acceptable would only be acceptable practice if the ketorolac injection actually produced some clinical benefit. In this small sample, the complete extirpation rate was lower in the ketorolac group than in the saline group and therefore no further work with this drug regime is planned by the authors.

The fact that incomplete extirpation of the pulp was associated with higher McGill pain scores at baseline is interesting and could indicate further areas of study. Regular patients in general dental practice will be well known to the dentist in terms of their level of trait anxiety and treatment can therefore be adjusted to compensate for this. In a hospital or emergency situation, the dentist will not know the patient's level of trait anxiety or have any objective measure of their level of pain. From previous research it appears that few practitioners use dental anxiety questionnaires in their assessment of patients (Dailey *et al.* 2001). Pain questionnaires are not routinely used either but would seem to be of benefit in trying to identify potentially difficult patients to treat in an emergency situation.

Conclusions

The use of an intra-oral injection of ketorolac in the treatment of patients with irreversible pulpitis did not give any benefit in comparison with a placebo. In addition, the injection itself produced significant local discomfort in four of the five patients who received it. Because of this, the study was stopped early and the technique used here cannot be recommended as a treatment in this situation.

References

Buckley MMT, Brogden RN (1990) Ketorolac: a review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential. *Drugs* **39**, 86–109.

Dailey YM, Humphris GM, Lennon MA (2001) The use of dental anxiety questionnaires: a survey of a group of UK dental practitioners. *British Dental Journal* **190**, 450–3.

Edwards RW, Head TW (1989) A clinical trial of intraligamentary anesthesia. *Journal of Dental Research* **68**, 1210–4.

Melzack R (1987) The short-form McGill pain questionnaire. *Pain* **30**, 191–7.

Mietinnen OS, Nurminen M (1985) Comparative analysis of two rates. *Statistics in Medicine* **4**, 213–26.

Nusstein J, Reader A, Nist R, Beck M, Meyers WJ (1998) Anesthetic efficacy of the supplemental intraosseous injection of 2% Lidocaine with 1:100,000 epinephrine in irreversible pulpitis. *Journal of Endodontics* **24**, 487–91.

Parente SA, Anderson RW, Herman WW, Kimbrough WF, Weller RN (1998) Anesthetic efficacy of the supplemental intraosseous injection for teeth with irreversible pulpitis. *Journal of Endodontics* **24**, 826–8.

Penniston SG, Hargreaves KM (1996) Evaluation of periapical injection of ketorolac for management of endodontic pain. *Journal of Endodontics* **22**, 55–9.

Rogers MJ, Johnson BR, Remeikis NA, BeGole EA (1999) Comparison of effect of intracanal use of ketorolac tromethamine and dexamethasone with oral ibuprofen on post-treatment endodontic pain. *Journal of Endodontics* **25**, 381–4.

Stock CJR, Gulabivala K, Walker RT (2004) *Endodontics*, 3rd edn. London, UK: Elsevier Mosby.

Teixera LS, Demarco FF, Coppola MC, Bonow MLM (2001) Clinical and radiographic evaluation of pulpotomies performed under intrapulpal injection of anaesthetic solution. *International Endodontic Journal* **34**, 440–6.

COMMENTARY

The use of an intra-oral injection of ketorolac in the treatment of irreversible pulpitis – comment

The decision whether to publish the paper submitted by Mellor, Dorman and Girdler was a difficult one. A key requirement of the CONSORT guidelines for the reporting of clinical trials, which the journal has adopted, relates to adequacy and justification of sample size. The study as planned was very small and, had it completed as originally intended, might well have been judged unacceptable on these grounds. In the event, it was terminated pre-maturely because of unacceptability of the active treatment. This resulted in a still lower sample size, quite insufficient to support normal comparative analyses of efficacy.

Nevertheless we considered that it would be unethical to withhold publishing the conclusion that an

nonsteroidal anti-inflammatory drug injection should not be considered for such patients on account of its high unacceptability rate. The authors were invited to resubmit a shortened paper, with much of the detailed results and discussion removed. We asked them to consider the possibility that this treatment modality was merely wrong in detail – in which case a further study with a modified intervention regime could be worthwhile. They concluded that it was inappropriate in principle and not worth developing or evaluating further. We are publishing these findings, both to avoid patients being subjected to harmful interventions, and to discourage future researchers from seeking to reinvent this square wheel.

How should we as a profession have reached the conclusion that this treatment modality is unacceptable to patients? Not by a clinical trial designed to compare the active treatment against a matching placebo. A small, uncontrolled preliminary study would have reached the same conclusion. It may well have exposed the same number of patients to the inferior treatment, but would have avoided both the effort of performing a controlled trial and the resulting danger of misinterpretation of findings on efficacy. We strongly urge investigators to pilot their interventions and study procedures before embarking on what they hope will be a definitive study.

The concept of a pilot study has unfortunately suffered much misuse by the research community. A pilot study is a small study which is designed to assess the feasibility and acceptability of all stages of a proposed definitive study. We recommend to readers the excellent paper by Lancaster *et al.* (2004), freely downloadable in pdf format from Blackwell Synergy and ISI Web of Knowledge, which lists several valid and highly worthwhile objectives of pilot studies.

1. Obtain preliminary data on which to base a credible sample size calculation.

2. Test the integrity of the study protocol.

3. Test data collection forms or questionnaires.

4. Pilot the implementation of the randomization procedure.

5. Pilot the recruitment and consent processes, and estimate the likely attrition rate and how much the sample size should be increased accordingly.

6. Assess the acceptability of the intervention – including both willingness to consent to it and also (as would have been most relevant here) the degree to which it is tolerated.

7. Select the most appropriate primary outcome measure.

What a pilot study is not is a sneak preview of the results of the definitive study. Many research proposals reviewed by NHS Trust Research and Development Committees are incorrectly labelled as pilot studies. In reality they are studies of very inadequate size and hence inadequate power to support analyses seeking to detect a difference between treatment regimes. Nevertheless it is clear that the investigators intend to interpret the results very much as in a definitive clinical trial, by performing comparative statistical analyses, as if these were meaningfully interpretable in this situation. Unfortunately, they are not, but can be very misleading, particularly if hypothesis testing is regarded as the mainstay of drawing inferences. Hypothesis testing then incurs a high risk of dismissing potentially clinically important differences as 'nonsignificant'. Confidence intervals are less likely to be misinterpreted, but simply show that the degree of uncertainty is so large as to render the results of little use.

A hypothesis test assesses whether the evidence from the study is 'statistically significant', that is, sufficient to reject the null hypothesis (H_0) that two treatments do not differ in their effect on a particular outcome measure, in favour of the alternative hypothesis (H_1) that some difference exists. Many researchers simplistically dichotomise reality, interpreting 'significant' as 'real' and 'non-significant' as 'null'. In hypothesis testing two types of errors arise. A type I error, or false positive result, occurs when the null hypothesis is true but the data prompt its rejection. A type II error, or false negative result, occurs when a real difference fails to be detected as statistically significant. The type I error rate is usually regarded as fixed at some conventional value, usually 5%, known as the alpha level. Statistical power is the complement of the type II error rate. When it is low, there is a high chance that a real difference will not be demonstrated; this kind of research can only be described as 'hit-and-miss'. Moreover, often a multiplicity of hypothesis tests may be performed, or considered: there may be several outcome variables, several response times at which results are collected, and several possible pairs of groups to compare. In this situation it is only to be expected that some hypothesis tests will yield 'significant' differences, even if overall the null hypothesis is true and the two treatments are identical in effect. For these reasons, an interpretation based primarily on hypothesis testing is generally inadequate, in any case, but especially for small, underpowered studies. This underlines why hypothesis testing is particularly

unhelpful in underpowered studies, including many incorrectly described as 'pilot studies'.

For an intelligent appraisal of the results it is important to consider the size of effect shown also, with confidence intervals to express the degree of uncertainty resulting from the limited sample size. For example, the proportion of patients treated with ketorolac whose extirpation was incomplete was three times that on placebo. This relative risk is based on only 10 procedures; hence the confidence interval is very wide, extending from 0.58 to 19.4. The fact that the lower limit is <1 implies that technically the increased risk on ketorolac would not be judged to be statistically significant, but that would be a most unhelpful summary of the data. Rather, the extremely wide confidence interval appropriately expresses the consequence of the small number of procedures on which the relative risk is based: in the long run the active treatment is unlikely to halve the risk of incomplete extirpation, and may increase it grossly. The study is still too small to warrant firm quantitative conclusions, but the results are expressed in a more cogent manner.

For a simple introduction to the principles of confidence intervals and sample size planning and some straightforward calculation methods, albeit targeted at another dental specialty, see Newcombe (2000a,b, 2001).

R. G. Newcombe
Professor of Medical Statistics,
Wales College of Medicine,
Cardiff University,
Heath Park, Cardiff,
UK

References

Lancaster GA, Dodd S, Williamson PR (2004) Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice* **10**, 307–12, Downloadable from http://www.blackwell-synergy.com/.

Newcombe RG (2000a) Statistical applications in orthodontics. Part I. Confidence intervals: an introduction. *Journal of Orthodontics* **27**, 270–72.

Newcombe RG (2000b) Statistical applications in orthodontics. Part II. Confidence intervals for proportions and their differences. *Journal of Orthodontics* **27**, 339–40.

Newcombe RG (2001) Statistical applications in orthodontics. Part III. How large a study is needed? *Journal of Orthodontics* **28**, 169–72.