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

Jennifer N Stinson PhD1,2,3,4, Anna Huguet PhD5, Patrick McGrath PhD5, Brittany Rosenbloom BA(Hons)1, Charlene Soobiah BSc(Hons)1, Meghan White BA(Hons)1, Geraldine Coburn MSc1

Background: While paper headache pain diaries have been used to determine the effectiveness of headache treatments in clinical trials, recent advances in information and communication technologies have resulted in the burgeoning use of electronic diaries (e-diaries) for headache pain.

Objective: To qualitatively review headache e-diaries, assess their measurement properties, examine measurement components and compare these components with recommended reporting guidelines.

Methods: The databases Medline, the Cumulative Index to Nursing and Allied Health Literature, Embase, PsychInfo, the Education Resources Information Centre and ISI Web of Science were searched for self-report headache e-diaries for children and adults. A total of 21 publications that involved e-diaries were found; five articles reported on the development of an e-diary and 16 used an e-diary as an outcome measure in randomized controlled trials or observational studies. The diary measures’ components, features and psychometric properties, as well as the quality of evidence of their psychometric properties, were evaluated.

Results: Five headache e-diaries met the a priori criteria and were included in the final analysis. None of these e-diaries had well-developed evidence of reliability and validity. Three e-diaries showed evidence of feasibility. E-diaries with ad hoc measures developed by the study investigators were most common, with little to no supportive evidence of reliability and/or validity. Compliance with the reporting guidelines was variable, with only one-half of the e-diaries measuring the recommended primary outcome of headache frequency.

Conclusions: Specific recommendations regarding the development (including essential components) and testing of headache e-diaries are discussed. Further research is needed to strengthen the measurement of headache pain in clinical trials using headache e-diaries.

Key Words: Adults; Children; Electronic pain diaries; Headache diary; Psychometric properties

Primary headache disorders (eg, migraine, tension-type headaches) are frequent in the general population. Headache prevalence varies considerably through the lifespan, typically increasing in childhood and youth, and remaining relatively stable and high over the third to fifth decades of life before markedly declining in both sexes. Headaches can negatively impact all aspects of quality of life (psychological [1-3], work/school [4-8] and social functioning [9,10]), resulting in costly disability. Paper pain diaries have been used for several decades in diverse patient populations for many purposes. In the headache patient population, they have primarily been used to assess the impact of headaches and to evaluate the efficacy of headache treatments. However, paper headache diaries exhibit several important limitations, including participant noncompliance and inaccuracies in data entry (11). Noncompliance in participants has been found to be due to the fact that paper diaries are bulky and cannot be used easily or discreetly in the workplace or school. Inaccuracies result from false entries or multiple entries being made at a more convenient time to compensate for noncompliance (11).

When pain diaries (paper or electronic) are used to assess the efficacy of treatments for recurrent and chronic pain conditions, ideally, they should conform to the recent recommendations regarding the core outcome domains suggested for use in clinical trials in children (Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [Ped-IMMPACT]) and adults (IMMPACT) with recurrent and chronic pain (12,13). The eight core outcomes include pain, physical functioning, emotional functioning, role functioning, pain interference, anxiety and depression, social functioning, and patient satisfaction.

©2013 Pulsus Group Inc. All rights reserved
Qualitative review of electronic headache diaries

More specifically, efficacy studies of headache treatments should conform to clinical trial guidelines for both pharmacological (14) and behavioral (15) headache trials, which recommend headache frequency as the primary outcome measure. Recommended secondary outcomes include other headache characteristics (headache index, duration, severity), functional status, headache-related disability and quality of life (see Table 1 for a summary of these guidelines). Furthermore, there should be evidence that these diaries are reliable and valid, so that valid conclusions can be drawn regarding the efficacy of pharmacological and nonpharmacological treatment strategies.

During the past decade, there has been a growing trend toward measuring recurrent and chronic pain using real-time data capture (RTDC) approaches that use electronic pain diaries (e-diaries) (11,16). E-diaries have the advantages of maximizing participant compliance in completing pain ratings as well as the validity of those ratings because e-diaries can be accessed from any computer or handheld device (17). Although headache e-diaries are becoming more common, little is known about their components and features, psychometric properties and feasibility.

| TABLE 1 | Guidelines for headache and chronic pain clinical trials |
|-----------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Guidelines | Primary outcomes | Recommended measures (when provided) | Secondary outcomes | Recommended measures (when provided) |
| Guidelines for pharmacological headache trials (14) | Percentage of patients pain-free at 2 h | Number of patients pain-free at 2 h with no use of rescue medication or relapse within 48 h | Headache intensity | Headache relief |
| | Frequency | Number of attacks per four-week period | Time to meaningful relief | Global evaluation of medication |
| | | | Functional disability | Presence of photophobia |
| | | | Patients’ preference for treatment | Consistency of effect |
| | | | | |
| Migraine | Frequency | Number of attacks per four-week period | Headache intensity | Headache duration |
| | | | Drug consumption for symptomatic treatment | Patient preferences |
| | | | | Same as headache |
| Episodic | Frequency | Number of attacks per month | Headache index | Headache duration |
| | | | Peak headache severity | Functional status |
| | | | Quality of life | Medication use |
| | | | Patient preferences | Psychological symptoms |
| | | | | Missed work/school |
| | | | | Headache-specific quality of life |
| | | | | |
| Chronic | Frequency | Headache days per month | Headache index | Number of severe headache days per month |
| | | | Headache duration | Peak headache severity |
| | | | Functional status | Quality of life |
| | | | Patient preferences | Psychological symptoms |
| | | | | Same as episodic |
| Guidelines for chronic pain clinical trials (12,13) | Pain intensity | 11-point (0–10) numerical rating scale of pain intensity | Physical functioning | Multidimensional Pain Inventory |
| | Emotional functioning | | Role functioning | Interference Scale OR Brief Pain Inventory interference items |
| | Symptoms and adverse events | | Global judgement of satisfaction with treatment | Beck Depression Inventory OR Profile of Mood States |
| | Global judgement of satisfaction with treatment | | Sleep | School attendance, PedMIDAS, PedsQL* |
| | Sleep | | Economic factors | Passive capture of spontaneously reported adverse events and symptoms and use of open-ended prompts |
| | Economic factors | | | Sleep diary, Sleep Habits Questionnaire* |

*Pediatric recommendations. PedMIDAS Pediatric Migraine Disability Assessment; PedsQL Pediatric Quality of Life Inventory
Search strategy for identification of studies

Included studies were accessed primarily through a search of Medline (1947 to September 9, 2009), the Cumulative Index to Nursing and Allied Health Literature (1981 to September 9, 2009), Embase (1947 to September 9, 2009), PsychINFO (1947 to September 9, 2009), the Education Resources Information Centre (1966 to September 9, 2009) and ISI Web of Knowledge (1950 to September 9, 2009) databases. The search terms included MeSH headings, subjects, text words, wild cards and/or keywords relevant to the following terms: “headache”, “migraine”, “cephalagia”, “electronic diary or diary”, “pen-and-paper diary or diaries”, “personal digital assistant”, “decision-making support”, “ecological momentary assessment” and “real time data capture” were used. Reference lists from all identified appropriate research and review articles were examined. Non-English language abstracts were excluded due to the cost of translation. The exclusion of non-English language research studies was minimal (552 of the total of 7102). Two librarians (at two different institutions) independently conducted the literature searches.

Methods of the review

Two reviewers (BR, CS) screened all identified titles and abstracts for relevance and assessed potentially relevant studies for inclusion independently (n=7102). Abstracts were included in the review if they assessed headache and used an e-diary as a primary outcome measure. Non-English language abstracts, guidelines, reviews and unpublished manuscripts were immediately excluded from the review (Figure 1). Level of agreement between the two reviewers (BR, CS) was 98%, and any disagreements were resolved by discussion with a third reviewer (GC). A systematic approach to data extraction was used to produce a descriptive summary of the psychometric findings and feasibility of the headache e-diaries. The two reviewers first developed and tested an extraction database. For all abstracts meeting the inclusion criteria (n=344), the full article was found and subsequently entered into the database by the two reviewers; level of agreement was 100%. Outcome data extraction focused on the measures of age range, type of headache pain and the diary's psychometric properties, including reliability, validity, responsiveness, interpretability and feasibility (compliance, acceptability and technical problems). Table 2 includes operational definitions of these measurement terms (18-24). Data were also collected on the components assessed in the diary, features of the diary (eg, alarms) and type of e-diaries used (eg, handheld device, desktop computer, etc).

Finally, the overall level of psychometric evidence supporting each e-diary was explored using two complementary methods. Initially, two raters (BR, CS) applied the Cohen criteria (25) to evaluate the e-diaries based on the psychometric criteria outlined above, which are described as essential in the Standards for Educational and Psychological Testing (26). According to these established criteria, measures can be classified as ‘well-established’, ‘approaching well-established’ or ‘promising’ (Appendix I [25]). Additionally, the two raters (BR, CS) also used the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) checklist (27), a four-step tool, to evaluate the methodological quality of the studies designed to assess the psychometric properties of these e-diaries. Specifically, the COSMIN checklist explores whether the following properties have been assessed in the study: internal consistency, reliability, measurement error, content validity, construct validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, interpretability and generalizability (27). The level of agreement between the raters was 100% when applying both the Cohen criteria and the COSMIN checklist.

METHODS

All published peer-reviewed English-language research studies examining the psychometric properties and feasibility of self-reported headache pain e-diaries as a main outcome measure in children and adults were considered for inclusion in the present review. Unpublished manuscripts, reviews, guidelines, commentaries and other descriptive articles were excluded. Published abstracts were not included because the information provided in abstracts is limited and frequently not peer reviewed. Outcome measures other than headache e-diaries, such as surveys, questionnaires and clinical interviews, were excluded. Studies published in languages other than English were also excluded due to time and financial constraints (translation costs) and no attempt was made to locate unpublished material or to contact authors of unpublished studies.

RESULTS

A total of 7102 abstracts were identified from the electronic searches, of which 386 abstracts were removed due to manuscript type (guidelines, reviews, unpublished manuscripts). After accounting for non-English language publications, 552 abstracts were also removed (of the non-English language publications, 27 contained diary information; however,
it was unclear from the abstracts whether they were referring to elec-
tronic or paper diaries), leaving 6164 abstracts for further considera-
tion. Of these, 3925 abstracts were removed based on their study population
(not headache) and 1895 were excluded because they were not diary
studies, leaving 344 abstracts for inclusion. Included articles were further
analyzed to yield 323 studies involving paper diaries and 21 e-diary stud-
ies. A list of the excluded articles is available from the primary author
on request.

On initial review, it was observed that there were two distinct
groups of articles: those using an e-diary to capture outcome data
(n=16), which will be referred to as ‘outcome studies using e-diaries’
(studies 6 through 21) (28-43), and those reporting on the develop-
ment and/or psychometric testing (n=5) of headache e-diaries, which
will be referred to as ‘evaluation studies of e-diaries’ (studies 1
through 5) (44-48). The study characteristics of these five evaluation
studies of e-diaries are summarized in Table 3. Each of these studies will
be referred to as Studies 1 through 5 in the results. However, due to the
limited information on the headache e-diaries reported in the out-
comes studies, only brief summaries of the characteristics of these
diaries were possible (Table 4).

**EVALUATION STUDIES OF E-DIARIES**

**Study characteristics**

Reviewed studies were conducted between 2004 and 2005, with four in
the United States and one in the Netherlands. A randomized contro-
rolled trial (RCT) study design was used in two of the five studies (studies 2,
4); the remaining three studies implemented a prospective observational
study design (one with iterative cycles). In the two RCTs, control group
participants used paper diaries with identical outcome measures to their
electronic counterparts. Study duration ranged from one week to four
months (mean ± SD 6.2±7.22 weeks) with participants logging one
to five diary entries per day (1.95±1.74 entries per day). The sampling
protocol in four of the five studies was time based or at prespecified times
each day (studies 2, 3, 4, 5); the fifth study (study 1) featured an event-
based sampling protocol in which participants made diary entries when
they experienced headaches (study 1).

**Characteristics of study participants**

The number of participants using the e-diaries ranged from five to
94 and, collectively, all studies involved a total of 272 participants
(213 headache patients). Three of the evaluation studies included only
participants with headache (n=118; studies 1, 3 and 5); the remaining
evaluation studies featured participants with headache (n=94), juven-
ilide arthritis (n=50) and sickle cell disease (n=9; studies 2 and 4). Three
of the evaluation studies involved only female participants between
18 and 55 years of age (studies 1, 3 and 5). The remaining two studies
specifically involved youth (eight to 16 years of age [studies 2 and 4]).
Four evaluation studies reported participant ethnicity (studies 1, 2, 3 and 4),
and in each of these studies at least 65% of participants were Caucasian.
The participants also varied in type of headache. Three studies involved participants with migraine (one with or without
aura, one without aura, and one not specified; studies 1, 3 and 5); the
remaining two evaluation studies involved participants with recurrent
headache (study 2) and headache not specified (study 4). One study
involved additional, more specific criteria for inclusion, recruiting only
participants who reported at least two headache attacks in the previous
month (study 3). Only one of the five reviewed evaluation studies used a
community sample (study 3); the other four studies (studies 1, 2, 4 and 5)
featured clinical samples.

**Outcome measures**

The headache e-diaries varied with regard to the components assessed, as
shown in Table 2. Sensory characteristics of the headaches were

---

**TABLE 2: Operative definition of terms**

<table>
<thead>
<tr>
<th>Term</th>
<th>Operative definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>The reproducibility of a measure over different occasions. Is concerned with minimizing sources of random error so that measures are reproducible (18). In general, acceptable reliability coefficients for research and clinical purposes are ≥0.7 and ≥0.9, respectively (18,19).</td>
</tr>
<tr>
<td>Inter-rater (interobserver) reliability</td>
<td>The agreement between different raters/observers of an observational measure of pain (18)</td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>The agreement between observations within the same individuals on at least two occasions (18)</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>A type of reliability that includes the mean of the correlation of scores from a measure with the scores of all the items in the measure (18)</td>
</tr>
<tr>
<td>Validity</td>
<td>Used to assess whether the scale is measuring what it is intending to measure (18)</td>
</tr>
<tr>
<td>Face validity</td>
<td>Whether the pain scale includes appropriate items that appear to measure what it is proposing to measure (18)</td>
</tr>
<tr>
<td>Content validity</td>
<td>The assessment of whether the items in the pain measure include the appropriate information and content (18)</td>
</tr>
<tr>
<td>Criterion</td>
<td>Includes concurrent validity and predictive validity. In general, correlations between the new measure and the gold standard should be at least r0.3–0.5. The magnitude of the coefficients are hypothesis dependent but should not be so high as to make the new measure redundant. In predictive validity, the correlation of the measure to the criterion variable is determined at a later time (18).</td>
</tr>
<tr>
<td>Construct</td>
<td>Determines the validity of abstract variables, such as pain, that cannot be directly observed. These constructs are assessed by their relationships with other variables (18,20).</td>
</tr>
<tr>
<td>i. Convergent validity</td>
<td>Evaluates how well items on a pain scale correlate with other measures of the same construct or related variables. In general, correlations between the measure and another measure of the same construct should be r0.3–0.5; however, the magnitude of the coefficients are hypothesis dependent (18).</td>
</tr>
<tr>
<td>ii. Discriminant validity</td>
<td>Evaluates how items in a pain scale correlate with other measures that are unrelated. In general, correlations between the measure and another unrelated measure should be r0.3; however, the magnitude of the coefficients are hypothesis dependent (18).</td>
</tr>
<tr>
<td>Responsivity</td>
<td>Measures whether the measure is able to identify changes in pain over time that is clinically important to patients. An acceptable effect size should be r0.05; however, the effect size is hypothesis dependent (21,22).</td>
</tr>
<tr>
<td>Interpretability</td>
<td>The meaningfulness of the scores obtained from the measure (20).</td>
</tr>
<tr>
<td>Feasibility</td>
<td>How easily the measure can be scored and interpreted (23).</td>
</tr>
<tr>
<td>Compliance</td>
<td>The number of observed over the expected number of diary entries completed for the study period.</td>
</tr>
<tr>
<td>Acceptability</td>
<td>What the end-users liked and disliked about the electronic headache diary.</td>
</tr>
</tbody>
</table>

Adapted from Stinson et al (24)
TABLE 3: Study characteristics

<table>
<thead>
<tr>
<th>Study number</th>
<th>Author (reference), year (country)</th>
<th>Target population; sample size; mean age, years (range or SD)</th>
<th>Study design and goal</th>
<th>Type of electronic device; characteristics of diary</th>
<th>Variables (outcome measures used in diary capture)</th>
<th>Sampling protocol; number of diary entries per day; length of diary use</th>
<th>General features of assessment tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Goldberg et al (44), 2007 (USA)</td>
<td>Adult females; n=20; 37 years of age (range 21–47)</td>
<td>Prospective observational To evaluate an electronic diary as a diagnostic tool to study headache and premenstrual symptoms</td>
<td>Handheld device (Proventra Biometrics, USA); Set number of direct questions uses branching logic to trigger subsequent questions</td>
<td>Presence or absence of headache symptoms Pain severity (0–10 NRS) Pain localization (ad hoc*) Pain quality (ad hoc*) Associated symptoms (ad hoc*) Functioning interference (ad hoc*) Premenstrual symptoms (ad hoc*)</td>
<td>Event-based; mean 1.75 per day; three months</td>
<td>Acceptability: Technical issues resulted in low satisfaction with e-diary – this was assessed from the perspective of researchers. Technical problems: 35% of the total number of entries for the study had abnormal session endings; 34.5% of diary entries did not record an answer to the first question</td>
</tr>
<tr>
<td>2</td>
<td>Lewandowski et al (45), 2009 (USA)</td>
<td>Children with either recurrent headache (n=54), juvenile chronic arthritis (n=30); SCD (n=9); 12.5±2.5 years of age</td>
<td>RCT To compare diary and retrospective reports of pain</td>
<td>Handheld PDA (Jornada 548, Hewlett Packard, USA); Specific sequence of questions with a built-in response loop</td>
<td>Pain intensity (Faces Pain Scale [39]) Activity limitations (Children’s Activity Limitations Interview [40])</td>
<td>Time-based; once per week</td>
<td>Compliance: 98% compliance (e-diary users completed a mean of 6.89 days of diary reports over a seven-day period versus 4.97 days for paper diary users)</td>
</tr>
<tr>
<td>3</td>
<td>Moloney et al (46), 2009 (USA)</td>
<td>Women, 18–55 years of age; n=94; 37.5 years of age (range 21–54)</td>
<td>Prospective observational To explore feasibility and acceptability of Internet-based headache diary</td>
<td>Personal desktop computer; Paper and pencil headache diary transformed for Internet using javascript and ColdFusion (Adobe Systems Inc, USA)</td>
<td>Triggers Probdomal symptoms and their severity Use of self-treatment Probability of occurrence of a headache (ad-hoc*; Likert-type scales, multiple-choice and open-ended questions)</td>
<td>Time-based; four months; one per day</td>
<td>Compliance: 75% of all diary pages were completed within two days Acceptability: Participants reported that the diary website was user-friendly and easy to navigate, with straightforward instructions. Technical problems: 21% reported occasional difficulties with computer or website access; entries posted in a later time zone than that of the study were logged as late entries and error messages were sent</td>
</tr>
<tr>
<td>4</td>
<td>Palermo et al (47), 2004 (USA)</td>
<td>Children, 8–16 years of age, with headache (n=40) or juvenile idiopathic arthritis (n=20); 12.3±2.4 years of age</td>
<td>RCT To compare compliance, accuracy and acceptability of electronic versus paper diaries</td>
<td>Handheld PDA (Jornada 548, 32 MB RAM, colour screen [Hewlett Packard, USA]) using the Windows CE operating system (Microsoft Corporation, USA); Specific sequence of questions with a built-in response loop</td>
<td>Daily pain occurrence (ad hoc*) Pain localization (validated body outline [41]) Pain intensity (Faces Pain Scale [39]) Pain duration (ad hoc*) Emotional upset (ad hoc*) Activity limitations (Children’s Activity Limitations Interview [40]) Somatic symptoms (Children’s Somatization Inventory, six core items) (42)</td>
<td>Time-based; once per day; one week</td>
<td>Compliance: Compliance was significantly higher for e-diary users, with 83.3% of children completing all e-diary entries versus 46.7% of paper diary users Acceptability: 83.3% of e-diary users found the diary ‘very easy or easy to remember to fill out’; 72.2% found the e-diary ‘no bother at all to fill out daily’; 61.1% of e-diary users indicated they ‘very much liked the way the daily diary looked’, no significant difference in participants’ acceptability of either paper or electronic diary format Technical problems: Power failure and PDA malfunction caused data from five participants to be unusable; three stylus pens broken/lost; two AC adapters required repair</td>
</tr>
</tbody>
</table>
captured in varying degrees across the five studies. Pain intensity/severity was captured in all five studies. Pain location was recorded in studies 1, 2 and 4, while headache duration was recorded only in study 4. Premonitory symptoms and triggers were investigated in two of the five studies (studies 3 and 5), while one study investigated associated symptoms (study 1) and another investigated somatic symptoms (study 4). Physical activity limitation was captured in studies 2 and 4, whereas study 1 investigated the interference of headache with overall functioning (usual activities at home, work and/or school). Only two of the five studies recorded what steps participants were taking to self-manage or cope with their headaches (studies 3 and 5). Other outcomes included emotional upset (study 4), premonstrual symptoms (study 1) and predicting the likelihood of headache development (study 3). Two of the studies featured ad-hoc measures that were developed by study investigators (studies 1 and 3), one study featured electronic versions of validated paper measures (study 2) and two studies used a mixture of both ad-hoc and validated paper measures (studies 4 and 5). The validated paper measures transformed for the e-diaries included: Faces Pain Scale (49), Children’s Activity Limitations Interview (50), validated body outline for pain localization (51) (study 4), Children’s Somatization Inventory (52) (study 4), and the International Headache Society’s (IHS) International Classification of Headache Disorders diagnostic criteria (53) to diagnose headache type (study 5).

Diary characteristics
The types of e-diaries used in the studies were somewhat homogenous, as shown in Table 2, with four of the five evaluation studies featuring e-diaries on a handheld device (studies 1, 2, 4 and 5), while one was delivered over the Internet on a computer (study 3). The handheld devices used were the Jornada 548 personal digital assistant (PDA) (Hewlitt-Packard, USA) (studies 2 and 4), the PalmOne Treo 600 (Palm Inc, USA) with a dedicated server, runs on Linux supported by: Apache Web Server; Java; PostgreSQL; Tomcat (study 1), and the Nindo handheld device (Philips Electronics, Netherlands) (study 1). All handheld devices featured a specific question sequence and used question logic to trigger subsequent questions. Four studies provided detailed regarding data retrieval (studies 1, 3, 4 and 5). Handheld devices in studies 1 and 4 stored data to be uploaded to desktop computers once the diaries were returned to the research centre through a delimited text file (study 1) and synchronization cable (study 4). The handheld device in study 5 featured data transmission through the Internet, with diary answers being saved separately on a server to avoid data loss. Two of the handheld devices stored data in files that were immediately suitable for statistical packages (studies 1 and 5). The Internet delivered stored e-diary data to online servers that were immediately available to researchers (study 3).

Quality and overall psychometric evidence
The Cohen criteria (25) used to evaluate the quality and overall evidence supporting the psychometric testing of the e-diaries (Appendix I) revealed that none of the five e-diaries met any of the criteria, indicating that their psychometric properties were not described sufficiently and/or not reported at all. Overall, the five studies evaluated had poor quality or reporting according to the COSMIN checklist criteria (27). Of the 10 scales of the COSMIN, all five studies only achieved ratings on the interpretability scale, indicating that all other scales (internal consistency, reliability, measurement error, content validity, construct validity, hypothesis testing, cross-cultural validity, criterion validity and responsiveness) were inadequately reported and, therefore, could not be rated. Interpretability ratings for all five studies were poor.

Feasibility
Compliance: Compliance with the e-diaries was reported in four of the five evaluation studies (studies 2, 3, 4 and 5). Overall compliance was high, with a mean compliance rate of 84.1±9.88% for all four studies (range 75% to 98%). The two studies featuring an RCT design (studies 2 and 4) reported greater compliance with the e-diaries compared with paper diaries used by the control group. Study 4 reported a statistically significant difference (P<0.001), with 83% of e-diary users completing all diary entries compared with 47% of paper diary users.

Only one study provided device training (study 4) with scheduled in-home instruction on how to use the PDA. Four of the five studies provided details of added features to enhance compliance. Programmed audible alarms were used in three of the studies (studies 2, 4 and 5), with an option to convert to a visual reminder when an audible alarm was not convenient (church service, meeting or concert) in study 5. Study 4 also featured a visual reminder on the diary screen when the diary had not been completed the day before and one reminder phone call during the recording period. Monetary incentives were provided in two studies (studies 2 and 4): in study 2 participants received a $30 gift card as compensation; and in study 4 participants were given a $10 incentive when diaries were returned. Participants were also entered into a monthly draw for an additional $10 incentive for completing entries on all seven days. Finally, feedback was provided twice per day.
TABLE 4
Summary of outcome studies using e-diaries

<table>
<thead>
<tr>
<th>Study number</th>
<th>Author (reference), year (country)</th>
<th>Study design</th>
<th>Type of diary</th>
<th>Diary measure type</th>
<th>Diary measures: ad hoc/validated</th>
<th>Feasibility/compliance data</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Bjorling (42), 2009 (USA)</td>
<td>Prospective</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Mean compliance rate 62.5% (range 30% to 92%)</td>
</tr>
<tr>
<td>7</td>
<td>Kikuchi et al (39), 2007 (Japan)</td>
<td>Prospective</td>
<td>Wrist-watch</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Mean compliance rate 96% (range not provided)</td>
</tr>
<tr>
<td>8</td>
<td>Barton and Blanchard (41), 2001 (USA)</td>
<td>Prospective</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc HA Index</td>
<td>Not available</td>
</tr>
<tr>
<td>9</td>
<td>Giffin et al (43), 2003 (United Kingdom)</td>
<td>Prospective</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>10</td>
<td>Holroyd et al (40), 2007 (USA)</td>
<td>Prospective</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>11</td>
<td>Hu et al (37), 2008 (USA)</td>
<td>Multisite prospective observational</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>12</td>
<td>James (38), 1998 (Australia)</td>
<td>Double-blinded placebo-controlled crossover design</td>
<td>Light-weight portable computer</td>
<td>Primary</td>
<td>Ad hoc and modified versions of validated paper measures</td>
<td>Not available</td>
</tr>
<tr>
<td>13</td>
<td>Massey et al (34), 2009 (The Netherlands)</td>
<td>Prospective</td>
<td>Desktop computer</td>
<td>Primary</td>
<td>Ad hoc and modified versions of validated paper measures</td>
<td>Mean compliance rate 57%</td>
</tr>
<tr>
<td>14</td>
<td>Mathew et al (35), 2007 (USA)</td>
<td>Multicentre, double-blinded, placebo-controlled parallel-group (intervention)</td>
<td>Handheld</td>
<td>Primary and secondary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>15</td>
<td>Ng-Mak et al (33), 2008 (USA)</td>
<td>Multicentre prospective follow-up</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>16</td>
<td>Ng-Mak et al (32), 2007 (USA)</td>
<td>Multicentre observational follow-up</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>17</td>
<td>Sheftell et al (31), 2005 (USA)</td>
<td>Randomized, double-blinded, parallel-group (intervention)</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>18</td>
<td>Thorn et al (29), 2007 (USA)</td>
<td>Randomized controlled trial (intervention)</td>
<td>Desktop computer</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>19</td>
<td>Strom et al (28), 2000 (Sweden)</td>
<td>Randomized controlled trial (intervention)</td>
<td>Desktop computer</td>
<td>Primary</td>
<td>Ad hoc HA index</td>
<td>Not available</td>
</tr>
<tr>
<td>20</td>
<td>Van Gerven et al (30), 1996 (The Netherlands)</td>
<td>Double-blinded randomized parallel-group design (intervention)</td>
<td>Portable computer</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
<tr>
<td>21</td>
<td>Honkoop et al (36), 1999 (USA)</td>
<td>Prospective</td>
<td>Handheld</td>
<td>Primary</td>
<td>Ad hoc</td>
<td>Not available</td>
</tr>
</tbody>
</table>

Ad hoc measures were developed by authors for study purposes; HA headache

Acceptability (likes and dislikes): Three studies reported on the user acceptability of headache e-diaries (studies 3, 4 and 5). All three reported positive participant ratings of acceptability, indicating that the e-diaries were user-friendly, convenient and appealing. Quantifiable satisfaction data was reported in study 4, who found that 83.3% of participants that rated the acceptability of the e-diary found it ‘very easy or easy to remember to fill out’, 72.2% of participants found it ‘no bother at all to fill out daily’ and 61.1% ‘very much liked the way the daily diary looked’. While study 1 reported a low researcher satisfaction rating of the e-diary based on the high number of technical problems encountered in the study that resulted in lost data, no participant acceptability data were reported for this study.

Technical difficulties: Of the five reviewed evaluation studies, four reported technical problems. Studies 3, 4 and 5 reported minimal technical problems with participants experiencing occasional difficulty accessing the website (study 3); hardware problems including technical failure of electronic device, broken or lost stylus pens and AC adapters requiring replacement (study 4); minor internal problems with the PDA (study 5); and software problems obstructing the storage of feedback (first cycle of study 5). Major technical problems in study 1 that caused abnormal session endings in 35% of all diary entries and 34.5% of responses to the first question to be lost resulted in researchers reporting dissatisfaction with the e-diary.

OUTCOME STUDIES USING E-DIARIES

The search strategy yielded 16 studies published between 1996 and 2009 found to be using an e-diary to obtain outcome data (28-43) (Table 4). Of these, 10 (62.5%) were prospective observational studies (studies 6 through 11, 13, 15, 16, 21); the remaining six were interventional RCT studies. The devices used in the studies varied. The majority (62.5%) used a handheld device (studies 6, 8 through 11, 13 to 17, 21); however, other studies reported using a desktop computer (18.7%) (studies 13, 18 and 19),
a portable computer (12.5%) (studies 12 and 20) and a wrist watch computer (6.3%) (study 7). Reliability and validity were not reported for any of the diaries used in the outcome studies; however, despite this, 15/16 studies were using e-diaries to collect primary outcome data (studies 6 through 13, 15 to 21), with the final study using the e-diary to collect both primary and secondary outcome data (study 15). Based on the information available, the multidimensional measures used in the e-diaries were found to be ad-hoc for the most part (87.5%) (studies 6 through 11, 14 to 21), with two of the studies using a combination of ad hoc measures and modified versions of validated paper measure (studies 12, 13). Only three of these studies provided compliance data with moderate to high ratings reported (57% to 96%) (studies 6, 7, 13).

Conforming to headache and chronic pain clinical trial reporting recommendations

RCT evaluation and outcome studies were compared with the recommended headache and chronic pain guideline. None evaluated all of the outcomes outlined in any one set of guidelines; however, each assessed at least one (either primary or secondary) recommended outcome.

Pharmacological headache trial guidelines

Two RCT evaluation studies (studies 2 and 4) and the six RCT outcome studies using e-diaries (studies 12, 14, 17 through 20) were compared with the recommended pharmacological headache trial guidelines first published in 1991 (14). Frequency, the recommended primary outcome for pharmacological headache trials, was evaluated in four of the eight RCTs (studies 4, 12, 18 and 19), while one additional study (study 14) evaluated the other primary outcomes recommended for pharmacological trials: percentage of patients pain free at 2 h; and sustained pain free. Seven of the eight RCTs assessed for compliance to the pharmacological trial guidelines measured pain intensity, a recommended secondary outcome (studies 2, 4, 12, 17 through 20). Time to pain free, a recommended secondary outcome for pharmacological trials, was measured in study 17.

Behavioural headache trial guidelines

One RCT evaluation study (study 2) and the three RCT outcome studies (studies 14, 18 and 19) published after the behavioural headache trial guidelines were made available in 2005 (15) were assessed for behavioural trial guideline compliance. Frequency, the recommended primary outcome for behavioural headache trials, was evaluated in two of the four assessed RCTs (studies 18 and 19). Three of the four RCTs measured pain intensity, a recommended secondary behavioural outcome (studies 2, 18 and 19). Other secondary outcomes recommended by the behavioural headache trial guidelines included medication use, which was evaluated in two of the RCTs (studies 18 and 19); psychological symptoms, evaluated in two studies (studies 4 and 18); and headache duration, evaluated in two studies (study 2).

Chronic pain trial guidelines

Only one article, an RCT evaluation study (study 2) was published after the chronic pain guidelines became available and, therefore, is the only study compared with the IMMPACT guidelines (12,13). Study 2 measured pain intensity, a recommended primary chronic pain outcome.

DISCUSSION

The present study sought to report on the development and measurement properties of electronic pain diaries to capture the experience of headaches in children and adults and the components of these diaries and their conformity to recommended reporting guidelines for headache and chronic pain clinical trials. Despite the rapid growth in the use of electronic pain diaries in research, there is little evidence supporting the validity of e-diaries in this population. Most of the e-diaries were developed ad hoc by the investigators and only a few incorporated validated paper versions of self-reported pain outcome measures in terms of intensity, frequency, headache location and impact on functioning. There is some evidence regarding the feasibility of headache e-diaries in terms of high compliance and acceptability compared with paper diaries; however, there were several drawbacks noted in terms of technical issues reported across the five studies. Finally, the studies that used headache e-diaries as primary outcome measures did not adequately report on the essential features of the diary (validity, sampling procedure, etc.), were of poor methodological quality and did not meet the recommended reporting guidelines for headache clinical trials.

It is essential that investigations of headache treatments (especially clinical trials) use reliable and valid pain measures to ensure that the results are correctly interpreted. There appears to be the implicit assumption in these studies that if the measures are valid in paper format, the method of administering the measure is inconsequential; however, studies have found that this may not be the case (54,55). To ensure that valid and reliable measures are being used in clinical trials, when the data are not available on the psychometric properties of e-diaries, researchers need to determine the validity and reliability of the RTDC versions before using them in definitive randomized controlled trials. Researchers developing headache e-diaries should use guidelines for the development of patient-reported outcome measures such as the Food and Drug Administration guidance (56).

Feasibility is a critical component when establishing the properties of a new patient-reported outcome measure. In terms of e-diaries, feasibility has been conceptualized as compliance, acceptability and technical issues. In terms of compliance, e-diaries have been found to be superior to paper diaries. Similar to the findings in the present study, Dale and Hagen (57) found that when reviewing electronic versus paper diaries, e-diaries led to improved protocol compliance. While compliance was measured differently across all studies, the majority reported the e-diary users to have greater numbers of diary entries at greater frequency. Additionally, the time stamp features of the PDAs precluded falsification of PDA data records (58), a prudent issue when considering the almost 80% falsification rate of paper diary data reported by the Stone et al (11) study in their comparison of paper and e-diaries. However, it is not known how factors such as training, alarms and other reminders, and monetary incentives influence compliance.

Recent reviews regarding e-diaries have highlighted the advantages of this new mode of data capture compared with paper diaries such as overcoming recall bias, diary compliance and portability issues (59). Stone and Broderick (58) also suggest that e-diaries limit distortions to past experiences that are present in recall measures, particularly the changes in frequency of nonpain episodes. Given that headache frequency is the recommended primary outcome measure of both pharmacological and behavioural headache trials, e-diaries can improve the accuracy of this outcome by minimizing recall bias. However, e-diaries can be burdensome for the respondent due to the duration of monitoring (weeks to years), length of the assessment battery comprising each diary report and the device training required at the start of diary use. Additionally, the cost of the device itself may be prohibitive, especially in the context of clinical trials involving hundreds of participants (39). Furthermore, requiring participants to carry a device specific to the e-diary in addition to devices any they already carry (smartphone, pager, etc) may create an additional perceived burden. However, multiple hand-held devices (eg, iPhone [Apple Inc, USA]) that can support e-diaries are becoming increasingly popular, especially among youth and young adults with headaches (60). Future research to develop and validate diaries across a variety of devices/platforms (ie, iPhone, Android [Google Inc, USA], personal computers) that participants can download to their own device is warranted to increase adherence, feasibility and reduce burden.

While there are many reported advantages of e-diaries (11,16), there are several drawbacks that need to be acknowledged. The present review found that four of five studies reported some technical difficulties ranging from software to hardware issues and devices being lost. These technical issues can result in loss of data that may make it difficult to determine whether data are missing at random or systematically due to these technical issues. Similarly, in a systematic review of RCTs and quasirandomized controlled trials comparing PDA and paper diaries, Dale and Hagen (57) found that five of nine reviewed studies reported power problems and/or PDA malfunctions, resulting in data...
loss in three of the studies. These technical issues can also affect participant compliance and acceptability of this method of diary administration. These issues highlight the importance of conducting usability and feasibility testing before evaluating the measurement properties of these diaries or their use in clinical trials.

Another area that was deficient in the headache e-diaries in the present review was that the investigators did not follow the guidelines for reporting that have been proposed in studies using e-diaries. In addition to established psychometric measurement reporting guidelines for patient-reported outcome measures, e-diaries have additional testing components, such as usability and feasibility, that must also be reported. Stone and Shiffman (16) developed such a set of criteria, which they suggest must be explicitly described in any report of clinical trials using RTDC approaches using e-diaries. These criteria include the following dimensions: providing rationale for the momentary sampling design; specifying details of sampling procedures (including participants' discretion in responding, descriptions of missing or delayed data and suspension of prompts); description of how data acquisition questions that must be truncated to accommodate brief collection of data are designed and presented; compliance with sampling plan (particularly systematic flaws leading to missing data, hoarding or backfilling entries or fabricating reports); reporting of training and monitoring protocols; reporting on how the large amount of data are managed; and describing the approach to data analysis (16).

Stone and Shiffman (16) strongly suggest these criteria be clearly documented at the outset of study protocols for readers of the trial to make appropriate judgments about generalizability and validity.

There are unique issues when deciding what outcomes should be measured using e-diaries to capture headache symptoms in which the painful events are episodic in nature. According to the IHS, the most important outcome to measure in headache sufferers is headache frequency (14). Moreover, when using e-diaries, care needs to be taken to choose the most important outcome variables to include, especially when participants will be asked to complete the diary daily and over long periods of time (several months). Stone and Shiffman (16) recommend a maximum of 20 items in daily diaries. With this in mind, we would recommend including, in addition to headache frequency, headache intensity and impact of headache on sufferer’s daily functioning including physical, role and emotional functioning, and sleep. All of these variables are suggested by both IHS15 and IMMPACT/IMMPACT (12,13) groups. Other variables, such as side effects and treatment satisfaction, are important, but would be better captured using paper questionnaires at end of study to reduce patient burden and the potential for poor compliance. Furthermore, consensus on the validated paper-based measures for these domains and the resultant evaluation of their psychometric properties using e-diaries will facilitate the pooling of data and increase the quality of clinical headache trials.

Finally, the present review provides researchers with directions regarding future areas of research so that we can harness the full potential of this medium to capture patient-reported outcomes in headache investigations. First, research efforts need to be directed at establishing the psychometric properties of headache e-diaries as outlined in Table 2. Second, end-users should be involved in the development and testing of these diaries. Third, they need to be developed using a phased approach that includes examining their usability and feasibility before testing their psychometric properties (61,62). Finally, the diaries should also include the most important outcomes as recommended by the IHS15 and IMMPACT (12,13) statements to keep the diary entries brief (fewer than 20 questions) and minimize respondent burden.

Study limitations

The present study has several limitations. First, because few evaluations of headache e-diaries have been conducted over the past 15 years, our findings are based on a very small number of studies. Future reliability and validity investigations will be required to establish whether e-diaries are reliably efficacious for capturing headache characteristics and precursors. Due to the limited data reported in the studies included in the present review, we were only able to provide a qualitative synthesis. In addition, several factors may limit the generalizability including the fact that studies were limited to English and that the studies found were primarily conducted in North America.

CONCLUSIONS

The results of the present study indicate that, despite technical problems, researchers can expect higher levels of compliance and user acceptability when using headache e-diaries. However, further research focusing on the psychometric properties of e-diaries is required. It is important to ensure that the electronic questionnaires used are tested for validity and reliability against their paper counterparts to ensure that the psychometric properties are maintained. Furthermore, the items on headache e-diaries should include the most important outcomes (headache frequency, pain intensity, the impact of pain on aspects of physical, emotional and role functioning and sleep) to minimize burden and optimize compliance. While it is clear that there will be a role for paper diaries in research for the foreseeable future, e-diaries are an exciting medium that offer many advantages over traditional data collection methods.

APPENDIX I

Criteria for evidence-based assessment

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-established assessment</td>
<td>The measure must have been presented in at least two peer-reviewed articles by different investigators or investigatory teams.</td>
</tr>
<tr>
<td></td>
<td>Sufficient detail about the measure to allow critical evaluation and replication (eg, measure and manual provided or available upon request).</td>
</tr>
<tr>
<td></td>
<td>Detailed information (eg, statistics presented) indicating good validity and reliability in at least one peer-reviewed article.</td>
</tr>
<tr>
<td>Approaching well-established assessment</td>
<td>The measure must have been presented in at least two peer-reviewed articles, which may be by the same investigator or investigatory team.</td>
</tr>
<tr>
<td></td>
<td>Sufficient detail about the measure to allow critical evaluation and replication (eg, measure and manual provided or available upon request).</td>
</tr>
<tr>
<td></td>
<td>Validity and reliability information presented in either vague terms (eg, no statistics presented) or moderate values.</td>
</tr>
<tr>
<td>Promising assessment</td>
<td>The measure must have been presented in at least one peer-reviewed article.</td>
</tr>
<tr>
<td></td>
<td>Sufficient detail about the measure to allow critical evaluation and replication (eg, measure and manual provided or available upon request).</td>
</tr>
<tr>
<td></td>
<td>Validity and reliability information presented in either vague terms (eg, no statistics presented) or moderate values.</td>
</tr>
</tbody>
</table>

Adapted from Cohen et al (25)
REFERENCES


Qualitative review of electronic headache diaries

Pain Res Manag Vol 18 No 3 May/June 2013 151


