Osteoarthritis

A historic issue of the *Annals*: three papers examine paracetamol in osteoarthritis

R Neame, W Zhang, M Doherty

The CONSORT statement for the full reporting of clinical trials should be consistently applied

Although current European League Against Rheumatism (EULAR)[1] and American College of Rheumatology (ACR)[2] guidelines both support paracetamol (acetaminophen) as the first line oral analgesic for patients with knee osteoarthritis (OA), until now there has been a paucity of clinical trial data to confirm the efficacy of paracetamol in large joint OA. This issue of the *Annals* is unique in containing one meta-analysis and two large placebo controlled studies that examine evidence for the efficacy of paracetamol in OA. Never in the 127 years of paracetamol's existence have so much trial data on OA been reported.

Because of the marked placebo effect in pain trials[3] comparison with a placebo is clearly important to determine the true efficacy of any analgesic. Therefore it is surprising that despite the widespread use and support of paracetamol in OA until now there have been only four placebo controlled trials of this drug in OA. The first two studies reported in 1983[4] and in 1995[5] both demonstrated the superiority of paracetamol over placebo for pain relief in large joint OA using parallel group designs. However, the first study contained only 25 patients with knee OA followed up for just 6 weeks.[4] The second study was larger (60 patients) but was of just 1 week's duration and contained a mixed population with knee or hip OA.[5] A third parallel group study was reported in 2003.[6] This compared paracetamol, diclofenac, and placebo over a 12 week period in 82 patients with knee OA but was the first negative study for paracetamol, showing no difference between paracetamol and placebo using Western Ontario McMaster Osteoarthritis Index (WOMAC) assessments. Most recently, Pincus *et al* carried out a crossover study comparing paracetamol, celecoxib, and placebo taken for 6 week periods (PACES-a).[7] The authors found that paracetamol was effective for pain but was no better than placebo for total WOMAC scores.

In the light of these heterogeneous data, Zhang *et al* undertook a meta-analysis to aggregate the evidence that was available up until July 2003.[8] The analysis, published in this issue, shows that paracetamol gives pain relief in OA that is better than placebo (effect size 0.21, 95% confidence interval (CI) 0.02 to 0.41). This estimate was based only on the two studies[6 7] that provided pain intensity at both baseline and end point. Nevertheless, the finding is in accord with other evidence from pain studies and with clinical experience.

Two large randomised placebo controlled studies of paracetamol in large joint OA are now reported in this issue. They use different designs and interestingly find contrasting results, one being negative and one positive for paracetamol, so each will be discussed separately.

THE FRENCH STUDY BY MICELI-RICHARD *ET AL*[9]

This is a 6 week double blind parallel group trial of paracetamol versus placebo in 779 patients with knee OA. The primary end point was a 30% decrease in global knee pain during physical activity in the past 24 hours. The proportions meeting this criterion at the end of the study were the same for paracetamol (52.6%) and placebo (51.9%) so at first sight this is a very strong negative result suggesting that paracetamol is no better than placebo. However, there are several unusual features and caveats to this study, especially:

- The high pain scores at baseline
- The unexpectedly high placebo response, and
- The high dropout rate.

Of all the potential forms of bias that limit the validity and generalisability of clinical trial findings, selection bias is probably the most important. In the Miceli-Richard study, particularly, a number of aspects regarding the source

and characteristics of the participants merit careful consideration.

Firstly, the participants were recruited from general practice. Normally this should be beneficial and reduce the referral selection that accompanies many hospital based studies. However, there were fully 200 recruiting sites giving a very small average recruitment of just 3–4 patients for each practitioner. Apart from the difficulty of standardising the research conduct of 200 study personnel the generalisability of the findings is greatly diminished. It would have been far preferable, and more cost effective, to have enrolled a larger number of participants from a smaller number of centres, thus capturing a greater proportion of representative patients from the total available OA population at each site. Given the very high prevalence of knee OA in the community, the main justification for including so many centres would be to capture an *uncommon subset* of OA. However, this was not the stated purpose of the trial so the rationale remains speculative.

Secondly, the majority of patients appeared to be consulting with a significant exacerbation or ''flare'' of symptoms. They had high pain levels at entry, over half had pain at night, and 65% reported sudden increases of pain in the preceding 2 weeks. Yet the vast majority (86%) were taking no analgesics or non-steroidal anti-inflammatory drugs (NSAIDs) at enrolment. This may reflect the prescribing and the healthcare delivery system in France but contrasts with a UK community study that found very high analgesic usage by patients with knee OA.[10] Importantly, however, pain that is marked or extreme when measured at one time will be closer to its central tendency when measured at a subsequent time— the phenomenon of regression to the mean.[11] This is likely to explain the unusually high apparent ''placebo'' response and the fact that approximately 50% of patients in this trial improved regardless of treatment group—the natural history of their marked exacerbation was in favour of improvement. Repeat measures during a lead in period before baseline would enable estimation of pain variability and the magnitude of regression to the mean, but this is not widely practised in OA trials. To specifically avoid this problem some OA studies impose an upper limit of pain severity within the entry criteria. However, the current study had a minimum requirement only (30% on a 100 mm visual analogue scale) and the *mean* entry pain intensity was very high at 68%, which is close to the 70% upper limit imposed by some

OA studies. Thus a significant number of these patients would have been considered to be in too severe pain to be entered into some other OA trials.

A third problem was the high dropout rate (26% in the paracetamol group, 30% in the placebo group) despite the relatively short 6 week duration. Most dropped out after 1 week because of inefficacy, again reflecting the severity of the "flares". This rate may have been compounded by the absence of rescue analgesia—an interesting issue in that paracetamol itself is the usual escape analgesic for OA trials. Although an intention to treat strategy (last value carried forward) was employed, this degree of attrition reduces the power to detect a treatment effect.

Although the main outcome of the trial was negative, the authors suggest that paracetamol may be beneficial in a subgroup of patients with "non-inflammatory symptoms". It is, of course, debatable as to whether worsening of pain in the previous 2 weeks and pain at night truly reflect "inflammation" rather than worsening of pain due to mechanical or other factors. More importantly, however, although the attempt to identify predictors of response to treatments in OA is to be encouraged, the use of separate analyses for each treatment group to identify predictors of response in this study is questionable. Logistic regression should be applied to the total study population to examine possible interactions.

In summary, therefore, the high proportion of patients entering this multicentre study at a time of marked pain exacerbation is likely to make the study group unrepresentative of the usual clinical status in patients with knee OA and to have led to the apparent unexpectedly high placebo response and dropout rates. These in turn greatly reduced the possibility of showing a treatment effect. The lessons to be learnt from this unusual study for future OA trial design are clear, but the relevance of the study results for clinical management of common knee OA is questionable.

THE AMERICAN PACES STUDY BY PINCUS ET AL
This report presents the full data from PACES-a[7] together with the results of a second study with an identical crossover design but a different sequence of statistical testing for multiple outcomes (PACES-b).[9] Although three treatments were compared (paracetamol, celecoxib, placebo), each patient was randomised to two treatment periods in one of six sequences. In PACES-b, paracetamol was found to be better than placebo for pain relief as in PACES-a, but unlike

PACES-a paracetamol was also better than placebo for reduction in total WOMAC scores. It should be remembered that the WOMAC instrument inquires about the three domains of pain, stiffness, and function but that most of its 24 questions relate to function. Thus a total WOMAC score, even with weighting of the domains, may dilute pure analgesic efficacy. When examining potentially modest effect sizes for such an individual and multidimensional experience as pain, patient preference for one drug over another has advantages over a between-group comparison and may permit a more clinically meaningful comparison. In PACES-a 37% of participants preferred paracetamol, 28% preferred placebo, and 35% expressed no preference. In PACES-b the proportions were 48%, 24%, and 28%, respectively.

In general this a well designed and robust multicentre study. However, with respect to possible selection bias, it is unclear whether it was based in primary or secondary care and how many centres were involved in recruitment. It seems, however, that participants had moderately severe disease as reflected by their baseline pain, WOMAC, and radiographic scores. In contrast with the French study, the entry criteria included a lower (40%) and upper (90%) limit for pain and approximately 70% were taking NSAIDs or analgesics at screening. Whether any were in a "flare" or had clinical symptoms or signs to suggest inflammation is not reported, though there was an obvious increase in mean pain scores after the washout period to mean group scores predominantly in the lower to mid-60s, suggesting prior benefit from their oral drug treatment. Nevertheless, the dropout rates (26% and 27%) in this 13 week trial are slightly lower than those in the French 6 week trial. An unusual design feature of this study was the inclusion of the opioids codeine or tramadol as rescue analgesics, although these were taken by fewer than 5% of participants. Usually opioids are considered to be higher up the analgesic ladder than paracetamol or celecoxib and this may be the first example in an OA trial where a stronger analgesic than the study drug is used to retain participants within a trial. It is unfortunate that the data for patients with knee OA and hip OA were combined and not reported separately. The prevalence, risk factors, natural history, and outcome of hip and knee OA show a number of differences and it cannot be assumed that outcome from the same treatment will be identical.

Two important potential caveats of a crossover design are a carryover effect of a treatment benefit from one period to

another (for long acting treatments or insufficient intervals between treatment periods) and an order effect (where expectation of the second blinded treatment may be different from that of the first). Examination of the pain scores at the start of each period of PACES-a and PACES-b (table 2 of the paper) does suggest a period effect, with pain scores being uniformly lower at the start of the second period. However, this potential bias was adjusted using analysis of covariance. As with the French study it would be useful to have more details about patient selection and the clinical setting and geographical sites from where patients were enrolled into the two studies. Presumably block randomisation was used to minimise the effects of site differences, though this is not specified. This was a strictly double blind, double dummy design and, therefore, on the information provided, it is difficult to explain the different results in PACES-a and PACES-b, though variation in patient sampling and characteristics remains most likely.

WHAT DIFFERENCE DO THESE TWO STUDIES MAKE TO THE META-ANALYSIS BY ZHANG?
Given the two new trials, we reanalysed the data comparing paracetamol with placebo (fig 1). Owing to heterogeneity (Q = 8.27, p = 0.04), which means that the study result is markedly discordant from other studies, the results of the French study[9] cannot rationally be incorporated into the statistical pooling, especially considering the small number of available studies. If it were included, the random effects model would have to be used. When the result of the PACES-b are included the effect size for pain relief changes very little, but the confidence interval tightens (effect size 0.23, 95% CI 0.13 to 0.34). Thus the aggregated evidence is more confident in showing that paracetamol affords pain relief in large joint OA, albeit "small" according to Cohen's interpretation of effect sizes, where 0.2 is small, 0.5 is modest, and more than 0.8 is large.[13] With respect to overall WOMAC as an outcome, Zhang et al were able to include only two studies[6][7] in their meta-analysis, giving a pooled effect size of 0.14 (95% CI −0.06 to 0.34). Inclusion of the additional data from PACES-b again slightly changes this finding (effect size 0.16, 95% CI 0.06 to 0.28).

IS PARACETAMOL AS EFFECTIVE AS NSAIDS?
Initial trials comparing paracetamol with ibuprofen[14] and with naproxen[15] found that paracetamol had similar
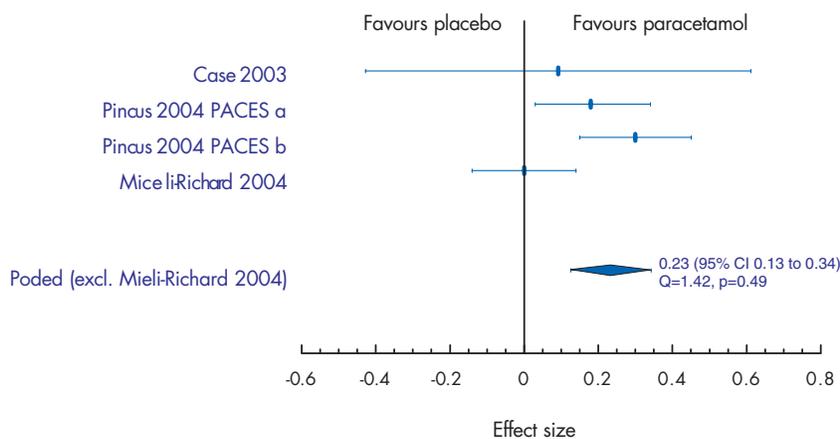
Figure 1 Effect size of pain reduction from baseline and 95% confidence interval.

efficacy to the NSAIDs, although comparison with diclofenac/misoprostol[16] found the NSAID superior for most outcomes. More recently, WOMAC pain relief, and improvement in physical function and stiffness were shown to be greater with two selective cyclo-oxygenase-2 inhibitors (coxibs) than with paracetamol.[17] Most recently, in both PACES trials, improvements in WOMAC scores and pain relief were better for celecoxib than for paracetamol.[12] Zhang et al included eight randomised controlled trials in their meta-analysis and found an aggregated effect size of 0.20 (95% CI 0.10 to 0.30) for pain relief with NSAIDs versus paracetamol. NSAIDs were also better than paracetamol for WOMAC outcomes (effect size 0.3, 95% CI 0.17 to 0.44). This enables us to gain a perspective of the relative efficacy of placebo, acetaminophen, and NSAIDs. Thus, the difference in pain relief between placebo and paracetamol is of similar magnitude to the difference between NSAIDs and paracetamol.

PROBLEMS IN MOVING FROM EVIDENCE BASE TO CLINICAL PRACTICE

Extrapolation of randomised controlled trials (RCTs) and other research data to the "real world" of clinical practice is more problematic than may appear at first sight.[18] Patients with OA in RCTs are usually subject to a large number of exclusions—for example, extreme age, comorbidity, concomitant therapy, bilateral equally painful knee OA, associated calcium pyrophosphate crystal deposition, or degree of radiographic change. Thus they select a small but homogeneous sample from the total population of people with OA and then often investigate the efficacy of a single treatment for a short time. Whether the same treatment works in more typical but "complex" patients is often unanswered by such studies and because many variables are eliminated by inclusion criteria possible predictors of response cannot easily be investigated. An alternative approach is to have very few exclusions, to sample as much of the population as possible, and to examine two or more treatments in a factorial design for a longer period. Although such studies require larger participant numbers, the results are more generalisable to the total OA population. Retaining the heterogeneity of patient variables permits determination of predictors of outcome, and a factorial design permits examination of additive treatment effects. In practice several concurrent treatments are included in most management plans and it is clinically relevant to have data on combined as well as individual treatments—for example, paracetamol alone, NSAID/coxib alone, and NSAID/coxib *plus* paracetamol. The latter situation even exists in many RCTs of monotherapy when paracetamol is used as escape analgesia.

In these respects it is somewhat disappointing that both of the new studies in this issue were of only 6 weeks' duration; that the American study excluded elderly subjects with significant comorbidity; that selection bias in the French trial limits the generalisability of the results to clinical practice or to a meta-analysis; and that predictors of response were inadequately examined in the French study and omitted altogether in the American study. In both trials there is a lack of information about key aspects of study design (for example, the method of selection and recruitment, the randomisation procedure), which creates difficulties in the interpretation of the results. The CONSORT agreement[19] is an attempt to improve the full and unambiguous reporting of key information in clinical trials and both studies have deficiencies in this respect. Although a number of outcome measures for clinical trials in OA have been agreed, there is still considerable debate as to which of these are most useful. This is reflected in the use of different end points in the two new studies. Such heterogeneity of outcome measures is confusing and hampers attempts to pool studies in meta-analyses.

Statistically significant end points must be differentiated from clinically significant changes. For instance, in the PACES trials, the difference in pain improvement on a scale from 0 to 100 between NSAID and paracetamol was 4.9 in PACES-a and 5.9 in PACES-b. Defining the level of change that is clinically significant is difficult, although it has been suggested that a 15% pain reduction or a 30% increase in function are likely to be clinically important.[20] Measurement of outcomes before and after intervention assumes that any change during that time period is due to the intervention. However, patients recalibrate and reconceptualise their pain and disability even over a short time (response shift).[21] This is hard to eliminate but at least can be examined and identified and may in part explain the heterogeneity of some RCT results.

CONCLUSIONS

- Despite a second negative clinical trial the aggregated research data still support paracetamol as being more effective than placebo in relieving pain of large joint OA.

- NSAIDs and coxibs show superior efficacy to paracetamol and are also effective for stiffness.

- Longer term studies of paracetamol, as well as many other treatments for OA, are still required.

- Participants recruited to clinical trials examining common treatments should be representative of patients with OA in general.

- Where possible examination of simple clinical predictors of treatment outcome should be incorporated more often into trial designs.

- The CONSORT agreement for the full reporting of clinical trials should be consistently applied by authors and editors to facilitate the interpretation of clinical trial data and to allow its generalisability to be properly judged.

• •

Authors' affiliations

W Zhang, M Doherty, Academic
Rheumatology, Clinical Sciences Building, City
Hospital Nottingham, NG5 1PB, UK
R Neame, Department of Rheumatology,
King's Mill Hospital, Sutton-in-Ashfield,
NG17 4JL, UK

Correspondence to: Professor M Doherty,
Michael.Doherty@nottingham.ac.uk

REFERENCES

1 **Jordan KM**, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, *et al.* EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the standing Committee for International Clinical Studies Including Therapeutic Trials. *Ann Rheum Dis* 2003;**62**:1145–55.
2 **ACR**. American College of Rheumatology Subcommittee on osteoarthritis guidelines. Recommendations for the medical manage-ment of osteoarthritis of the hip and knee: 2000 update, *Arthritis Rheum* 2000;**43**:1905–15.
3 **Turner JA**, Deyo RA, Loeser JD, Von Korff M, Fordyce WE. The importance of placebo effects in pain treatment and research. *JAMA* 1994;**20**:1609–14.
4 **Amadio P**, Cummings DM. Evaluation of acetaminophen in the management of osteoarthritis of the knee. *Curr Ther Res* 1983;**34**:59–66.
5 **Zoppi M**, Peretti G, Boccard E. Placebo-controlled study of the analgesic efficacy of an effervescent formulation of 500 mg paracetamol in arthritis of the knee or hip. *Eur J Pain* 1995;**16**:42–8.
6 **Case JP**. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis. *Arch Intern Med* 2003;**163**:169–78.
7 **Pincus T**, Fort JF, Mangal B, Koch G, Wolfe F, Moskowitz R, *et al.* Patient Preference for Placebo, Acetaminophen (paracetamol), or Celebrex Effectiveness Study (PACES-1): a double-blinded, randomised, cross-over clinical trial in patients with osteoarthritis of the hip or knee [abstract]. *Ann Rheum Dis* 2003;**62**(suppl I):73.
8 **Zhang W**, Jones A, Doherty M. Does paracetamol (acetaminophen) reduce the pain of osteoarthritis?—a meta-analysis of randomised controlled trials. *Ann Rheum Dis* 2004;**63**:901–7, doi: 10.1136/ard.2003.018531 [Published Online First: 5 March 2004].
9 **Miceli-Richard C**, Le Bars M, Schmidely N, Dougados M. Paracetamol in osteoarthritis of the knee. *Ann Rheum Dis* 2004;**63**:923–30.
10 **Jordan KM**, Sawer S, Coakley P, Smith HE, Cooper C, Arden NK. The use of conventional and complementary treatments for knee osteoarthritis in the community. *Rheumatology (Oxford)* 2004;**43**:381–4.
11 **Whitney CW**, Von Korff M. Regression to the mean in treated versus untreated chronic pain. *Pain* 1992;**50**:281–5.
12 **Pincus T**, Koch G, Lei H, Mangal B, Sokka T, Moskowitz R, *et al.* Patient Response for Placebo, Acetaminophen (paracetamol) or Celecoxib Efficacy Studies (PACES): two randomised, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Ann Rheum Dis* 2004;**63**:931–9.
13 **Cohen J**. *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
14 **Bradley JD**, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an anti-inflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. *N Engl J Med* 1991;**325**:87–91.
15 **Williams HJ**, Ward JR, Egger MJ, Neuner R, Brooks RH, Clegg DO, *et al.* Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. *Arthritis Rheum* 1993;**36**:1196–206.
16 **Pincus T**, Koch GG, Sokka T. A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. *Arthritis Rheum* 2001;**44**:1587–98.
17 **Geba GP**, Weaver AL, Polis AB, Dixon ME, Scnitzer TJ. Vioxx ACT (VACT) Group. Vioxx ACT (VACT) Group. Efficacy of rofecoxib, celecoxib, and acetaminophen in osteoarthritis of the knee: a randomised trial. *JAMA* 2002;**287**:64–71.
18 **Doherty M**, Dougados M. Evidence-based management of osteoarthritis: practical issues relating to the data. *Best Pract Res Clin Rheumatol* 2001;**15**:517–25.
19 **Altman D**, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, *et al.* The revised CONSORT statement for reporting randomized trials: elaboration and explanation. *Ann Intern Med* 2001;**134**:663–94.
20 **Dougados M**, LeClaire P, Van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip. *Osteoarthritis Cartilage* 2000;**8**:395–403.
21 **Carr A**. Problems in measuring or interpreting change in patient outcomes. *Osteoarthritis Cartilage* 2002;**10**:503–5.