, this phenomenon can contribute to a high placebo rate and attenuate the relative benefit of active treatment.

In headache chronification research, extreme score/group selection arises as a consequence of our frequency denominated classification system. The International Classification of Headache Disorders (ICHD) defines CM and CTTH in part based on a headache frequency of 15 or more days per month for at least 3 months. In the seminal studies examining rates

and predictors of headache chronification, the baseline surveys were stratified according to headache frequency. Scher et al.² used controls (< 104 days/year), intermediate frequency (105 to 179 days/year), and cases (> 180 days/year) to define frequency groups. Katsarava et al.⁸ utilized categories of low (0 to 4 days/month), intermediate (6 to 9 days/month), critical (10 to 14 days/month), and chronic daily headache (CDH; ≥ 15 days/month) to define frequency groups. Bigal et al.⁹ utilized the ICHD-2 criteria for episodic (< 15 days/month) and chronic (≥ 15 days/month) frequency groups, and both Bigal et al.⁹ and Scher et al.² also utilized actual baseline headache frequency as continuous predictors. In each of these elegant studies, these artificially created groups were used to calculate the risk of progressing to a more frequent state at the time of the next assessment period. While it would be difficult to study this issue without such a practice, this choice creates groups that exhibit extreme scores in relation to the population mean. Extreme scores by themselves can pose interpretive difficulties but are especially problematic when unreliability is inherent in headache frequency measurements.

Reliability (Unreliability)

In classical test theory, reliability is the consistency of a measurement when observed under similar conditions.¹⁵ Although there are different types and ways to measure reliability, a fundamental axiom is that a measurement (eg, a monthly headache frequency) is considered to consist of a 'true score' plus some amount of error.

$$\text{Observed score} = \text{True score} + \text{Error}$$

The 'true score' is the actual construct being measured and, depending on the precision the construct is afforded, is the reason to expect consistency across measurements.¹⁵ For example, an individual's true height should not change with two assessment periods separated by only 10 minutes. The 'error' represents any context or event that causes the height measurements to vary from one measurement occasion to another (eg, on the second occasion, the participant slouches or the measurer reads the height differently). When two measurements are taken and the true score is unchanged, the only variability that exists in the observed score is measurement error. In this way, the reliability of a measurement is the consistency of the observed measurement including any sources of error. Measurements that have little error variability are considered reliable in that they will yield a similar score under similar conditions across measurement occasions. Unreliable measurements will vary substantially from occasion to occasion even when the true score does not change at all.

In headache chronification research, as in research on practically any topic, the measurements of interest are imperfectly reliable in that they contain at least some measurement error. This is especially true in survey research of headache frequency because its quantification requires retrospection. Headache frequency is a rate (eg, headache days per time period) and is typically quantified by asking individuals on how many days over the last week or month they have had a headache. In each of the aforementioned seminal studies on headache chronification, participants were required to recollect (without the benefit of a headache diary) the number of days with headache they experienced during such a time frame. Such retrospection is challenging to complete with accuracy and can be influenced by

a host of factors including intervening life events, affective distress, time since last headache, length of the actual retrospective period (ie, longer recall periods lead to a more representative headache sample but less precision), and memory capacity. Difficulties also arise from the distinction between enumerating headache attacks versus days wherein any headache pain was experienced (ie, does waking with a headache that quickly resolves count as a headache day?). Making these frequency estimates for an event that happened some time ago requires the use of any number of cognitive heuristics (See: ¹⁶). A companion manuscript examines this issue in some detail to illustrate the error in measurements that can be expected when individuals are asked to report their headache activity over the past month. Due to these measurement limitations, headache frequency reporting is imperfectly reliable but even more problematic when assessed in a study that has initially relied on extreme score selection and two measurement occasions. When these factors are present, a phenomenon known as regression to the mean is a statistical inevitability.

Regression Artifact: Regression to the Mean

In longitudinal research, regression to the mean is a statistical artifact by which extreme scores measured at one time are more likely to be closer to (“regress to”) the population mean when measured on a second occasion.¹⁷ More generally, the phenomenon occurs whenever the level of one variable is used to predict the level of another variable and there exists an imperfect correlation between the two variables. First described by Galton,¹⁸ regression to the mean has been plaguing experimental researchers for over 100 years (See: ¹⁹⁻²¹). Group mean data regress toward the mean with increasing magnitude as the correlation between two measurement occasions decreases. As such, the effect is best described as an artifact in that it is ever present to some degree and easily misinterpreted as being something other than a statistical anomaly. As Campbell and Kenny assert, regression to the mean is a “fact” (p.19) that occurs every time there is less than perfect correlation between two measurement occasions.²⁰

To provide a straightforward graphical depiction of the process of regression to the mean, Figure 1 presents three scenarios with decreasing correlation between the Time 1 and Time 2 measures. As the degree of correlation between the two occasions decreases (ie, the similarity between the average individual’s Time 1 and Time 2 score decreases), the amount that extreme scores regress to the mean increases. Significant regression to the mean occurs even when the correlation between two occasions is quite high (such as $r = 0.80$ as depicted in Figure 1), a correlation akin to a common standard for ‘high’ test-retest reliability in assessment instruments (see: ²²). Indeed, when the distribution of scores is normal (Gaussian), the amount of regression to the mean is readily predicted given the correlation between the two time periods.

Unfortunately, in headache chronification research, studying the expected degree of regression artifacts is more challenging than for the graphical examples in Figure 1 because the underlying headache frequency distributions are not normally distributed in the population. In the general population, most individuals exhibit very infrequent headaches. For instance, the modal frequency of episodic migraine is 1-4 days per month (62.7%); only 13.8% have migraine 5-14 days per month.²³ Distributions such as these that deviate from a

Gaussian (normal) distribution are much less studied in terms of expected regression artifacts. The companion manuscript²⁴ displays the distinctly non-normal shape of the distribution of headache frequencies in the population. The study of regression artifacts in headache chronification research is complex because it is extremely difficult to distinguish between a likely regression artifact (ie, infrequent headaches appearing to become more frequent among an initially low-frequency headache group or frequent headaches appearing to become less frequent among an initially high-frequency group) and the possibility that headache is in fact a progressive disease (ie, headache frequency is genuinely changing over time) for some individuals. Any factor that diminishes the already-imperfect correlation between the two measurement occasions will serve to increase the degree of regression artifacts, including the chronification effect itself. In other words, chronification estimates and effects are at least partially confounded with the factors that would lead to spurious results.

Chronification series

Headache chronification studies have relied on longitudinal research designs with two measurement occasions, have utilized extreme group selection, rely on a headache frequency outcome measure that is imperfectly reliable, and, as a result, *must* be subject to observing some degree of regression to the mean. How much, if any, of the published headache chronification and remission rates can be attributed to these artifacts? Figure 2 demonstrates two competing hypotheses that can be informally postulated. Hypothesis 1 is that all of the observed variance in progression and remission rates is due to meaningful changes occurring within an individual over time (ie, headache is in fact a progressive disease). Hypothesis 2 is that the dynamics observed in the chronification and remission rates are due exclusively to regression artifacts in the sample and represent *only* variability observed in frequency counts that is a function of measurement error. Of course, there is an expansive middle ground between the two hypotheses. Partitioning the amount of observed variance in rates of chronification and remission attributable to regression artifacts is a critical yet previously overlooked step to learning more about headache as a potentially progressive disease.

This series of manuscripts will systematically examine this issue to estimate the expected degree of regression to the mean that should be observed under a variety of assumptions and conditions. In the first companion manuscript, we examine one very particular, and currently unrecognized, source of measurement error in headache progression studies (ie, the “heaping” of frequency count estimates.²⁴ In the second companion manuscript, a thorough simulation study is used to estimate how much regression to the mean is expected under various ranges of measurement error.²⁵ This simulation estimates the rates of chronification (ie, episodic headache frequency range changing to chronic headache frequency range) and remission (ie, chronic headache frequency range to episodic headache frequency range) that can be predicted from random measurement error alone. A final companion manuscript concludes by summarizing what was learned about threats to the interpretation of the headache chronification literature and provides several recommendations to improve our study of this area by reducing significant methodological threats.²⁶ Our hope is that by highlighting these salient and ever-present threats, we can begin to rethink our understanding of headache chronification and improve our research efforts in future designs.

Acknowledgments

Financial Support: Supported by **NIH/NINDS** R01NS065257.

References

1. Lipton RB, Bigal MG. Concepts and mechanisms of migraine chronification. *Headache*. 2008; 48:7–15. [PubMed: 18184280]
2. Scher AI, Stewart WF, Ricci JA, Lipton RB. Factors associated with the onset and remission of chronic daily headache in a population-based study. *Pain*. 2003; 106:81–89. [PubMed: 14581114]
3. Manack A, Buse DC, Serrano D, Turkel CC, Lipton RB. Rates, predictors, and consequences of remission from chronic migraine to episodic migraine. *Neurology*. 2011; 76:711–718. [PubMed: 21270413]
4. Stewart WF, Wood GC, Manack A, Varon SF, Buse DC, Lipton RB. Employment and work impact of chronic migraine and episodic migraine. *J Occup Environ Med*. 2010; 52:8–14. [PubMed: 20042889]
5. Stokes M, Becker WJ, Lipton RB, Sullivan SD, Wilcox TK, Wells L, Manack A, Proskorovsky I, Gladstone J, Buse DC, Varon SF, Goadsby PJ, Blumenfeld AM. Cost of health care among patients with chronic and episodic migraine in Canada and the USA: Results from the International Burden of Migraine Study (IBMS). *Headache*. 2011; 51:1058–1077. [PubMed: 21762134]
6. Serrano D, Manack AN, Reed ML, Buse DC, Varon SF, Lipton RB. Cost and predictors of lost productive time in chronic migraine and episodic migraine: Results from the American Migraine Prevalence and Prevention (AMPP) study. *Value Health*. 2013; 16:31–38. [PubMed: 23337213]
7. Wang SJ, Fuh JL, Lu SR, Liu CY, Hsu LC, Wang PN, Liu HC. Chronic daily headache in Chinese elderly: Prevalence, risk factors, and biannual follow-up. *Neurology*. 2000; 54:314–319. [PubMed: 10668689]
8. Katsarava Z, Schneeweiss S, Kurth T, Kroener U, Fritsche G, Eikermann A, Diener HC, Limmroth V. Incidence and predictors for chronicity of headache in patients with episodic migraine. *Neurology*. 2004; 62:788–790. [PubMed: 15007133]
9. Bigal ME, Serrano D, Buse D, Scher A, Stewart WF, Lipton RB. Acute migraine medications and evolution from episodic to chronic migraine: A longitudinal population-based study. *Headache*. 2008; 48:1157–1168. [PubMed: 18808500]
10. Bigal ME, Lipton RB. Modifiable risk factors for migraine progression. *Headache*. 2006; 46:1334–1343. [PubMed: 17040331]
11. Scher AI, Midgette LA, Lipton RB. Risk factors for headache chronification. *Headache*. 2008; 48:16–25. [PubMed: 18184281]
12. Lu SR, Fuh JL, Chen WT, Juang KD, Wang SJ. Chronic daily headache in Taipei, Taiwan: Prevalence, follow-up and outcome predictors. *Cephalalgia*. 2001; 21:980–986. [PubMed: 11843870]
13. Hagen K, Vatten L, Stovener LJ, Zwart JA, Krokstad S, Bovim G. Low socio-economic status is associated with increased risk of frequent headache: A prospective study of 22718 adults in Norway. *Cephalalgia*. 2002; 22:672–679. [PubMed: 12383064]
14. Whitney CW, Von Korff M. Regression toward the mean in treated versus untreated chronic pain. *Pain*. 1992; 50:281–285. [PubMed: 1280801]
15. Novick MR. The axioms and principal results of classical test theory. *J Mathematical Psychol*. 1966; 3:1–18.
16. Tversky, Amos; Daniel, Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*. 1973; 5:207–232.
17. Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol*. 1976; 104:493–498. [PubMed: 984023]
18. Galton F. Regression towards mediocrity in hereditary stature. *J Roy Anthropol Inst Great Brit Ireland*. 1886; 15:246–263.
19. Campbell, DT.; Stanley, JC. *Experimental and Quasi-Experimental Designs for Research*. Rand McNally; Chicago, IL: 1966.

20. Campbell, DT.; Kenny, DA. A Primer on Regression Artifacts. Guilford Press; New York, NY: 1999.
21. Rogosa, D. Myths about longitudinal research. In: Schaie, KW.; Campbell, RT.; Meredith, W.; Rawlings, SC., editors. Methodological Issues in Aging Research. Springer; New York, NY: 1988. p. 171-210.
22. Nunnally, JC. Psychometric Theory. 2nd ed. McGraw-Hill; New York, NY: 1978.
23. Lipton RB, Bigal ME, Diamond M, Freitag F, Reed ML, Stewart WF, AMPP Advisory Group. Migraine prevalence, disease burden, and the need for preventive therapy. *Neurology*. 2007; 68:343–349. [PubMed: 17261680]
24. Houle TT, Turner DP, Houle TA, Smitherman TA, Martin V, Penzien DB, Lipton RB. Rounding behavior in the reporting of headache frequency complicates headache chronification research. *Headache*. 2013; 53:908–919. [PubMed: 23721238]
25. Houle TT, Turner DP, Smitherman TA, Penzien DB, Lipton RB. Influence of random measurement error on estimated rates of headache chronification and remission. *Headache*. 2013; 53:920–929. [PubMed: 23721239]
26. Lipton RB, Penzien DB, Turner DP, Houle TT. Methodological issues in studying rates and predictors of migraine progression and remission. *Headache*. 2013; 53:930–934. [PubMed: 23721240]

Figure 1.

Several examples of regression to the mean. When there is a perfect correlation ($r = 1.0$) between two time measurement occasions (Time 1 and Time 2), there is no regression to the mean (standardized Z-scores remain consistent, top panel). However, when the two measurement occasions are imperfectly related (ie, an individual has a different score at Time 1 and Time 2), as will be the case when the variable being measured is imperfectly reliable, regression to the mean is inevitable. In the middle panel, Time 1 and Time 2 scores that are correlated $r = 0.80$ exhibit less regression to the mean than scores that are correlated $r = 0.50$ (standardized Z-scores regress to the mean, lower panel).

Figure 2.

This series on headache chronification will examine the current view of headache chronification (Hypothesis 1 [H1]) as a disease wherein a subset of individuals exhibit increased headache frequency (up arrow) on a second measurement occasion vs a novel hypothesis that posits that at least some degree of the headache chronification/remission rates are due to statistical artifacts (up and down arrows) and not durable change (Hypothesis 2 [H2]). T1 = Time 1; T2 = Time 2.

Table 1

Summary of Longitudinal Studies of Headache Chronification

Study Authors (Year)	Source of sample and sample size	Population	Follow-up time	Chronification rate	Remission rate
Wang et al. (2000)	Population-based N = 1,533	Chinese	2 & 4 years	NA	33%
Lu et al (2001)	Population-based N = 3,377	Taiwanese	2 years	NA	65%
Hagen et al. (2002)	Population-based N = 22,718	Norwegian	11 years	NA	NA
Scher et al. (2003)	Population-based, N = 1,932	American	11 months (average)	3%	57%
Katsarava et al (2004)	Clinic-based N = 532	German	1 year	14%	NA
Bigal et al. (2008)	Population-based N = 16,577	American	1 year	2.5%	14.7%