Meta-analysis: incidence of endoscopic gastric and duodenal ulcers in placebo arms of randomized placebo-controlled NSAID trials

Y.-H. YUAN*, †, C. WANG*, Y. YUAN* & R. H. HUNT*

*Farncombe Family Digestive Health Research Institute, Division of Gastroenterology, Department of Medicine, McMaster University Health Science Centre, Hamilton, Canada; †Department of Gastroenterology, 2nd Affiliated Hospital, Sun Yet-Sen University, Guangzhou, China

Correspondence to:
Prof. R. H. Hunt, Division of Gastroenterology (Rm 4W8A), Department of Medicine, McMaster University Health Science Centre, Hamilton, ON, Canada, L8N 3Z5.
E-mail: huntr@mcmaster.ca

Publication data
Submitted 10 December 2008
First decision 8 January 2009
Resubmitted 4 May 2009
Accepted 7 May 2009
Epub Accepted Article 11 May 2009

SUMMARY
Background
The safety of NSAIDs is often evaluated by comparison with placebo in clinical trials.

Aim
To investigate the incidence of gastric and duodenal ulcers (GDU) in placebo arms in NSAID trials over the last three decades.

Methods
Randomized placebo-controlled trials of oral NSAIDs from 1975 to 2006 were systematically reviewed. The pooled incidence of GDU in placebo arms was calculated and compared. Meta-regression was used to identify risk factors related to the incidence of the placebo ulcer at the study level.

Results
Thirty-six studies met inclusion criteria (duration of 6.5 days to 24 weeks). In total, 3.29% GDUs were reported in 36 placebo arms. The incidence of GDU in placebo arms was 0, 4.20% and 3.03% in the studies from 1975–1989, 1990–1999 and 2000–2006 respectively (P > 0.05). Eligible subjects with previous GI events and eligible subjects on co-therapy with low-dose aspirin/corticosteroids were associated with the increase in placebo ulcer incidence after adjusting for other factors.

Conclusions
The incidence of GDU in placebo arms has not changed significantly over the last three decades, although has decreased in the past 10 years. Studies show that previous GI events and co-therapy with low-dose aspirin/corticosteroids were associated with increasing GDU in placebo arms.

Aliment Pharmacol Ther 30, 197–209
INTRODUCTION

Non-steroidal anti-inflammatory drugs (NSAIDs) are widely used in the treatment of pain and inflammatory joint diseases. Their major disadvantages are the severe gastrointestinal (GI) adverse effects, including GI mucosal erosions, ulcer, and ulcer complications including haemorrhage and perforation, which may contribute to the high morbidity and mortality in NSAID users.\(^1\), \(^2\)

Endoscopic studies of gastric and duodenal damage are routinely used to assess the effects and the safety of NSAIDs [including nonselective NSAIDs, cyclooxygenase (COX)-2 selective and novel NSAIDs] in the upper GI tract, especially in phase I-II studies. The GI safety profile of NSAIDs is often evaluated by comparison with placebo in randomized controlled trials (RCTs). However, a ‘placebo effect’ does not mean that there is ‘no effect’ and placebo cannot therefore be considered as ‘no treatment’.\(^3\)–\(^5\) Moreover, approximately one-fifth of healthy subjects have some degree of adverse experiences when taking placebo.\(^6\) Therefore, placebo is important in these clinical trials and the difference in ulcer incidence between active treatment and placebo arms gives a reliable estimate of effect in the treatment arms. To determine the effect of background noise, any change in the incidence of gastric and duodenal ulcers (GDU) in placebo arms may influence our interpretation of the relative safety of an NSAID.

Over the past several decades, the prevalence of peptic ulcer disease has fallen significantly, attributed most commonly to the decreasing prevalence of Helicobacter pylori (H. pylori) infection in the general population.\(^7\) Helicobacter pylori infection works synergistically with NSAIDs in the causal development of GDU.\(^8\) Therefore, we might expect that the incidence of GDU in placebo groups in endoscopic trials would have also decreased correspondingly. However, it is not known whether such a change has occurred and how it could affect the relative safety of an NSAID.

In this systematic review, we examine the incidence of endoscopic GDU in subjects treated with placebo in RCTs comparing NSAIDs [including aspirin, nonselective NSAIDs, COX-2 selective inhibitors or COX-inhibiting nitric oxide donors (CINODs)] with placebo to assess the prevalence of GDU in placebo arms in RCTs over the last three decades and to identify factors which might directly influence the placebo-associated ulcer incidence.

METHODS

A computerized recursive literature search was performed in MEDLINE, PubMed and Cochrane databases for relevant RCTs published in the English language from 1966 to December 2006. Search keywords included: anti-inflammatory agents/drugs, non-steroidal OR NSAIDs OR cyclooxygenase 2 inhibitors OR COX-2 selective inhibitors OR coxibs OR nitric oxide donors OR CINOD OR aspirin, individual NSAIDs generic and brand names; AND peptic ulcer OR gastric ulcer OR duodenal ulcer OR stomach ulcer.

Inclusion and exclusion criteria

The inclusion criteria were: randomized placebo-controlled trials comparing orally prescribed NSAIDs with placebo in adults (≥16 years of age) with an endoscopic ulcer as one of the primary outcomes. Endoscopies had to be performed both before and after therapy. Treatment duration was required for longer than 5 days. Exclusion criteria were: studies without raw data for outcomes of interest; duplicate publications; studies without pre- or post-treatment endoscopy; placebo given in combination with other NSAIDs except low-dose aspirin; non-oral NSAID trials. Non-English language studies were excluded and abstracts were not included in the analysis, but all related abstracts were reviewed to search for published full studies.

Data extraction

Trial eligibility was determined independently by two of the authors and data were also independently extracted by two authors. Disagreement between reviews was resolved by consensus or discussion with a third reviewer (RH). Key extracted information included publication year, characterization of study subjects, study design, treatment regimen including...
placebo; pre- and post-endoscopy results in placebo arms, definition and location of mucosal damage and peptic ulcer. Pre-defined risk factors including mean age of subjects, eligible subjects with a history of previous GI event, co-therapy with low-dose aspirin and/or corticosteroids, treatment duration, baseline disease(s) and <10 erosions (defined as mucosal damage but not ulcers by the authors) at time of baseline endoscopy, all of which were extracted at the study level. For any relevant information not included in the reports, we attempted to contact the principal author of the article or the study sponsor.

**Quality assessment**

The quality of each study was evaluated according to the Jadad score with a score of 3 and above indicating ‘high quality’ in the 0 to 5 scoring system. Quality measures included randomization, blinding and withdrawals/drop outs. We did not exclude studies because of a low JADAD score.

**Statistical methods**

The crude pooled incidence with corresponding 95% confidence intervals (CI) of GDU in placebo arms was calculated as the proportion of subjects with ulcer in the intention-to-treat (ITT) population. ITT subjects were defined as all subjects who were randomized and took at least one study drug. Corrected chi-squared or two-tailed Fisher's exact test was used when appropriate to compare the crude pooled GDU incidence for placebo arms within subgroups (details below). Relative risk (RR) with corresponding 95% CI was calculated using GraphPad Instat (GraphPad Software Inc, San Diego, CA, USA). The weighted pooled proportion of GDU in placebo arms was also estimated by weighting the sample size in each study. Cochran's chi-square test (Q test) for heterogeneity was carried out for a pooled proportion calculation where \( P < 0.10 \) indicated significant heterogeneity between studies and the \( I^2 \) value was also calculated and significant heterogeneity defined as \( I^2 > 25\% \). If the \( P \) value for the Q test was >0.10 but \( I^2 > 25\% \), significant heterogeneity was also judged to be present. The fixed effects model was used when homogeneity between studies was seen and the random effects model was used when significant heterogeneity between studies was seen (\( P < 0.1 \)), by using StataDirect 2.6.5 (Cheshire, UK).

Univariate analysis was used to compare the crude pooled placebo ulcer incidence in different pre-defined stratified subgroups at the study level regardless of other risk factors, including publication year, study design (crossover or not), single blind or double blind study, age (≥ 60 years); baseline gastroduodenal mucosal damage; baseline disease with rheumatoid (RA) or osteoarthritis (OA) or healthy volunteer; treatment duration (≥ 28 days); previous GI event(s); co-therapy with aspirin and/or corticosteroids. Weighted pooled ulcer incidence was compared by the test of noncombinability of studies (Cochrane Q test) using StataDirect 2.6.5. To identify the possible risk factors related to the crude ulcer incidence in placebo arms, we assessed these risk factors (except blinding, which related to publication year) at the study level by weighted meta-regression using STATA 9.0 (College Station, TX, USA). A two-tailed \( P \) value less than 0.05 was considered to be significant in all analyses except for the test for heterogeneity; 0.5 was used to replace a zero event in the placebo arm arms to adjust the continuity correction when performing the meta-regression.

**RESULTS**

Our search strategy identified 202 published and 21 unpublished titles and 187 were excluded for not meeting our inclusion criteria. Thirty-six placebo-controlled RCTs meeting inclusion criteria were identified (Table 1)\(^{15–50}\) in which, 3226 subjects received placebo (2860 ITT subjects) and 9870 subjects received active treatments. In total, there were 36 placebo arms and 106 treatment arms (including NSAIDs and non-NSAIDs arms) in papers published between 1975 and 2006, with a treatment duration ranging from 6.5 days to 24 weeks.

The age of subjects ranged from 18 to 88 years (mean age ranged from 21 to 69.3 years). Five studies were of crossover design\(^{15, 17, 29, 45, 46}\) with a 2–4 week washout period between active drugs and placebo; normal oesophageal and gastroduodenal mucosa or ≤10 erosions were required before starting any treatment in these studies. Seven studies were single blind RCTs\(^{15–19, 21, 22}\) and six of them with a study quality score of 2; one score of 3;\(^{22}\) all other studies were double blind RCTs with a Jadad score above 3. The study score was not regarded as an exclusion criterion considering that the earlier studies were less likely to be double-blinded.
Table 1. Characters of the included randomized, placebo-control endoscopic NSAID trials

<table>
<thead>
<tr>
<th>Refs</th>
<th>Study design</th>
<th>Enrolled subject</th>
<th>Treatment during</th>
<th>GDU events in placebo arms (n/N)</th>
<th>Definition of ulcer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanza, 1975, USA, SC</td>
<td>SB, crossover</td>
<td>Healthy subjects</td>
<td>7 days</td>
<td>0/4</td>
<td>Not given for ulcer in Lanza 4-point scale</td>
</tr>
<tr>
<td>Lanza, 1979, USA, SC</td>
<td>SB</td>
<td>Healthy Subjects</td>
<td>7 days</td>
<td>0/5</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1979, USA, SC</td>
<td>SB</td>
<td>Healthy subjects</td>
<td>7 days</td>
<td>0/5</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1980, USA, SC</td>
<td>SB, crossover</td>
<td>Healthy subjects</td>
<td>7 days</td>
<td>0/5</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1981, USA, SC</td>
<td>SB</td>
<td>Healthy Subjects</td>
<td>7 days</td>
<td>0/5</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1982, USA, SC</td>
<td>DB</td>
<td>Healthy subjects</td>
<td>10 days</td>
<td>0/12</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1983, USA, SC</td>
<td>SB</td>
<td>Healthy subjects</td>
<td>7 days</td>
<td>0/20</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1984, USA, SC</td>
<td>SB</td>
<td>Healthy subjects</td>
<td>14 days</td>
<td>0/15</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Lanza, 1987, USA, SC</td>
<td>DB</td>
<td>Healthy male</td>
<td>7.5 days</td>
<td>0/12</td>
<td>Lanza 4-point scale, invasive ulcers of any size were rated 4</td>
</tr>
<tr>
<td>Cryer, 1990, USA, SC</td>
<td>DB</td>
<td>Healthy subjects</td>
<td>7.5 days</td>
<td>0/6</td>
<td>Lanza 5-point scale, grade 5 was ulceration</td>
</tr>
<tr>
<td>Marini, 1993, Italy, SC</td>
<td>DB</td>
<td>Dyspeptic Subjects</td>
<td>7 days</td>
<td>0/10</td>
<td>5 grades scoring system, ulcer was graded as 5</td>
</tr>
<tr>
<td>Laine, 1995, USA, SC</td>
<td>DB</td>
<td>Healthy subjects</td>
<td>28 days</td>
<td>1/16</td>
<td>Ulcer was defined as a break of any size with clear-cut depth</td>
</tr>
<tr>
<td>Patoia, 1996, Italy, SC</td>
<td>DB</td>
<td>Healthy male</td>
<td>28 days</td>
<td>0/13</td>
<td>Lanza 4-point scale, ulcer was score 4</td>
</tr>
<tr>
<td>Bocanegra 1998, USA, MC</td>
<td>DB</td>
<td>OA</td>
<td>6 weeks</td>
<td>3/80</td>
<td>7-point scale, ulcer was graded as 7</td>
</tr>
<tr>
<td>Lanza, 1998, USA, SC</td>
<td>DB, crossover</td>
<td>Healthy Subjects</td>
<td>7 days</td>
<td>0/24</td>
<td>4 grades scoring system, ulcer was graded as 4</td>
</tr>
<tr>
<td>Lipscomb, 1998, UK, SC</td>
<td>DB</td>
<td>Healthy Subjects</td>
<td>28 days</td>
<td>0/11</td>
<td>4 grades scoring system, ulcer was graded as 4</td>
</tr>
<tr>
<td>Simon, 1998, USA, MC</td>
<td>DB</td>
<td>Healthy Subjects</td>
<td>6.5 days</td>
<td>0/32</td>
<td>Lanza 7-point scale, ulcer was defined as any lesion of any size with unequivocal depth</td>
</tr>
<tr>
<td>Simon, 1999, USA+ Canada, MC</td>
<td>DB</td>
<td>RA</td>
<td>12 weeks</td>
<td>4/99</td>
<td>Ulcer was defined as any break in the mucosa at least 3 mm in diameter with unequivocal depth</td>
</tr>
<tr>
<td>Agrawal, 1999, USA, MC</td>
<td>DB</td>
<td>OA</td>
<td>6 weeks</td>
<td>16/334</td>
<td>Ulcer was defined as any break in mucosa ≥ 3 mm in diameter with unequivocal depth</td>
</tr>
<tr>
<td>Lanza, 1999, USA, SC</td>
<td>DB</td>
<td>Healthy Subjects</td>
<td>7 days</td>
<td>0/50</td>
<td>5-grade scoring system, ulcer was defined as white-based mucosal breaks with unequivocal depth</td>
</tr>
<tr>
<td>Laine, 1999, USA, MC</td>
<td>DB</td>
<td>OA</td>
<td>24 weeks</td>
<td>11/158</td>
<td>Ulcer was defined as a mucosa break ≥ 3 mm with unequivocal depth</td>
</tr>
<tr>
<td>Hawkey, 2000, MN (USA, CA Argentina, Prue)</td>
<td>DB</td>
<td>OA</td>
<td>24 weeks (placebo arms only 12 weeks)</td>
<td>5/182</td>
<td>Ulcer was defined as a mucosa break ≥ 3 mm with unequivocal depth</td>
</tr>
</tbody>
</table>

**Note:**
- OA: Open Access
- RA: Randomized Access
- MC: Multicenter
- DB: Double Blind
- SC: Single Center
- GDU: Gastric Duodenal Ulcer
- N: Number of subjects
- Refs: References

**Definition of Ulcer:**

- **Lanza 4-point scale:** Ulcers of any size were rated 4.
- **Lanza 5-point scale:** Ulceration was defined as a lesion which produces an actual crater.
- **5 grades scoring system:** Ulcer was defined by grading as 5.
- **4 grades scoring system:** Ulcer was defined as any lesion of any size with unequivocal depth.
All studies included only healthy volunteers until 1993, but one trial studied non-ulcer dyspepsia patients in 1993 and was not included in our calculations when we performed subgroup analysis for the baseline disease. OA/RA patients were enrolled in RCTs after 1998. Eight studies permitted patients to use...
low-dose aspirin (≤ 325 mg/day) and/or corticosteroids (≤ 10 mg/day) as a baseline treatment. Thirty studies permitted patients with previous upper GI events to enter studies. Thirteen studies allowed eligible patients to have more than one but less than 10 mucosal erosions at baseline. Mucosal damage was recorded and scored in 25 studies, with scales ranging from 4 points to 7 points. Most studies used the Lanza 4 point scale, but this score has been changed or modified with time. Some studies clearly stated that ulcers of any size were rated as 4, 5 or 7 which were the highest scores in the scales. As subjects treated with placebo were all scored as 0 or 1 in this period, we presumed that no ulcer was detected in the placebo arms in studies from 1975 to 1989. In studies from 1990, GDUs were separately noted, although the scoring systems were still used in some of the studies and it was possible to calculate the GDU incidence since that time. In addition, the definition of peptic ulcer was not clear until 1995, when it was defined as a break of any size with clear-cut depth in the mucosa. Other papers defined an ulcer as any lesion ≥ 3 mm in size with unequivocal depth and erosion was defined as discontinuity in the mucosa but without depth. We considered any break in the mucosa at least 3 mm in diameter with unequivocal depth as an ulcer in our analysis for studies after 1999.

Incidence of GDU in the placebo arms

By ITT analysis, 3.29% (95% CI 2.66–4.01%, 94/2,860) (weighted incidence was 3.38%, 95% CI 2.75–4.07, test for heterogeneity \( P = 0.35, I^2 = 6.8\% \)) endoscopic GDUs were reported in placebo treated subjects during the past three decades. Nine studies, between 1975 and 1989, were the earliest studies and constituted one group and all came from the same Lanza study centre with the same inclusion criteria, although only 83 subjects received placebo. Studies were then divided into three time periods: 1975–1989 (9 studies); 1990–1999 (12 studies); 2000–2006 (15 studies). The incidence of endoscopic GDU in placebo arms was 0% (95% CI 0–4.35%, 0/83, weighted incidence = 2.36%, 95% CI 0.27–6.43%), 4.20% (95% CI 2.94–5.80%, 35/833, weighted incidence = 4.34%, 95% CI 3.07–5.81%) and 3.03% (95% CI 2.32–3.90%, 59/1,944, weighted incidence = 2.82%, 95% CI 1.89–3.93%) respectively regardless of risk factors. No significant difference was seen between the three time periods when crude pooled GDU incidences were compared, although the placebo ulcer incidence declined numerically after 2000 when compared with the period 1990–1999 (RR = 1.25, 95% CI 0.96, 1.64, \( P > 0.05 \)). No significant difference was seen for the weighted pooled GDU incidences when compared between the three time periods (\( P > 0.05 \)) (Table 2).

In the univariate analysis, regardless of other risk factors, no significant difference was seen in the crude pooled placebo GDU incidence between cross-over studies and parallel-group RCTs (\( P = 0.295 \)), or single blind studies and double blind studies (\( P = 0.288 \)). The crude pooled GDU incidence in placebo arms was significantly increased in studies in which the subjects mean age was > 60 years (4.29% vs. 1.91%, \( P = 0.001 \)); with a treatment duration ≥ 4 weeks (3.69% vs. 0.95%, \( P = 0.006 \)); in studies which enrolled OA/RA patients vs. those which enrolled only healthy volunteers (3.76% vs. 1.04%, \( P = 0.004 \)); in studies with eligible subjects with < 10 gastroduodenal mucosal erosions at baseline vs. normal mucosa (3.74% vs. 1.04%, \( P = 0.004 \)) and in studies in which eligible subjects had a history of GI events (3.93% vs. 0.84%, \( P < 0.001 \)). The incidence of GDU was not significantly different between OA patients and RA patients (4.16% vs. 3.22%, \( P = 0.532 \)). There was no significant difference in the incidence of GDU in placebo arms in studies which permitted co-therapy with low-dose aspirin (≤ 325 mg/day) or corticosteroids (≤ 10 mg/day) compared to studies without co-therapy (3.40% vs. 3.16%, \( P = 0.804 \)) (Table 2). When the weighted pooled ulcer incidences were compared, a significant difference was seen between the subgroup analyses of mean age and previous GI history (Table 2). As no raw data were available for individual patients, statistical analysis was only possible at the study level.

Meta-regression was conducted to detect factors significantly associated with GDU in placebo arms at the study level, when study design, mean age, treatment duration, baseline disease, previous GI events, co-therapy with low-dose aspirin/corticosteroids and baseline mucosal damage were considered in the regression model. Co-therapy with low-dose aspirin/corticosteroids and previous GI events were significantly associated with placebo GDU in meta-regression when adjusting for other factors (Table 3).
Table 2. Incidence of gastroduodenal ulcer (GDU) in placebo arms in different subgroups

<table>
<thead>
<tr>
<th>Studies with</th>
<th>Subgroups (n = arms)</th>
<th>Crude pooled GDU incidence, n/N (%), 95% CI</th>
<th>Weighted pooled GDU incidence % (95% CI)</th>
<th>RR (95% CI), P†</th>
<th>P‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication year</td>
<td>1975–1987 (n = 9)</td>
<td>0/83 (0, 0–4.45)</td>
<td>2.36 (0.27–6.43)</td>
<td>0, 0.066a</td>
<td>0.250a</td>
</tr>
<tr>
<td></td>
<td>1990–1999 (n = 12)</td>
<td>35/833 (4.20%, 2.94–5.80)</td>
<td>4.34 (3.07–5.81)</td>
<td>0, 0.202b</td>
<td>0.781b</td>
</tr>
<tr>
<td></td>
<td>2000–2006 (n = 15)</td>
<td>59/1944 (3.03%, 2.32–3.90)</td>
<td>2.82 (1.89–3.93)§</td>
<td>1.25 (0.96–1.64), 0.149c</td>
<td>0.081c</td>
</tr>
<tr>
<td>Mean age</td>
<td>&gt;60 years (n = 9)</td>
<td>71/1654 (4.29%, 3.37–5.13)</td>
<td>4.40 (3.47–5.44)</td>
<td>1.32 (1.17–1.49), 0.001</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>≤60 years (n = 27)</td>
<td>23/1206 (1.91%, 1.21–2.85)</td>
<td>2.21 (1.46–3.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding</td>
<td>Single blind (n = 7)</td>
<td>93/2758 (3.37%, 2.73–4.12)</td>
<td>3.43 (2.79–4.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Double blind (n = 29)</td>
<td>94/2801 (3.36%, 2.72–4.10)</td>
<td>3.40 (2.76–4.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-over study</td>
<td>Yes (n = 5)</td>
<td>1/102 (0.98%, 0.02–5.34)</td>
<td>2.08 (0.25–5.62)</td>
<td>0.29 (0.04–2.09), 0.295</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>No (n = 31)</td>
<td>93/2758 (3.37%, 2.73–4.12)</td>
<td>3.43 (2.79–4.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment duration</td>
<td>&gt;4 weeks (n = 16)</td>
<td>90/2438 (3.69%, 2.98–4.52)</td>
<td>3.50 (2.48–4.69)§</td>
<td>1.13 (1.08–1.18), 0.006</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>≤4 weeks (n = 20)</td>
<td>4/422 (0.95%, 0.26–2.41)</td>
<td>1.98 (0.89–3.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OA/RA patients</td>
<td>n = 12*</td>
<td>89/2368 (3.76%, 3.03–4.60)</td>
<td>3.56 (2.46–4.86)</td>
<td>1.15 (1.09–1.21), 0.004</td>
<td>0.080</td>
</tr>
<tr>
<td>Healthy volunteers</td>
<td>n = 23</td>
<td>5/482 (1.04%, 0.34–2.40)</td>
<td>2.03 (0.99–3.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous history of</td>
<td>Permitted (n = 13)</td>
<td>89/2267 (3.93%, 3.16–4.81)</td>
<td>3.96 (3.03–5.00)§</td>
<td>1.20 (1.14–1.27), &lt; 0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GI events</td>
<td>Not permitted (n = 23)</td>
<td>5/593 (0.84%, 0.27–1.96)</td>
<td>1.53 (0.71–2.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concomitant use low</td>
<td>Permitted (n = 8)</td>
<td>52/1531 (3.40%, 2.55–4.43)</td>
<td>3.49 (2.50–4.65)</td>
<td>1.04 (0.86–1.25), 0.804</td>
<td>0.742</td>
</tr>
<tr>
<td>dose aspirin ≤325 mg/d</td>
<td>Not permitted (n = 28)</td>
<td>42/1329 (3.16%, 2.29–4.25)</td>
<td>3.25 (2.37–4.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>or corticosteroids ≤10 mg/d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10 mucosa erosions in</td>
<td>Permitted (n = 13)</td>
<td>89/2381 (3.74%, 3.01–4.58)</td>
<td>3.54 (2.46–4.80)§</td>
<td>1.14 (1.09–1.20), 0.004</td>
<td>0.087</td>
</tr>
<tr>
<td>baseline endoscopy</td>
<td>Not permitted (n = 23)</td>
<td>5/479 (1.04%, 0.34–2.42)</td>
<td>2.05 (0.99–3.47)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


* One arm [25] included dyspeptic patients (n = 10) which was not included this subgroup analysed.
† Crude pooled incidences were compared. RR, relative risk, CI, confidence interval.
‡ Weighted pooled incidences were compared. No adjustment was made for multiple comparisons.
§ Data were pooled by random effects model because of significant heterogeneity (I² > 25%). Other data were pooled by fixed effects model because of statistical homogeneity.
OA, osteoarthritis; RA, rheumatoid arthritis.
Four studies did not provide raw GDU data for endoscopic gastric ulcer (GU) and duodenal ulcer (DU) in the placebo arms.\textsuperscript{32, 42, 46, 50} Within the 11 studies that provided raw data for both GU and DU after 1990, GU was 4-fold higher than DU in the placebo arms (3.48% vs. 0.80%). Both the GU and DU incidence slightly declined after 2000 when compared with the GU and DU incidence between 1990 and 1999 (4.25% vs. 3.17%; 1.19% vs. 0.63%, \( P > 0.05 \), respectively).

The first study that checked \textit{H. pylori} status before treatment was published in 1995.\textsuperscript{26} In total, 19 studies checked \textit{H. pylori} status before subjects were enrolled since 1995, while two studies enrolled only \textit{H. pylori} negative subjects according to the authors' protocol\textsuperscript{30, 44} and two studies did not provide the \textit{H. pylori} infection rate.\textsuperscript{26, 49} Five studies provided the \textit{H. pylori} infection rate both before and after treatment,\textsuperscript{32, 33, 37, 47, 50} while another ten studies only provided the \textit{H. pylori} infection rate at baseline. The \textit{H. pylori} infection rate at baseline ranged from 19% to 61%.\textsuperscript{40, 48} For reasons of lack of raw data, no information was obtained to make possible any analysis of the relationship between \textit{H. pylori} infection and the trend of the GDU incidence on placebo.

DISCUSSION

Upper GI endoscopy has become an important tool for evaluating the safety of NSAIDs. Although ulcers and erosions may exist transiently, they can still be predictive of clinical outcomes\textsuperscript{51} and endoscopic ulcer is an important surrogate endpoint because the absence of lesions can support a positive interpretation of clinical safety obtained in randomized controlled trials.\textsuperscript{52}

The incidence of endoscopic ulcer decreased in recent years in many countries because of \textit{H. pylori} eradication.\textsuperscript{7, 53} However, questions remain with respect to ulcer prevalence, including whether there has been a change in the placebo associated incidence of ulcers in NSAID trials over the past three decades. Our studies found that the incidence of GDU in placebo arms has not significantly changed over the last three decades, but did decrease numerically in the last 10 years, as determined by either the crude pooled ulcer incidence or weighted pooled proportion estimate calculation (Figure 1).

It is well known that placebo treatment is associated with a variety of effects in humans.\textsuperscript{5, 54} Adverse events were reported in between 19%\textsuperscript{6} and 60%\textsuperscript{55} of healthy volunteers and this number was more frequent after repeated dosing\textsuperscript{6} and the proportion of healthy volunteers with adverse events could change with time.\textsuperscript{55} The duodenal ulcer-healing rate is up to 45% while taking placebo.\textsuperscript{56} The placebo effect contributes to the observed treatment effect and differs according to clinical conditions and interventions.\textsuperscript{57} Therefore, the effects of a placebo should be considered before the effects of an active treatment in placebo-controlled

| \textbf{Coefficient} | \textbf{Std. Error} | \textbf{Z} | \( P > |Z| \) | \textbf{95% Confidence intervals} |
|----------------------|---------------------|--------------|-------------|---------------------------------|
| Study design (cross-over or parallel-group) | 0.003 | 0.02 | 0.18 | 0.86 | -0.03, 0.04 |
| Subjects (RA/OA patients or healthy volunteers/dyspepsia) | -0.03 | 0.05 | -0.57 | 0.57 | -0.14, 0.07 |
| Duration (\( \geq \)4 or <4 weeks) | -0.0003 | 0.0002 | -1.68 | 0.09 | -0.0006, 0.00004 |
| Mean age (\( \geq \)60 years or <60 years) | 0.005 | 0.008 | 0.65 | 0.52 | -0.01, 0.02 |
| Co-therapy (with or without low-dose aspirin/corticosteroids) | -0.03 | 0.01 | -2.25 | 0.03 | -0.05, -0.004 |
| GI previous history | 0.06 | 0.02 | 3.08 | 0.002 | 0.02, 0.10 |
| Baseline mucosa damage (<10 erosion) | 0.02 | 0.05 | 0.44 | 0.66 | -0.08, 0.13 |
| Constant | 0.02 | 0.008 | 2.74 | 0.006 | 0.006, 0.04 |

\( Z \) value (Standard score) shows how many standard deviations above (positive) or below (negative) the mean our data value is. Bold represents significant risk factors.
clinical trials can be properly assessed. The more the placebo mimics the active treatment, the less the difference that will be detected between placebo and an active treatment. As a result of the process of informed consent before initiating a clinical trial, the subject is aware of the pharmacological classes of drugs and thus the effects and side effects of the active treatment and any unpleasant previous experiences the patient may have had with a drug may also play a role. Therefore, placebo adverse effects are often active disease-related. The magnitude of the relative therapeutic gain or toxic effect of any new agent in comparison with placebo varies with the response of study participants to the placebo.

Peptic ulcer is a multifactorial disease and H. pylori infection, psychosocial, socioeconomic, environmental, genetic and behavioural factors also contribute to ulcer formation. Placebo has been observed to ‘treat’ peptic ulcer effectively in some trials, although endoscopic ulcer is also seen in placebo-treated patients in NSAID trials. The reason why placebo, as a non-active drug, may be associated with endoscopic ulcer might be partly because of some background factors including the type of patient, physicians involved in care, relationship between patient and physician and the shape and colour of the drug, etc. In addition, old age, baseline disease and concomitant active medications are risk factors for peptic ulcer.

In our studies, unexpectedly, no ulcer was detected in subjects treated with placebo in NSAID trials before 1990, which might be because of a quirk of the scoring system before 1990 and after 1995. A review from a single author also showed virtually no damage in the gastric mucosa in those treated with placebo with an average mean Lanza score of 0.24 from all the endoscopic studies in a single centre between 1975 and 1983. A further review in 1989 from this same author reported that during a 7-year period in a total of 1064 normal volunteers, only one ulcer was recorded out of 202 placebo users after 7 days (0.5%), but no further information was available. It is possible that because the subjects were all young, healthy volunteers in these NSAID trials and only a small number of subjects were treated in each arm (4–20 per arm) over a short treatment time (less than 14 days), this does not provide enough power to detect endoscopic ulcers in trials earlier than 1990s. In addition, single-blind and/or cross-over designs may not be the best way to investigate GI safety with NSAIDs because of possible bias and too few patients in each arm in these trials, which were not likely to detect uncommon adverse effects. We believe that the placebo-associated endoscopic incidence of GDU was probably underestimated during that period.

As the strict selection of subjects does not reflect the real world population, studies performed only in healthy volunteers may seriously limit the generalizability of these studies to the population. The finding that the more ‘high risk’ subjects who were entered and the larger the NSAID trials with a longer treatment duration resulted in a higher incidence of GDU in the placebo arms after 1990 is not unexpected. As shown in our results, the incidence of GDU in placebo arms increased to 3.38% after 1990. Our results also suggest that the incidence of GDU in placebo treated arms numerically but not significantly decreased in the last two decades (from 4.20% in 1990–1999 to 3.03% in 2000–2006). In a Danish cohort study, the incidence of both complicated and uncomplicated GDU decreased from 1993 to 2002, but the proportion of NSAIDs-related peptic ulcer increased and also both first-time GU and DU diagnosis increased with age. As more elderly, OA/RA patients and ‘high risk’ patients were enrolled in trials after 2000, the incidence of placebo ulcer in these RCTs does not represent the ulcer epidemiology of the general population.

Several risk factors are associated with the development of GDU in NSAID users, including age above 65 years, prior GI symptoms or ulcer related events, use of multiple NSAIDs (including aspirin), high dose of NSAID, concomitant use of corticosteroids and/or anticoagulants, treatment duration; serious systemic illness, duration and severity of RA or OA. Our univariate analysis of the crude pooled ulcer incidence at the study level found that a mean age ≥60 years, treatment duration ≥4 weeks, inclusion of
showed no correlation between baseline *H. pylori* status with GDU. Data from these clinical trials are insufficient to confirm the ulcerogenic nature of *H. pylori* in the placebo arms. One can hypothesize that the slight decrease in the incidence of placebo ulcer over the past 6 years might be associated with the decreased prevalence of *H. pylori* infection in the general population. However, we could not retrieve enough information about the prevalence of *H. pylori* infection and the time trends in our included studies. The *H. pylori* infection rates were variable in the studies which we included, even during the same decade. Moreover, the *H. pylori* status of the enrolled subjects may not represent the epidemiology of *H. pylori* infection in the general population at that time.

Limitations of our study should be considered. As we have undertaken meta-analysis, bias may certainly exist, especially as we included studies over a 30-year period and with differing study designs and we only searched for papers published in the English language. Clinical heterogeneity in the study population included different baseline risk factors for ulcer, which may influence the study results. For example, the definitions of ulcer were not standardized and mucosal damage was recorded as Lanza scores only until 1995 after which it was a modified Lanza score with time. In the early time period, studies were not powered adequately and were of single-blind design and detected endoscopic ulcer according to their small size. The small number of ulcer events might have prevented us from detecting the real trend of the incidence of placebo ulcer between time periods. Some, but not all, investigators checked for known risk factors for upper GI damage such as a history of ulcer disease, or co-medications. On the other hand, many other factors could have an impact on the results. For instance, gastric toxicity can persist for a long time after previous exposure to NSAIDs and may sensitize the subject to further gastric mucosal damage. Endoscopic procedures cannot be repeated frequently and so any mucosal damage can only be visualized at a specific point in time; therefore, when the endoscopy was performed, some subjects might have healed mucosal damage. Furthermore, a lesser therapeutic effect in placebo arms might have caused more patients to drop out of the trial and consequently affected the original study quality and also influenced our study results. Hence, results from the various studies of this kind are difficult to combine or compare. Finally, because of the nature of this study, the
effect of the risk factors for ulcer could only be assessed at the study level rather than at the subject level. Moreover, information is lacking on possible confounding factors and therefore the data should be interpreted with caution. Although meta-regression can be used to estimate the treatment-covariate interactions using published data, it is also known to lack statistical power and may be prone to bias,\textsuperscript{25} just as other meta-analysis techniques.

As we know, the pure and simple truth is rarely pure and never simple. Changes in baseline patient characteristics and study design have an important implication for interpretation of future studies in this area and more reliable results should be drawn in the future.

In summary, the incidence of endoscopic GDU in placebo arms of NSAID trials has not significantly changed over the last three decades, but numerically decreased in the last 10 years of analysis. Multiple risk factors are associated with an increased incidence of ulcer in placebo users. After adjusting for other factors, eligible subjects with previous GI events and co-therapy with low-dose aspirin and/or corticosteroids were associated with an increased incidence of placebo ulcer. Although H. pylori infection has declined in developed countries in recent years, unexpected GDUs are still detected in placebo arms of RCTs in the past decade, which suggests that placebo control is still required when evaluating the GI safety of NSAIDs treatment.

ACKNOWLEDGEMENTS

Declaration of personal and funding interests: None. We acknowledge Jiaqing Huang, MD, PhD for the initial data and Marroon Thabane, MSc for the help with statistics.

REFERENCES


52 Lanza F, Sahba B, Schwartz H, et al. The upper GI safety and tolerability of oral alendronate at a dose of 70 milligrams once