Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? - a meta-analysis of randomised controlled trials

Weiya Zhang, Adrian Jones, and Michael Doherty

DISCLAIMER
The initial version of ARD Online First articles are papers in manuscript form that have been accepted and published in ARD Online but they have not been copy edited and not yet appeared in a printed issue of the journal. Copy editing may lead to differences between the Online First version and the final version including in the title; there may also be differences in the quality of the graphics. Edited, typeset versions of the articles may be published as they become available before final print publication.

Should you wish to comment on this article please do so via our eLetter facility on ARD Online (http://ard.bmjjournals.com/cgi/eletter-submit/ard.2003.018531v1)

DATE OF PUBLICATION
ARD Online First articles are citable and establish publication priority. The publication date of an Online First article appears at the top of this page followed by the article's unique Digital Object Identifier (DOI). These articles are considered published and metadata has been deposited with PubMed/Medline.

HOW TO CITE THIS ARTICLE

Online First articles are posted weekly at http://ard.bmjjournals.com/onlinefirst.shtml
Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? – a meta-analysis of randomised controlled trials

Weiya Zhang a, Adrian Jones b, Michael Doherty a

a. Academic Rheumatology, University of Nottingham, Clinical Sciences Building, City Hospital, Nottingham NG5 1PB
b. Rheumatology Unit, Nottingham City Hospital, Nottingham NG5 1PB.

Correspondence to Dr Weiya Zhang
Tel: +44 (0)115 8404 733
Fax: +44 (0)115 8404 732
Email: weiya.zhang@nottingham.ac.uk
Abstract

**Objective:** To assess the best available evidence for efficacy of paracetamol (acetaminophen) in the treatment of osteoarthritis (OA).

**Design:** Systematic review and meta-analysis of randomized controlled trials (RCTs)

**Data sources:** Medline, EMBASE, Scientific Citation Index, CINAHL, Cochrane Library and conference abstracts from the British Society for Rheumatology, the European League Against Rheumatism, the American College of Rheumatology and the Osteoarthritis Research Society International.

**Subjects:** 10 RCTs including 1712 patients with either symptomatic OA of the knee (6 trials), hip/knee (3 trials) or multiple joints (1 trial).

**Main outcome measures:** (i) Effect size (ES) for pain, stiffness and functional scores from baseline to endpoint; (ii) rate ratio and number needed to treat (NNT) for clinical response rate and patient preference to treatment.

**Results:** The results showed paracetamol was effective in relieving pain due to OA (ES=0.21, 95% confidence interval (CI) 0.02, 0.41). NSAIDs were superior to paracetamol for pain relief (ES=0.20, 95%CI 0.10, 0.30). Clinical response rate was higher with NSAIDs than with paracetamol (RR=1.24, 95% CI 1.08, 1.41), and the number of patients who preferred NSAIDs was more than twice that preferring paracetamol (RR=2.46, 95% CI 1.51, 4.12). However, NSAIDs were associated with more frequent gastrointestinal discomfort than paracetamol (RR=1.35, 95% CI 1.05, 1.75).

**Conclusion:** Paracetamol is an effective agent for pain relief due to OA. Although safer, it is inferior in efficacy to NSAIDs. However, these data suggest that for safety reasons paracetamol should be the first line therapy with NSAIDs reserved for those who failed to respond.

**Key words:** paracetamol, osteoarthritis, meta-analysis
Introduction

Paracetamol (acetaminophen) has been used as an analgesic for over 120 years. Although the exact site and mechanism of action is not clearly defined, it appears to produce analgesia by elevation of the pain threshold predominantly through a central rather than peripheral mechanism. It has a narrow therapeutic window but in recommended doses (1g three to four times daily) is very safe. Its favourable efficacy, excellent safety, widespread availability and low cost together appear to justify its position as the world market leader for analgesics.

Osteoarthritis (OA) is by far the commonest joint disease. OA of the knee, the principal large joint to be targeted by OA, results in disabling symptoms in an estimated 10% of UK people older than 55 years, a quarter of whom are severely disabled. The risk of disability attributable to knee OA alone is as great as that due to cardiac disease and greater than that due to any other medical disorder in the elderly. Current European evidence-based recommendations for management of knee OA devised by the European League against Rheumatism (EULAR) recommend paracetamol as “the oral analgesic to try first and, if successful, the preferred long-term oral analgesic”. However, only one small randomised placebo controlled trial of paracetamol in knee OA was found within the period of the EULAR literature review to support this statement. A Cochrane systematic review in 2002 similarly found just the one randomised controlled trial (RCT) to answer whether paracetamol is more effective than placebo for OA, but did find RCT evidence for superiority of non-steroidal anti-inflammatory drugs (NSAIDs) over paracetamol. However, although some RCTs comparing NSAIDs to paracetamol in OA find NSAIDs to be superior, others report equal efficacy. More recently, the efficacy of paracetamol for knee OA has been seriously questioned by another placebo controlled trial that reported no difference between paracetamol and placebo. This study comes at a time when coxibs are being heavily promoted in terms of equal efficacy but improved gastrointestinal safety relative to non-selective NSAIDs. In America particularly, confidence in both the efficacy and safety of paracetamol compared to NSAIDs and coxibs has come under challenge.

Therefore we have undertaken a meta-analysis of RCTs including both placebo controlled designs and head to head comparisons of paracetamol versus NSAIDs to determine the efficacy of paracetamol in the treatment of OA.
Methods

Retrieval of published studies

Reports of RCTs of paracetamol versus placebo and NSAIDs versus paracetamol were identified through a systematic literature search consisting of:

[1] an electronic search of MEDLINE, EMBASE, CINAHL, Scientific Citation Index and Cochrane Library for the period 1966 to 31st July 2003

[2] searches of reference lists of original reports and review articles, retrieved through the electronic searches

[3] searches for conference abstracts in the last two years via established international societies of rheumatology, such as the British Society for Rheumatology, EULAR, the American College of Rheumatology and the Osteoarthritis Research Society International.

The medical subject heading (MeSH) search used in MEDLINE, EMBASE and CINAHL consisted of three steps, each contained any possible MeSH relevant to the target condition [OA], study drug [paracetamol], and study method [randomised controlled trial]. All MeSHs were exploded. The three steps were then combined to produce citations associated with RCTs of paracetamol versus placebo and NSAIDs versus paracetamol in the treatment of OA. Key word search was undertaken in the Scientific Citation Index and Cochrane Library using the words osteoarthritis and paracetamol/acetaminophen. Titles and abstract were reviewed for possible RCTs and hard copies of the publication were obtained for further scrutiny.

Inclusion and exclusion criteria

Only RCTs comparing paracetamol with placebo or NSAIDs were included. To facilitate interpretation, only studies undertaken in OA (radiographic evidence or ACR clinical criteria for OA) or pain associated with OA were included. Studies in other conditions such as rheumatoid arthritis, non-OA joint pain, pain due to tooth extraction, surgery and injury were excluded.

Quality assessment

Quality of studies was assessed based on randomisation, masking and withdrawal. However, we did not allocate any additional score to an RCT according to whether it described the method of randomisation. In our view, this is a feature of the reporting of the trials and allocation of additional points may be arbitrary. A randomised study was defined as one in which the investigators reported it as being randomised without necessarily defining the randomisation method explicitly since in the past this was not a requirement in the reporting of RCTs. Masking was differentiated as double-blind, single-blind and open-label. Parallel and crossover designs were also categorised. Percentage of withdrawals was calculated. The impact of these quality components to our meta-analysis was assessed using sensitivity analysis.

Data extraction

Two of us (WYZ, AJ) undertook data extraction independently. Any disagreement was resolved by discussion. A customised form was used to record the authors of the study, the year of publication, design of trial (double-blind or single blind, parallel or crossover), location of trial, length of study, number of subjects, patient age,
gender, site of OA (hand, hip, knee or multiple joints), baseline and endpoint scores for pain, stiffness and function, clinical response rate and patient preference rate. In addition, we recorded the proportion of withdrawals and the number of patients reporting gastrointestinal discomfort, nausea, headache and dizziness.

**Outcome measures**

The primary outcome measure for our analysis was pain reduction from baseline. In addition, we looked at other outcome measures, such as the change in total WOMAC (Western Ontario and McMasters Universities OA Index) scores, stiffness and functional scores. Clinical response rate and patient preference were also examined. Clinical response rate was defined as the percentage of patients reporting at least moderate to excellent or greater than 50% pain relief or symptomatic improvement. Patient preference was defined as the proportion of patients preferring the treatment.

**Statistical analysis**

We abstracted the mean and standard deviation of the scores for WOMAC, pain, stiffness and function at baseline and endpoint from individual studies to calculate the mean reduction and the standard deviation of the reduction. The difference of the reduction and its standard deviation between the interventions was calculated for each individual study. The standard difference or effect size (ES) was then calculated using Hedges unbiased approach. The weighted pooled ES was estimated using our previously described method.

The rate ratio (RR) was estimated for the dichotomous efficacy data such as clinical response rate and patient preference. Relative risk was estimated for adverse effect, such as gastrointestinal discomfort, nausea, headache and dizziness. In all cases, Rothman’s method was used for interval estimation of the individual rate ratio or relative risk, and intention to treat analysis was used. In the weighted pooling of rate ratio or relative risk, the method of DerSimonian and Laird was used. The number needed to treat (NNT) and its 95% confidence intervals (95%CI) were estimated as described by Cook et al.

A random effects model was used if trials were heterogeneous on the basis of the Q statistic for heterogeneity and the reason of heterogeneity could not be identified.
Results

1. Characteristics of trials

The literature search based on the search strategies produced 420 citations, including 238 from EMBASE, 88 from MEDLINE, 14 from CINAHL, 78 from Scientific Citation Index, and 2 from conference abstracts. After deleting duplications, 323 citations remained for further scrutiny. Of the 323 citations, 29 potential RCTs associated with paracetamol in the treatment of OA were identified.\(^7\text{--}\text{13}, \text{23-44}\) 10 of them met our inclusion, including 2 placebo controlled trials,\(^7, \text{27}\) 2 placebo and NSAIDs controlled trials\(^13, \text{25}\) and 6 head to head (NSAIDs to paracetamol) trials\(^8-\text{10}, \text{23-24}, \text{26}\) (Table 1). A search in the Cochrane Library produced 101 hits regarding osteoarthritis and paracetamol, including 17 systematic reviews and 84 registered clinical trials. These were scrutinised but no additional trials were identified.

### Table 1. Literature search and retrieval of relevant randomised controlled trials (RCTs)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number of studies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw citations from all sources</td>
<td>422</td>
<td></td>
</tr>
<tr>
<td>Overlaps / duplications</td>
<td>-99</td>
<td></td>
</tr>
<tr>
<td>Non-RCTs</td>
<td>-294</td>
<td></td>
</tr>
<tr>
<td>RCTs reviewed for inclusion criteria</td>
<td>29</td>
<td>7-13, 23-44</td>
</tr>
<tr>
<td>Non-placebo and non-NSAID comparisons</td>
<td>-5</td>
<td>28-32</td>
</tr>
<tr>
<td>Paracetamol combination vs. other agents</td>
<td>-12</td>
<td>33-44</td>
</tr>
<tr>
<td>Duplicated publications</td>
<td>-2</td>
<td>11-12</td>
</tr>
<tr>
<td>RCTs met inclusion criteria</td>
<td>10</td>
<td>7-10, 13, 23-27</td>
</tr>
<tr>
<td>Paracetamol vs. placebo</td>
<td>2</td>
<td>7, 27</td>
</tr>
<tr>
<td>Paracetamol vs. NSAIDs vs. placebo</td>
<td>2</td>
<td>13, 25</td>
</tr>
<tr>
<td>Paracetamol vs. NSAIDs</td>
<td>6</td>
<td>8-10, 23-24, 26</td>
</tr>
</tbody>
</table>

The RCTs included were mainly undertaken in the United States except for one in the UK\(^9\) and another in Italy.\(^27\) They were stated as randomised controlled trials but no further details of random allocation (e.g., allocation sequence and concealment), including those\(^13, \text{23-26}\) published after the CONSORT statement\(^45\): 5 were double blind parallel studies;\(^9-\text{10}, \text{13}, \text{23, 27}\) 3 were double blind crossover studies;\(^7, \text{24-25}\) one used an “n of 1” design;\(^8\) and one RCT gave no further definition of blindness (abstract only available)\(^26\) (Table 2). Withdrawals varied from 0% to 40%.
Table 2. Characteristics of randomised controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Design</th>
<th>OA</th>
<th>Duration</th>
<th>Mean Age (range/±SD)</th>
<th>Male/female</th>
<th>%BPS</th>
<th>Comparison (mg/day)</th>
<th>n1</th>
<th>n2</th>
<th>ES±SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paracetamol vs. placebo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amadio 1983 7</td>
<td>DB-C</td>
<td>knee</td>
<td>6 weeks</td>
<td>64 (43-38)</td>
<td>3/22</td>
<td>-</td>
<td>Par (4000)</td>
<td>25</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>Zoppi 1995 27</td>
<td>DB-P</td>
<td>knee/hip</td>
<td>1 week</td>
<td>56 (20-70)</td>
<td>23/37</td>
<td>65</td>
<td>Par (3000)</td>
<td>30</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td><strong>NSAIDs vs paracetamol vs. placebo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 2003 13</td>
<td>DB-P</td>
<td>knee</td>
<td>12 weeks</td>
<td>62 (40-75)</td>
<td>41/41</td>
<td>34</td>
<td>Dic (150)</td>
<td>25</td>
<td>29</td>
<td>0.36±0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo 25</td>
<td>29</td>
<td>28</td>
<td>0.09±0.27</td>
</tr>
<tr>
<td>Pincus 2003 25</td>
<td>DB-C</td>
<td>hip/knee</td>
<td>6 weeks</td>
<td>63</td>
<td>195/329</td>
<td>51</td>
<td>Cele (200)</td>
<td>178</td>
<td>168</td>
<td>0.14±0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td>168</td>
<td>169</td>
<td>0.23±0.11</td>
</tr>
<tr>
<td><strong>NSAIDs vs paracetamol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradley 1991 10</td>
<td>DP-P</td>
<td>knee</td>
<td>4 weeks</td>
<td>56 (±11.6)</td>
<td>47/137</td>
<td>50</td>
<td>Ibu (1200)</td>
<td>62</td>
<td>61</td>
<td>0.04±0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geba 2002 23</td>
<td>DB-P</td>
<td>knee</td>
<td>6 weeks</td>
<td>63 (39-91)</td>
<td>121/261</td>
<td>60</td>
<td>Rof (12.5)</td>
<td>96</td>
<td>94</td>
<td>0.14±0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo 96</td>
<td>95</td>
<td>94</td>
<td>0.46±0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rof (25)</td>
<td>97</td>
<td>94</td>
<td>0.16±0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 1994 8</td>
<td>n of I</td>
<td>any</td>
<td>6 weeks</td>
<td>64 (38-85)</td>
<td>5/20</td>
<td>-</td>
<td>Dic (100)</td>
<td>15</td>
<td>15</td>
<td>0.31±0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (2000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo 15</td>
<td>15</td>
<td>15</td>
<td>0.28±0.13</td>
</tr>
<tr>
<td>Pincus 2001 24</td>
<td>DB-C</td>
<td>hip/knee</td>
<td>6 weeks</td>
<td>61 (±19.6)</td>
<td>67/160</td>
<td>53</td>
<td>Dic (150)</td>
<td>112</td>
<td>115</td>
<td>0.28±0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Par (4000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo 112</td>
<td>115</td>
<td>115</td>
<td>0.08±0.45</td>
</tr>
<tr>
<td>Shen 2003 26</td>
<td>RCT</td>
<td>knee</td>
<td>3 months</td>
<td>60</td>
<td>44/134</td>
<td>53</td>
<td>Nap (750)</td>
<td>73</td>
<td>75</td>
<td>0.32±0.17</td>
</tr>
<tr>
<td>Williams 1993 9</td>
<td>DB-P</td>
<td>knee</td>
<td>2 years*</td>
<td>60 (33-85)</td>
<td>44/134</td>
<td>53</td>
<td>Par (2600)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OA=osteoarthritis
DB-C=double-blind crossover, DB-P=double-blind parallel, RCT=randomised controlled trial, no further information for blindness
%BPS=percentage of baseline pain score relative to maximum pain score on scale 0%=no pain, 50%=moderate pain, 100%=severe pain
Par=paracetamol, Dic=diclofenac, Cele=celecoxib, Ibu=ibuprofen, Rof=rofecoxib, Nap=naproxen
ES=effect size. SE standard error
n1=number of patients in the treatment group, n2=number of patients in the control group
* Only 6-week results were available for the primary outcome (pain at baseline and endpoint). The 2-year study mainly looked at the withdrawals due to lack of efficacy and side effects
Of the 10 trials, 6 were conducted in patients with knee OA, 3 included subjects with hip and knee OA, and one was undertaken in patients with OA at any joint (single or multiple joints affected). Prior to entry to the trials, patients were asked to stop any previous treatments for OA symptoms and some had 2-7 week pre-washout periods. Except for one study, patient baseline pain severity was 50% to 65% (Table 2) and no trial had a requirement for a worsening pain score after stopping previous treatments. The median length of study was 6 weeks (range 1 week to 2 years). For the two-year study, only 6-week efficacy data were available to derive the primary outcome, i.e., pain score at baseline and endpoint. The follow-up period up to two years mainly looked at number of patients withdrawn and reasons of withdrawals. Of the 10 trials, 7 used a fixed dose of paracetamol 4000 mg per day and 3 used 2000, 2600 and 3000 mg per day respectively (Table 2). Doses of NSAIDs varied according to individual NSAIDs and their usual dosage. Mean ages were relatively young and ranged from 55 to 63 years. All but one study included an excess of women. Pain was used as the primary outcome for efficacy in all studies and was measured using either the WOMAC pain scale or a single visual analogue 0-100 mm scale. Where available, baseline and endpoint pain scores were abstracted for further analysis (Table 2).

2. Efficacy

Pain reduction

The effect sizes in pain reduction with paracetamol versus placebo, and NSAIDs versus paracetamol are shown (Figure 1). Two placebo controlled trials provided pain intensity at both baseline and endpoint. Analgesic effect was therefore estimated using pain reduction from baseline. Of these two trials, one demonstrated that paracetamol was no better than placebo whereas the other showed a statistically significant superiority of paracetamol over placebo. The pooled effect size (ES) was 0.21 (95% confidence interval (CI) 0.02, 0.41, p=0.02). 8 head to head comparisons of NSAIDs versus paracetamol were analysed. While some of these showed superiority of NSAIDs over paracetamol, others demonstrated equal reductions in pain for both agents. Nevertheless, the pooled ES was 0.20 (95% CI 0.10, 0.30, p=0.000), indicating that NSAIDs are better than paracetamol in relieving pain due to OA (Figure 1).

Overall WOMAC

Two placebo controlled trials provided overall WOMAC scores. Both showed no statistically significant difference between paracetamol and placebo (pooled ES 0.14, 95% CI -0.06 to 0.34) (Table 3). In contrast, NSAIDs showed a statistically significant superiority over placebo (pooled ES 0.34, 95% CI 0.14, 0.54) or paracetamol (pooled ES 0.3, 95% CI 0.17-0.44) (Table 3).
Table 3. Effect sizes of overall WOMAC and other outcome measures

<table>
<thead>
<tr>
<th>Outcome improved</th>
<th>N</th>
<th>ES</th>
<th>95% CI</th>
<th>$\chi^2_{\text{heter}}$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paracetamol vs. placebo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC</td>
<td>2</td>
<td>0.14</td>
<td>-0.06, 0.34</td>
<td>1.20 (1)</td>
</tr>
<tr>
<td>Pain</td>
<td>2</td>
<td>0.21</td>
<td>0.02, 0.41*</td>
<td>0.24 (1)</td>
</tr>
<tr>
<td><strong>NSAIDs vs. placebo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC</td>
<td>2</td>
<td>0.34</td>
<td>0.14, 0.54**</td>
<td>0.04 (1)</td>
</tr>
<tr>
<td>Pain</td>
<td>2</td>
<td>0.34</td>
<td>0.14, 0.54**</td>
<td>0.53 (1)</td>
</tr>
<tr>
<td><strong>NSAIDs vs. paracetamol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC</td>
<td>4</td>
<td>0.30</td>
<td>0.17, 0.44**</td>
<td>5.43 (3)</td>
</tr>
<tr>
<td>Pain</td>
<td>11</td>
<td>0.20</td>
<td>0.10, 0.30**</td>
<td>7.68 (10)</td>
</tr>
<tr>
<td>Stiffness</td>
<td>5</td>
<td>0.34</td>
<td>0.19, 0.50**</td>
<td>5.51 (4)</td>
</tr>
<tr>
<td>Function</td>
<td>6</td>
<td>0.22</td>
<td>0.08, 0.35**</td>
<td>7.37 (5)</td>
</tr>
</tbody>
</table>

*p ≤ 0.05; **p ≤ 0.01; N = number of comparisons; ES = effect size; CI = confidence interval; $\chi^2_{\text{heter}}$ (df) = chi-square for heterogeneity (degree of freedom).

Clinical response rate and patients preference

Two placebo-controlled trials\(^7, 27\) were available for this outcome measure (Figure 2). Both trials demonstrated paracetamol was better than placebo, but the results were heterogeneous ($Q = 4.93; p = 0.03$). Clinical response rate ratios were 16 (2.32, 110.45; $p = 0.02$) and 1.67 (1.00, 2.76; $p = 0.05$) respectively. In contrast, the trials comparing NSAIDs and paracetamol were homogeneous. The results showed that NSAIDs were statistically superior to paracetamol. The pooled clinical response rate ratio was 1.24 (95% CI 1.08, 1.41, $p = 0.001$). The NNT was 8 (95% CI 5, 19, $p < 0.001$), i.e., 8 patients needed to be treated before one drug (NSAID) shows clear benefit over the other (paracetamol), as judged by moderate to excellent pain relief.

Three trials\(^8, 24-25\) also examined patient preference to NSAIDs or paracetamol when both treatments were taken in turn in either a crossover design\(^24-25\) or an n of 1 design.\(^8\) The results showed that more patients preferred NSAIDs (61%) to paracetamol (20%). The pooled rate ratio was 2.46 (95% CI 1.51, 4.12, $p < 0.001$) and the NNT was 3 (95% CI 2, 7, $p < 0.001$), i.e., on average, treating 3 patients would lead to one patient preferring NSAIDs to paracetamol. The results were homogeneous irrespective of paracetamol doses of 2 g and 4 g daily. The percentage of patients preferring paracetamol was similar to that preferring neither treatment (18%). The pooled rate ratio was 0.96 (0.79, 1.32).

3. Side-effects

Comparison of side effects showed that paracetamol had a similar safety profile to placebo, whereas NSAIDs caused more gastrointestinal (GI) discomfort (defined as any of the following GI events: abdominal pain, GI distress, nausea, vomiting, dyspepsia, or diarrhoea) than paracetamol. The relative risk (RR) of GI discomfort with NSAIDs versus paracetamol was 1.35 (95% CI 1.05, 1.75) (Table 4). Breakdown of NSAIDs into conventional NSAIDs and coxibs provided further detail of this difference: while the conventional NSAIDs had a higher risk of GI discomfort (RR=1.39, 95%CI 1.07, 1.80), coxibs had a similar risk to paracetamol (RR=0.65, 95%CI 0.17, 2.52).
Table 4. Comparison of risk of side effects (95% confidence intervals) with paracetamol versus placebo, and NSAIDs versus paracetamol

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>GI discomfort</th>
<th>Nausea</th>
<th>Headache</th>
<th>Dizziness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paracetamol vs. placebo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude rate</td>
<td>5/55 vs. 6/55</td>
<td>1/25 vs. 0/25</td>
<td>2/55 vs. 2/55</td>
<td>1/55 vs. 7/55</td>
</tr>
<tr>
<td>Weighted relative risk</td>
<td>0.80 (0.27, 2.37)</td>
<td>3.00 (0.13, 70.30)</td>
<td>1.06 (0.16, 6.95)</td>
<td>0.36 (0.04, 2.96)</td>
</tr>
<tr>
<td><strong>NSAIDs overall vs. paracetamol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude rate</td>
<td>108/704 vs. 82/702</td>
<td>29/491 vs. 23/492</td>
<td>27/581 vs. 32/580</td>
<td>5/288 vs. 3/282</td>
</tr>
<tr>
<td>Weighted relative risk</td>
<td>1.35 (1.05, 1.75)*</td>
<td>1.26 (0.73, 2.18)</td>
<td>0.85 (0.52, 1.40)</td>
<td>1.57 (0.36, 6.85)</td>
</tr>
<tr>
<td><strong>Conventional NSAIDs vs. paracetamol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude rate</td>
<td>105/416 vs. 76/420</td>
<td>15/203 vs. 8/210</td>
<td>5/293 vs. 8/298</td>
<td>/</td>
</tr>
<tr>
<td>Weighted relative risk</td>
<td>1.39 (1.07, 1.80)*</td>
<td>1.94 (0.84, 4.48)</td>
<td>0.67 (0.23, 1.96)</td>
<td>/</td>
</tr>
<tr>
<td><strong>Coxibs vs. paracetamol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude rate</td>
<td>3/288 vs. 6/282</td>
<td>14/288 vs. 15/282</td>
<td>22/288 vs. 24/282</td>
<td>5/288 vs. 3/282</td>
</tr>
<tr>
<td>Weighted relative risk</td>
<td>0.65 (0.17, 2.52)</td>
<td>0.92 (0.45, 1.89)</td>
<td>0.91 (0.52, 1.60)</td>
<td>1.57 (0.36, 6.85)</td>
</tr>
</tbody>
</table>

*p<0.05. GI=gastrointestinal. NSAIDs=non-steroidal anti-inflammatory drugs

4. Sensitivity analyses

Sensitivity analyses were undertaken to determine whether the quality of trials influences the results. Trials designed as double-blind parallel, double-blind crossover and other designs such as n of 1 and RCT with no further information of blindness were stratified and the effect size (ES) of pain reduction and 95% confidence intervals were compared. There were no statistically significant differences between strata (95% CIs overlapped) although the stratification did sometimes affect significance levels within each stratum. For example, although double-blind parallel and crossover designs produced statistically significant ES and other designs produced non-statistically significant ES, there were no statistically significant difference among three designs, as their 95% confidence intervals were overlapped (Table 5). In addition, we divided the studies into two strata according to withdrawal rate (<10% and 10% or more). The results showed that the effect sizes were not statistically different between the two strata (Table 5).

Table 5. Sensitivity analysis with respect to the quality of study

<table>
<thead>
<tr>
<th>Quality of study</th>
<th>Effect size of pain reduction and 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paracetamol vs. placebo</td>
</tr>
<tr>
<td>Study design</td>
<td></td>
</tr>
<tr>
<td>Double-blind parallel</td>
<td>0.09 (-0.43, 0.61)</td>
</tr>
<tr>
<td>Double-blind crossover</td>
<td>0.23 (0.02, 0.44)*</td>
</tr>
<tr>
<td>Others</td>
<td>-</td>
</tr>
<tr>
<td>Withdrawal rate</td>
<td></td>
</tr>
<tr>
<td>&lt; 10%</td>
<td>-</td>
</tr>
<tr>
<td>≥ 10%</td>
<td>0.21 (0.02, 0.41)*</td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.01
Discussion

This is the first comprehensive systematic literature review of RCTs of paracetamol in the treatment of OA. Compared with the recent Cochrane review,6 three more placebo controlled RCTs have been retrieved13, 25, 27 providing evaluation of significantly more RCT evidence in OA for this widely used analgesic. Our meta-analysis confirms that paracetamol is effective in relieving pain due to OA. The effect size of 0.21 is moderate according to Cohen’s definition46 but is statistically significant. We also calculated clinical response rate and showed that paracetamol has a higher response rate than placebo. However, because the definitions of this outcome may be different for different trials, i.e., either moderate to excellent pain relief or clinical symptom improvement, the results are heterogeneous and not relevant for pooling. The analysis does not show any statistically significant difference between paracetamol and placebo for other outcomes such as overall WOMAC score, supporting the construct that paracetamol is a simple analgesic rather than an anti-inflammatory agent and may have limited effects on other aspects of OA symptomatology.

Unlike paracetamol, NSAIDs relieve not just pain due to OA but other outcomes such as total WOMAC, stiffness and physical function. Patients obtain better pain relief with NSAIDs than with paracetamol. The effect size is 0.20, similar to that obtained by comparing paracetamol to placebo. In addition, clinical response rate is higher with NSAIDs than with paracetamol and more patients prefer NSAIDs to paracetamol in the short term. However, paracetamol does appear to be better tolerated with a side-effect profile similar to placebo. In contrast, NSAIDs are associated with significantly increased rates of GI discomfort. A meta-analysis of 16 clinical trials, 23 case-control and 9 cohort studies provides even more compelling evidence for the increased rate of GI side effects of NSAIDs.47 However, more information concerning long-term relative safety and tolerability of NSAIDs and paracetamol is required.

The safety of paracetamol at currently recommended doses has been challenged by two recent studies that suggest a dose-dependent increased risk of GI bleeding. One is a case record linkage study using an automated UK database48 and the other a retrospective cohort analysis from Canada.49 However, both studies contain major confounding factors and one48 found an increased risk of dyspepsia but not serious GI events. Both are contrary to a large body of epidemiological evidence showing no GI risk from paracetamol14, 50, including a meta-analysis based on individual patient data derived from contemporaneous case-control studies examining serious upper GI bleeding50. Although not directly addressing safety, our analysis indicates that paracetamol is safe in the treatment doses up to 4 g daily from 1 week to 2 years, at least in the setting of an RCT. However, the GI discomfort in this review is defined as any GI events, such as abdominal pain, GI distress, nausea, vomiting, or dyspepsia. These are different from the serious GI events such as GI bleeding or other GI ulcer complications, which cannot be determined in this group of short-term randomised controlled trials without endoscopic assessment. Several short-term (<7 days) endoscopic trials have been undertaken to investigate the effects of paracetamol, NSAIDs and/or placebo for these outcomes in healthy volunteers.51-53 None of them demonstrate that the GI toxicity of paracetamol exceeds placebo, though that of NSAIDs clearly does. However, there is no trial evidence to confirm the longer–term (e.g., 3 months and
over) GI safety of paracetamol. Also, we still do no know whether it is more toxic to the population at high risk, such as the elderly with OA. Further well designed long-term studies in patients with OA may be helpful to ascertain the risk.

The analysis has some limitations. First, like many other systematic reviews, we are unable to retrieve unpublished trials so that publication bias may be a factor. Second, the use of studies published only in abstract form are incomplete reports and we were unable to fully assess the quality of the studies. This may cause information bias. Third, in converting outcome measures, particularly for clinical response rate, the variation of the outcome definitions from study to study may affect the results. We therefore could not pool the results if there was any heterogeneity among the trials for these outcome measures. This is why we have chosen effect size as our primary outcome measure since it neglects the different scales and presents a standardized difference between two groups. A caveat to many studies of OA is that most trials are only short-term and more long-term efficacy studies are required in a condition that causes long-term pain and disability.

In conclusion, this analysis confirms that paracetamol is effective in relieving the pain of large joint OA. NSAIDs have a higher effect size than paracetamol for pain relief and in addition help other symptoms of OA such as stiffness. However, weighed against this relative difference in efficacy is the excellent safety record of paracetamol. The selection of treatments depends on a balance of factors of which efficacy is just one. Other factors include safety, tolerability, availability, cost and patient acceptance. However, given its favourable safety profile within the treatment dose of up to 4 g per day, this largest meta-analysis of paracetamol confirms significant efficacy for pain relief in OA. It therefore supports the recommendation that paracetamol be considered the first-line oral analgesic in the management of OA.

Funding: Arthritis Research Campaign: D0565, D0593

Competing interests: None declared
Figure legends

Figure 1. Effect size of pain reduction from baseline and 95% confidence interval

Figure 2. Response rate ratio and 95% confidence interval
References

1. Raffa RB, Stone DJ, Tallarida RJ. Unexpected and pronounced antinociceptive synergy between spinal acetaminophen (paracetamol) and phentolamine. Eur J Pharmacol 2001;412:R1-2


12. Bradley JD, Katz BP, Brandt KD. Severity of knee pain does not predict a better


Figure 1. Effect size of pain reduction from baseline and 95% confidence interval

Paracetamol vs. placebo
- Case 2003
- Pincus 2003
- Pooled

NSAIDs vs. paracetamol
- Bradley 1991
- Case 2003
- Geba 2002
- March 1994
- Pincus 2001
- Pincus 2003
- Shen 2003
- Williams 1993
- Pooled

- 0.21, 95% CI 0.02, 0.41, p=0.02
- Q = 0.24, p = 0.62

- 0.20, 95% CI 0.10, 0.30, p=0.00
- Q = 7.68, p = 0.66

Effect size

Favours placebo/paracetamol  Favours paracetamol/NSAIDs
**Figure 2.** Response rate ratio and 95% confidence interval

- **Paracetamol vs. placebo**
  - Amadio 1983
  - Zoppi 1995

- **NSAIDs vs. paracetamol**
  - Bradley 1991
  - Geba 2002

- **Pooled**

  - Rate ratio (log scale): 1.24, 95% CI 1.08, 1.41
  - $Q=4.93, p=0.16$
Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomised controlled trials
Weiya Zhang, Adrian Jones and Michael Doherty

Ann Rheum Dis  published online March 5, 2004

Updated information and services can be found at:
http://ard.bmj.com/content/early/2004/03/05/ard.2003.018531.citation

These include:

Email alerting service  Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/