Agreement Between a Daily Electronic Headache Diary and Self-Report Questionnaire from a Dose-Response Randomized Controlled Trial of Spinal Manipulation Therapy for Cervicogenic Headache

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Linda M. Hanson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Roni L. Evans PhD, DC, MS

July 2016

Acknowledgements

I would like to express sincere gratitude to my thesis advisor Roni Evans, PhD, MS, DC for her council, generous contribution of time and expertise, and encouragement throughout the life cycle of this project. Also, heartfelt thanks to Gert Bronfort, PhD, DC and Craig Schulz, DC, MS for their thoughtful feedback, humor, and encouraging support. Big cheers to my colleagues and dear friends, Brent Leininger DC, MS and Corrie Vihsadt MS, L.Ac.-- I am grateful for your methodological input, edits, and the many laughs, debates, and memories shared throughout our program.

This thesis project is based on data collected in a randomized controlled trial of which the successful completion was dependent on many. Huge thanks to the Wolfe-Harris Center for Clinical Studies team (2011-2015) and to our collaborators and friends at the Center for Outcome Studies in Portland, OR, for your diligence and dependability, careful attention to detail, willingness to arrive early and stay late, adherence to study protocols and for the extraordinary care and respect you showed to all study participants. Thank you.

I wish to acknowledge the Co-Principal Investigators of the trial for which data from this project was drawn: Drs. Gert Bronfort and Mitch Haas. Special thanks to the funding agency, the National Institutes of Health/National Center for Complementary and Integrative Health (R01AT006330, R-25AT003582) and the NCMIC Foundation for their generous financial support and continued commitment to lifelong learning.

Dedication

This thesis is dedicated to my sisters, Abby and Ivy Hanson whose humor, tenacity and zest for life continue to encourage and inspire me daily.

Abstract

Objective: To determine if a minimally burdensome, easy to administer self-report questionnaire based on recall can be used interchangeably with a gold-standard headache diary to measure headache outcomes in an adult cervicogenic headache (CGH) population.

Design: Secondary analysis of outcomes collected in a prospective, parallel group, observer blinded dose-response randomized controlled trial of spinal manipulation for CGH (R01AT006330).

Setting: General population in Minneapolis/St. Paul metropolitan region (MN, USA)

Participants: 18 years of age and older with a history of chronic CGH (5 headache days per month for three months) and a pain intensity of ≥ 3 (0-10) who are otherwise in good health at baseline.

Measurements: A daily, electronic headache diary (gold standard) and self-report questionnaire based on recall are used to ascertain headache frequency, measured in days, and intensity, measured using the 11 point Likert scale (0-10), collected over four weeks at baseline.

Analysis: Baseline characteristics are summarized using descriptive statistics. The Bland Altman method is used to assess agreement, including limits of agreement (mean difference ± 2SDs). Linear regression is used to evaluate the presence of proportional bias. A two-tailed t-test is used as a measure of inference for mean differences between measures.

Results: 87 participants are included in this analysis. The mean difference (SD) and limits of agreement (LOA) for CGH frequency and intensity are 0.77 (4.3) days (LOA: -7.6 - 9.1) and 0.14 (0.8) points (LOA: -1.43 - 1.70), respectively. Linear regression shows evidence of proportional bias for headache intensity (β=0.286, 95% CI 0.01-0.27, p=0.000). Group differences between the questionnaire and diary were not statistically significant: frequency t(86) = 1.69, p=0.09 and intensity t(86)=1.6, p=0.11.

Conclusions: There is a lack of agreement between the questionnaire and electronic diary for measuring headache frequency and intensity. It is not recommended to use the questionnaire in lieu of the gold-standard daily headache diary for measuring headache outcomes in clinical research. A self-report questionnaire based on recall may be appropriate however to inform CGH management in clinical settings.

Table of Contents

List of Tables

List of Figures

Background

Headache, although seemingly benign, is a common and disabling chronic pain condition affecting persons of all ages, races, and socioeconomic classes worldwide.(1) About half to three quarters of the global adult population experience headache during their lifetime, and one half on an annual basis. (1, 2) Sufferers report increased disability, (2) diminished quality of life, (1) stressed personal relationships, (3) and fear of stigmatization. (1) Further, the societal burdens credited to headache are huge. Total headache expenditures are estimated at 20 billion dollars annually in the US (4) which are attributed to the direct costs associated with diagnosis and management and indirect costs from millions of lost work and school days each year.(4)

Cervicogenic headache (CGH) is a secondary headache condition characterized by head pain from an osseous or soft tissue disorder in the cervical spine. (5) Some sufferers also experience decreased cervical range of motion and tenderness in the cervical paraspinal soft tissues as precursors to their headache. (6) The estimated lifetime prevalence of CGH is about 4.1% (7), and the point prevalence ranges from 0.4% to 4.6% (7-9) likely due to variations in the headache's definition. The quality of life impact of CGH is substantial, with burdens similar to episodic tension-type and migraine without aura, the two most common primary headache conditions. (1, 10)

Substantial consumer utilization of complementary and alternative medicine (CAM) in the US has been observed, (11) particularly for neck pain and headaches.(12) Of the estimated 34% of US adults who utilize CAM annually, spinal manipulative therapy (SMT) is among the most common.(11) There is a growing body of literature that supports the use of SMT for headache, and to date, no other intervention has been shown to be superior to manipulation for CGH

management. (13, 14) Despite evidence of efficacy, there remains little consensus on the appropriate dose of SMT needed to achieve maximal benefit, (15, 16) thus necessitating the need for more rigorous, methodologically sound research.

Measurement of Headaches

Headaches are subjective experiences whereby correct diagnosis and appropriate management are dependent on the patient's accurate reflection of their headache experience. Headache characteristics and outcomes (e.g., headache frequency and intensity) are routinely collected for diagnosis and to measure treatment effects in clinical research (5). Various self-report mechanisms, such as patient history, patient-oriented outcome questionnaires, and headache diaries are commonly used. Self-report is an easy, inexpensive method for obtaining headache information, avoiding the use of expensive and often inconclusive equipment and serum markers. Self-report approaches are subject to limitations however, which may affect methodological and data quality.

Questionnaires

Self-report questionnaires are routinely used to collect health outcomes, (17) and generally require persons to reflect back on a period of time. In general, they are easy to administer, minimally burdensome, and familiar to most people. Their low cost nature allows for greater reach to larger samples and compared to other types of data collection, such as one-on-one qualitative interviews, they can make data entry and analysis a smoother, less time-consuming process. Despite these advantages, there exists inherent limitations with their use. Recall bias is potentially, a major confounder. Participants are asked to retrospectively recall information about their headache condition days, weeks, and even months preceding the questionnaire's administration. Also, recent unpleasant experiences are likely to stand out to the respondent and

influence his/her response, (18, 19) which calls into question the accuracy of recall and the validity of data collected. A person's affect and any encounters experienced during the period of recall may also influence responses. (20) Social desirability bias, the tendency for respondents to answer questions in a fashion intended to be perceived favorably by others, can also influence the validity of study results. This, however, is a concern in all types of data acquisition, including diaries.

Diaries

Self-report headache diaries (e.g., paper and electronic) have been used to evaluate headache impact and treatment efficacy for years. (21) Diaries are used for data collection in approximately twenty-five percent of all phase II, III, and IV drug trials. (17) In headache research, daily diaries documenting headache severity, frequency and other outcomes are the recommended gold-standard for assessing efficacy.(22) The purpose of daily entries is to minimize or circumvent potential biases in retrospective self-report.

Paper Diaries

For years, paper diaries have been used in diverse patient populations. (21) The advantages for using them are similar to those of self-report questionnaires. They are inexpensive, easy to develop and administer, and require minimal training compared to other forms of data collection, such as computerized devices (Table 1). Patient compliance with paper diaries, however, is poor at best (21, 23) and control over when and how these are completed is limited. (23) Stone et al. investigated compliance in an adult chronic pain population (N=84) using an electronic diary compared to a paper diary with electronic instrumentation to detect when the paper diary was accessed. Participants self-reported a 90.5% compliance rate; however, actual compliance was only 10.9%, and seventy-five percent of subjects did not access their paper diary on at least one

day despite reporting they had. (17) Collins et al. found that forty percent of headache patients were not compliant with their daily headache diary recordings (N=124), and the most common violation was related to retrospective recall. (24) Instead of completing the diary daily over the course of a week, patients go back in time and complete the diary on one given day. In contrast, forward completion occurs when outcomes are recorded for headache days that have not yet occurred. Failure to complete entries daily calls into question the validity of data collected. Similar to self-report questionnaires, recall bias is also a potential major limitation here, because patients have select memory for more ominous episodes which subsequently influence how they respond. (19) Furthermore, large sums of data can be collected using paper diaries, which can result in time-consuming, (25) resource intensive data entry due to the sheer number of data points (25, 26). Paper diaries are also susceptible to data errors (21, 26) due to illegible (poor penmanship) and/or illogical (e.g., a headache patient records a severe headache intensity rating on a day when they deny having a headache) entries.

Electronic Diaries

An improvement in capturing self-reported information, including subjective physical, physiological, and behavioral data is the ecological momentary assessment (EMA) approach, pioneered by Schwartz and Stone.(27) EMA allows patients to report subjective symptoms in real time, at multiple intervals in their natural environment, thus minimizing the period of recall (28, 29), and improving compliance and data validity. (23) Originally, EMA used paper diaries but with advances in computer and mobile phone technology (i.e., e-mail, text messaging, and smart phone applications) and increased access to these platforms, electronic diaries offer a novel delivery platform for the EMA approach.

Electronic diaries are increasingly employed to collect health outcomes (30) and are used in a wide range of clinical research. (19, 31, 32) Computerized EMA diaries are now the recommended gold-standard in pain medicine, (23) and electronic diaries offer many important advantages for data collection (Table 1). Data quality is the most crucial concern in new data capture systems, (31) and a major improvement among electronic platforms is their ability to produce more accurate data. (19, 23) Electronic daily diaries collect information about a patient's health status, in real-time, or close to when an event occurred, which minimizes the recall period and the potential for self-report bias. (19) Also, it has been shown that electronic diaries improve diary completion and foster patient compliance (17, 23, 30) compared to paper diaries. The convenience of portable handheld computer technology allows patients to access their diary anywhere anytime, and the preferred electronic device can be programmed to remind patients to make an entry, similar to setting an alarm. Date and time stamping also specifies when participant entries are made and provides the means for confirming patient compliance independent of patient report. (26) A systematic review on compliance with momentary pain measurement using electronic platforms also highlights additional mechanisms that improve compliance: providing patients with a user manual, financial compensation for diary completion, and electronic messages from study staff. (30) Notably, some of these efforts could improve adherence with paper diaries. Other studies have shown that pain patients willingly accept and prefer electronic mechanisms to paper diaries (25, 32) because of their user-friendly, convenient and appealing nature, (25) all of which can be important motivators for patient compliance. (31) Furthermore, electronic diaries offer improvements in data management. Specifically, they can be manipulated to collect data during specified time periods, preventing unnecessary data collection (outside of designated time frames) and handling. (31, 32) Software can be designed to prevent illogical responses and retroactive data entry (31) and programmed to merge with other

clinical databases for analysis, thus eliminating the need for resource intensive data entry. Finally, electronic diaries allow investigators to monitor patients' conditions in real-time and provide immediate feedback regarding patients' responses to study treatment.

These important advances may improve methodological rigor; however, electronic diaries are not immune to challenges. It has been our experience and that of others, that considerable administrative and financial resources are required to maintain quality.(18) Research and informational technology (IT) staff are essential for ensuring patients are compliant with daily entries and for troubleshooting and mitigating technology malfunctions, including software and hardware failures, respectively. (21, 30, 33) Also, electronic diaries query patients frequently for extended periods of time, which can be inconvenient (e.g., while travelling internationally with scanty Internet access) and burdensome to participants. (30) Fewer and more concise diary questions may curtail this burden. Also, the use of electronic devices may prove difficult for some populations, such as the elderly, those with poor hand-eye coordination, the blind, and those who experience 'computer anxiety.' (30) Finally, technological malfunctions, including failure for databases to record participant responses, website accessibility issues, and personal data assistant malfunction have been reported. (21)

In summary, choosing between different self-report methods for headache data acquisition can have important implications in terms of data validity, patient burden, resources, and costs. Therefore, it is critical to understand how the different self-report methods for measuring headaches compare.

Comparison of Questionnaires and Diaries for Measuring Headaches

Researchers have compared self-report questionnaires based on recall and daily diaries in headache and other pain populations. (25, 32, 34-41) To our knowledge, six studies compared these in neck-related, tension-type, and migraine headache in adult and pediatric populations. (35-40, 42) Four of these used a daily paper diary (35, 37-39) and because of this approach, concerns about the validity of the data collected, particularly because of documented non-compliance with paper platforms should be raised, irrespective of what they found. (17) van den Brink's group failed to include the specific diary type, (42) and only one of the six studies, led by Heyer and colleagues, states using the recommended electronic platform, in this case a daily Internet diary.(36) Heyer et al., compared 90-day and 30-day recall of pediatric migraine disability elements and headache frequency among pediatric patients and their parents (N=52). They found patients recall is better than their parents for all outcomes and headache disability and frequency improves at 30 days compared to ninety.(36) No other studies used an electronic method to collect daily headache outcomes.

The method posed by Bland and Altman is cited as being the most appropriate and popular approach for assessing agreement between two measures, (43, 44) because it considers systematic differences. Of the five studies that compared diaries and questionnaires in headache populations, only one study used this analysis. (38) This study, done by McKenzie et al (N=209 adult migraineurs), found that headache frequency reported on the questionnaire based on recall was equally reliable compared to a daily diary, but headache severity on questionnaires tended to be higher than what was reported on the diary. (38) Blizzard and colleagues assessed agreement between a self-report questionnaire and a 31-day, prospective calendar diary in 100 patients with neck-related headaches, a headache condition similar to this agreement study. They found

moderate agreement (k=0.66) between the two measures for headache frequency; (35) and while a kappa statistic is an appropriate measure when assessing agreement for categorical variables, it does not provide necessary information about the systematic differences between continuous measures. Notably, McKenzie and Blizzards' groups used paper diaries, and not electronic platforms to collect daily headache outcomes. Because of their approach, the conclusions that can be drawn from their work are limited.

To our knowledge, no studies have assessed agreement between electronic diaries and self-report questionnaires in patients with cervicogenic headache using the recommended Bland Altman agreement analysis. In order to maintain methodological rigor and efficiency, it is imperative that we identify and implement valid data collection tools, which balance patient compliance with patient burden.

This thesis aims to answer the question: can a self-report questionnaire based on recall be used interchangeably with a daily electronic headache diary to ascertain headache outcomes in adults with cervicogenic headache?

Table 1: Advantages & Limitations of Electronic and Paper Platforms for Data Collection

Electronic Platforms	**Paper Platforms**
Example • SMS • E-mail	Example • Self-report questionnaire • Paper diary
Advantages ▪ Real-time data collection ▪ Reliable and valid data acquisition ▪ Improved compliance ▪ Less subject to recall bias ▪ Daily monitoring of patient status ▪ Less data entry and handling ▪ Integrate with clinical databases	Advantages ▪ Easy to create ▪ Easy to administer ▪ Inexpensive ▪ Minimal training required
Limitations ▪ Resource intensive (start-up costs, IT monitoring & troubleshooting, daily monitoring) ▪ Technology training required (participant and staff) ▪ Difficult for some patient populations	Limitations ▪ Poor compliance ▪ Subject to recall bias ▪ Date and time stamping not possible ▪ Illegibility issues (poor handwriting) ▪ Illogical responses (inappropriate responses) ▪ Data management

Specific Aims

Aim 1: To determine if a self-report questionnaire based on recall can be used interchangeably with a daily electronic diary for the measurement of cervicogenic headache frequency using the Bland Altman analysis for agreement.

Aim 1 Hypothesis 1: Based on the work by Babel et al., Eich et al., Clark et al., Stone et al., and McKenzie et al., we hypothesize that the self-report questionnaire and electronic headache diary will not agree sufficiently to be used interchangeably in headache clinical research. (29, 38, 45-47)

Aim 2: To determine if a self-report questionnaire based on recall can be used interchangeably with a daily electronic dairy for the measurement of cervicogenic headache intensity using the Bland Altman analysis for agreement.

Aim 2 Hypothesis 1: Based on the work by Babel et al., Eich et al., Collins et al., Stone et al., and McKenzie et al., we hypothesize that the self-report questionnaire and electronic headache diary will not agree sufficiently to be used interchangeably in headache clinical research. (29, 38, 45-47)

Methods

This agreement study is a secondary analysis of de-identified data collected in a two site prospective, parallel group, observer-blinded, randomized clinical trial evaluating the dose-response relationship of chiropractic spinal manipulation (SMT) for chronic cervicogenic headaches (CGH) in 256 adults. (16) Fifty percent of the participants are enrolled at Northwestern Health Sciences University in Bloomington, MN and the other half at the University of Western States in Portland, OR. The trial is funded by the National Center for Complementary and Integrative Health (R01AT006330) and is registered on https://clinicaltrials.gov/ (Identifier:

NCT01530321). Approval for the parent trial is granted by the Institutional Review Boards from Northwestern Health Sciences University, the University of Western States and the University of Minnesota. IRB approval for this agreement study is provided by the University of Minnesota (1503E64481).

Population

Participants in the parent trial are recruited from the general population in Minneapolis, Minnesota, Portland, Oregon and their surrounding metropolitan communities using post-card mailers, social media, and online advertisements. Participants from the Minnesota site only are included in this analysis.

Eligibility

Participants who respond to recruitment solicitation are initially screened for eligibility on the phone using a computer-assisted questionnaire. Those eligible are scheduled for a clinical baseline evaluation which includes informed consent, a personal health history, physical exam and cervical imaging to determine study eligibility. The eligibility criteria are as follows:

Parent Trial: Generally healthy men and women, ages 18 and older with at least 5 cervicogenic headaches (5) a month for three months are invited to participate in the parent trial. Participants with concomitant tension-type headaches (chronic or episodic) and/or infrequent episodic migraine with or without aura occurring on less than 1 calendar day per month in the last year are also included. Persons with contraindications to study treatments (e.g., inflammatory arthritis, severe osteoporosis, cancer) and other headache conditions (e.g., medication overuse, headaches associated with temporomandibular disease, chronic migraine) are excluded. Participants are also excluded if they take daily prescription or non-prescription pain medication for any condition,

had recent or ongoing care for neck pain and/or headaches by a licensed health care provider and for failing to complete at least 24 out of 28 days on their daily headache diary (i.e., paper or electronic diary) prior to enrollment. Table 2 provides a full description of the eligibility criteria.

Agreement Study: Participants who are eligible for enrollment into the parent trial and compliant with the electronic diary during the four-week baseline period are included in this secondary analysis. Electronic diary compliance is defined as having completed the first two questions, and the third if applicable (see Data Collection-Electronic Diary), on at least 24 of the last 28 days. The 24 out of 28 day threshold was originally defined as 'compliant' in the parent trial a priori by the investigative team. Participants are excluded from this analysis if they used a paper diary in lieu of an electronic option at any point during the baseline period.

Table 2: Eligibility Criteria

Inclusion	Exclusion
Parent Trial	
18 years of age	Contraindications to study treatment
CGH ≥ 5 days per month for 3 months	Other headache conditions*
CGH pain intensity ≥3 (0-10 scale)	Daily pain medication
SMT candidate	Diary non-compliance (<24/28 days)
English literate	Treatment for headache/neck pain in past 3 months
Agreement Study	
Eligible for enrollment in parent trial at the Minnesota site.	<24/28 days completed on e-diary
	Use of a paper diary
	UWS (Portland, OR) participant
*Episodic or chronic tension type headache and migraine with or without aura <1 day per month in the last year are included	

Definition of Cervicogenic Headache

Currently, the International Classification of Headache Disorders (ICHD) defines cervicogenic headache based on the following diagnostic criteria:(5)

A. Any headache fulfilling criterion C.

B. Clinical, laboratory and/or imaging evidence of a disorder or lesion within the cervical spine or soft tissues of the neck, known to be able to cause headache.

C. Evidence of causation demonstrated by at least two of the following:

   i. Headache has developed in temporal relation to the onset of the cervical disorder or appearance of the lesion

   ii. Headache has significantly improved or resolved in parallel with improvement in or resolution of the cervical disorder or lesion

   iii. Cervical range of motion is reduced and headache is made significantly worse by provocative maneuvers

   iv. Headache is abolished following diagnostic blockade of a cervical structure or its nerve supply

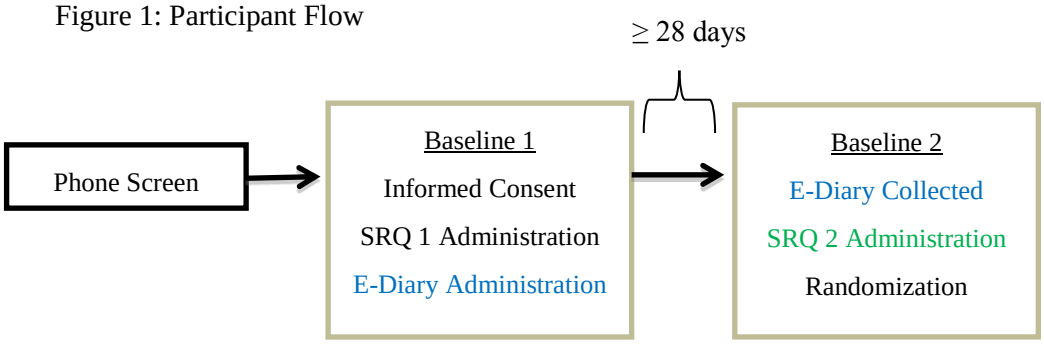
D. Not better accounted for by another ICHD-3 diagnosis.

In the parent trial, the ICHD 2nd edition is used for headache diagnosis. (48) Criterion D in the 2nd edition is not used to determine eligibility in the parent trial, because it cannot be assessed in prospective efficacy studies. Further, the term "neck-related headache" is used with patients when referring to cervicogenic headaches in the parent trial and this agreement study to simplify terminology.

Outcome Measures/Data Collection

Prior to enrollment in the parent trial, participants attend two baseline screening appointments (Baseline 1 and Baseline 2) four to eight weeks apart. Self-report questionnaires are administered and collected at both baseline visits independent of investigator influence. Data from all questionnaires are verified for completion by research staff and double data entered manually into REDCap (REDCap Software - Version 5.3.1 - © 2015 Vanderbilt University). (49) For the purposes of this agreement study, data from the self-report questionnaire administered at Baseline 2 is used. The period of recall collected on this questionnaire is the same time frame for which the electronic diary is completed. Headache diaries (electronic or paper) are administered at the first baseline visit. They are monitored over the four week baseline period, collected and verified for eligibility at Baseline 2.

Figure 1: Participant Flow



Outcomes

Demographics: age, gender, race, ethnicity, marital status, employment status, and education are collected on the baseline 1 questionnaire.

Headache frequency: is measured in days with headache over four weeks using a self-report questionnaire and the daily electronic diary. The International Headache Society recommends using frequency as the primary efficacy measure in headache trials. (22) Frequency is the primary outcome in the parent trial. (16)

Headache intensity: is measured with the 11-point numerical scale (0-10, in which 0=no headache and 10= as bad as could be) using both self-report questionnaires and an electronic diary. This valid and reliable scale is recommended by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) group.(50) Intensity is a secondary outcome in the parent trial. (16)

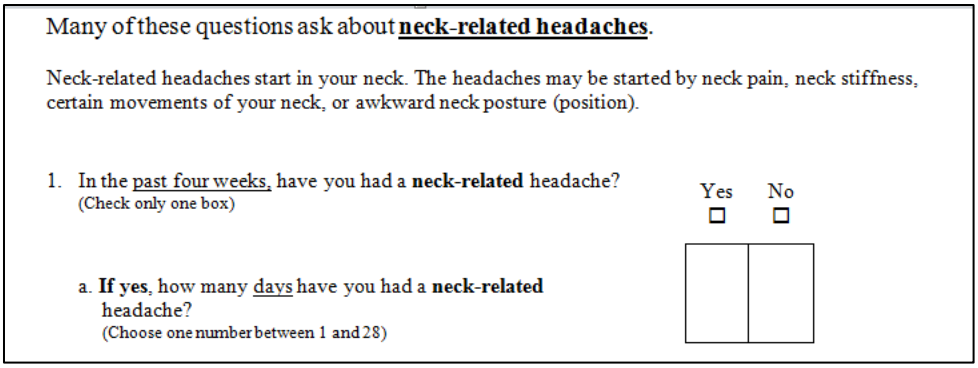
Data Collection

Self-Report Questionnaire (SRQ)

Participants are provided a definition of "neck-related headache" at the beginning of the questionnaire: *"Neck-related headaches start in your neck. The headaches may be started by neck pain, neck stiffness, certain movements of your neck, or awkward neck posture (position)."*

Headache Frequency: the questionnaire asks participants whether or not they had a neck-related headache in the last four weeks (mark yes or no), and if yes, the number of days with a neck-related headache (choose one number between 1 and 28) see Figure 2.

Figure 2: Self-Report Questionnaire (frequency)


Many of these questions ask about **neck-related headaches**.

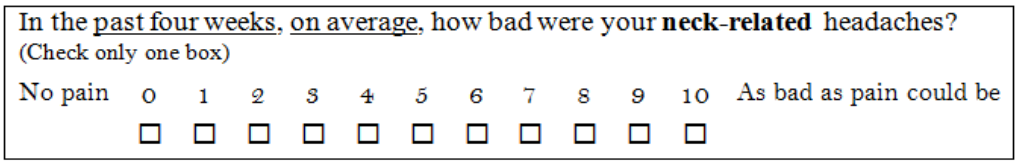
Neck-related headaches start in your neck. The headaches may be started by neck pain, neck stiffness, certain movements of your neck, or awkward neck posture (position).

1. In the past four weeks, have you had a **neck-related** headache? (Check only one box) — Yes ☐ No ☐

   a. **If yes**, how many days have you had a **neck-related** headache? (Choose one number between 1 and 28)

Headache Intensity: the questionnaire asks participants to rate their average neck-related headache pain over the past four weeks using the 11-point numerical rating scale (Figure 3). (51)

Figure 3: Self-Report Questionnaire (intensity)


In the past four weeks, on average, how bad were your **neck-related** headaches? (Check only one box)

No pain 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ As bad as pain could be

Electronic Diary (e-Diary)

Headache outcomes are also collected via a daily electronic diary using a 'Short Message Service" (SMS) (© SMS-Track ApS. https://sms-track.com/), which includes daily emails or text messages. The electronic diary is the preferred method of data collection in the parent trial because of its ability to date and time stamp when patients' entries are made. (23)

At Baseline 1, participants are given the option of choosing e-mail or text messaging; research staff define a "neck-related headache" (see definition under "Data Collection- Self Report Questionnaire"), review the diary questions and anchors to ensure patients understood what is being asked, and provide specific instructions based on the participants preferred electronic choice. Diaries are sent daily (8pm CST) between Baseline 1 and 2 (for at least 4 weeks), and

participants are given 24 hours to respond. Those who fail to respond within the designated 24-hour time frame cannot retroactively make an entry. Research staff monitor the electronic diaries daily for completion and provide reminder text messages, emails, and phone calls to those who are not compliant. Responses are captured in a SMS-Track database and stored on a secure, HIPAA compliant server maintained at NWHSU. Participants without daily Internet or mobile device access are provided a paper diary to complete. Paper diaries are also used in the parent trial as a back-up to the electronic diary if participants miss their electronic entry or in the case of a technological malfunction. For the purposes of this secondary analysis, outcomes data collected via the paper diary are not included due to the nature of the research question, and because it cannot be confirmed whether the paper diaries were completed daily as intended according to the study protocol. The following is a brief description of the electronic diary options available to patients:

Electronic Diary--Email: One e-mail is sent each day (8pm CST) to the participants preferred e-mail address. Four questions are included in each email:

1. Did you take any medication for a neck-related headache today?
2. Did you have a neck-related headache today?
3. If yes, how bad was your neck-related headache today?
4. Did you have a headache not related to your neck today?

For the purposes of this analysis, questions pertaining to headache frequency (question 2) and intensity (question 3) are explored. Questions 1 and 4 will be addressed in a separate analysis. Participants respond to the questions and submit their responses. A thank you notification is sent back immediately confirming the data was received and captured in the database. Participants without an unlimited texting plan are encouraged to use the e-mail diary option. Research staff

also confirm that participants have access to their email in the evenings and over the weekend to ensure diaries can be completed outside of school or the workplace.

Electronic Diary--Text Messaging: The same four questions (see Electronic Diary—Email) are sent via four separate automatic text messages. The first question is sent to the participant's mobile device at 8pm CST. Subsequent questions in the form of text messages are delivered once a response to the previous question is received. SMS does not notify participants if their messages failed to send; however, some mobile devices have a notification mechanism in place. Participants who prefer the electronic diary but do not have access to e-mail in evenings, on weekends or holidays, are encouraged to use the text messaging option. Table 3 compares the frequency and intensity questions for the questionnaire and diary.

Table 3: Electronic Headache Diary and Self-Report Questionnaire Questions

Outcome	Self-Report Questionnaire	Anchor	Electronic Diary	Anchor
Frequency	Q. In the past four weeks, have you had a neck-related headache?	Yes or No (1-28 days)	Q. Did you have a neck-related headache today?	Yes or No (1-28 days)
Intensity	Q. In the past four weeks, on average, how bad were your neck-related headaches?	0-10	Q. If yes, how bad was your neck-related headache today?	0-10

Sample Size

This is a convenience sample driven by the eligibility criteria chosen a priori for this agreement study (Table 2); consequently the sample size was determined by the number of eligible participants (n=87). Participants enrolled at the Minnesota site only are included in this investigation, because at the time of this analysis, the Portland site was enrolling participants in the parent trial. Also, the Portland site did not incorporate electronic diaries at the beginning of

the study and relied heavily on paper diaries. As the focus of this project is to assess agreement between self-report questionnaires and electronic diaries, the Portland site is not included.

Statistical Analysis

Data analysis is performed using the Statistical Package for Social Sciences (SPSS) (Version 22.0. Armonk, NY), MEDCALC (Version 15.10 © 1993-2015 MedCalc Software bvba) and Microsoft Excel for Mac 2011 (Version 14.4.8). Descriptive statistics summarize clinical and baseline characteristics and includes group means and standard deviations (SD).

First, scatter plots are included to preliminarily examine the relationship between the questionnaire and diary along with a line of best fit. Second, a method for assessing agreement recommended by Bland and Altman (43, 44) is used to address specific aims 1 and 2. This analysis is widely cited (44) and accepted as the standard approach to measuring agreement. Differences between the two measurements (self-report questionnaires and electronic diaries) and their mean differences are calculated for each patient. Bland Altman plots for frequency and intensity are used: difference scores, or the amount of disagreement, are plotted on the y-axis, and mean differences, or the systematic difference between the measures, on the x-axis for all 87 patients. (43) Upper and lower 95% limits of agreement (LOA) are computed: group mean difference (headache outcome) ± 1.96 (SD of the difference) tell us how far apart the self-report questionnaire and electronic diaries are likely to be for the majority of participants. (52, 53) These limits are estimates of the values which apply to the population. 95% confidence intervals assess the uncertainty of the LOA estimates: (mean difference) ±1.96 (SD) ± t ($\sqrt{}$ ($3SD^2$ /n)). Group mean differences and 95% LOAs are plotted. Finally, a linear regression model is used to assess

for proportional bias, or the propensity of the two measures to not agree throughout the range of the measurements. All tests are performed with an alpha equal to 0.05.

Supplementary Analysis

The Bland Altman agreement analysis does not make inference regarding mean differences between the two measures. Therefore, a two-tailed t-test is used to determine if mean headache outcomes (i.e., frequency and intensity) measured with the self-report questionnaire and electronic diary are different. Also, histograms of the differences between both measures are constructed to examine frequency distributions for both outcomes and to confirm the normality assumption (Appendix A Figures 8 & 9).

Results

128 participants are enrolled in the parent trial at the Minnesota site. Of these 128, 87 people (68%) meet the inclusion criteria for this agreement study. 41 participants (32%) were ineligible, because they used a paper diary at some point during the four week baseline period: 24 (19%) participants disqualified because they completed <24 of 28 days on the e-diary and used a paper diary as a backup, and 17 (13%) participants are excluded, because they used a paper diary in lieu of an electronic diary.

Table 4: Baseline Demographics

Characteristic	**N=87**	**N=128**
Mean age, y (SD)	42.8 (11.1)	44.9 (12.5)
Women, N (%)	71%	68%
Men, N (%)	29%	32%
White race, N (%)	77 (89)	112 (88)
Black race, N (%)	2 (2)	3 (2)
Asian race, N (%)	4 (5)	6 (5)
>1 race, N (%)	3 (3)	5 (4)
Other, N (%)	1 (1)	2 (1)
Hispanic, N (%)	3 (2)	5 (4)
Married, %	70%	64%
Employed,%	86%	87%
College Graduate,%	69%	66%

The mean age of the sample is 42.8 (SD 11.1), and participants are largely female (71%), white (89%), non-Hispanics (98%). Similar characteristics are observed in the parent trial, which is comprised of mostly white (88%), non-Hispanic (96%) women (68%) with a mean age of 44.9 years (SD 12.5). Table 4 compares the demographic characteristics for enrollees in this agreement study and the parent trial.

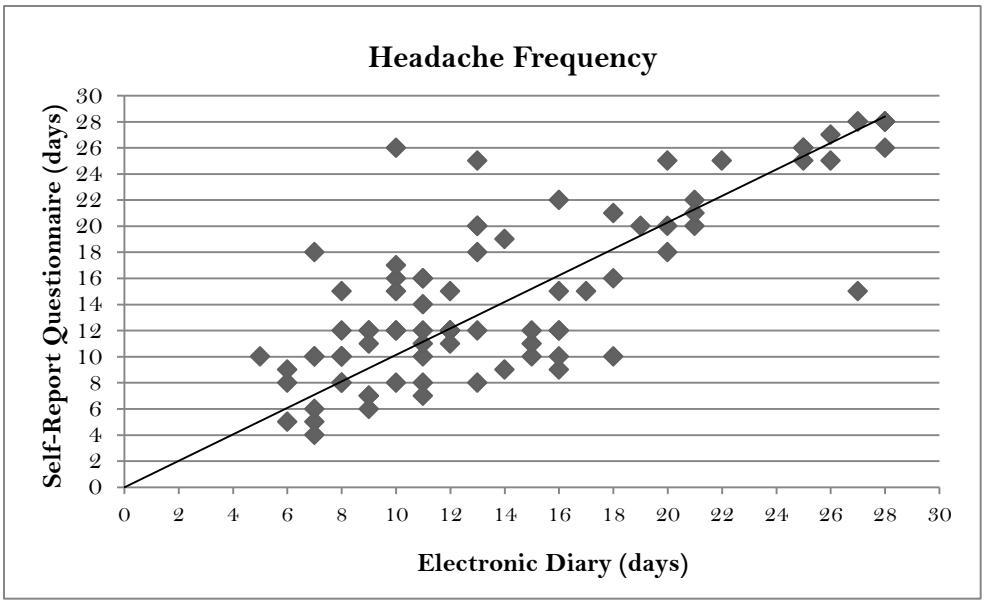


Figure 4: Scatter plot of headache frequency measured with the self-report questionnaire and daily electronic diary in 87 adults.

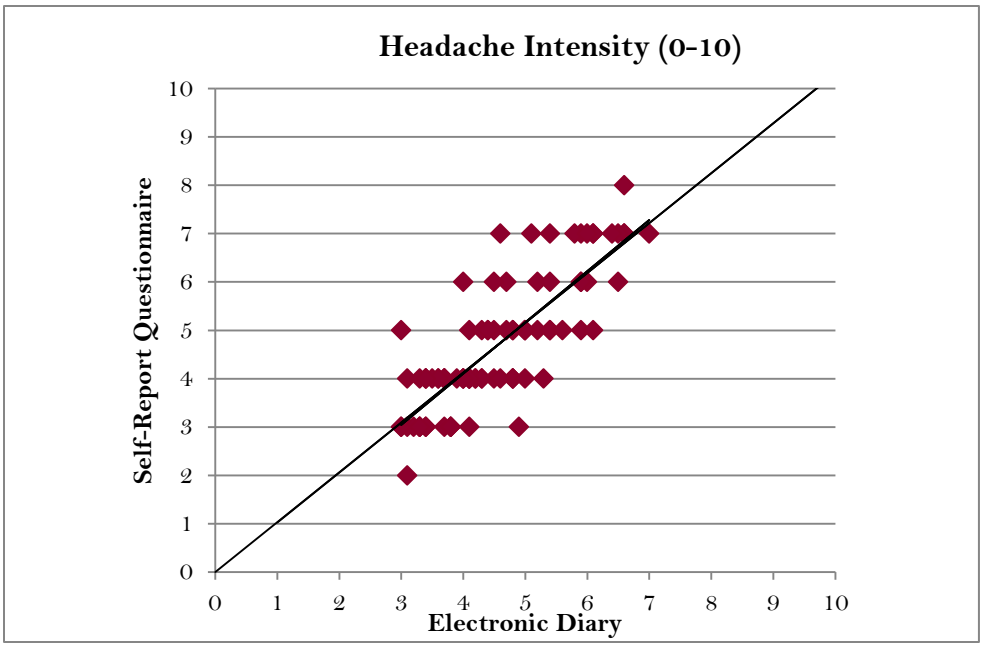


Figure 5: Scatter plot of headache intensity measured with the self-report questionnaire and daily electronic diary in 87 adults.

The scatter plots for frequency and intensity (Figures 4 & 5) show a positive association between the questionnaire and diary for both outcomes. Here, there is obvious disagreement between the two measures. Data points above the line of equality indicate the outcome is reported higher on the questionnaire compared to the diary and vice versa. Frequency plots suggest patients report more headache days and higher pain intensity on the questionnaire (Figures 4 & 5 and Appendix A- Figures 8 &9).

Frequency: Patient's self-reported cervicogenic headache frequency is higher on the self-report questionnaire compared to the daily electronic headache diary by an average of 0.77 days, as indicated by the solid dark blue line (y=0.8, Figure 6). The standard deviation of the differences in headache frequency is 4.25. For approximately 95% of participants, their differences in frequency will be within two standard deviations of the mean difference. The red dashed lines in Figure 6 show these 95% limits of agreement. Differences between the self-report questionnaire and electronic diary range from 9.1 days (95% CI 7.55 to 10.65) to -7.6 days (95% CI -9.1139 to -6.0060), a span of 16.7 days which suggests that large differences in measurements were seen. The orange, dashed line intersecting the y-axis at zero indicates perfect agreement. Figure 6 shows greater variability between the two measures for lower headache frequencies and less spread for almost daily headaches. However, data points are scattered evenly around zero which suggests there is no consistent bias of one approach versus the other. A simple regression line fitted to the Bland Altman plot is not significantly different from zero (β=0.034, 95%CI 0.63-0.91, p=0.635-- Table 6).

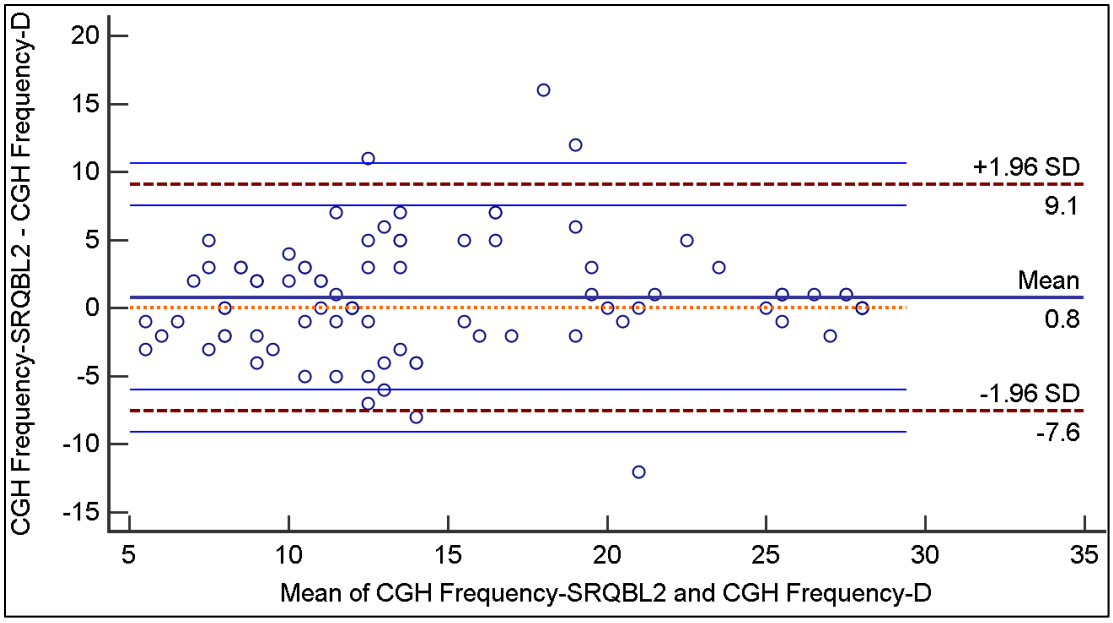


Figure 6: Bland Altman plot of difference in cervicogenic headache frequency measured in days (self-report questionnaire minus 4-week daily electronic headache diary) against the mean of the two measures. Mean difference in headache frequency was rounded to (0.8)

Intensity: Patient's self-reported cervicogenic headache intensity is greater on the self-report questionnaire compared to the electronic headache diary by 0.14 points on average, as indicated by the dark blue line (y= 0.14, Figure 7). The standard deviation of the differences in headache intensity is 0.8. For approximately 95% of participants, their differences in cervicogenic headache intensity will be within two standard deviations of the mean difference. The red dashed lines in Figure 7 show these 95% limits of agreement. Differences between the self-report questionnaire and electronic diary range from 1.7 points on the 11-point numerical rating scale (95% CI 1.4081 to 1.9912) and -1.4 (95% CI -1.7177 to -1.1346), a span of 3.1 points.

CGH intensity recorded via the electronic diary was lower than the SRQ for mild to moderate intensities and higher on the questionnaire for moderate to severe CGH intensities. A regression line fitted to the Bland Altman plot was significantly different from zero, which confirms the presence of proportional bias (β=0.286, 95%CI 0.01-0.27, p=0.000 see Table 6).

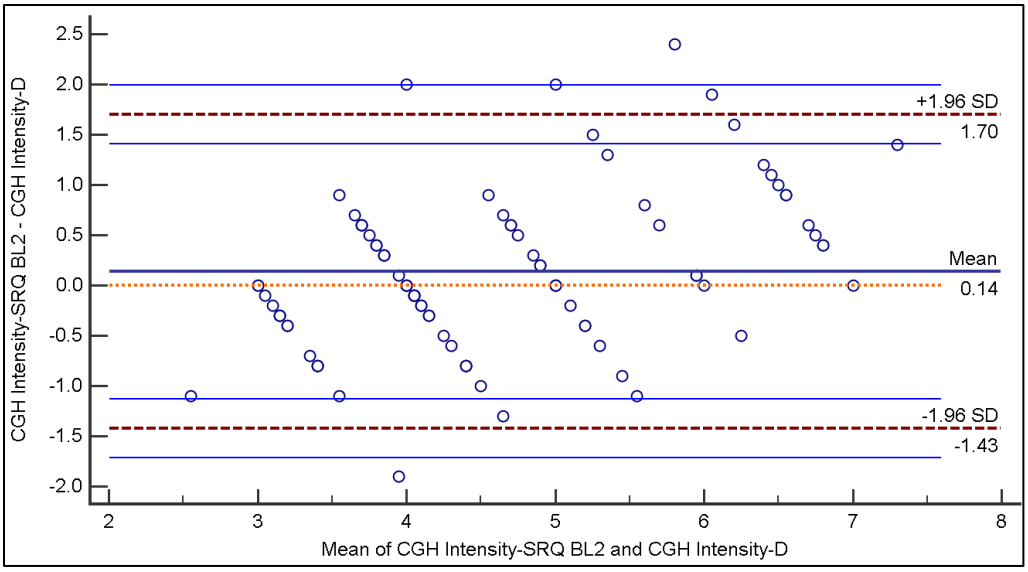


Figure 7: Bland Altman plot of difference in cervicogenic headache intensity, 0-10 pain scale, (self-report questionnaire minus 4-week daily electronic headache diary) against the mean of the two measures.

Table 5: Bland Altman Analysis: mean difference, 95% limits of agreement

	CGH Frequency (days) SRQ BL2-Diary	CGH Intensity (0-10) SRQ BL2-Diary
Mean Difference	0.77	0.14
Mean 95% CI	-0.14 to 1.68	-0.03 to 0.31
Standard Deviation	4.25	0.79
Upper Limit	9.1	1.7
95% CI UL	7.55 to 10.65	1.41 to 1.99
Lower Limit	-7.6	-1.43
95% CI LL	-9.11 to -6.01	-1.72 to -1.13

Table 6: Linear Regression Analysis

Dependent Variable	B Coefficient (SE)	95% CI	P-Value
Frequency (SRQ-Diary)	0.034 (0.071)	0.63-0.91	0.635
Intensity (SRQ-Diary)	0.286 (0.068)	0.01-0.27	0.000

Supplemental Results

Descriptive statistics for cervicogenic headache frequency and intensity are displayed in Table 7. Mean headache frequency collected on the self-report questionnaire and electronic headache diary is 15.3 (SD 6.7) and 14.5 (SD 6.8), respectively. This difference was not statistically significant ($t(86) = 1.69$, $p=0.09$). Average cervicogenic headache intensity collected on the self-report questionnaire and electronic diary was 4.71 (SD 1.4) and 4.6 (SD 1.1), respectively. The difference was also not significant ($t(86) =1.6$, $p=0.11$

Table 7: Headache Frequency & Intensity Group Means

Characteristic	Diary	SRQ	T-Value (df)	p-value
Frequency, mean (SD)	14.5 (6.8)	15.3 (6.7)	1.7 (86)	0.09
Intensity, mean (SD)	4.6 (1.1)	4.7 (1.4)	1.6 (86)	0.11

Discussion

The goal of this research is to determine if an inexpensive, easy to administer, minimally burdensome self-report questionnaire can be used interchangeably with a gold-standard daily headache diary to collect headache outcomes in an adult cervicogenic headache population. This thesis examines the level of agreement between these two measures over four weeks for headache frequency and intensity. In order to assess interchangeability, a detailed examination of the limits of agreement from the Bland Altman analysis is used. Also, the context in which the data is applied is also considered. For the purposes of this discussion, context can be viewed broadly in terms of clinical research and clinical practice implications.

Bland and Altman argue that it is unlikely for two measures to have identical results for all individuals and therefore agree exactly (mean difference = 0). (53) They recommend assessing how much one method (new method) differs from the other (gold-standard), in this case the self-report questionnaire from the electronic diary using the systematic difference and limits of agreement. If their difference is not enough to cause problems or affect the interpretation of study results and/or the clinical interpretation, then it is fair to assume the new method can be used in lieu of the gold-standard, or they can be used interchangeably. The amount of

discrepancy deemed reasonable or acceptable to not cause problems is subject to interpretation and the context for which that data applies. (53)

Clinical Research Implications

The limits of agreement provide important information for deciding whether the questionnaire is an appropriate substitute for the gold-standard daily, electronic headache diary. For headache frequency, the questionnaire based on recall can overestimate headache frequency by as many as 9 headache days or underestimate up to 8 days, a span of 17 days (95% limits of agreement 9.1002 to -7.56 days). Limits of agreement for headache intensity show that the questionnaire can overestimate headache intensity by as many as 1.7 points or underestimate by 1.4 points, a range of 3.1 points (95% limits of agreement 1.6997 to -1.4261). Due to the variability that exists between the questionnaire and diary for both outcomes, sample integrity could be jeopardized due to the misclassification of patients. The implications of this misclassification in research can be serious and irreversible. For example, a hypothetical participant could indicate on the questionnaire they had 10 headache days in the last four weeks. At face value, assuming the participant is otherwise eligible, this person qualifies based on the eligibility criteria detailed in Table 2. But, using the limits of agreement, in reality, the patient could have had 19 headache days or 2 headache days. If the patient's true headache frequency was only 2 days, they should have been excluded from participation. If a second hypothetical participant indicated they only had 4 headaches in the last four weeks on the questionnaire, they would be disqualified (Table 2). But, in reality they may have had up to 13 days and thus should have been included. Erroneous eligibility determination has important consequences. As mentioned above, sample integrity may be threatened and begs an important question for researchers—are you studying who you think you are studying? Also, recruitment and enrollment is notoriously difficult in randomized

studies, and erroneous exclusion should be avoided. Failure to meet recruitment goals can have resource and power implications and thus lead to false conclusions about treatment effects. (54) Similar principles apply for erroneous eligibility determination using headache intensity. Further, the International Headache Society suggests using a responder rate as a secondary efficacy measure, which is defined as the percentage of subjects in a treatment group with at least a 50% improvement in the efficacy measures (22). In this sample, participants would require a frequency of at least 18 days and an intensity of 3.4 points using the questionnaire to ensure a 50% reduction was not due to measurement error. Measurement error can result in an increase in the variance of the estimated treatment effects.

Moreover, there is evidence of proportional bias for headache intensity, that is, it appears the two methods do not agree through the range of measurements (Figure 7). In this sample, the electronic diary estimates pain severity higher compared to the questionnaire for mild to moderate pain (0-5 out of 10) levels. Participants with moderate to severe pain (5-10, out of 10), report more average pain on the self-report questionnaire based on recall compared to average pain collected on the daily diary. Elevated pain levels on retrospective questionnaires compared to daily or current pain estimates have been seen in adult and pediatric headache populations and in patients who experienced painful medical procedures. (38, 40, 42, 46, 55, 56)

To prevent or minimize the misclassification of patients at baseline and systematic measurement error, we recommend a questionnaire based on recall *not* be used in lieu of an electronic daily headache diary for the measurement of cervicogenic headache frequency and intensity in clinical research. Instead, using the gold-standard electronic daily headache diary for measuring these

outcomes, which aligns with the findings and recommendations of the International Headache Society and other researchers (38) is recommended. Also, our observation that the diary and questionnaire do not agree through the range of measurements for severity, also informs this recommendation.

Clinical Implications

Clinical practice settings are more immune to the variability between the limits of agreement especially for some headache types. For example frequency is not a criterion for cervicogenic headache diagnosis using IHS diagnostic classification, and to our knowledge, treatments are not tailored solely to the frequency of this condition. For this reason, when cervicogenic headaches are suspected, an inexpensive, minimally burdensome questionnaire based on recall to ascertain frequency of cervicogenic headaches may be appropriate in clinical practice. However, for other headache types, daily electronic headache diaries might prove more useful. The IHS diagnostic criteria include specific criteria regarding headache frequency for accurate diagnosis for several headache conditions, including medication overuse and daily headache. (5) In addition, some diagnostic protocols are dependent on a diagnosis that relies entirely on headache frequency. For instance, frequent-episodic and chronic tension-type headache are differentiated by the number of headache days and chronicity. Treatments are then tailored based on the correct distinction between the two, thus having a valid means for collecting headache frequency is important. Additionally, inaccuracies in patient-rated intensity may not be as consequential in clinical practice as in clinical research. As is the case with frequency, diagnosis and treatment decisions for cervicogenic headache are not dependent on headache severity; because of this, intensity based on recall on a questionnaire, for example, may be adequate. Severity becomes especially important for determining treatment effectiveness for a particular patient. If this is the case, it

may be advantageous for patients to monitor headache outcomes using the more accurate, daily electronic diary.

Discussion of supplemental results

Participants report more headache days on the self-report questionnaire compared to the electronic diary by an average of 0.77 days (95% CI -0.14, 1.68). There are plausible explanations for the small, non-statistically significant difference (Table 7) between the two measures. Cervicogenic headaches are not always discrete events with an obvious start and end; thus, it may be difficult for sufferers to reliably recall the frequency of their headache episodes. Further, cervicogenic headache patients report greater headache intensity on the questionnaire compared to the electronic daily diary by an average of 0.14 points on the 11-point numerical rating scale (95% CI -0.03 to 0.31). While the trend for the questionnaire to report higher frequency than the diary is consistent with previous research done by McKenzie et al., it is inconsistent in terms of magnitude. (38) McKenzie and colleagues found average headache severity on a 0-3 point pain scale was significantly higher on the questionnaire (1.84) compared to the diary (1.63) ($P<0.001$) in an adult migraine population. (38) In a pediatric headache population, van den Brink et al (42) examined headache recall by comparing a retrospective questionnaire to a daily, 4 week prospective headache diary. For headache frequency and duration there was no difference between the two measures (frequency: $P=0.21$) and (duration: $P=0.00$); however, the questionnaire significantly overestimated severity (65 versus 37 on the 100-mm visual analog scale, $P=0.00$). The type of diary used (paper or electronic) was not specified. Similarly, Peters et. al showed a retrospective questionnaire yielded significantly higher pain compared to a computerized diary assessment over 4 weeks in adults with generalized idiopathic pain

conditions, including neck pain, back pain, and shoulder pain (4.0 vs. 2.8, paired t = 6.21, df =62, P<0.001). (40)

Research on pain memory suggests that past pain influences subsequent pain experiences, (47) and pain patients are more likely to inflate past pain experiences, such as pain intensity, (45) especially if they are in pain at the time of recall. (18, 46, 57) This may explain why headache intensity was reported slightly higher on the questionnaire in this agreement study, albeit not significant, and in other research conducted in headache and pain populations. (38, 40, 42) We do not know if participants were in pain at the time the questionnaires were completed, but all participants included in this analysis had a history of chronic cervicogenic headache with at least mild intensity (3 out of 10, on the 11 point numerical rating scale for pain) averaged over one month prior to questionnaire completion. Further, retrospective assessments introduce bias, which generally results in an overestimation of pain. (58) Several small studies have explored headache memories, and the general consensus is that headache severity recall is overestimated in adults and children. (47)

Stone and Schwartz pioneered the ecological momentary assessment (EMA), an approach that may enhance data validity in conjunction with electronic platforms. (27) Multiple entries throughout the day, in real-time, and in the patient's natural environment using electronic devices minimizes recall bias and selective memory for more ominous or salient experiences or episodes. (19) This may be particularly useful for headache intensity, which is not necessarily static, but instead may fluctuate throughout the day. Several pain ratings throughout the day may provide a more accurate picture of the subjective pain experience versus one pain rating collected daily.

Limitations

This agreement study uses a small convenience sample, which limits the generalizability of study results. However, it is similar demographically to those enrolled in the parent trial in Minnesota suggesting this sample is representative of the larger population of adults enrolled in the parent study. Also, this study may not be adequately powered to detect statistically significant differences between population means for frequency and intensity. Further, similar to other agreement studies, we were unable to determine a priori what constitutes a clinically meaningful or important difference for each outcome, in this case headache frequency and intensity, between the two measures. In this analysis, a priori estimates for these measures were not defined, because to our knowledge, absolute minimally important change estimates have not been established for headache, as is the case for other chronic pain conditions. (59, 60) Minimal clinically important change (MCIC) for non-specific, non-radicular neck pain on the 0-10 NRS scale, have been established. (60) By definition, cervicogenic headache is precipitated by dysfunction in the cervical spine, which may include neck pain. Therefore, it may be appropriate to use the MCIC for neck pain to interpret the limits of agreement and the degree of discrepancy reasonable to use the data collection measures interchangeably. MCIC values vary depending on the methods used to determine MCIC and the variability in the population, and thus should be interpreted and used with caution. (60-62) It is assumed that the daily tracking of headache outcomes on the electronic diary could artificially improve the accuracy of the participant's one time recall of their headache frequency and intensity on the questionnaire. In this study, participants were not asked whether or not diary completion influenced them directly while completing the questionnaires. Future studies could ask patients (quantitatively or qualitatively) if the diary influenced their ability to recall headache outcomes.

Future Research

More research is needed to confirm these findings in cervicogenic and other headache populations. Also, few studies have examined how agreement changes overtime. Finally, little is known about how the design and delivery of technological platforms impacts user compliance and psychometric properties.

Conclusion

Daily electronic headache diaries are the preferred method for collecting cervicogenic headache outcomes in clinical research. In clinical practice, it may be appropriate to use self-report questionnaires for collecting headache outcomes to inform cervicogenic headache diagnosis and management. However, the limitations of self-report should be considered.

Authorization

The principal investigator for the parent trial (R01AT006330) Minnesota site, Gert Bronfort PhD, DC, authorized use of this data set.

References

1. Atlas of headache disorders and resources in the world 2011. A collaborative project of World Health Organization and Lifting The Burden. Geneva: WHO; 2011.
2. Stovner L, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, et al. The global burden of headache: a documentation of headache prevalence and disability worldwide. Cephalalgia. 2007;27(3):193-210.
3. Lipton RB, Bigal ME, Kolodner K, Stewart WF, Liberman JN, Steiner TJ**.** The family impact of migraine: population-based studies in the USA and UK. Cephalalgia. 2003;23(6):429-40.
4. Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R**.** Lost productive time and cost due to common pain conditions in the US workforce. Jama. 2003;290(18):2443-54.
5. The International Classification of Headache Disorders, 3rd edition (beta version). Cephalalgia. 2013;33(9):629-808.
6. Sjaastad O, Fredriksen TA, Pfaffenrath V**.** Cervicogenic headache: diagnostic criteria. The Cervicogenic Headache International Study Group. Headache. 1998;38(6):442-5.
7. Sjaastad O, Bakketeig LS**.** Prevalence of cervicogenic headache: Vaga study of headache epidemiology. Acta Neurol Scand. 2008;117(3):173-80.
8. Nilsson N**.** The prevalence of cervicogenic headache in a random population sample of 20-59 year olds. Spine (Phila Pa 1976). 1995;20(17):1884-8.
9. Sjaastad O, Fredriksen TA**.** Cervicogenic headache: criteria, classification and epidemiology. Clin Exp Rheumatol. 2000;18(2 Suppl 19):S3-6.
10. van Suijlekom HA, Lame I, Stomp-van den Berg SG, Kessels AG, Weber WE**.** Quality of life of patients with cervicogenic headache: a comparison with control subjects and patients with migraine or tension-type headache. Headache. 2003;43(10):1034-41.
11. Clarke TC BL, Stussman BJ, Barnes PM, Nahin RL. Trends in the use of complementary health approaches among adults: United States, 2002-2012. National Health Statistics Reports. 2015;79:1-16.
12. Barnes PM, Bloom B, Nahin RL**.** Complementary and alternative medicine use among adults and children: United States, 2007. Natl Health Stat Report. 2008(12):1-23.
13. Bronfort G, Haas M, Evans R, Leininger B, Triano J**.** Effectiveness of manual therapies: the UK evidence report. Chiropr Osteopat. 2010;18:3.
14. Clar C, Tsertsvadze A, Court R, Hundt GL, Clarke A, Sutcliffe P**.** Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal

conditions: systematic review and update of UK evidence report. Chiropr Man Therap. 2014;22(1):12.

15. Haas M, Spegman A, Peterson D, Aickin M, Vavrek D**.** Dose response and efficacy of spinal manipulation for chronic cervicogenic headache: a pilot randomized controlled trial. Spine J. 2010;10(2):117-28.
16. Hanson L, Haas M, Bronfort G, Vavrek D, Schulz C, Leininger B, et al. Dose-response of spinal manipulation for cervicogenic headache: study protocol for a randomized controlled trial. Chiropr Man Therap. 2016;24:23.
17. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR**.** Patient compliance with paper and electronic diaries. Control Clin Trials. 2003;24(2):182-99.
18. Bolger N, Davis A, Rafaeli E**.** Diary methods: capturing life as it is lived. Annu Rev Psychol. 2003;54:579-616.
19. Burton C, Weller D, Sharpe M**.** Are electronic diaries useful for symptoms research? A systematic review. J Psychosom Res. 2007;62(5):553-61.
20. Bradburn NM, Rips LJ, Shevell SK**.** Answering autobiographical questions: the impact of memory and inference on surveys. Science. 1987;236(4798):157-61.
21. Stinson JN, Huguet A, McGrath P, Rosenbloom B, Soobiah C, White M, et al. A qualitative review of the psychometric properties and feasibility of electronic headache diaries for children and adults: where we are and where we need to go. Pain Res Manag. 2013;18(3):142-52.
22. Bendtsen L, Bigal ME, Cerbo R, Diener HC, Holroyd K, Lampl C, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. Cephalalgia. 2010;30(1):1-16.
23. Salaffi F, Sarzi-Puttini P, Atzeni F**.** How to measure chronic pain: New concepts. Best Pract Res Clin Rheumatol. 2015;29(1):164-86.
24. Collins FJT, JK;**.** Reliability and Standardization in the Assessment of Self-Reported Headache Pain. Journal of Behavioral Assessment. 1979;1(1):73-86.
25. Salaffi F, Gasparini S, Ciapetti A, Gutierrez M, Grassi W**.** Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. Rheumatology (Oxford). 2013;52(11):2062-70.
26. Green AS, Rafaeli E, Bolger N, Shrout PE, Reis HT**.** Paper or plastic? Data equivalence in paper and electronic diaries. Psychol Methods. 2006;11(1):87-105.
27. Schwartz JE, Stone AA**.** Strategies for analyzing ecological momentary assessment data. Health Psychol. 1998;17(1):6-16.

28. Friedberg F, Sohl SJ. Memory for fatigue in chronic fatigue syndrome: the relation between weekly recall and momentary ratings. Int J Behav Med. 2008;15(1):29-33.
29. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. Pain. 2004;107(1-2):61-9.
30. Morren M, van Dulmen S, Ouwerkerk J, Bensing J. Compliance with momentary pain measurement using electronic diaries: a systematic review. Eur J Pain. 2009;13(4):354-65.
31. Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak. 2006;6:23.
32. Salaffi F, Gasparini S, Grassi W. The use of computer touch-screen technology for the collection of patient-reported outcome data in rheumatoid arthritis: comparison with standardized paper questionnaires. Clin Exp Rheumatol. 2009;27(3):459-68.
33. Moloney MF, Aycock DM, Cotsonis GA, Myerburg S, Farino C, Lentz M. An Internet-Based Migraine Headache Diary: Issues in Internet-Based Research. Headache: The Journal of Head and Face Pain. 2009;49(5):673-86.
34. Andrasik F, Holroyd KA. Reliability and concurrent validity of headache questionnaire data. Headache. 1980;20(1):44-6.
35. Blizzard L, Grimmer KA, Dwyer T. Validity of a measure of the frequency of headaches with overt neck involvement, and reliability of measurement of cervical spine anthropometric and muscle performance factors. Arch Phys Med Rehabil. 2000;81(9):1204-10.
36. Heyer GL, Perkins SQ, Rose SC, Aylward SC, Lee JM. Comparing patient and parent recall of 90-day and 30-day migraine disability using elements of the PedMIDAS and an Internet headache diary. Cephalalgia. 2014;34(4):298-306.
37. Larsson B, Fichtel A. Headache prevalence and characteristics among adolescents in the general population: a comparison between retrospect questionnaire and prospective paper diary data. J Headache Pain. 2014;15:80.
38. McKenzie JA, Cutrer FM. How well do headache patients remember? A comparison of self-report measures of headache frequency and severity in patients with migraine. Headache. 2009;49(5):669-72.
39. Niere KJA. Measurement of headache frequency, intensity and duration: comparison of patient report by questionnaire and headache diary. Physiother Res Int. 2004;9(4):149-56.

40. Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PF, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain. 2000;84(2-3):181-92.
41. Stewart WF, Lipton RB, Kolodner KB, Sawyer J, Lee C, Liberman JN. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. Pain. 2000;88(1):41-52.
42. van den Brink M, Bandell-Hoekstra EN, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. Headache. 2001;41(1):11-20.
43. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
44. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. PLoS One. 2012;7(5):e37908.
45. Clark DM, Teasdale JD. Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. J Abnorm Psychol. 1982;91(2):87-95.
46. Eich E, Reeves JL, Jaeger B, Graff-Radford SB. Memory for pain: relation between past and present pain intensity. Pain. 1985;23(4):375-80.
47. Babel P. Memory of pain and affect associated with migraine and non-migraine headaches. Memory. 2015;23(6):864-75.
48. The International Classification of Headache Disorders: 2nd edition. Cephalalgia. 2004;24 Suppl 1:9-160.
49. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.
50. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1-2):9-19.
51. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain. 1986;27(1):117-26.
52. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet. 1995;346(8982):1085-7.
53. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-60.

54. Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S**.** Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. BMJ Open. 2012;2(1):e000496.
55. Redelmeier DA, Kahneman D**.** Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain. 1996;66(1):3-8.
56. Lewandowski AS, Palermo TM, Kirchner HL, Drotar D**.** Comparing diary and retrospective reports of pain and activity restriction in children and adolescents with chronic pain conditions. Clin J Pain. 2009;25(4):299-306.
57. Schmier JK, Halpern MT**.** Patient recall and recall bias of health state and health status. Expert Rev Pharmacoecon Outcomes Res. 2004;4(2):159-63.
58. Sorbi MJ, Peters ML, Kruise DA, Maas CJ, Kerssens JJ, Verhaak PF, et al. Electronic momentary assessment in chronic pain I: psychological pain responses as predictors of pain intensity. Clin J Pain. 2006;22(1):55-66.
59. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976). 2008;33(1):90-4.
60. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC**.** Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976). 2007;32(26):3047-51.
61. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-34.
62. Schuller W, Ostelo RW, Janssen R, de Vet HC**.** The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. Health Qual Life Outcomes. 2014;12:53.

Appendix A

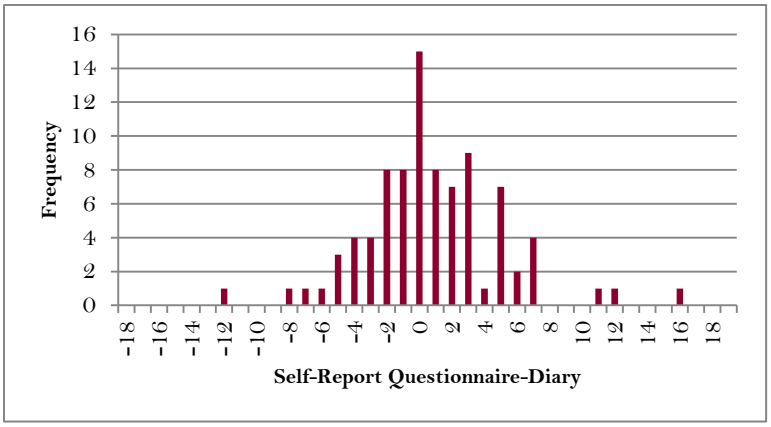


Figure 8: Histogram of differences in headache frequency measured with a self-report questionnaire and daily electronic headache diary over four weeks (N=87)

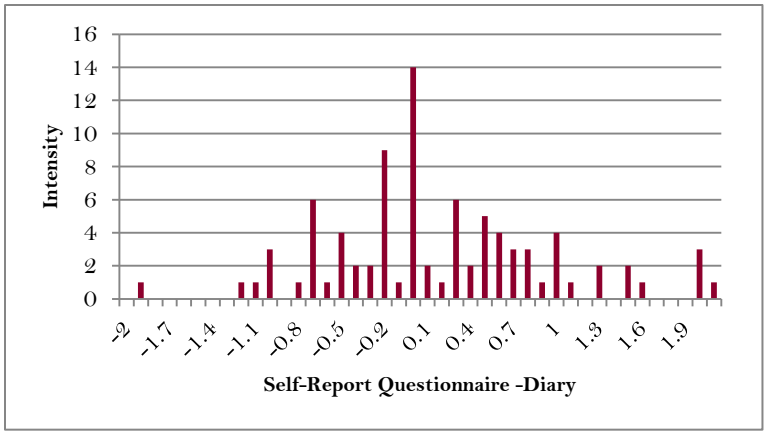


Figure 9: Histogram of differences in headache intensity measured with a self-report questionnaire and daily electronic headache diary over four weeks (N=87)