Observations on spontaneous improvement in patients with podagra: implications for therapeutic trials of non-steroidal anti-inflammatory drugs

N. BELLAMY¹, W. WILSON DOWNIE² & W. WATSON BUCHANAN³

¹Department of Medicine, University of Western Ontario, London, Ontario, Canada, ²Warner Lambert Co., Morris Plains, New Jersey U.S.A. and ³Division of Rheumatology, McMaster University Health Sciences Centre, Chedoke-McMaster Hospitals, Hamilton, Ontario, Canada

1 The natural history of acute gouty arthritis was studied in 11 volunteers with podagra.

2 Two patients withdrew from the study on day 4 because of severe persistent pain. Of the remaining patients all showed some improvement in pain by day 5 and in swelling by day 7. Tenderness improved in seven patients by day 7 but two continued to experience the same amount of discomfort as at trial entry. In spite of these improvements only three patients noted resolution of their pain during the study period.

3 These data indicate that while the majority of patients show spontaneous improvement, resolution is unlikely over a period of 7 days without the use of effective non-steroidal anti-inflammatory medication.

4 Documentation of the natural history of the acute gouty attack may assist clinical investigators in interpreting the results of uncontrolled evaluations of non-steroidal anti-inflammatory drugs.

Keywords acute gouty arthritis historical controls clinical therapeutic trials non-steroidal anti-inflammatory analgesics

Introduction

It is commonly accepted that for the majority of patients podagra is self-limiting, albeit painful disorder in which spontaneous recovery can be expected (Wyngaarden & Holmes, 1979). While historical references attest to the finite course of this common disorder, few document the rate at which improvement occurs. In the execution of clinical trials in patients with gouty arthritis, one of two designs is commonly followed (Rosenbloom *et al.*, 1985). The first employs an unblinded quasi-experimental one-group design in which all patients receive the therapeutic agent of interest and there is no comparison or control group. In the second, a double-blind parallel

design is adopted in which patients are randomly allocated to treatment with one of two pharmacologically active agents. Irrespective of the design employed, the results of such trials invariably demonstrate resolution of the acute attack. However given historical evidence for spontaneous improvement, this outcome is only to be expected and therefore interest should be focused on the rate at which improvement is achieved rather than on its actual attainment. No data currently exist to our knowledge which document the outcome of untreated patients with gouty arthritis. The study reported was undertaken in order to define the natural out-

Correspondence: Professor W. W. Buchanan, Division of Rheumatology, McMaster University Health Sciences Centre, Chedoke-McMaster Hospitals, 1200 Main Street West, Room 2F10, Hamilton, Ontario, Canada L8N 3Z5

come of a group of patients who may be regarded as phantom controls in the untreated half of a conventional controlled study of a non-steroidal anti-inflammatory drug *vs* no treatment.

Methods

Patients were eligible for the study if they presented with classical features of acute podagra, had experienced prior attacks of gouty arthritis and were known to be or have been hyperuricaemic. All patients agreed willingly to participate in the study after a full discussion of its nature. Following admission, patients were permitted to continue their prior urate-lowering drugs, but did not receive any anti-inflammatory or analgesic agents. Patients were however provided with bedrest and could elect to withdraw from the study at any time and receive therapeutic doses of indomethacin together with analgesics as required. Patients were assessed daily for 7 days. Pain, tenderness, juxta-articular skin temperature (heat), swelling and erythema were each scored on Likert-type adjectival scales (nil = 0, mild = 1, moderate = 2, severe = 3, very severe = 4). Since such ordinal level data have been shown to be amenable to analysis traditionally reserved for interval level data (Gaito, 1980), a group mean and standard deviation were calculated for each variable on each day of the study. In accordance with statistical doctrine however the within-group response was analyzed using non-parametric statistical techniques (Conover, 1980). In order to account for any patients withdrawing from the study, the response data were analyzed in two ways. The first analysis was based on available data only, i.e. patients contributed no further data to the analysis after withdrawal. It was reasoned however that if several patients with severe disease withdrew at some point in the study that an apparent improvement in the group mean might occur even though all remaining patients were unchanged. Therefore in the second analysis, the termination values of the five variables of interest from any withdrawal were extrapolated as that patient's values for subsequent study days. This manoeuvre would effectively negate any bias which was simply generated by patient withdrawal and provide a more conservative estimate of the response.

Results

All 11 patients who entered the study were male and all had a prior history of hyperuricaemia.

Two had gouty tophi, five had past urinary calculi and two had both tophi and calculi. The mean age of the study group was 55 years (range = 43–67). All patients had experienced prior attacks of gouty arthritis (range = 1–20 attacks, median = 4) extending over 1 to 7 years (mean = 4.1 years). Nine patients were receiving allopurinol, one probenecid and the final patient was not taking any prophylactic therapy at the time of the study. The most recent attack of podagra had been present for between 1 and 5 days (mean = 2.8 days) prior to enrollment. At entry, mean scores $\pm$ 1 s.d. on the five variables were as follows: pain = 3.73 ± 0.47 , tenderness = 3.91 ± 0.30 , heat = 2.73 ± 0.79 , swelling = 2.18 ± 0.60 and erythema = 2.09 ± 0.70 .

Nine patients completed the 7 days of the study, the remaining two patients being withdrawn on day 4 and responding to therapeutic doses of indomethacin. As previously noted in the first analysis only actual data were used whereas in the second analysis withdrawal scores were perpetuated as that patient's subsequent score for each day until the completion of the study. Results are expressed as change from baseline in terms of mean scores $\pm$ 1 s.d. for each day of the study in Figures 1, 2, 3, 4 and 5. An increase of one or more points on a variable has been termed worse and a decrease of one or more points termed better.

Two patients noted improvement in pain by day 2, the remaining seven who completed the study achieving improvement by day 5. No patient deteriorated with respect to pain during the study. The two who were withdrawn had experienced very severe sustained pain since the day of entry. Tenderness had improved in only one patient by day 2 and in one case there was transient deterioration until day 3. Of the nine patients who completed the study, tenderness improved in seven. Swelling improved in one patient by

Figure 1 Pain. Group response showing daily mean and s.d. using actual (—) and extrapolated (---) data ($P \leq 0.05$ for comparison with baseline: actual = *, extrapolated = †).

Figure 2 Tenderness. Group response showing daily mean and s.d. using actual (—) and extrapolated (---) data $P \leq 0.05$ for comparison with baseline: actual = *, extrapolated = †).

Figure 3 Heat. Group response showing daily mean and s.d. using actual (—) and extrapolated (---) data $P \leq 0.05$ for comparison with baseline: actual = *, extrapolated = †).

Figure 4 Swelling. Group response showing daily mean and s.d. using actual (—) and extrapolated (---) data $P \leq 0.05$ for comparison with baseline: actual = *, extrapolated = †).

day 2, the majority of patients improving by day 5. In one patient who was withdrawn joint swelling increased, while in another swelling was both severe and persistent. Juxta-articular skin temperature failed to decline in any patient until day 3 but improved by day 4 in all patients who completed the study. In no case was any deterior-

Figure 5 Erythema. Group response showing daily mean and s.d. using actual (—) and extrapolated (---) data $P \leq 0.05$ for comparison with baseline: actual = *, extrapolated = †).

ration noted. Erythema failed to decline until day 3 and in one case actually worsened. However by day 3 the majority of patients improved and by day 5 all patients who completed the study had improved.

A score of 0 or 1 on any variable may be reasonably regarded as representing the point of resolution of the acute attack. Of the five variables assessed, pain reduction was considered to be the most important clinical response. When assessed using the aforementioned criteria, the following numbers of patients demonstrated resolution of the acute attack by day 7: pain = 3, tenderness = 0, heat = 9, swelling = 9, and erythema = 9 patients.

In a clinical trial, a comparison of baseline with subsequent values provides a statistical test of the within-group response. In order to assess whether the spontaneous response in these gouty subjects reached statistical significance and might therefore be confused with the response of a group of patients undergoing active treatment, we compared baseline with subsequent values using both the Wilcoxon and Mann Whitney U statistical tests (Colton, 1974). As noted previously two analytic philosophies were compared. Thus in Figures 1, 2, 3, 4 and 5, the values generated by actual data are shown in the solid lines and those generated using the extrapolation procedure for day 4 onwards shown in dotted lines. The second analytic strategy invariably resulted in higher means on each variable for days 4 to 7 of the study and in general to larger standard deviations. With two exceptions the method of analysis did not affect the statistical interpretation of the data set. Thus analysis of both actual (unpaired) data by the Mann Whitney U test and of extrapolated (paired) data using the Wilcoxon test showed statistically significant differences between baseline and days 4, 5, 6 and 7 of the study for each of the five variables

(Figures 1, 2, 3, 4 and 5). The only differences observed were with respect to heat which statistically improved by day 3 when assessed by the Wilcoxon but not by the Mann Whitney U statistic (Figure 3) and in tenderness in which significant improvement was noted at days 4 and 5 by the Mann Whitney U but not by the Wilcoxon statistic (Figure 2). Overall therefore it can be stated that based on actual data, a sustained, clinically important and statistically significant improvement in all variables occurred as early as the fourth day following enrollment in the study.

Discussion

This study was prompted by the necessity to obtain accurate data on patients with acute gouty arthritis from whom anti-inflammatory analgesic therapy was withheld. Although for obvious reasons the number of patients enrolled in this study was small, it is clear that the majority of patients improved within 7 days of admission to the study. Furthermore group responses were statistically significant from the fourth day after enrollment onwards. Patients had experienced podagra from between 1 and 5 days prior to enrollment. Since this is often the case in patients admitted to drug trials, no attempt was made to define the natural history from the onset of the attack, but only the 'natural history' from the point at which the patient would have been

eligible for inclusion in such a study. It has been noted that although the majority of patients did improve over a 7 day period only less than a third of patients can be expected to note resolution of their major symptom, pain, during the same period.

Although the use of concurrent randomized placebo treated or untreated control subjects facilitated a more rigorous methodological design, the routine use of such a manoeuvre cannot be ethically justified. It is in such a situation that historical controls may play an important role in the acquisition of new knowledge (Bailar *et al.*, 1984) regarding the potency of anti-inflammatory agents. It is clear from the present data that spontaneous response in gout is substantial and care must be exercised in interpreting the results of therapeutic trials.

In conclusion, we believe that the data presented in this paper will provide a useful historical reference for those carrying out therapeutic trials of anti-inflammatory drugs in acute gouty arthritis. It is difficult today to justify the use of a placebo control group, when proven remedies for this condition are available. By referring to the data in this report, it is possible to compare the performance of a new agent against the natural history of the disease and thereby estimate its true efficacy.

We wish to thank the Arthritis Society of Canada and Warner Lambert Laboratories for financial support.

References

- Bailar II, J. C., Louis, T. A., Lavori, P. W. & Polansky, M. (1984). Studies without internal controls. *New Engl. J. Med.*, **311**, 156-162.
- Colton, T. (1974). *Statistics in medicine*, pp. 219-227. Boston: Little Brown & Company.
- Conover, W. J. (1980). *Practical nonparametric statistics*, pp. 213-338. New York: John Wiley & Sons.
- Gaito, J. (1980). Measurement scales and statistics: resurgence of an old misconception. *Psych. Bull.*, **87**, 564-567.
- Rosenbloom, D., Brooks, P. M., Bellamy, N. & Buchanan, W. W. (1985). Critical evaluation of anti-rheumatic drug therapy: A review of clinical trials. In *Clinical trials in the rheumatic diseases*. New York: Praeger Publishers.
- Sackett, D. L. & Ghent, M. (1979). Controversy in counting and attributing events in clinical trials. *New Engl. J. Med.*, **301**, 1401-1412.
- Wyngaarden, J. B. & Holmes, E. W. (1979). Clinical gout and pathogenesis of hyperuricaemia. In *Arthritis and allied conditions*, ed. McCarty, D. J., pp. 1193-1228. Philadelphia: Lea & Febiger.

(Received 5 January 1987,
accepted 13 February 1987)