Lack of benefit for prophylactic drugs of tension-type headache in adults: a systematic review

Arianne P Verhagen\textsuperscript{a,\textasteriskcentered}, Léonie Damen\textsuperscript{a}, Marjolein Y Berger\textsuperscript{a}, Jan Passchier\textsuperscript{b} and Bart W Koes\textsuperscript{a}

\textsuperscript{a}Department of General Practice and \textsuperscript{b}Department of Medical Psychology and Psychotherapy, Erasmus Medical Center, Rotterdam, The Netherlands.

\textsuperscript{\textasteriskcentered}Correspondence to Arianne P Verhagen, Department of General Practice, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; E-mail: a.verhagen@erasmusmc.nl

Received 7 January 2009; Revised 13 October 2009; Accepted 30 November 2009.

Objective. To assess the efficacy and tolerability of prophylactic drugs for chronic tension-type headache (TTH) in adults.

Methods. We searched several databases from inception to August 2009. We selected randomized trials that reported the effects of prophylactic drugs in patients with TTH, with a pain measure (intensity, frequency, duration, improvement or index) as outcome measure. Two authors independently assessed risk of bias and extracted data from the original reports. A data synthesis was carried out according to the type of medication.

Results. We included 44 trials (3399 patients), of which 15 (34.1\%) were considered to be of low risk of bias. Main types of medications studied were antidepressants, muscle relaxants, benzodiazepines and vasodilator agents. Overall, antidepressants were no more effective than placebo, and there were no significant differences between different types of antidepressants. There was conflicting evidence about the effectiveness of benzodiazepines and vasodilator agents compared with placebo. Furthermore, there was limited evidence that propranolol had negative effects on depression in TTH patients, when compared with placebo or biofeedback. There was no evidence concerning the effectiveness of muscle relaxants alone or 5-HT receptor agonist compared with placebo.

Conclusions. Overall, antidepressants were no more effective on headache intensity or frequency and analgesic use than placebo. Propranolol seemed to have negative effects on depression in TTH patients when compared with placebo or biofeedback. No evidence was found for the use of muscle relaxants alone or 5-HT receptor agonist.

Keywords. Family medicine, meta-analysis, preventive medicine, randomized controlled trial, systematic review.

Introduction

Tension-type headache (TTH), also known as tension headache or muscle contraction headache, is the most commonly experienced type of headache. Population-based studies suggest 1-year prevalence rates of TTH of 35–40\% in adults.\textsuperscript{1,2} Chronic TTH has been defined in the classification of the International Headache Society (IHS) as headache frequency of at least 15 days a month during at least 6 months. The headache is usually pressing/tightening in quality, mild or moderate in severity, bilateral and does not worsen with routine physical activity.\textsuperscript{3,4} In addition, no more than one additional clinical feature of migraine (nausea, photophobia or phonophobia) is permitted and no vomiting.\textsuperscript{4}

Although TTH is common, the pathophysiology and likely mechanism remain unclear. The pathophysiology is considered to be multifactorial, involving factors from the central and peripheral nervous system as well as environmental factors.\textsuperscript{3}

The uncertainty of the pathogenesis is reflected in the variety of prophylactic drugs available. Prophylactic drugs are those drugs that are taken every day, regardless whether currently headache is being experienced, with the aim of preventing headache attacks. The most common used prophylactic drugs are tricyclic antidepressant agents (TCA), other antidepressants and muscle relaxants, but benzodiazepines and vasodilator agents are also prescribed.

Because of the variety in prophylactic drugs, it is unclear for the clinician that which prophylactic drugs to prescribe. Recently, several (systematic) reviews became available which summarized the efficacy and tolerability of antidepressants as prophylactic drugs.\textsuperscript{3,5}
treatment in TTH.\textsuperscript{6–8} One review evaluated the efficacy of TCAs and selective serotonin re-uptake inhibitors (SSRIs) in patients with chronic headaches but did not differentiate between patients with migraine or TTH.\textsuperscript{6} A Cochrane review evaluated the efficacy of only SSRIs in patients with migraine and TTH separately.\textsuperscript{7} The authors concluded that in patients with TTH, SSRIs are no better than placebo’s and somewhat less efficacious than TCAs, but with less side effects. Recently, a narrative review has been published stating that preventive treatments in patients with TTH were on average not effective.\textsuperscript{8} No clear systematic overview exists on all prophylactic drugs in adults with TTH. Therefore, the objective of this review was to summarize the evidence concerning the efficacy and tolerability of prophylactic drugs compared with placebo, other drugs or other prophylactic treatments in adult patients with TTH.

Methods

Search strategy

Medline, Pubmed, Cinahl, Cochrane and Embase were searched from inception to August 2009 and the Cochrane Controlled Trials Register, Cochrane Library (issue 2, 2009) using the terms ‘tension-type headache’, ‘tension headache’, ‘stress headache’ or ‘muscle contraction headache’ together with the search strategy for identifying randomized controlled trials described by Robinson and Dickerson.\textsuperscript{9} Additional strategies for identifying trials included searching the reference lists of review articles and included studies.

Study selection

We selected only randomized clinical trials (RCTs) evaluating oral prophylactic drugs used in the management of TTH compared with placebo, other drugs or non-pharmacological treatments in adults (≥18 years). Studies should use reasonable eligibility criteria designed to distinguish TTH from migraine. The use of a specific set of diagnostic criteria (e.g. IHS and Ad Hoc)\textsuperscript{3,10} was not required, but TTH diagnoses had to be based on at least some of the distinctive features of TTH, e.g. bilateral in location, no nausea or vomiting, mild or moderate intensity or no exacerbation by exercise. Studies with at least one of the following outcome measures were included: headache measure (intensity, frequency, duration, improvement or index), analgesic use, depression or adverse events. No language restriction was applied.

Two authors independently screened titles and abstracts of studies identified by the literature search. All potentially relevant studies were retrieved as full papers and then again independently reviewed by two authors for eligibility.

Risk of bias assessment

Two authors independently rated the risk of bias of the included trials using the Delphi list.\textsuperscript{11,12} The Delphi list is a generic methodological criteria list developed by international consensus and consists of the following 9 items: (1) randomization; (2) adequate allocation concealment; (3) groups similar at baseline; (4) specification of eligibility criteria; (5) blinding of outcome assessor; (6) blinding of care provider; (7) blinding of patient; (8) presentation of point estimates and measures of variability and (9) intention-to-treat analysis. One extra item was added: (10) withdrawal/dropout rate unlikely to cause bias because this was considered to be relevant. All selected criteria were scored as yes, no or don’t know. In case of a disagreement between the two authors, consensus was used to resolve disagreement. When consensus could not be reached, a third author made the final decision (MB or AV).

An overall score was computed by counting the number of positive scores. All studies receiving a score of six or more were regarded as of low risk of bias.\textsuperscript{13}

Data extraction

One author performed the extraction of data from the original reports and this was checked by a second. Disagreements were resolved by consensus. Extracted information included demographic data, detailed description of the intervention and control (i.e. dose given and study duration), outcome measures and information on adverse effects. Outcome measures of interest were headache intensity, frequency and duration, recovery, depression and analgesic use. A GP and a pharmacologist decided in consensus to combine the drugs into six main categories based on pharmacological working mechanism or effect on patients: antidepressants (TCA, SSRI and other non-tricyclic antidepressants), muscle relaxants, benzodiazepines, vasodilator agents, 5-HT receptor agonist or other drugs.

Data analysis

First, we evaluated the inter-observer reliability of the risk of bias assessment using kappa statistics. Next, we calculated for each study standard mean differences (SMD) with 95% confidence interval (CI) for continuous outcomes or relative risks (RR) with 95% CI in case of dichotomous variables. We presented data as treatment success, indicating that a RR >1 and a SMD >0 represent a better outcome for the first mentioned intervention group.

Statistical pooling of the results was the preferred analysis. In case of clinically heterogeneity concerning the patient population, and the type of intervention and control interventions, we analysed the results using a rating system with levels of evidence.\textsuperscript{13} The evidence was judged to be strong when multiple trials
with low risk of bias produced consistent findings. Results were considered consistent if ≥75% of the studies reported similar results on the same outcome measure. It was judged to be moderate when one RCT with low risk of bias or multiple RCTs with high risk of bias produced generally consistent findings. Evidence was considered to be limited when no RCTs were found or when the authors provided insufficient data for analysis.

We performed sensitivity analysis in studies having adequate power (at least 25 subjects per study arm), using IHS or Ad Hoc criteria for patient selection and with low risk of bias.

Results

Search results

A total of 2439 publications were identified by our search strategy (see figure 1: flow chart). Of these, 44 trials were included in this review of which six papers concerned double publications. Thirteen studies used a crossover design.

Description of studies

Details of the included studies are presented in Tables 1 and 2.

Participants. The number of included participants in each trial ranged from 16 to 375 (mean 77.3), with a total of 3399 patients included. Most studies were small; out of 44 studies, 25 included study arms with <25 subjects (of which three studies had <10 subjects in each arm), whereas only three studies included >50 subjects in each study arm. The mean percentage of participants who dropped out was 18.1% (range 0–52%). Age of the participants ranged from 18–87 years. Twenty-six trials used the criteria of the IHS to classify TTH, five trials used the Ad Hoc Committee’s criteria, whereas the remaining 13 studies used varying definitions.

Intervention. The mean (± SD) total study duration was 12.8 (±7.2) weeks, split in a mean baseline period of 2.3 (±1.2) weeks, a mean treatment period of 8.2 (±4.7) weeks and a mean follow up period of 1.9 (±4.4) weeks; 66% of the studies, however, did not include a follow-up period. Eighteen studies compared
<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Results (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashina et al.14/Bendtsen et al.15,16 RCT-CO; QS: 6 (items: 1, 4, 5, 7, 8, 10)</td>
<td>Study duration: 32 weeks (4 weeks baseline, 8 weeks treatment, 2 weeks wash-out, 8 weeks treatment, 2 weeks wash-out, 8 weeks treatment)</td>
<td>TTH (IHS): N = 40, 6 dropouts; female 64.7%; mean age: 40.7 years. Setting: outpatient headache clinic, Denmark</td>
<td>I1: AD amitriptyline 75 mg (n = 34) Week 1: 25 mg. Week 2: 50 mg. Weeks 3–8: 75 mg</td>
<td>HA intensity: duration × intensity HA intensity 1p Likert scale HA duration, frequency Analgesic use: number Adverse events: stated</td>
<td>HA intensity: I1 versus C: SMD = 0.06 (–0.42; 0.53); I2 versus C: SMD = 0.17 (–0.3; 0.65); I1 versus I2: SMD = 0.06 (–0.3; 0.59) HA duration: I1 versus C: SMD = 0.22 (–0.26; 0.7); I2 versus C: SMD = 0.01 (–0.46; 0.49); I1 versus I2: SMD = 0.21 (–0.27; 0.68) Analgesic use: no differences Adverse events: I1: n = 15; I2: n = 15; C: n = 15 No data available.</td>
</tr>
<tr>
<td>Bendtsen and Jensen23 RCT-CO; QS: 7 (items: 1, 2, 4, 5, 7, 8, 10)</td>
<td>Study duration: 22 weeks (4 weeks baseline, 8 weeks treatment, 2 weeks wash-out, 8 weeks treatment)</td>
<td>TTH (IHS): N = 24, 2 dropouts; female 45.5%; mean age: 45 years. Setting: outpatient headache clinic, Denmark</td>
<td>I1: AD mirtazapine 15 mg (n = 22) C: placebo (n = 22)</td>
<td>HA improvement: area under the curve &gt;30%. HA severity: 11p Likert scale. Analgesic use: number. Adverse events: stated.</td>
<td>HA intensity: I2 versus C: SMD = 0.17 (–0.41; 0.75) I3 versus C: SMD = 0.0 (–0.57; 0.57) I2 versus I3: SMD = 0.18 (–0.35; 0.75) Adverse events: I1: n = 15; I2: n = 14; I3: N = 11; C: N = 10</td>
</tr>
<tr>
<td>Bendtsen et al.27 RCT; QS: 4 (items: 1, 4, 8, 10)</td>
<td>Study duration: 12 weeks (4 weeks run-in, 8 weeks treatment)</td>
<td>CTTH (IHS): N = 93; 9 dropouts; female 50%; mean age: 39.3 years. Setting: outpatient headache clinic, Denmark</td>
<td>I1: AD mirtazapine 4.5 mg + ibuprofen 100 mg. (N = 23) I2: AD mirtazapine 4.5 mg + placebo. (N = 23) I3: ibuprofen 400 mg + placebo. (N = 24). C: placebo. (N = 23)</td>
<td>HA intensity: VAS HA frequency: days/4 weeks. Analgesics intake: days/4 weeks; doses/4 weeks. Adverse events: stated.</td>
<td>HA intensity: I2 versus C: SMD = 0.17 (–0.41; 0.75) I3 versus C: SMD = 0.0 (–0.57; 0.57) I2 versus I3: SMD = 0.18 (–0.39; 0.75) Adverse events: I1: N = 15; I2: N = 14; I3: N = 11; C: N = 10</td>
</tr>
<tr>
<td>Boz et al.38 RCT; QS: 5 (items: 1, 3, 4, 8, 10)</td>
<td>Study duration: 16 weeks (4 weeks run-in, 12 weeks treatment)</td>
<td>CTTH (IHS). N = 90; female 87.8%; mean age 39.1; 6 dropouts. Setting: outpatient headache clinic, Turkey</td>
<td>I: AD amitriptyline. (N = 46). First week 10 mg, thereafter 25 mg. C: AD sertraline 50 mg. (N = 44).</td>
<td>HA intensity: VAS HA frequency: days/4 weeks HA duration: hours/day Analgesic use. Adverse events: stated.</td>
<td>HA intensity: I1 versus C: SMD = 1.12 (0.68; 1.57) HA frequency: I1 versus C: SMD = 0.16 (–0.25; 0.58) HA duration: I1 versus C: SMD = 0.33 (–0.10; 0.76) Only point estimates presented. Adverse events: I: n = 46 versus C: n = 3</td>
</tr>
<tr>
<td>Boline et al.39 RCT; QS: 6 (items: 1, 2, 3, 4, 8, 9)</td>
<td>Study duration: 12 weeks (2 weeks baseline, 6 weeks treatment, 4 weeks follow-up)</td>
<td>TTH (IHS): N = 150, 24 dropouts; female 61.1%; mean age: 41.7. Setting: advertisement, Minneapolis, USA</td>
<td>I: AD amitriptyline 30 mg (n = 56) C: spinal manipulation (n = 70) 2 × 20 minute/week</td>
<td>HA intensity: 6p Likert scale HA frequency: Functional status: SF-36. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: 5p scale. Adverse events: stated.</td>
</tr>
<tr>
<td>Carasso et al.40 RCT; QS: 3 (items: 1, 7, 10)</td>
<td>Study duration: 12 weeks (baseline, 12 weeks treatment)</td>
<td>TTH; N = 31, 3 dropouts in C; female: 54.8%; age range: 35–70 years</td>
<td>I: AD amitriptyline (30–110 mg) (n = 13) C: AD clomipramine (20–75 mg) (n = 15)</td>
<td>HA improvement: 3p scale. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 1.24 [0.67; 2.3] Adverse events: I: severe sedative effect (90% within first 2 weeks, 40% till end of study). No data available.</td>
</tr>
<tr>
<td>Cerbo et al.41 RCT; QS: 1 (item: 1)</td>
<td>Study duration: 12 weeks (2 weeks baseline, 6 weeks treatment, 4 weeks follow-up)</td>
<td>TTH; N = 52, 13 dropouts; female (%): N/S; age range: 15–69 years</td>
<td>I: AD amitriptyline 10 mg (n = 11) C1: AD amitriptyline 25 mg (n = 14) C2: AD amitriptyline 50 mg (n = 14)</td>
<td>HA intensity: 3p scale HA index Adverse events: N/S</td>
<td>HA improvement: duration × intensity HA intensity 1p Likert scale HA duration, frequency Analgesic use: number Adverse events: stated</td>
</tr>
<tr>
<td>Study</td>
<td>Methods</td>
<td>Participants</td>
<td>Interventions</td>
<td>Outcomes</td>
<td>Results (95% CI)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>De Benedittis and Massei</td>
<td>Study duration: 12 weeks (2 weeks baseline, 6 weeks treatment, 4 weeks follow-up)</td>
<td>Mixed; N = 50, 19 dropouts; female: 74.2%; mean age: 41.9 years. Setting: pain unit, Italy</td>
<td>I: AD 5-hydroxytryptophan 400 mg (L5-HTP) (n = 31) C: placebo (n = 31)</td>
<td>HA intensity: 3p scale HA index: HA density Depression: ZSDS; HDRS. Adverse events: stated.</td>
<td>HA index: I versus C: SMD = 0.08 [-0.42; 0.58] HA density: I versus C: SMD = 0.04 [-0.46; 0.53] Adverse events: I: n = 6 C: n = 7</td>
</tr>
<tr>
<td>Denaro et al.</td>
<td>Study duration: 13 weeks (1 week washout)</td>
<td>TTH (Ad hoc); N = ?, N completed 20; female: 60.0%; age range: 21–48 years</td>
<td>I1: AD mianserin 30 mg (n = 20) I2: VD clonidine 0.150 mg (n = 20) C: placebo (n = 20)</td>
<td>Histamine threshold Overall rating: Global impression: 3p scale. HA intensity: 5p scale; VAS. Depression: BDI Adverse events: stated.</td>
<td>Overall rating: I versus C: RR = 6.0 [1.97; 18.25] Adverse events: I: n = 21 versus C: n = 7</td>
</tr>
<tr>
<td>Fogelholm and Murros</td>
<td>Study duration: 14 weeks (2 × 6 weeks treatment, 2 weeks wash-out)</td>
<td>TTH (HIS); N = 78, 25 dropouts; female: 69.8%; mean age: 43.6</td>
<td>I: AD amitriptyline 75 mg (n = 24) C: placebo (n = 29)</td>
<td>HA duration. Analgesic use. Adverse events: stated.</td>
<td>HA duration: I versus C: SMD = 0.28 [-0.26; 0.82] Analgesic use: I versus C: SMD = 0.21 [-0.33; 0.75] Adverse events: I: n = 15 C: n = 8</td>
</tr>
<tr>
<td>Gobel et al.</td>
<td>Study duration: 6 weeks (6 weeks treatment)</td>
<td>TTH (IHS); N = 41, 5 dropouts; female: 80.5%; mean age: 32.3 years. Setting: research clinic, USA</td>
<td>I: AD amitriptyline 50–75 mg (n = 17) C: cognitive–behavioural therapy (n = 19) 3 × 60 minutes in Weeks 1, 5 and 8 × 15 minutes telephone conversation in Weeks 2 and 6. C: placebo (n = 20)</td>
<td>HA improvement: 4p scale HA index: 10p scale Depression: BDI Anxiety: STPI Analgesic use Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 0.48 [0.15; 1.56]; HA index: I versus C: SMD = −0.58 [-1.25; 0.09] Depression: I versus C: SMD = −0.07 [-0.73; 0.58] Analgesic use: I versus C: SMD = 0.62 [-0.05; 1.29] Adverse events: I: n = 10 (mild 6, moderate 2, substantial 2) versus C: n = 0</td>
</tr>
<tr>
<td>Holroyd et al.</td>
<td>Study duration: 16 weeks (4 weeks baseline, 12 weeks treatment)</td>
<td>TTH; N = 203, 59 dropouts; female: 76.4%; mean age: 37.0 years. Setting: primary practice referrals and advertisements, USA</td>
<td>I1: AD amitriptyline 12.5–50 mg (n = 44) (if 50 mg not tolerated than nortriptyline 25–75 mg) I2: amitriptyline and stress management (n = 40) C1: stress management (n = 34) 3 × 60 minutes C2: placebo (n = 26)</td>
<td>HA improvement: (&gt;50% reduction) HA index: mean ratings HA intensity: 10p scale Analgesic use Adverse events: stated.</td>
<td>HA improvement: I1 versus C2: RR = 1.29 [0.74; 2.27]; I1 versus C1: RR = 1.09 [0.65; 1.82]; I2 versus C2: RR = 2.2 [1.35; 3.57]; I2 versus C1: RR = 1.85 [1.2; 2.85] HA index: reduction I1: 37.7% versus C1: 34.7% versus I2: 64.2% versus C2: 29.2% Adverse events: I: 80% versus C3: 30.0%</td>
</tr>
<tr>
<td>Study</td>
<td>Methods</td>
<td>Participants</td>
<td>Interventions</td>
<td>Outcomes</td>
<td>Results (95% CI)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Langemark et al. et al.</td>
<td>RCT; QS: 5 (items: 1, 4, 5, 7, 8)</td>
<td>Study duration: 6 weeks (43 days treatment). TTH (Ad hoc 1962); N = 114, 32 dropouts; female (%): N/S; age range: 18-69 years. Setting: headache clinic, Denmark</td>
<td>I: AD clomipramine 75–150 mg (n = 26) I2: AD mianserin 30–60 mg (n = 22) C: placebo (n = 34)</td>
<td>HA improvement: (&gt;50% reduction). HA intensity: VAS. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I1 versus C: RR = 1.18 [0.74; 1.88]; I2 versus C: RR = 1.1 [0.68; 1.79]; I1 versus I2: RR = 1.07 [0.67; 1.71] HA intensity: I1 versus C: SMD = 0.51 [0.10; 1.91]; I2 versus C: SMD = 0.57 [0.07; 1.06] I1 versus I2: SMD = 0.05 [-0.48; 0.57]</td>
</tr>
<tr>
<td>Langemark and Olesen</td>
<td>RCT (CO for non-responders); QS: 3 (items: 1, 5, 7)</td>
<td>Study duration: 16 weeks (4 weeks baseline, 16 weeks treatment). TTH (HHS); N = 50, 13 dropouts; female: 60%; age range: 20–70 years. Setting: headache clinic, Denmark</td>
<td>I: AD paroxetine 30 mg (n = 18) C: sulpiride 400 mg (n = 19)</td>
<td>HA improvement: 5p scale. HA intensity: 4p scale. Analgesic use. Adverse events.</td>
<td>HA improvement: I versus C: RR = 0.86 [0.34; 2.19] Analgesic use: stated.</td>
</tr>
<tr>
<td>Manna et al.</td>
<td>RCT; QS: 6 (items: 1, 3, 4, 5, 7, 10)</td>
<td>Study duration: 12 weeks (4 weeks baseline, 8 weeks treatment). TTH (HHS); N = 50, 10 dropouts; female: 62.5%; mean age: 36.3 years.</td>
<td>I: AD fluvoxamine 100 mg (n = 20) C: AD mianserin 60 mg (n = 20) I1: AD mianserin 30 mg (n = 20) I2: VD clonidine 0.15 mg (n = 20) C: placebo (n = 20)</td>
<td>HA intensity: 4p scale Depression: ZSDS; HDRS. Adverse events: stated. HA intensity: VAS. Depression: HDRS. Adverse events: N/S.</td>
<td>No data available. Adverse events: I: n = 9, C: n = 6. No data available.</td>
</tr>
<tr>
<td>Martucci et al.</td>
<td>RCT-CO; QS: 3 (items: 1, 4, 8)</td>
<td>Study duration: 42 weeks (1-week baseline, 13 weeks treatment, 1-week washout, 13 weeks treatment, 1-week washout, 13 weeks treatment). TTH (Ad hoc 1962); N = 7, N completed 20; female: 60.0%; mean age: 34.5 years</td>
<td>I1: AD mianserine 30 mg (n = 20) I2: VD clonidine 0.15 mg (n = 20)</td>
<td>HA improvement: mean weakly HA index. HA index: frequency × severity. Depression: ZSDS. Adverse events: N/S.</td>
<td>HA improvement: I2 52%; I1: 60%; C1: 48%; C6: 18% HA index: I1 versus C6: SMD = 4.14 [3.28; 4.99]; I1 versus C1: SMD = 1.85 [1.25; 2.44]; I2 versus C6: SMD = 4.31 [3.47; 5.14]; I2 versus C1: SMD = 1.83 [1.26; 2.39] I2 versus I1: SMD = 0.21 [-0.69; 0.26] Depression: I1 versus C6: SMD = -0.53 [-1.02; 0.03]; I1 versus C1: SMD = 0.24 [-0.26; 0.74] I2 versus C6: SMD = -0.78 [-1.25; -0.3]; I2 versus C1: SMD = -1.0 [-1.5; -0.5]; I2 versus I1: SMD = -1.10 [-1.61; -0.59]</td>
</tr>
<tr>
<td>Mathew</td>
<td>RCT; QS: 2 (items: 1, 8)</td>
<td>Study duration: 32 weeks (4 weeks baseline, 28 weeks treatment). TTH/mixed; N = 375, 86 dropouts; female: 95.5%; mean age: 40.2 years.</td>
<td>I: AD amitriptyline 25–75 mg (n = 31) I2: VD propranolol 60–160 mg (n = 38)</td>
<td>HA improvement: mean weakly HA index. HA index: frequency × severity. Depression: ZSDS. Adverse events: N/S.</td>
<td>HA improvement: mean weakly HA index. HA index: frequency × severity. Depression: ZSDS. Adverse events: N/S.</td>
</tr>
<tr>
<td>Mitsikostas et al.</td>
<td>RCT; QS: 3 (items: 1, 4, 10)</td>
<td>Study duration: 16 weeks (4 weeks baseline, 12 weeks treatment). TTH (IHS); N = 58, 9 dropouts; female: 62.1%; mean age: 42.8 years. Setting: headache service, Greece</td>
<td>I: AD amitriptyline 50 mg (n = 27) C: 5-HTA buspirone 30 mg (n = 22)</td>
<td>HA improvement: (≥50% reduction in HA pain index). HA intensity: 11p Likert scale Depression: HDRS. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 0.87 [0.54; 1.4] Adverse events: I: n = 14; C: n = 21.</td>
</tr>
</tbody>
</table>
### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Results (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mørland et al.</td>
<td>Study duration: 20 weeks (9 weeks treatment, 2 weeks wash-out, 9 weeks treatment).</td>
<td>Mixed; N = 23, 9 dropouts; female: 91.3% Setting: Italy</td>
<td>I: AD doxepin 100 mg (n = 14); C: placebo (n = 14)</td>
<td>HA intensity: 3p scale. HA index: frequency × severity. Analgesic use Adverse events: stated.</td>
<td>Only point estimates presented. Adverse events: I: n = 18; C: n = 5 Depression: 8 weeks I versus C: SMD = 0.22 [-0.42; 0.86]; follow-up I versus C: SMD = 0.07 [-0.57; 0.77] Adverse events: I: n = 17; C: n = 19</td>
</tr>
<tr>
<td>Nappi et al.</td>
<td>Study duration: 20 weeks (4 weeks baseline, 12 weeks treatment, 4 weeks follow-up)</td>
<td>TTH (IHS); N = 7, N completed 38; female: 78.9%; mean age: 38 years. Setting: Italy</td>
<td>C: AD amitriptyline 50 mg (n = 19) I: AD ritanserin 10 mg (n = 23); 9 dropouts; female: 91.3% Setting: Italy</td>
<td>HA intensity: 4p scale. HA index. Depression: DRS. Adverse events: stated.</td>
<td>No data available.</td>
</tr>
<tr>
<td>Oguzhanoglu et al.</td>
<td>Study duration: 12 weeks (2 weeks baseline, 6 weeks treatment, 8 weeks follow-up)</td>
<td>TTH (IHS); CTTH: N = 14, 1 dropout; ETTH: N = 21, 2 dropouts; female: 90.6%; mean age: 38 years. Setting: Germany, Austria and Switzerland</td>
<td>I: AD amitriptyline 25–75 mg (n = 40). I2: AD amitriptylineoxide (n = 44)</td>
<td>HA improvement: (&gt;50% reduction). HA intensity: VAS. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I1 versus C: RR = 1.02 [0.45; 1.95]; I2 versus C: RR = 1.52 [0.85; 2.74]; I1 versus I2: RR = 0.67 [0.38; 1.19] Depression: I1 versus C: SMD = 0.34 [–0.15; 0.83] Analgesic use: I versus C: SMD = 0.09 [0.09; 1.08] Adverse events: I: n = 6; C: n = 3</td>
</tr>
<tr>
<td>Pfaffenrath et al.</td>
<td>Study duration: 12 weeks (4 weeks baseline, 12 weeks treatment, 8 weeks follow-up)</td>
<td>TTH (IHS); N = 197, 48 dropouts; female: 55.8%. Setting: Germany, Austria and Switzerland</td>
<td>I: AD amitriptyline 25–75 mg (n = 40). I2: AD amitriptylineoxide (n = 49). C: placebo (n = 51)</td>
<td>HA improvement: (&gt;50% reduction). HA intensity: VAS. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: SMD = 1.06 (0.61; 1.51) Depression I versus C: SMD = 0.34 [–0.15; 0.83] Analgesic use: I versus C: SMD = 0.09 [0.09; 1.08] Adverse events: I: n = 6; C: n = 3</td>
</tr>
<tr>
<td>Rampello et al.</td>
<td>Study duration: 16 weeks.</td>
<td>CTTH (IHS), comorbidity of depression. N = 88; female 62.5%; mean age 39.07 years. Setting: not stated, Italy</td>
<td>I: AD amitriptyline 25–50 mg. (N = 44) First week 25 mg. Week 2–16 75 mg C: AD citalopram 20 mg (N = 44).</td>
<td>Days with TTH per month. Depression (HDRS) Adverse events: stated.</td>
<td>HA improvement: I versus C: SMD = 2.08 [1.39; 2.77] Depression: I versus C: SMD = 0.89 [0.38; 1.4] HA frequency: I versus C: SMD = 0.34 [–0.15; 0.83] Analgesic use: I versus C: SMD = 0.59 [0.09; 1.08] Adverse events: I: n = 6; C: n = 3</td>
</tr>
<tr>
<td>Ribeiro et al.</td>
<td>Study duration: 12 weeks (2 weeks baseline, 8 weeks treatment, 2 weeks follow-up)</td>
<td>TTH (IHS); N = 78, 13 dropouts; female: 87.7%; mean age: 39.8 years</td>
<td>I: AD l-5-hydroxytryptophan (5-HTP) 300 mg (n = 34) C: placebo (n = 31)</td>
<td>HA improvement: 6p scale. HA severity: 4p scale. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: SMD = 2.74 [1.41; 4.07] Depression: I versus C: SMD = 0.34 [–0.15; 0.83] Analgesic use: I versus C: SMD = 0.09 [0.09; 1.08] Adverse events: I: n = 6; C: n = 3</td>
</tr>
<tr>
<td>Saper et al.</td>
<td>Study duration: 16 weeks (4 weeks baseline, 12 weeks treatment)</td>
<td>TTH; N = 64, 10 dropouts; female: 81.5%; mean age: 35.5. Setting: headache centres, USA</td>
<td>I: AD fluoxetine 20–40 mg (n = 30) C: placebo (n = 24)</td>
<td>HA intensity: 5p scale; VAS. HA free days. Depression: BDI. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 2.24 [0.94; 5.34] HA improvement: I versus C: RR = 4.5 [1.08; 18.77] HA index: I versus C: SMD = 1.66 [1.01; 2.3] Analgesic use: I versus C: SMD = 2.08 [1.39; 2.77] Adverse events: I: n = 6; C: n = 4.</td>
</tr>
<tr>
<td>Singh and Misra</td>
<td>Study duration: 10 weeks (2 weeks baseline, 4 weeks treatment, 4 weeks follow-up)</td>
<td>TTH (IHS); N = 60, 10 dropouts; mean age: 28.7 years</td>
<td>I: AD sertraline 100 mg (n = 25) C: placebo (n = 25)</td>
<td>HA improvement: 4p scale. HA index: frequency × severity × duration. Depression: HDRS. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 4.5 [1.08; 18.77] HA index: I versus C: SMD = 1.66 [1.01; 2.3] Analgesic use: I versus C: SMD = 2.08 [1.39; 2.77] Adverse events: I: n = 6; C: n = 4.</td>
</tr>
<tr>
<td>Study</td>
<td>Methods</td>
<td>Participants</td>
<td>Interventions</td>
<td>Outcomes</td>
<td>Results (95% CI)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sjaastad 36 RCT-CO QS: 2</td>
<td>Study duration: 9 weeks (4 weeks treatment, 1 week wash-out, 4 weeks treatment)</td>
<td>TTH; N = ?, N completed 16; female: 81.3%; mean age: 35.5 years</td>
<td>I: AD fluoxetine 400 mg (n = 16); C: placebo (n = 16).</td>
<td>HA intensity: 3p scale. Depression: ZSDS. Analgesic use. Adverse events: stated. Total tenderness. Score: 4p Likert scale. Self-evaluating anxiety scale. Self-evaluating depression scale. Adverse events: stated.</td>
<td>HA days: I 41.4% versus C 91.6%. Adverse events: I: n = 12; C: n = ? Only point estimates presented. Adverse events: I: N = 2; C: N = 0</td>
</tr>
<tr>
<td>De Tommaso et al. 54 RCT; QS: 2 (items: 1, 4)</td>
<td>Study duration: 9 weeks</td>
<td>CTTH (IHS). N = 18; female 55.6%; mean age 37.3 years. Setting: Headache Centre of the Neurology Clinic of Bari University, Italy</td>
<td>I: AD amitriptyline 10 mg. C: intra-oral appliance of prosthesis.</td>
<td>HA intensity: 3p scale. Depression: ZSDS. Analgesic use. Adverse events: stated.</td>
<td>Only point estimates presented.</td>
</tr>
<tr>
<td>Walker et al. 55 RCT; QS: 2 (items: 1, 4)</td>
<td>Study duration: 12 weeks (2 weeks baseline, 6 weeks treatment, 4 weeks follow-up)</td>
<td>TTH (IHS). N = 37, 12 dropouts; female 81.1%; mean age 35 years. Setting: neurology outpatient clinics</td>
<td>I: AD fluoxetine 20 mg (n = 12); C: AD desipramine 75 mg (n = 13)</td>
<td>HA improvement: 3p scale. HA intensity: VAS. Depression: HADS; MADRS. Functional status: SF-36. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 1.63 [0.83; 1.18]. Adverse events: I: n = 3; C: n = 8.</td>
</tr>
<tr>
<td>Worz and Scherhag 56 RCT; QS: 4 (items: 1, 3, 4, 8)</td>
<td>Study duration: 8 weeks (2 weeks baseline, 6 weeks treatment)</td>
<td>TTH; N = 54, 14 dropouts; female 57.5%; mean age 42.5 years. Setting: pain clinics and pain practices.</td>
<td>I: AD amitriptyline 25-50 mg (n = 21); C: AD doxepin 25-50 mg (n = 19).</td>
<td>HA improvement: (reduction VAS ≥50%). HA intensity: 4p scale. HA days. Depression: von Zerssen. Analgesic use. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 0.99 [0.58; 1.68]. HA intensity: I versus C: SMD = 0.17 [-0.45; 0.79]. HA days: I versus C: SMD = 0.36 [-0.27; 0.98]. Depression: I versus C: SMD = 0.11 [-0.51; 0.73]. Analgesic use: I versus C: SMD = 0.12 [-0.5; 0.74]. Adverse events: I: n = 13; C: n = 12. Clinical global improvement I versus C: SMD = 0.24(-0.28; 0.76) Days with HA I versus C: SMD = 0.26 (-0.26; 0.78) Index I versus C: SMD = 0.47 (-0.05; 0.99). Adverse events: I: N = 10; C: N = 5.</td>
</tr>
<tr>
<td>Zissis et al. 57 RCT; QS: 10</td>
<td>Study duration: 16 weeks (4 week screening, 12 weeks treatment)</td>
<td>TTH (all subtypes IHS). N = 60; female 81.7%; mean age 40.8 years; 20 dropouts. Setting: not stated, Greece</td>
<td>I: AD venlafaxine XR 75-150 mg. C: placebo.</td>
<td>Clinical global improvement. Days with HA. Hours of HA. Total headache intensity index. Adverse events: stated.</td>
<td>Clinical global improvement I versus C: SMD = 0.24(-0.28; 0.76) Days with HA I versus C: SMD = 0.26 (-0.26; 0.78) Index I versus C: SMD = 0.47 (-0.05; 0.99). Adverse events: I: N = 10; C: N = 5.</td>
</tr>
</tbody>
</table>

AD, antidepressant; BDI, Beck depression inventory form; C, control; CO, cross over; CTTH, chronic tension type headache; ETTH, episodic tension type headache; HA, headache; HADS, Hospital Anxiety and Depression Scale; HDRS, Hamilton depression rating scale; I1, intervention1; I2, intervention2; MADRS, Montgomery and Asberg Depression Rating scale; NS, not stated; NRS, numerical rating scale; p, point; QS, quality score; STPI, State Trait Personality Inventory; TTH, tension type headache; VAS, visual analogue scale; VD, vasodilator agents; ZSDS, Zung Self-rating Depressive Scale.
<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Participants</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Results [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bram[^60] RCT; QS: 5 (items: 1, 2, 6, 7, 10)</td>
<td>Study duration: 2–4 weeks (baseline, 2–4 weeks treatment)</td>
<td>TTH; N = ?, N completed 80.</td>
<td>I: MR MH532 (gamaquil) (n = 68). C: placebo (n = 12).</td>
<td>HA improvement: 2p scale Pain intensity. Number of HA days per month. Duration of HA. Adverse events: stated.</td>
<td>HA improvement: I versus C: RR = 0.68 [0.39; 1.19]</td>
</tr>
<tr>
<td>Bettucci et al[^61] RCT; QS: 2 (items: 1, 8)</td>
<td>Study duration: 17 weeks (4 weeks observation period, 13 weeks treatment).</td>
<td>CTTH (IHS). N = 18; female 72.2%; mean age 35.2. Setting: Headache Centre; Department of Neurology, Italy</td>
<td>I: MR tizanidine 4 mg + AD amitriptyline 20 mg. (N = 9). Tizanidine 4 mg/day in the first 21 days. C: AD amitriptyline 20 mg (N = 9).</td>
<td>HA improvement: 2p scale Pain intensity. Number of HA days per month. Duration of HA. Adverse events: stated.</td>
<td>Pain I1 versus I2: SMD = 0.0 (-0.92; 0.92). HA days/4 weeks I1 versus I2: SMD = -0.19 (-1.11; 0.74). HA duration: I1 versus I2: SMD = 0.34 (-0.60; 1.27). Adverse events: I: N = 3; C: N = 4.</td>
</tr>
<tr>
<td>Medvedeva et al[^67] RCT; QS: 3 (items: 1, 4, 8)</td>
<td>Study duration: not stated</td>
<td>CTTH. N = 126; female 65.1%; age range 18–66 year. Setting: Russian Scientific Centre for Surgery, Russian Academy of Medical Sciences, Moscow.</td>
<td>I: mexidol 200 mg (4 ml) + standard treatment (N = 64). C: MR tizanidine (8 mg/day), vinpocetin (15 mg/day), fluoxetine (20 mg/day), sessions manual acupressure.</td>
<td>HA intensity I versus C: SMD = -2.20 (-2.63; -1.75)</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Methods</td>
<td>Participants</td>
<td>Interventions</td>
<td>Outcomes</td>
<td>Results [95% CI]</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Murros et al. (^5)</td>
<td>RCT; QS: 6</td>
<td>TTH (IHS); N = 185, 15 dropouts; female: 74.6%; mean age: 44.0 years</td>
<td>I1: MR tizanidine 6 mg (n = 56). I2: MR tizanidine 12 mg (n = 49). C: placebo (n = 55).</td>
<td>HA intensity: VAS. Depression: BDI. Analgesic use. Adverse events: stated.</td>
<td>HA intensity:</td>
</tr>
<tr>
<td></td>
<td>(items: 1, 2, 3, 4, 8, 10)</td>
<td>Study duration: 8 weeks (2 weeks baseline, 6 weeks treatment)</td>
<td></td>
<td></td>
<td>I1 versus C: SMD = 0.04 [-0.34; 0.41]; I2 versus C: SMD = -0.01 [-0.4; 0.37] HA free days: I1 versus C: SMD = 0.17 [-0.2; 0.55]; I2 versus C: SMD = -0.15 [-0.54; 0.23] Adverse events: I1/2: n = 27; C: n = 9</td>
</tr>
<tr>
<td>Paiva et al. (^6)</td>
<td>RCT; QS: 6</td>
<td>TTH (Ad hoc 1962); N = 36, 4 dropouts; female: 75.0%; mean age: 37.5 years. Setting: Headache Clinic, Portugal</td>
<td>I: BD diazepam (n = 8). C1: BD diazepam placebo (n = 8). C2: biofeedback true (n = 8). C3: biofeedback false (n = 8) 3 x 30 minutes/week.</td>
<td>HA intensity: frequency. Adverse events: N/S.</td>
<td>HA intensity:</td>
</tr>
<tr>
<td></td>
<td>(items: 1, 5, 6, 7, 8, 10)</td>
<td>Study duration: 12 weeks (4 weeks baseline, 4 weeks treatment, 4 weeks follow-up)</td>
<td></td>
<td></td>
<td>I versus C1: SMD = 0.77 [-0.25; 1.78]; I versus C2: SMD = 0.22 [-0.77; 1.2]; follow-up I versus C1: SMD = 0.33 [-0.66; 1.32] I versus C2: SMD = -1.64 [-0.46; -2.82] HA frequency: I versus C1: SMD = 0.66 [-0.34; 1.67]; I versus C2: SMD = -0.39 [-1.38; 0.61] follow-up I versus C1: SMD = -0.53 [-0.37; 1.64]; I versus C2: SMD = -1.19 [-0.1; -2.27]</td>
</tr>
<tr>
<td>Shukla et al. (^7)</td>
<td>RCT-CO; QS: 6</td>
<td>TTH (IHS); N = 32, 4 dropouts; female: 50.0%; mean age: 22.8 years. Setting: Headache Clinic, Lucknow</td>
<td>I: VD nifedipine 5 mg (n = 28). C: placebo (n = 28).</td>
<td>Overall response: HA index severity x duration. HA intensity: 4p scale. Analgesic use. Adverse events: stated.</td>
<td>Overall response:</td>
</tr>
<tr>
<td></td>
<td>(items: 1, 4, 5, 7, 8, 10)</td>
<td>Study duration: 10 weeks (2 weeks baseline, 4 weeks treatment, 4 weeks treatment)</td>
<td></td>
<td></td>
<td>I versus C: RR = 0.67 [0.32; 1.38] Analytical use: 4 weeks I versus C: SMD = 0.05 [-0.48; 0.57] Adverse events: I: n = 4; C: n = 1</td>
</tr>
<tr>
<td>Shukla et al. (^8)</td>
<td>RCT-CO; QS: 6</td>
<td>TTH (IHS); N = 62, 14 dropouts; female: 71.0%; mean age: 26.9 years. Setting: Headache Clinic, Lucknow</td>
<td>I: BD alprazolam 0.75 mg (n = 48). C: placebo (n = 48).</td>
<td>HA improvement: &gt;50% reduction in HA frequency/week. Overall response: 3p scale. Depression: HDRS. Analgesic use: Adverse events: stated.</td>
<td>HA improvement:</td>
</tr>
<tr>
<td></td>
<td>(items: 1, 4, 5, 6, 7, 8)</td>
<td>Study duration: 16 weeks (2 weeks baseline, 4 weeks treatment, 2 weeks washout, 4 weeks treatment, 4 weeks follow-up)</td>
<td></td>
<td></td>
<td>I versus C: RR = 2.0 [1.05; 3.81] Overall response: I versus C: RR = 2.0 [0.65; 6.2] Analytical use: I versus C: SMD = 0.16 [-0.24; 0.56] Adverse events: I: n = 8; C: n = 8</td>
</tr>
<tr>
<td>Worz et al. (^9),(^10)</td>
<td>RCT; QS: 3</td>
<td>TTH (IHS); N = 111, 58 dropouts; female: 60.4%; mean age: 42.1 years. Setting: neurological centres</td>
<td>I: analgetica flupirtine (n = 30). C: placebo (n = 23).</td>
<td>Overall response: 4p scale. HA intensity: 5p scale. Depression: von Zerssen. Analgesic use. Adverse events: stated.</td>
<td>HA intensity:</td>
</tr>
<tr>
<td></td>
<td>(items: 1, 4, 8)</td>
<td>Study duration: 8 weeks (2 weeks baseline, 6 weeks treatment)</td>
<td></td>
<td></td>
<td>I versus C: SMD = -0.64 [-1.2; -0.09] Depression: I versus C: SMD = -0.27 [-0.81; 0.28] Adverse events: I: n = 10; C: n = 9</td>
</tr>
</tbody>
</table>

AD, antidepressant; BDI, Beck depression Inventory form; C, control; CO, cross over; EMG, electromyography; HA, headache; HDRS, Hamilton depression rating scale; I1, intervention1; I2, intervention2; MADRS, Montgomery and Asberg Depression Rating scale; MR, muscle relaxant; p, point; QS, quality score; SMD, standardized mean difference; TTH, tension type headache; VD, vasodilator agents.
one or more types of antidepressants with placebo,14,17,21,25–28,32,33,36,44,46,51–53 three studies compared muscle relaxants with placebo,29,58,60 three studies compared benzodiazepine with placebo,27,32,35,59 four studies compared vasodilator agents with placebo 27,32,35,46 and one study compared flupenthixol with placebo.31,35,59 Most interventions varied with respect to types of drugs, e.g. various types of antidepressant (e.g. amitriptyline and citalopram), dose and frequency of intake.

Outcome measures. No two studies defined their outcomes the same way, but all reported some measure of headache (frequency, intensity, duration or index), analgesic use (n = 21), depression (n = 21) or adverse events (n = 29). All studies used diaries to assess outcomes. Using this diary, amongst others, the headache measures frequency, intensity and duration were scored on a Likert-scale. In several studies, clinical improvement was calculated, meaning that the patients’ headache declines by 50% or more.

The methodological score [quality score (QS); with positive items in parentheses] is presented in the tables. The inter-observer reliability of the risk of bias assessment was high (kappa statistic 0.83). There was an initial disagreement between the two authors in 8.3% of the criteria. After the consensus meeting, no disagreement persisted. The median QS was 4.2 (range 1–10). Using a cut-off point of 6 of 10 criteria, 15 of 44 studies (34.1%) were considered to be with low risk of bias.14,17,25,28,30,34,35,39,45,51–53,57–59 The most prevalent methodological shortcomings were a concealed randomization method (unclear 82% and negative 2%), blinding of the care provider (unclear 75% and negative 9%) and intention-to-treat analysis (unclear 25% and negative 66%).

Effectiveness of prophylactic drugs

Antidepressants. Versus placebo. Eighteen studies compared one or more types of antidepressants with placebo, including eight studies with low risk of bias.14,17,25,28,51–53,57 Five studies did not provide adequate data on outcome measures.25,27,32,33,36 Some studies have two or more comparisons with placebo. Overall, most comparisons (81.3%) did not show statistical significant effects of antidepressant versus placebo on headache and no differences between studies with low or high risk of bias. In the four studies with low risk of bias, adequate power and patient selection,14,17,51,53 two studies21,53 showed significant differences in favour of antidepressants concerning headache improvement, but not on headache frequency, whereas two other studies14,17 showed no significant differences between the groups concerning headache duration or intensity.

Concerning amitriptyline separately, four of five studies (of which two with low risk of bias) did not show any difference. Several studies reported data on adverse events, which were often minor and comparable in the antidepressant and placebo groups (see Table 1).

Versus antidepressants. Twelve studies compared different types or dosages of antidepressants. Amitriptyline was compared with citalopram,14,50 clomipramine,40 amitriptyline,41 ritanserin,48 fluoxetine,49 amitriptylinoxide,21,22 doxepin56 or sertraline.38 Eight studies provided data on headache frequency, intensity, duration or index and all but two38,50 found no significant differences. The only study with low risk of bias and adequate power found no significant differences in headache intensity or duration.14

Versus other treatments. Amitriptyline was compared with other medications: ibuprofen,27 clonidine,27,32 propranolol,46 sulpiride44 or 5-HTA buspirone47; all showing no statistically significant or clinically relevant differences. Amitriptyline was also compared with spinal manipulation,39 cognitive behavioural therapy,42 stress management,43 biofeedback46 or intra-oral orthosis.54 No significant differences were found between amitriptyline and cognitive behaviourual therapy or stress management. The headache index was significantly reduced in the amitriptyline group when compared with biofeedback.46 Only one study at low risk of bias with adequate power showed no significant differences between amitriptyline and spinal manipulation.39

Overall, the use of antidepressants was not clearly more effective than placebo or other drugs. In the subgroup of larger and better studies, there was conflicting evidence about the efficacy of antidepressants compared with placebo, but no differences were found between antidepressants and other drugs or amitriptyline and spinal manipulation.

Muscle relaxants. Versus placebo. Three studies compared tizanidine with placebo. In two studies with three comparisons, of which one study with low risk of bias and adequate power (two comparisons), we found no significant differences in headache intensity.58 One study (n = 18) evaluated the additional effectiveness of tizanidine on an antidepressant but found no significant differences between the two groups.61 Only one study mentioned an adverse event (tiredness), which was comparable between intervention and placebo groups.58

Versus muscle relaxants. Two different doses of tizanidine (6 mg and 12 mg) were compared in one study with low risk of bias and adequate power and reported no significant difference.58

We concluded that the use of muscle relaxants was not more effective than placebo in the prevention of TTH.
Benzodiazepines. Versus placebo. Three studies compared a benzodiazepine with placebo, of which two provided data. One study with low risk of bias and adequate power reported that alprazolam had significant headache improvement, but the difference in overall response or analgesic use was not statistically significant. The adverse events were only mild and comparable between alprazolam and placebo.

Versus behavioural treatment. One study with a very small sample size compared diazepam with biofeedback and found no significant differences post-treatment but found significant differences in favour of biofeedback at 1-month follow-up.

The ‘levels of evidence’ approach indicated that there was conflicting evidence about the effectiveness of benzodiazepines compared with placebo.

Vasodilator agents. Versus placebo. Four studies compared vasodilator agents (clonidine, propranolol and nifedipine) with placebo, two of which provided adequate data. One study with low risk of bias and adequate power found no significant differences between nifedipine and placebo in headache complaints and analgesic use. Another study not only reported a significantly greater headache improvement but also a significantly higher depression score with propranolol compared with placebo. Adverse events were reported once and were only minor and comparable between groups.

Versus antidepressants. Vasodilator agents were compared with antidepressants in three studies. Just one study reported sufficient data and found no significant difference in headache improvement, while depression score was significantly higher with propranolol compared with amitriptyline.

Versus behavioural treatment. One study compared propranolol with biofeedback. After treatment, the depression score was significantly higher, while there was a significantly better headache index in the propranolol group compared with biofeedback.

There was conflicting evidence about the effectiveness of vasodilator agents compared with placebo. There was limited evidence concerning the negative effects of propranolol on depression in TTH patients, when compared with placebo or biofeedback.

5-HT receptor agonist. No studies compared a 5-HT receptor agonist with placebo. Buspirone was compared with amitriptyline in one study showing no significant differences in headache improvement.

There was no evidence concerning the effectiveness of 5-HT receptor agonists for TTH.

Other drugs. Versus placebo. One study compared flupenthixol with placebo, but no data were available to calculate effect estimates.

Versus other medication. One study compared tiapride with antidepressants + anxiolitica and analgetica and found significantly more people recovered in the group treated with tiapride compared with both other groups. One other study compared flupenthixol with diazepam (benzodiazepine) but did not provide data.

There was limited evidence that tiapride is more effective when compared with analgesics or antidepressants + anxiolitica.

Discussion

Effectiveness of prophylactic drugs

Overall, the efficacy or effectiveness of antidepressants was evaluated in 31 studies and did not show to be more effective on headache intensity, frequency, duration or improvement and analgesic use when compared with placebo or other medication. No clear differences between different types of antidepressants were found. In the subgroup of larger and better studies, there was conflicting evidence about the efficacy of antidepressants compared with placebo, but we found no differences between antidepressants and other drugs or between amitriptyline and spinal manipulation.

Other prophylactic drugs were evaluated less frequently. Muscle relaxants were also not more effective than placebo in the prevention of TTH. At present, we found conflicting evidence concerning the efficacy of benzodiazepines and vasodilator agents when compared with placebo. There was one study with high risk of bias that found that propranolol (vasodilator agents) resulted in higher depression scores in TTH patients, when compared with placebo or biofeedback. Lastly, we found no evidence (no studies found) concerning the effectiveness of 5-HT receptor agonist compared with placebo for TTH.

Strengths and weaknesses

Apart from two reviews on antidepressants, we are unaware of any prior systematic reviews or meta-analyses that have assessed the efficacy and tolerability of prophylactic drugs for chronic TTH in adults. We conducted the review procedures according to the high Cochrane standards and demonstrated a high degree of agreement in our eligibility decisions and risk of bias assessments. Our review succeeded in identifying a relatively large number of relevant trials.

Although systematic reviews offer the least biased method of summarizing research literature, our review should be considered with the following limitations in mind. We decided not to contact the authors for additional information because the majority of the trials included in this review were published before 1995 and authors or data would probably be hard to find. This might have influenced the risk of bias assessment,
resulting in misclassification of trials according to their methodological rigor.

Most studies were small; the smallest included only eight subjects in each study arm. This meant that most studies did not have enough power to detect statistical significant differences. When studies are underpowered only a meta-analytic approach could overcome this problem. Because of heterogeneity, we were unable to perform statistical pooling, but sensitivity analysis including only the studies with adequate power, patient selection and low risk of bias did not clearly change the results. When we considered a RR >2.0 or <0.5 and a SMD >0.5 as a clinically relevant result, some drugs [maprotiline (antidepressant) and alprazolam (benzodiazepine)] reached clinically relevance each in one study. It might be worthwhile to evaluate these drugs in trials with larger sample sizes. Apart from the small sample sizes, most studies had short medication periods and follow-up, probably too short to evaluate actual prophylaxis and the possible risk of dependence.

Also, most studies measured ‘headache improvement’ as the main outcome measure. These studies defined that only people with >50% improvement were considered clinically improved, which may be a large improvement. The Philadelphia panel advises cut-off scores for clinically relevant differences in musculoskeletal diseases of 15–20% improvement. For clinicians, headache improvement is a clinically relevant outcome measure but lower cut-off points for recovered and not recovered might be considered.

Furthermore, most studies used a headache diary each using different Likert-scales (varying from 3- to 11-point), which variously combined headache frequency, severity or duration into a headache index. None of these headache diaries or indices has been validated.

Adverse events were frequently assessed. Most studies described the adverse events as mild and save, but the simple description of the kind and number of adverse events often gives insufficient insight into the severity and impact of the events reported. Most of the drugs have only been evaluated in one or two studies, which may limit the generalizability of the findings. Also, comparing drugs with non-pharmacological treatment options e.g. behavioural treatment options or spinal manipulation was not evaluated properly.

For primary headaches, such as TTHs, there are no biological markers and therefore their diagnosis is made based on the diagnostic criteria of the IHS, which were updated in 2004. Nowadays, this diagnosis can be made with relatively high precision. However, 26 studies included in this review used the criteria of the IHS to classify chronic TTH, five studies used the Ad Hoc criteria, leaving 13 studies not using predefined criteria for their selection of the study population. This might raise problems because it remains unclear whether all included patients actually suffered from TTH; and therefore, might influence the outcome of the studies. For future trials, it is important that authors adhere to predefined diagnostic criteria.

Finally, this review shows that many RCTs on the efficacy of prophylactic drugs in TTH have methodological shortcomings. Most authors failed to explicitly specify the method or person responsible for the treatment allocation and the blinding procedure. In many studies, authors stated that the trial had a double-blind procedure. However, we scored the criteria blinding of outcome assessor, blinding of care provider and blinding of patient ‘unclear’, when the blinding procedure was not explicitly reported (i.e. identical looking tablets). Only a few studies performed an intention-to-treat analysis, meaning that most analyses were restricted to complete cases only, indicating a possible overestimation of the effect estimate. Favourably, studies should be performed and reported according to the Consolidated Standards of Reporting of Trials (CONSORT) statement to improve the quality of trials reports.

Recommendations
For the clinician, no strong recommendations can be made concerning the prescription of effective prophylactic drugs in TTH. At this moment, there is no clear evidence available to support (or refute) the use of any prophylactic drugs. This means that clinicians are not guided in their treatment decisions in an evidence based way and should rely on their clinical expertise.

Antidepressants are the most widely prescribed medication, and the effectiveness is widely evaluated for different indications apart from depression. Even in the treatment of depression, its effectiveness is disputed. We found that antidepressants are evaluated in 31 studies and did not show to be clearly more effective on headache improvement and analgesic use when compared with placebo or other medication. The difference in effect between prophylactic drugs and non-pharmacological treatment options is only evaluated in four studies. The non-pharmacological treatments were biofeedback, stress management, spinal manipulation or cognitive behavioural therapy, but not exercise treatment for instance. This needs to be evaluated further.

Further research on prophylactic treatment is of paramount importance. Recommendations for research indicate that there is a clear need for large rigorous trials evaluating preferably the most promising treatment options (with clinically relevant effect estimates) found in this review. A comparison between drugs and frequently prescribed non-pharmacological treatments should be considered. Attention should be paid to the use of the IHS diagnostic criteria for patient selection and to the CONSORT statement when reporting the results.
In conclusion, the evidence regarding the efficacy or effectiveness of prophylactic drugs for TTH showed that the currently available drugs were not clearly better than placebo.

Acknowledgements

Marjolein van Heest, Derek van der Have, two physiotherapy students who helped updating the latest version of the manuscript. Bart Snoeren was the pharmacologist who helped with the classification of medication. Authors contribution: APV and BWK designed the study. LD, APV and MYB performed study selection and risk of bias assessment. LD and APV performed the analysis, MYB and JP were content experts. LD and APV wrote the manuscript and all authors contributed to the final text and approved it.

Declaration

Funding: Netherlands Organization for Health Research and Development (ZONMw) (940-31-067). Ethical approval: None. Conflict of interest: None.

References

Prophylactic drugs in headache


