Acute pain: combination treatments and how we measure their efficacy

H. J. McQuay1 *, K. H. Poon1 2, S. Derry1 and R. A. Moore1

1Pain Research and Nuffield Department of Anaesthetics, University of Oxford, The Churchill, Headington, Oxford OX3 7LJ, UK. 2Department of Anaesthesia, National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074, Singapore

*Corresponding author. E-mail: henry.mcquay@pru.ox.ac.uk

Perioperative analgesic strategies are frequently tested using analgesic consumption as an outcome measure. This outcome measure is intuitive and superficially attractive, but has not been evaluated rigorously. Flaws in its use may be one explanation of continuing controversies surrounding the efficacy of analgesic strategies tested by this method. We contend that the analgesic consumption outcome measure is valid only when treatment groups achieve similar pain scores. A meta-analysis of perioperative gabapentin was used to test this hypothesis. Eighteen trials were identified, which were of sufficient methodological quality to include in the analysis. Trials reporting similar pain scores in treatment groups were classified as Category A and dissimilar scores as Category B. There were seven Category A trials: four reported reduced analgesic consumption with gabapentin compared with placebo, at one or more time points, and three found no difference. There were 11 Category B trials, all of which reported a decrease in analgesic consumption with gabapentin compared with placebo, at one or more time points. Analgesic consumption after gabapentin was similar for different postoperative analgesics. Sedation, dizziness, and vomiting were significant problems in pooled analysis. Analysis according to similarity of pain scores did not clarify whether perioperative gabapentin is useful in perioperative care. More rigorous examination of analgesic consumption as an outcome measure is needed, to establish whether achieving similar pain scores is as important as this paper claims and to determine those features of the analgesic delivery system, adverse effects, and other factors which may interfere with analgesic consumption as an outcome measure.

Br J Anaesth 2008; 101: 69–76

Keywords: analgesics opioid, morphine; analgesics, postoperative; pain, acute; surgery, postoperative period; surgery, preoperative period
those using patient-controlled analgesia (PCA) devices. Robustness of the system is assumed, but, to our knowledge, it has not been rigorously tested, especially in the pre-emptive or perioperative analgesic setting. If the analgesic consumption design is less than robust, then misleading conclusions may be made about the true efficacy of interventions such as pre-emptive analgesia, intra-articular morphine injection in the knee, and preoperative coxibs or gabapentin.

The analgesic consumption design used in many of these studies is shown in outline form in Figure 1, for a perioperative analgesic. Although pain scores may be recorded, it is the variation in analgesic consumption between the two treatment groups that is used as the primary outcome measure.

Within the context of a double-blind randomized trial, the patient is given a test treatment before or immediately after surgery. The amount of analgesia the patient needs over the ensuing hours or days, the analgesic consumption, is measured. This can be done in a variety of ways, via high technology such as PCA or, more simply, by recording the amount of tablets used. Pain scores are obtained at various intervals. An effective intervention should result in lower analgesic consumption than an ineffective intervention to achieve the same pain intensity level.

The crucial assumption here is that the pain scores in the two treatment groups should be the same at an appropriate time after operation, with regard to the characteristics of the intervention, like speed of action and duration. Several other assumptions are made:

(i) Patients will titrate themselves to a state of comfort, where low pain levels are traded against increased adverse events, like nausea.
(ii) In two similar groups of patients, with similar perioperative interventions, the average amount of postoperative analgesia consumed will be the same.
(iii) Use of an effective perioperative analgesic will reduce the postoperative analgesic consumption required to achieve the same state of comfort.

Although analgesic consumption will differ, the pain scores recorded will be the same.

These assumptions have not, to our knowledge, been tested in circumstances where a known effective perioperative analgesic intervention has been used. This is important, for if we use the method to test interventions of unknown efficacy, we are left uncertain as to the interpretation of the results.

One example is where postoperative pain scores in two differently treated groups of patients are reduced in one group where the analgesic consumption is also reduced. It could be that the intervention was so good that it disturbed the equilibrium of postoperative comfort that we have assumed to apply. A problem with this is that we do not know how robust the experimental system is; it may be very robust, in which case only an exceptional analgesic intervention would overwhelm it or it may be so fragile that even a trivial analgesic intervention would do so.

This poses the question of validity in trials of analgesic consumption. Our assumption is that a trial is valid when the postoperative pain scores are the same. Conversely, we assume that different postoperative pain scores signify either an invalid study or an overwhelmingly good analgesic intervention.

We hypothesize that trials that are able to achieve similar pain intensity scores with different intervention groups within a clinically and biologically plausible time period should in theory constitute stronger evidence. Failure to do so might reflect an inherent flaw in the study design, affording less credibility. We decided to test this hypothesis by performing a systematic review of perioperative gabapentin to see if we can differentiate the true clinical effects of gabapentin from a failed study design.

**Systematic review of perioperative gabapentin**

The Cochrane Library, PubMed, and EMBASE were used to identify RCTs on perioperative gabapentin up to May 2006. The search strategy used the words gabapentin, postoperative, and surgery. Reference lists of the retrieved articles were also searched, and we contacted Pfizer for any knowledge of other papers or for company search results. Trials were included if they were randomized, double-blind, active or placebo controlled, had at least 10 subjects per study group, and reported both analgesic consumption and pain scores. Perioperative is defined as within 24 h pre- and anytime postoperative. Only trials studying preoperative gabapentin alone were included; if gabapentin was part of a multimodal technique, the trial was excluded.

Trial quality was assessed using a validated three-item scale with a maximum quality score of five. A point was awarded even if there was no statement on withdrawals and dropouts if it was implicit from the report and data presentation that all patients enrolled subsequently...
appeared in the data analysis section. Trial validity was assessed using a validated 16-point scale. Quality and validity assessments were made independently by at least two reviewers and verified by one other reviewer. Disputes were settled by consensus. Studies with a low quality score, which we defined as less than 3 out of the maximum of 5, were excluded from analysis.

Classification of trials
Devising our own classification, we divided the trials into Category A and Category B (Table 1). Trials with similar pain scores between the groups within a predetermined time interval were classified as Category A. Those reporting different pain scores between the groups within the same time window were classified as Category B.

On the basis of our assumptions, Category A trials fit the expected outcome of the analgesic consumption model and are regarded as more robust trials. Category B trials may be regarded as less robust trials.

The predetermined time interval depends on the treatment under investigation. It is essentially based on the expected onset and duration of the treatment effect, assuming that the treatment is efficacious. In the case of gabapentin, the elimination half-life is 5–7 h and clinically the usual recommended dosing interval is 8 h. The likely time during which the clinical effect of gabapentin should be maximally present is 4–12 h after surgery. By 4 h after operation, patients should have titrated themselves to a state of comfort. Beyond 24 h, no residual effect of single-dose gabapentin was expected.

Further classification
We postulated that disparity in pain score similarity or difference within and/or between the 4–12 and 24 h recordings may make the trial less credible. We thus further subdivided Categories A and B (Table 1). Category A1 trials consistently report similar pain scores at 4–12 h and at 24 h. Category A2 trials have similar pain scores at 4–12 h but different pain scores at 24 h. Category B1 trials report different pain scores at 4–12 h but the same pain scores at 24 h. Category B2 trials report consistently different pain scores.

<table>
<thead>
<tr>
<th>Table 1 Classification of trials according to postoperative pain scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Pain intensity</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Data collection and analysis
We collected information on analgesic consumption, pain scores at rest and, when available, adverse effects and pain scores on movement. We also collected other relevant details of the trials, such as number of patients, operation type, and postoperative analgesia used. We calculated analgesic consumption with gabapentin compared with control at 24 h and also at 4, 8, and 12 h, whenever data were available. Our prior definition of a positive outcome was a trial where gabapentin significantly reduced postoperative analgesic consumption over placebo whereas a negative outcome trial did not.

In the analysis of analgesic consumption, a trial was considered positive if there was a positive outcome at any time point. We identified sedation, dizziness, nausea, and vomiting as adverse events relevant to our analysis. The relative risk and number-needed-to-harm were calculated with 95% confidence intervals. Relative risk was calculated using a fixed effects model, with no statistically significant difference between treatments assumed when the 95% confidence intervals included unity. We added 0.5 to treatment and comparator arms of trials in which at least one arm had no event. Number-needed-to-harm was calculated by the method of Cook and Sackett using the pooled number of observations only when there was a statistically significant difference of relative risk.

Results
We found 20 relevant published trials using our search strategy. One trial was excluded as it had a quality score of 1 out of 5. Another trial used gabapentin as part of a multimodal technique. The remaining 18 trials (1217 patients) were divided into Category A (seven trials, 416 patients) or B (11 trials, 801 patients; Fig. 2). All Category A trials that could be subclassified were A1 (five trials, 291 patients). There were two Category A trials which reported data only up to 4 and 8 h and thus could not be subclassified (125 patients). There were two Category B1 trials (90 patients) and nine B2 (711 patients). Details of included trials and outcomes are in Supplementary files 1–5. All trials were described as randomized, double-blind and used a placebo, with one trial using oxazepam as an ‘active placebo’ (classified as A1). These were all treated as placebo comparisons. All trials, except one, administered one or more doses of gabapentin before operation, with five of these continuing gabapentin after operation. The trial which used a single postoperative dose compared the outcomes with a single preoperative dose and placebo. Only data from the preoperative group were used for analyses. The quality and validity scores of the included trials were good with no differences between Category A and Category B trials. The median quality score was 5 (range 3–5) and the
median validity score was 13.5 (range 10–15). Trials were small with a mean of 34 and a median of 25 patients per group, with one larger trial with 153 patients per arm. All the operations were performed under general anaesthesia, with one exception involving minor ENT procedures performed under monitored anaesthetic care. Other operations included mastectomy, hysterectomy, laparoscopic cholecystectomy, donor nephrectomy, thyroidectomy, and various orthopaedic procedures. Distribution of operation type was broadly similar in Categories A and B. We defined baseline pain as the first pain scores at rest after operation. These were more than 30/100 mm VAS in all but two of the trials. One trial did not report baseline pain; the other reported 28/100.

Adverse effects
The results of relevant adverse effects were pooled from all trials when available. This revealed a statistically significant increase in the incidence of sedation and dizziness and a statistically (relative risk with 95% confidence interval was calculated using the fixed effects model and was considered significant when the 95% CI did not include one) significant decrease in the incidence of vomiting but not nausea with gabapentin compared with control group (Table 2).

Analysis by category
In Category A trials, the pain scores in the placebo groups at 4 h were all <30/100 mm VAS or equivalent. This was achieved even though baseline pain scores were greater than 30/100 mm with one exception. This may reflect an effective analgesic delivery system, which a valid analgesic consumption outcome measure should have. On the contrary, most Category B trials failed to achieve a similar reduction in pain score at 4 h. In some trials, it was not achieved even at later time points, and this may reflect a failing of analgesic delivery. This suggests that Category B trials were less robust than Category A (Supplementary file 4).

Of the seven Category A trials, four reported reduced analgesic consumption with gabapentin compared with placebo at one or more time points (including the trial with oxazepam as an active placebo) and three trials reported no difference between gabapentin and placebo (Fig. 3). The weighted mean analgesic consumption for gabapentin compared with placebo (24 h where available, or longest time) was 71% in Category A trials. All 11 Category B trials reported a decrease in analgesic consumption with gabapentin at one or more time points (Fig. 3). The weighted mean analgesic consumption of gabapentin compared with placebo was 59% for Category B trials. Combining all 18 trials, the weighted mean consumption was 62%, a reduction in analgesic consumption in the gabapentin group of 38% (Supplementary file 3).

Discussion
The gold standard for robust evidence of clinical efficacy is the RCT, and this review considers the RCTs for perioperative gabapentin. The discussion begins by looking at the analgesic consumption method as the outcome measure in these trials and concludes by looking at the issue of perioperative gabapentin.
sensitive with mild baseline pain or does it need moderate consumption methodology has not been sub-

improvement that has happened with the nurse observer method?14

The sensitivity of the con-

jected to the continual iterative quality control and morphine. Consumption methodology has not been sub-

factors influencing the analgesic consumption as an outcome measure, the accumu-

ated evidence is not conclusive. Some trials show efficacy, some do not, and convincing dose–response data are notable by their absence. There are at least two explanations why the data may be inconclusive. The first is that the analgesic interventions are not effective, and that the study design is flawed and capable of producing false positive results. The second explanation, corollary of the first, is that the interventions are effective, but flawed study design permits false negative results. The trials in general meet the quality standards for inclusion in systematic reviews, so that the problem is not simply due to lack of randomization or flawed double-blind.

We decided to take the thread common to these studies, the use of analgesic consumption as an outcome measure, to see if problems with this as an outcome could contribute to the conflicting evidence from seemingly well-designed trials.

It is clear that there is no good body of data establishing the validity of analgesic consumption as an outcome measure in analgesic investigations. We have assumed that it works, but we do not know whether, for instance, consumption methodology could provide reliable dose–response for a known effective perioperative analgesic like morphine. Consumption methodology has not been subjected to the continual iterative quality control and improvement that has happened with the nurse observer methods over the past 60 yr. The sensitivity of the consumption method is not known; for instance, is the method sensitive with mild baseline pain or does it need moderate or severe baseline pain to ensure sensitivity, as is the case with the nurse observer method?14

In the absence of such a body of knowledge, an underpinning assumption of consumption methodology must be that patients in different treatment groups will consume postoperative analgesic according to their levels of pain. This titration, reducing pain intensity to a state of comfort, should result in similar pain scores across the treatment groups. If it does not result in similar scores, then there is something wrong with the titration. This conclusion is independent of the method of analgesic delivery but applies equally to all.

The titration, the analgesic consumption, might also be affected by the pharmacology of the drug whose effect is being tested. If that drug caused excess sedation, altered cognitive function, or motor problems, those effects might impinge on analgesic consumption, for example, the subject making fewer demands on PCA, and be misinterpreted as analgesia.

The analgesic consumption could itself be problematic. If the subjects are unable to tolerate more than a certain amount of PCA, for instance because of nausea, then subjects who have received an ineffective test drug or a placebo may be unable to reduce their pain scores as far as they might wish, and be unable to reduce their pain scores as low as subjects who received an effective analgesic plus the same limited amount of PCA. This limitation of the ability to reduce pain scores to the desired level would make the assay, the comparison of the analgesic consumption in test and control groups at similar pain levels, invalid.

A compounding difficulty in these analgesic studies is the circumstance where the analgesic intervention being tested is given before it is known that the subject has sufficient pain to allow measurable decrease. Examples are pre-emptive analgesic studies or studies of intra-articular morphine injection. Low postoperative analgesic consumption might be interpreted as evidence of analgesic efficacy of the test intervention when in reality the subject had minimal pain to begin with. This difficulty does not appear to be a problem in the perioperative gabapentin studies, which show evidence of at least moderate levels of baseline pain (Supplementary file 4).

**Perioperative gabapentin**

Using the endpoint of similar pain scores in the treatment groups, we classified the perioperative gabapentin trials into Categories A and B, where evidence from Category B trials was considered less robust than evidence from Category A trials. All three negative trials were in Category A (Fig. 3).

A general rule of systematic reviews would be that trials of lower quality report greater efficacy of an intervention than do trials of higher quality. The finding that all three negative trials were in Category A may reflect this general
relationship between quality and efficacy. It also suggests that the question of the efficacy of perioperative gabapentin is not yet answered clearly.

**Is this a sensitivity issue?**

The method may fail because it lacks sensitivity, unable to differentiate treatments of different efficacy. A meta-analysis of perioperative use of acetaminophen, NSAIDs, and coxibs failed to demonstrate a significant difference in PCA morphine consumption, despite the known differing efficacy in nurse observer studies. We were also drawn by the similarity in the reduction in analgesic consumption for perioperative coxibs (27–41%) and perioperative gabapentin (35%). Coxibs at adequate dose are known effective postoperative analgesics, unlike gabapentin where it could be argued that there is no expectation that gabapentin should be effective. It could be that gabapentin is as good as these non-opioid analgesics in providing postoperative analgesia or that the consumption methodology has failed to detect differences in efficacy because it is just not sensitive enough.

**Is any perioperative analgesic effect of gabapentin unique to morphine?**

Gabapentin was found to enhance the analgesic effect of morphine in healthy volunteers. This was not due to altered pharmacokinetics of morphine, and the actual mechanism remained unclear. Gabapentin also showed an opioid-sparing effect when used with morphine to treat patients with neuropathic pain in the chronic pain setting, corroborating the results of the experimental pain study. This may, or may not, be relevant to postoperative pain. A significant proportion of the perioperative gabapentin trials used morphine as the postoperative analgesic. It was possible that the observed reduction in analgesic consumption was a reflection of the morphine–gabapentin combination rather than a real analgesic effect of gabapentin. To see if this was the case, we analysed the data by comparing the analgesic consumption between trials according to the postoperative analgesic (Fig. 4). There was no apparent difference between trials using morphine, fentanyl, other opioids and a single trial using NSAID. Any effect of gabapentin does not appear to be specific to morphine.

**Adverse effects**

Information about adverse effects is important in the interpretation of results in trials using analgesic consumption as an outcome. Sedation is of particular concern as a potential confounder because it could lead to reduced consumption, with that reduction subsequently misinterpreted as analgesic efficacy. Table 2 shows a statistically significant sedative effect of gabapentin compared with placebo. Most of this significance comes from just two trials, one of which is by far the biggest, and uses a low dose (300 mg) of gabapentin. It seems odd that no significant sedative effect compared with placebo was reported in the other trials using doses up to four times greater.

Nausea and vomiting may also influence postoperative analgesic consumption. If increasing opioid consumption results in more nausea and vomiting, it is likely that patients will limit their opioid consumption even though their pain scores are still high. Our analysis revealed a statistically significant increase in the incidence of vomiting in placebo groups. One could argue that this might lead to the placebo groups trading higher pain scores for less vomiting.

Unfortunately, no individual trial was adequately powered to detect a difference in adverse events and even with pooling of the trials, there were only 300–500 patients per treatment arm. The statistical significant differences for vomiting and dizziness were also borderline. Together with poor reporting, we cannot say with any confidence whether or not vomiting, dizziness, or sedation were indeed confounding factors in analgesic consumption.

**Conclusion**

The contention is that the analgesic consumption outcome measure, comparing consumption in different treatment groups after test and control interventions, is valid only when the groups have achieved similar pain scores. Indeed, the null hypothesis, no difference between treatment and control, can only be proven when the pain scores...
and consumption are the same in the treatment and control groups.

Having applied these thoughts to the 18 admissible trials of perioperative gabapentin, we are not clear about whether or not perioperative gabapentin is a useful part of perioperative care. The fact that the bulk of the trials were in the less valid Category B, where the pain scores in the treatment and control groups did not come down to similar levels, should make the reader sceptical.

There is a clear need for more rigorous examination of analgesic consumption as an outcome measure, to establish whether the achievement of similar pain scores is as important as this paper claims and to determine the functioning of analgesic consumption as an outcome measure.

Supplementary material
Supplementary material is available at British Journal of Anaesthesia online.

Funding
The Oxford Pain Relief Trust provided support for the work, and K.H.P. was supported by National University Hospital, Singapore for a clinical fellowship.

References


