A COMPARISON OF ASPIRIN AND PARACETAMOL
With a Note on Method

BY
JAMES PARKHOUSE AND PAULINE HALLION
Nuffield Department of Anaesthetics, Radcliffe Infirmary, Oxford

SUMMARY
In a double-blind comparison of aspirin, paracetamol and a placebo in postoperative orthopaedic patients a nurse-observer was able to distinguish both active drugs easily from the placebo on first dose information and less easily with second doses. A sequential analysis of between-dose preferences in individual patients was less successful. Few side effects were noted. Some of the problems of subjective scoring of pain and its relief are discussed.

This investigation followed the study of aspirin and nepenthe which was recently reported (Parkhouse, Collie and Wood, 1967). A wholetime nurse-investigator (P.H.) was employed; she spent a practice period with the previous nurse-investigator (V.W.) before beginning the study.

METHOD
Aspirin 600 mg, paracetamol 1 g and a placebo were compared. Each medication was made up in the form of two gelatine capsules; all the capsules were tasteless and identical in appearance.

The nurse-investigator studied a total of 103 patients, exclusively in an orthopaedic hospital. All the patients were suffering from postoperative pain of the type for which aspirin or a similar drug would normally have been given.

Because of our previous experiences with a three-dose study the present investigation was designed on a two-dose basis; each patient received two of the three medications. The experimental design was balanced over the intended total of 120 patients; in fact, it was not possible to complete this number of patients before the investigator left Oxford.

Data were recorded as in the previous investigation (Parkhouse, Collie and Wood, 1967), both the patient’s and the investigator’s opinion being noted at each hour. Statistical analysis of the results was again undertaken by computer. In addition, a sequential analysis was made; this was based upon a comparison between doses in each individual patient. To give the greatest number of available cases, the higher score according to either the patient or the investigator or both was considered as a preference, regardless of order of drug administration. In one case the patient’s preference disagreed with the investigator’s, the difference in total score for the two drugs being one on each assessment; this case was regarded as a tie.

RESULTS
Of the 103 patients studied only 64 required a second dose. Tables I and II show the number of patients who received each medication, with the mean age, mean weight and sex distribution for each group and for each dose.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First medication.</strong></td>
</tr>
<tr>
<td>Number of patients</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Paracetamol</td>
</tr>
<tr>
<td>Aspirin</td>
</tr>
<tr>
<td>Placebo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second medication.</strong></td>
</tr>
<tr>
<td>Number of patients</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Paracetamol</td>
</tr>
<tr>
<td>Aspirin</td>
</tr>
<tr>
<td>Placebo</td>
</tr>
</tbody>
</table>
A COMPARISON OF ASPIRIN AND PARACETAMOL

FIG. 1
First-dose data. Mean total relief as assessed by investigator and patient, with standard errors of means, for numbers of patients indicated in brackets.

Placebo Aspirin Paracetamol

FIG. 2
Second-dose data. Mean total relief as assessed by investigator and patient, with standard errors of means, for numbers of patients indicated in brackets.

Placebo Aspirin Paracetamol
Sequential comparison between medications (see text).

First-dose data. Investigator's assessment of mean pain relief at each hour, in patients to whom a subsequent drug had not yet been given (see text). Numbers of patients in brackets; some standard errors shown.
Table III shows the results of t-test comparisons between medications for both first and second doses, and for both patients' and investigator's assessments. Figure 1 shows the mean total relief score for each medication, with standard error, for first doses; figure 2 shows the same information for second doses. The results of sequential analysis are shown in figure 3.

The only side effect noted was "mild" drowsiness on 15 occasions and "moderate" drowsiness on 1 occasion. It was recorded after paracetamol 7 times, aspirin 6 times and placebo 3 times.

Both active drugs could be distinguished easily from the placebo on first medication data. With second doses aspirin could still clearly be distinguished from placebo but the difference between paracetamol and placebo was not statistically significant; it must be remembered that the numbers of patients in each group were appreciably smaller at the second dose.

There was no statistically significant difference between the two active drugs; from the first-dose data it appeared that 1 g of paracetamol was slightly more effective than 600 mg of aspirin; from the second-dose data it appeared that the aspirin was slightly more effective than the paracetamol.

As a second dose, both paracetamol and placebo were rather more effective following aspirin as a first dose than following placebo or paracetamol respectively; there is a suggestion, therefore, that the aspirin may have had some residual effect although these differences in second-dose response were not statistically significant.

On sequential analysis only 13 preferences were available for comparing aspirin and paracetamol and this was not enough to enable a statistical decision to be made. There were 22 preferences for the paracetamol/placebo comparison and the active drug was clearly superior. In the remaining comparison 26 preferences were available; with a restricted design based on \( \theta = 0.85 \) there was no statistically significant difference between aspirin and placebo, while with the difference stipulation \( \theta = 0.80 \) no conclusion could be reached. In plain language this means that a study designed to find out whether or not there was a fairly large difference between aspirin and a placebo would have reached the conclusion that there was no such difference. A study designed to detect a somewhat smaller difference between the two medications would have failed to give an answer with the number of cases available. Since order of administration was not taken into account in charting the results sequentially it is worth noting that, taking all three comparisons together, the second medication was regarded as better on 40 occasions and the first on only 21.

**DISCUSSION**

This study shows that with careful technique and training a wholetime nurse observing orthopaedic patients in the postoperative period can show a satisfactory distinction between a mild analgesic and a placebo, but not between approximately comparable doses of different mild analgesics. The use of second-dose data, whether because of loss of cases or for some other reason, gives a less clear distinction. The study may thus be said to confirm many of the conclusions of the previous investigation (Parkhouse, Collie and Wood, 1967).

The number of cases not requiring a second dose was a disappointment. It illustrates once again the difficulty of predicting the course of a study of this kind and deciding beforehand on the most appropriate design and method of statistical analysis. Sequential analysis alone, based on preferences between doses, would have more
time-consuming and would have required a greater number of patients than a single-dose t-test analysis; in addition there would have been a very considerable wastage of information from patients needing only one dose. It could also be said that in the case of one comparison, that between aspirin and placebo, sequential analysis alone would have given a completely opposed answer to the t-test and a much less satisfactory one. Even a more cautious choice of sequential design would have had only the negative advantage of enabling no decision to be made. This is, however, not so much criticism of the sequential method itself as of the way it was applied and the data it was applied to: it perhaps adds weight to the belief (Sunshine et al., 1964) that in this type of study “cross-over” data are a less sensitive index than first-dose data alone; it certainly suggests that order effects should be taken into account, thus further adding to the number of patients needed to reach a conclusion.

No side effects of any consequence were noted and it is no surprise that this should be so in a series of this size. The reported incidence of symptoms after any drug is likely to vary greatly with different observers, as has been shown to be the case with different physicians (Joyce, 1962) and psychiatrists (Reynolds et al., 1965). Mild drowsiness, as a case in point is as likely to be “in the eye of the beholder” as in the mind of the patient.

**COMMENTS ON METHOD**

Some general comments on this type of investigation may perhaps be appended, with particular reference to the interpretation of pain and relief scores.

*Subsequent doses and their significance.*

Common sense suggests that if a patient requires only one dose of an analgesic his pain is self-limited, whereas if he requires a subsequent dose the underlying pain has persisted. It would, on these grounds, be reasonable to expect a higher incidence of “apparent placebo reactors” among those patients who require no subsequent medication, and thus perhaps greater difficulty in discriminating between medications. In our Studies One and Two (Parkhouse, Collie and Wood, 1966) there were a few obvious cases in which a very large relief score was obtained after a placebo and no subsequent medication was needed; it was difficult to escape the feeling that in these cases pain had disappeared, or greatly diminished, spontaneously. These cases were, of course, included in the final analysis but not without our feeling some doubt as to whether they might more reasonably have been excluded. The number of drop-outs after first medication in Studies One and Two was too small to permit further study of the problem; there were rather more drop-outs after second medication and table IV of the present paper shows that there was evidence of a difference in response to the second medication according to whether or not a third medication was required.

Although the large number of drop-outs after first medication in the present study was disappointing it provided an opportunity to look once more at this question. Table V shows the first-dose mean total relief scores for the three medications in those patients who required, and did not require, a second medication. It is immediately obvious that there is an enormous difference in the mean relief score for each medication in these two groups. On the other hand, the distinction between medications is not appreciably clearer in one group than the other. Furthermore the number of patients requiring a second medication is very different in the case of the placebo and the two active drugs. Again, this is explicable on commonsense grounds: if a patient has a pain which persists for several hours he may well be content with a single dose of aspirin or paracetamol, but he is much less likely to be satisfied with a placebo and will often demand a further medication in the course of 1 or 2 hours. The need for a subsequent drug is, to some extent, a measure of the efficacy of the first drug.

It is now possible to examine what happens when results are combined for those patients who require a subsequent medication and those who do not. Suppose, for instance, that only one-quarter of the patients who receive an active drug require a second medication while three-quarters of the patients who receive a placebo require a second medication. When the data are pooled, the effect for the active drug will be to add a small number of patients with a low mean score (i.e. those having a second medication) to a large
A COMPARISON OF ASPIRIN AND PARACETAMOL

TABLE IV

Study Two*: second medication. Investigator's assessments (V.W.)

<table>
<thead>
<tr>
<th></th>
<th>Patients having third medication</th>
<th>Patients not having third medication</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean initial pain (SE)</td>
<td>Mean total relief (SE)</td>
<td>Mean initial pain (SE)</td>
</tr>
<tr>
<td>Aspirin and nepenthe</td>
<td>4.82 ± 0.71 (17)</td>
<td>7.90 (10)</td>
<td>5.96 ± 0.60 (27)</td>
</tr>
<tr>
<td>Aspirin</td>
<td>4.49 ± 0.58 (18)</td>
<td>6.00 (7)</td>
<td>4.92 ± 0.48 (25)</td>
</tr>
<tr>
<td>Nepenthe</td>
<td>3.41 ± 0.62 (17)</td>
<td>5.37 (8)</td>
<td>4.04 ± 0.47 (25)</td>
</tr>
<tr>
<td>Placebo</td>
<td>1.46 ± 0.38 (13)</td>
<td>6.00 (4)</td>
<td>2.53 ± 0.64 (17)</td>
</tr>
</tbody>
</table>

*Parkhouse, Collie and Wood (1967).

TABLE V

First medication. Patients' assessments.

<table>
<thead>
<tr>
<th></th>
<th>Patients having second medication</th>
<th>Patients having no second medication</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean initial pain (SE)</td>
<td>Mean total relief (SE)</td>
<td>Mean initial pain (SE)</td>
</tr>
<tr>
<td>Paracetamol</td>
<td>2.47 (17)</td>
<td>2.29 (17)</td>
<td>2.38 (34)</td>
</tr>
<tr>
<td>Aspirin</td>
<td>2.48 (21)</td>
<td>2.20 (15)</td>
<td>2.36 (36)</td>
</tr>
<tr>
<td>Placebo</td>
<td>2.54 (26)</td>
<td>2.00 (7)</td>
<td>2.42 (33)</td>
</tr>
<tr>
<td></td>
<td>1.62 ± 0.44 (26)</td>
<td>6.72 ± 0.41 (7)</td>
<td>2.70 ± 0.51 (33)</td>
</tr>
</tbody>
</table>

number of patients with a high mean score (i.e. those having no subsequent medication), while in the case of the placebo a large number of patients with a low mean score (i.e. those having a subsequent medication) will be added to a small number of patients with a high mean score (i.e. those having no subsequent medication). The result will be to magnify the difference between active drug and placebo. In practice, the distinction between patients who require subsequent medication and those who do not is arbitrary. The true clinical situation presents a continuous range of patients whose drug requirements vary from those demanding a subsequent medication within 30 minutes to those who require a subsequent drug only 15 or 24 hours later. One of the many factors on which this requirement will depend is the effectiveness of the first drug.

Although the above discussion provides a sound theoretical basis for including in statistical analysis those patients who require no subsequent medication, it still remains true that an even greater difference between active drug and placebo would result from excluding the patients who have a flagrantly gross "relief" from placebo and no further drug requirement. Apart from the practical difficulty of knowing where to draw the line, it must be accepted that the spontaneous remission of pain is a factor likely to operate just as much in the case of those patients who have received an active drug. Common sense cannot justify the exclusion of any case.

The importance of initial pain scores.

An important difference between the patient's assessment and the investigator's is that the former depends upon the initial level of pain, before medication, while the latter does not. Before a drug is given the patient is asked whether he would regard his pain as mild, moderate or severe, and this is entered on the data card as 1, 2 or 3. Thereafter, at hourly intervals, the patient is asked the same question and his reply is entered in the same way. The patient's assessment of relief is then derived by subtracting each of these hourly scores from the initial score; total
relief is obtained by summing these differences. The investigator’s assessment represents simply the investigator’s opinion, at each hour, of the degree of pain relief, this being graded as poor, moderate or good and correspondingly entered as 1, 2 or 3.

When a patient’s pain is severe to begin with there is usually little difference between these scores; but it sometimes happens that a patient whose initial pain is moderate, or only mild, appears to have obtained excellent relief from a drug. This is one of the reasons why investigators, as mentioned in the previous paper (Parkhouse, Collie and Wood, 1967), like to have the opportunity of recording their own impression in addition to the patient’s statement. A large difference can now occur: to take the extreme case, if the patient’s pain score before medication is only 1, the only improvement that is possible on his assessment is a complete disappearance of pain (scored as 0 = none) and even if this improvement were to persist for 6 hours the total relief score would only be 6. The investigator, however, may consider relief to be good at each of these 6 hours, thus entering 3 on each occasion and producing a total relief score of 18.

Reverting to the discussion above, it might be expected that patients requiring no subsequent medication would, on the whole, show a lower initial pain score than patients requiring subsequent medication. The appropriate column in table V shows that in the present study this was true. This raises a further point in connection with the scoring system. At first sight it is not surprising that patients with a lower degree of pain should show a higher degree of relief from a given drug; but as far as the patient’s assessment is concerned this is the opposite of what would be expected from the nature of the scoring system: when a patient has only a low initial pain intensity it is impossible for him to show as great a degree of relief as when his initial pain is severe. Table V was compiled from the patients’ assessments, rather than the investigator’s, and it is clear that any restriction imposed by the scoring system upon the measurement of relief from relatively mild pain was entirely insufficient to prevent the demonstration of an overwhelmingly greater relief in those patients whose initial pain scores were lower.

**The meaning of hourly scores.**

Some further thought may now be given to the meaning of individual hourly scores, for pain intensity and relief, as opposed to total scores.

To restate the situation: a drug is given; some patients continue to experience appreciable pain and request another drug within the first hour or two; some patients are content for a longer period of time; some never require another drug. Thus, at each hour after drug administration there will be a different number of patients who are still “scoring”. The problem is how to assess these hourly scores and how to make provision for the patients who have fallen out because a subsequent drug has been given.

The patient’s assessment is a pain score. When a subsequent drug has been given the pain score can arbitrarily be returned to its starting point, on the assumption that since effective relief has come to an end pain, for purposes of computation, should be regarded as having returned to its original level. Alternatively all remaining hourly scores for pain can be entered as 3. The investigator’s score is a relief score and this will usually have become 0 by the time a subsequent drug is given. The remaining hourly scores can be entered as 0 on the reasonable assumption that relief from this drug has ceased for evermore. Again, however, these are arbitrary and not real scores.

The most obvious way to treat hourly scores is to derive a mean at each hour from all patients, including those who have received a subsequent drug and for whom arbitrary scores have been entered. This creates a danger of overlooking variations in response to a drug, which may be a real disadvantage when, for example, pharmacogenetic differences are suspected. To give an illustration: suppose that three-quarters of a group of patients respond immediately to a drug while the one-quarter have a delayed response. Hourly relief scores for the majority of patients may read

\[
3 \quad 3 \quad 3 \quad 1 \quad 0 \quad 0
\]

while hourly relief scores for the minority read

\[
1 \quad 2 \quad 3 \quad 3 \quad 2 \quad 1.
\]

If all patients are included, at all hours, the mean relief scores will read

\[
2.50 \quad 2.75 \quad 2.20 \quad 1.50 \quad 0.50 \quad 0.25
\]
This gives the impression of a smoothly graded response in which the minority reaction is buried; for the later hours it is impossible to know whether all patients are still scoring but poorly, or only a few patients are still scoring but scoring well. If this method of dealing with hourly scores is adopted it is therefore necessary to state the number of patients contributing real scores at each hour, alongside the mean for all patients.

An alternative method of treatment is to set a time limit below which no patient will be included. It may be decided, for instance, that unless a patient achieves two or three real scores before receiving a subsequent drug he will be excluded from the study. We think this unreasonable, from the argument stated above, that the need for a further drug reflects the value of the drug already given. Even the requirement that one real score should be achieved is suspect, for it must be remembered that some patients will receive a placebo as a first medication and many of these patients will obtain no appreciable relief. If the doubtful concept of the placebo reactor (Parkhouse, 1963) is invoked the result of imposing a minimum "survival time" before a subsequent drug is given is to "screen in" the placebo reactors and to leave out of the study some of the most valuable patients, namely those who have appreciable pain and obtain no relief from the placebo. Sunshine et al. (1964) made the condition that patients must wait at least 3 hours before having a subsequent medication in order to be eligible for inclusion in their study; although the number of drop-outs on this account was small (Sunshine, personal communication) the potential disadvantages of the requirement are sufficiently clear.

The final possibility is to calculate mean hourly results only from those patients who are still achieving real scores. This will distinguish the quality of the relief obtained at each time interval from the number of patients achieving relief. Again, the number of patients contributing to each hourly score should be noted. In the hypothetical example given above, for instance, in which three-quarters of the patients obtained immediate relief and one-quarter obtain delayed relief, the mean hourly scores allowing for drop-outs would be

2.5 2.75 2.25 1.5 2.0 1.0

and the late rise due to the unusual response of a minority of patients can clearly be seen.

We compute mean hourly scores allowing for the drop-out of those patients who have received a subsequent drug (Parkhouse, 1967). Results in the present study, for first medication data, are shown in figure 4. The picture bears little superficial resemblance to the more usual hourly score representations in which all patients are included—for example figure 1 of Studies One and Two (Parkhouse, Collie and Wood, 1967). It is noteworthy that from the third hour onwards the number of patients remaining is no greater after paracetamol than after a placebo, although the mean score of the few remaining patients who received paracetamol is high. After aspirin more than one-third of the patients have not yet received a subsequent drug at 6 hours, and it is only after the fifth hour that the mean relief score for aspirin begins to fall. These observations lend support to the suggestion that a residual effect of aspirin may have been detectable in the second medication data. An interesting trend which is also revealed is the increase in relief following placebo with time, while the number of patients showing relief falls sharply. It may be supposed, once again, that there are some patients whose pain persists and some whose pain improves spontaneously during the period of assessment. The unmodified pain level, during these 6 hours, could thus be represented in some cases as

3 3 3 3 3 3

and in others as

3 2 2 1 1 0

Patients in the first group will tend to drop out after the first hour or two since they have received no relief from the placebo and have been given a further medication, while patients in the second group will show a relief score, even if the placebo has no effect, of

0 1 1 2 2 3

which corresponds with what is seen in figure 4. The same effect would not be expected to appear so clearly after an active drug since the drop-out tendency would be less marked: if a relief effect due to aspirin, for example, of

2 2 2 2 1 1
were superimposed on the two situations depicted above the result in the first group of patients would be a relief score of

2 2 2 2 1 1

and in the second group of

2 3 3 3 3 3,

resulting in a pooled mean hourly score of

2.0 2.5 2.5 2.5 2.0 2.0,

assuming equal numbers of patients in the two groups.

CONCLUSION

The temptation to "over-interpret" results in this type of investigation has repeatedly been stressed (Keats, 1966) but it is nevertheless important to try and understand the ways in which experimental design, recording of data and statistical analysis of results may influence the interpretation of a clinical phenomenon such as pain. Only in this way can a valid comparison be made between different investigations reported in the literature. There is no ideal means of dealing with subjective phenomena and it is for each investigator to choose a method according to his needs and his beliefs. The most important realization is that all such considerations are secondary to the requirement that the data themselves should be honestly and carefully recorded.

ACKNOWLEDGEMENTS

We are grateful to the surgical and nursing staffs of the Nuffield Orthopaedic Centre for their co-operation; to the Bristol-Myers company of New York for financial assistance; to Savory and Moore for supplying the capsules; to the Director of the Computer Laboratory at the Imperial College of Science and Technology for the use of computer time; to Mr. H. H. Johnson and Mr. C. D. James for computer programming; and to Mr. D. R. Golding and Miss Smith of the Oxford Regional Hospital Board Statistical Unit for key-punching I.B.M. cards.

REFERENCES


UNE COMPARAISON ENTRE L'ASPIRINE ET LE PARACETAMOL

AVEC UNE NOTE METHODOLOGIQUE

SOMMAIRE

Par une comparaison en double-aveugle entre l'aspirine, le paracétamol et un placebo dans la période post-opératoire chez des malades orthopédiques une infirmière-observatrice était aisément capable de distinguer les deux drogues actives du placebo après la première dose, mais la distinction était plus difficile après la seconde dose. Une analyse séquentielle pour mettre en évidence la préférence des différents malades pour l'une ou l'autre des deux drogues actives avait moins de succès. On a noté peu d'effets secondaires. Quelques problèmes touchant l'évaluation subjective de la douleur et de sa disparition sont discutés.

VERGLEICH VON ASPIRIN UND PARACETAMOL

MIT EINER BEMERKUNG ZUR METHODIK

ZUSAMMENFASSUNG