TREATMENT AND ILLUSTRATIVE ANSWERS

E. G. L. Bywaters

We cannot at present cure rheumatoid arthritis in the same sense that penicillin cures syphilis; the treatments hitherto available and those coming from the drug houses each week are numerous, sometimes dubious, often complementary and essentially palliative. They are intended to decrease the damage done to the joints by the disease, so that when it remits, as it frequently does, the patient is left without severe or crippling deformity. In general, treatment is not yet satisfactory and the physician has therefore an obligation to try new treatments provided he knows that they are not harmful or appreciably worse than existing methods.

Treatments utilized today, most of them still needing adequate assessment, include bed rest, splintage, exercise and physiotherapy in various forms, analgesic drugs, antiphlogistic drugs, muscle building anabolic drugs, surgical procedures and even morale builders like visits to the doctor or iodine lockets. In addition there are empirical treatments such as gold or antimalarials, often introduced on an erroneous hypothesis.

It is thus essential that one of these agents should be tested at a time with a uniform background of ancillary or complementary measures such as I have listed.

Furthermore with this chronic and fluctuating disease the therapeutic agent must be tested over a period long enough to reveal its usefulness as a long-term procedure; this will usually mean outpatient supervision and an inevitable increase in lapse rate. A long period of assessment is necessary not only to obviate the normal fluctuation of the disease but also because of habituation to the drug and to reveal the manifestations of drug toxicity, which in many instances have proved the limiting factor in clinical practice.

There are two major designs of trial we have used: one where each patient experiences in turn (but in random order) each type of treatment or, more commonly, one where several groups of

patients, each selected at random and matched as groups or in pairs for all important factors, are each given different treatments. These two designs may be combined.

While all trials should be conducted "blind" – that is the patient should not know whether he is getting the test treatment or the control – this sometimes presents difficulties, as when in one hospital ward one patient is treated by strict bed-rest and another with the same condition by active exercise. The *double* blind trial, where neither the patient nor the treating doctor knows which drug is being used, is better because it also eliminates doctor bias. It is being currently used in the assessment of gold therapy.

Allocation to treatment must be randomized either individually or in matched pairs. Stratification is an advantage and ensures that random allocation to test and control drug forms groups which contain roughly equal numbers of subgroups defined according to age, sex or any other important factor which may be thought to have an effect upon therapeutic response. The most important factor governing outcome in rheumatoid arthritis is whether the patient comes to hospital during the first year of his disease or after it. In a multicentre trial it is important that each centre should have a roughly equal number in the test and control group.

The participating clinicians have to decide on one of two types of treatment schedule – fixed dose or variable dose – and to decide to stick to it.

1. Fixed dosage is the tidiest plan, since the treating doctor can be blind in the sense used above. Its disadvantages are that all patients are not on optimal dosage and that some may have to be withdrawn because of toxicity.

2. A variable dosage trial where the drug is adjusted to the patient's optimum requirements is a more natural method, where the physician can steer his usual course between overdosage and therapeutic level. With many modern drugs the toxicity-efficacy ratio is very variable from patient to patient and may entirely determine the value of a particular drug for one patient or for the whole group.

A variable dosage scheme suffers from two defects. Firstly, in comparing a new drug with an older one like aspirin the treating physician cannot be blind because he must be aware of the specific toxic symptoms of both which determine the upper level of his variable dosage: thus a double blind trial with independently variable

dosage of both control and test drug is only possible if the assessor is brought in from outside and measures responses without reference to the treating doctor. Only for a trial of placebo against test drug is it possible to be blind. Here each patient is treated as if on test therapy from the viewpoint of dose variation.

Secondly, an even worse defect is that the test in a variable dosage scheme is not primarily of the drug but of the skill of the physician in adjusting the dosage. This is not a reproducible phenomenon and any one physician might be better at or more interested in adjusting the dose of drug A than that of drug B. If the trial is conducted in a number of different centres these different skills or biases may cancel out or merge in the residual variability, but if the trial is conducted by one physician only, even though assessment is done by someone else, the results may not be reproducible by another physician in another centre.

Table I summarizes the advantages and disadvantages of trials in rheumatoid arthritis conducted with fixed dosage and variable dosage.

TABLE I

CONTROLLED TRIALS IN RHEUMATOID ARTHRITIS

	Fixed Dosage	*Variable Dosage*
Advantages	Trial can be "double blind" Less dependent on individual physician's skill and bias Can be done in one centre	More natural: tests drug at a level in each patient which is considered optimal by physician in charge Less withdrawal due to toxic complications
Disadvantages	Drug not used at optimum level in all patients Some may have to be withdrawn because of toxic side-effects	Impossible to run double blind unless (*a*) Outside assessor (tests physician's skill and must be multicentre). (*b*) Control drug is placebo; all patients treated as if on test drug (better multicentre).

The next two topics I want to raise for discussion are concerned with the analysis of the trial. In any long-term trial carried on, as we

have done, over a period of three years, some patients will emigrate, others will dislike their doctors and go elsewhere, some may die, others will recover and refuse treatment, still others may have their treatment changed to something else on the grounds that the evil we know is better than that we know not. How should these withdrawals be treated? The ideal thing is to have none, but that is only possible in a trial lasting a few hours. We have used two methods.

1. We have tried to define carefully in advance under what circumstances a patient should be withdrawn from consideration, and have confined our analysis to those still remaining in the trial and on the specified treatment at each annual point. If this is done it is essential that comparisons should be made of the starting state of each residual group at each point of time. Careful consideration must be given to the reasons for withdrawal within each treatment group. Thus an equal number in each group could be withdrawn, but drug A withdrawals could all be because having got better they failed to attend, and all drug B withdrawals could have been changed to drug A. A high completion rate should be built into the dosage of the trial, but this is never completely achieved.

2. In the second method we have analysed the group as a whole at each point of time, excluding only those who could not be assessed due to death or non-attendance, irrespective of whether they have remained on the specified therapy. My own feeling is that this is essentially a trial less of continued maintenance therapy than of one particular limited course of treatment and its residual effects. There are objections to both these methods.

Finally, I think I should point out that when we come to analysis, if we have 20 different and unconnected criteria of improvement, a significant difference between the treated and control group is likely to occur by chance (according to our definition of chance) in at least one criterion even if there is no difference between the treatments. This is impossible to allow for statistically, since amongst 20 criteria there are all degrees of interrelatedness. I do not think that artificial indices of improvement, obtained by adding together (with appropriate weightings) various different criteria, serve any purpose other than to muddle the whole affair.

Finally I turn, as a practical example, to a trial with which I was most intimately concerned, and which can be described under the following headings.

This was a multicentre trial, under the auspices of the Empire

Rheumatism Council, and it was most important that all centres used the same criteria. It has been helpful to appoint a co-ordinator specifically to travel round from unit to unit to ensure comparability between units.

We selected people with undoubted rheumatoid arthritis, for at least one year, excluding other diseases, those too young or too old to co-operate, and those who were unable to walk, because of difficulties in assessment. We endeavoured to improve the completion rate by excluding also those who might be liable to withdrawal from the trial on account of pre-existing hypertension, gastric symptoms, or mental instability.

One hundred patients were taken into the trial – 50 patients on aspirin at a specified variable dosage and 50 on cortisone, again at a specified variable dosage (Table II). If the dosage is variable it is most important to specify within what limits it is to be varied, why it is to be varied, and the rule governing withdrawal.

TABLE II

E.R.C. CORTISONE-ASPIRIN TRIAL 1955: TREATMENT

Basic régime of splints and physiotherapy plus aspirin

Start: 4 g./day in five divided doses for one week or until max. effect, then adjust up or down between 1·0 and 8 g./day to minimum dose which maintains max. function without side-effects.

Other equivalents (e.g. Ca Aspirin) may be used.

or

plus Cortisone.

Start: 25 mg. three times daily p.o. until max. effect, then reduce by 12·5 mg./day at intervals of not less than one week till lowest dose on which patient is largely free from symptoms.

If maintenance needs 100 mg./day for over one month, withdraw patient from trial. If dose can be reduced to 25 mg./day, discontinue therapy after one month, restarting if necessary at 75 mg./day within one month or at 25 mg./day after that time.

Assessment was of function in four grades, of employability in four grades, and a subjective assessment by the patient in terms of "100 per cent well". The number of joints involved was also noted and the meaning of "involvement" was specified. We also compared X-rays, sedimentation rates, haemoglobin, and the number of complications occurring due to therapy (Table III).

TABLE III

E.R.C. CORTISONE-ASPIRIN TRIAL: ASSESSMENTS

Employment (4 grades)	I Fit for previous job. II Previous job modified. III Changed to lighter job. IV Unfit for Employment.
Function	I Normal.
1. Objective	II Moderate limitation. III Serious limitation. IV Helpless.
2. Subjective	100%, 75%, 50%, 25% or 1% fit.
No. of joints affected out of 68 possible	Any two of the following three: 1. Tenderness and pain. 2. Limitation. 3. Swelling.
X-ray of paired joints (hands).	
E.S.R., Westergren, mm./hr.	
Haemoglobin g.%.	
Complications.	

These assessments were entered on a pro forma for each patient. The essential readings were at 0, 6, and 12 months and thereafter six-monthly, but spaces were left for intermediate readings because these are very useful in checking the critical measurements and these patients had in any event to be seen monthly for purely therapeutic reasons.

The next important thing was the comparability of the two groups at entry into the trial. Once a patient was found to be suitable for the trial he was allocated to one or the other group at random with stratification for sex and centre. This allocation was done centrally and not in the particular unit. Comparability existed in respect of measured criteria and was similar initially whether we considered the starters, the 77 who were available at the end of the first year, or the 53 available at the end of the third year. Besides cortisone and aspirin each group had the same basic adjuvant therapy – splints and physiotherapy. The similarity at entry also extended to employment status and functional status, both objective and subjective.

Over the three years a total of 47 patients were withdrawn – 21 from the cortisone group and 26 from the aspirin group (Table IV).

These withdrawals were caused by death or intercurrent disease, because patients failed to attend, because of toxic complications needing change of therapy, or because of deterioration again needing change of therapy. It will be seen that the number of people changing therapy was about equal in both groups. The complica-

TABLE IV

E.R.C. CORTISONE/ASPIRIN TRIAL, WITHDRAWALS DURING THREE YEARS

Reasons for withdrawal	*First year*		*Second year*		*Third year*		*Total*	
	C	*A*	*C*	*A*	*C*	*A*	*C*	*A*
"Toxicity"*	4	5	2	0	2	0	8	5
Deterioration and/or change of therapy	6	1	0	3	3	4	9	8
Non-attendance	2	5	0	0	0	5	2	10
Intercurrent disease	0	0	0	1	1	1	1	2
Death (non-related)	0	0	0	1	1	0	1	1
Total	12	11	2	5	7	10	21	26

* *Cortisone:* Hypertension 2; perforated peptic ulcer, indigestion, headaches, psychosis, oedema of legs, and healed T.B., one each.

Aspirin: Indigestion, 5.

tions were slightly more frequent in the hormone group and were rather more serious. These withdrawals meant that we had for analysis a decreasing number of patients each year and the trial could not be conducted beyond the third year.

Apart from complications needing withdrawal, minor complications were few, but slightly more common in the hormone treated group. They included indigestion and hypertension (Table V).

At the end of the one, two, and three years there was some general

TABLE V

E.R.C. CORTISONE/ASPIRIN TRIAL COMPLICATIONS

(excluding those severe enough to need withdrawal from the trial)

Cortisone		*Aspirin*	
Indigestion	7	Indigestion	3
Depression	2	Nausea and tinnitus	2
Mild intercurrent infection	4*	Mild intercurrent infection	4
Oedema	2		
Hypertension	4		

* In addition, one case developing pulmonary tuberculosis is listed under withdrawals due to intercurrent disease.

improvement in both groups of patients (Table VI). This improvement was statistically significant in regard to subjective assessment, the number of joints affected which fell from roughly 20 to about 10, to the E.S.R., which fell, and to haemoglobin, which rose. There was little difference between the hormone group and the aspirin group; although haemoglobin improved significantly at the first year in the hormone group compared with the aspirin group, the position was reversed at the third year. By and large there was little difference between the two groups in regard to results.

TABLE VI

E.R.C. CORTISONE-ASPIRIN TRIAL: RESULTS

Means at		*Start*	*1 year*	*2 years*	*3 years*
Employment status I – IV	C	2·4	2·2	2·1	2·2
	A	2·4	2·1	2·2	2·1
Functional capacity I – IV	C	2·3	2·1	2·1	2·1
	A	2·3	2·1	2·2	2·2
Subjective assessment, 100% well – 1% well	C	49	(69)	(67)	(70)
	A	60	(69)	65	65
No. of joints affected per person	C	20·5	(11·2)	(13·1)	(10·2)
	A	18·0	(9·3)	(10·0)	(10·2)
E.S.R. mm./hr.	C	40·2	(23·4)	(23·9)	30·9
	A	35·8	(24·3)	29·4	30·0
Haemoglobin g.%	C	12·6	(13·3)	12·8	12·6
	A	12·8	13·1	12·4	(13·5)
Daily dosage mg.	C	75·0	68·7	67·5	67·1
gr.	A	58·7	58·6	54·7	52·5
Number left in trial	C	49	38	36	29
	A	50	39	34	24

Means which are significantly different from those at start of trial are in parentheses.

The X-ray changes showed some degree of worsening in both groups over the period of years, but the only significant difference was in the second year, when rather more patients in the aspirin group had new joints involved than in the cortisone group. This difference was not apparent at the third year. Thus I would emphasize that the comparisons made in this trial, and in any other trial, hold good only for the particular conditions of that trial and the particular selection of patients made for it. It is a commonplace when you have published such a trial to receive letters of criticism saying if you had done the trial in a different way you would have had different results. The proof of this is up to the critic, and the only way of learning is the hard way, to do it. In this respect it is most important in a multicentre trial to acknowledge fully the work of all colleagues whenever the results are shown.

The trial I have quoted most extensively from was conducted by the Empire Rheumatism Council at the following centres, with the participation of—

Drs. O. Savage and J. H. Glyn (Middlesex Hospital),
Prof. R. E. Tunbridge (Leeds),
Drs. J. J. R. Duthie and J. D. E. Knox (Edinburgh),
Prof. E. G. L. Bywaters, Drs. Barbara Ansell, G. R. Fearnley and A. Aranoff (Hammersmith Hospital),
Drs. E. T. D. Fletcher, J. H. Jacobs and Evelyn Hess (Royal Free Hospital),
Drs. G. D. Kersley and L. Mandel (Bath),
Drs. H. F. West and Margaret Fisher (Sheffield),
Drs. W. S. C. Copeman, O. Savage and Mary Joule (West London Hospital),
Drs. F. Dudley Hart, J. Golding and D. Burley (Westminster Hospital).

Dr. Lewis-Faning was responsible for the statistical analysis and Dr. Ivor Williams for the X-ray analysis.