The Intradermal Tuberculin Test as a Research Tool*

ITS PLANNING, EXECUTION, ANALYSIS AND INTERPRETATION

C. A. E. G. DE VILLE DE GOYET, *Tuberculosis Research Unit of the South African Medical Research Council, Onderstepoort, Tvl*

SUMMARY

Epidemiological studies and controlled comparative trials in tuberculosis require precise measurement of delayed hypersensitivity to tuberculin, and only the intradermal test can provide this accuracy. The correct procedures for random sampling, random allocation of treatments and blindness of assessment are discussed. Some basic principles of planning, collecting and recording data, analysing and interpreting of results, are illustrated by two examples.

S. Afr. Med. J., **47**, 142 (1973).

The tuberculin test is a most important procedure in any tuberculosis control scheme, and fortunately, it is also a very reliable biological test. Two developments have changed the emphasis. Today's statistical analysis of biological data requires measurements, and this is well possible with the Mantoux test. Secondly, with uniformity of substance and procedure, it is possible to compare different populations and different control schemes on national and international levels. The Mantoux test is used world-wide for surveys of prevalence and incidence of tuberculosis, and in the control of vaccination programmes. Because of the realization of widespread nonspecific sensitization, it is now often used in a comparative test of mammalian and another PPD. The significance of the dose and of the size of the reaction, has been recognized in most countries. Using a standard technique and standard substance does not exempt the medical researcher or epidemiologist from meeting some basic statistical and scientific requirements. The intradermal Mantoux test, with a low dose of tuberculin (2 TU),[1] is the most exacting method—and for this very reason, the most rewarding for the scientist.

A complete experience in the solving of a problem involves the following steps, and I am quoting Toole:[2]

'1. Becoming aware of a problem and formulating it.
2. Planning an investigation.
3. Collecting and recording the data.
4. Organizing and presenting the data in tables and graphs.
5. Analysing the data to discover important facts and relationships.
6. Interpreting the results of the analysis, making estimates of the parameters in the 'population' (universe) and stating the reliability of the estimates.
7. Making a report on the project.'

The second and third steps determine the over-all standard of a study.

Two different statistical situations will be considered—an epidemiological survey and a comparative trial.

Random Sampling and Random Allocation of Treatments

In an epidemiological survey we intend to obtain the best estimate of the prevalence of infection, by measuring the level of delayed hypersensitivity to a specific tuberculin in a 'population'. In a statistical context, population (universe) means the total group of humans, animals or other elementary units under study. In drawing and studying a sample from this population, we *estimate* some important characteristics or parameters of this population. An estimate of a parameter can be obtained from any sample, whatever its size, but no sample can bring more than an estimate of a parameter. A statement of uncertainty is an intrinsic part of research. This uncertainty can be estimated in terms of probability based on mathematical theory, but only if the sample is a random sample, i.e. every elementary unit of the population has the same chance to be selected. The experts are unanimous that one cannot rely on samples selected purposely, and the only safe practice is to have the sample chosen by the operation of chance. Persons encountered by a tuberculin test team in a kraal, or the schoolchildren present at school on the day of a skin test, are not representative samples of the kraal or the children population, respectively. (In the first case, the absence of the men away at work, and absenteeism due to sickness or unequal opportunities to attend school in the second case, are a few of the factors which can bias the results.) The scientist must ensure that all steps are taken to avoid any distortion or bias in the sample, caused by any known or unknown factor.

In a comparative trial, one or more treatments are compared with a standard or control group. (Human PPD may be compared to avian PPD, 2 units of PPD compared to 5 units of PPD, tuberculin hypersensitivity

*Date received: 1 August 1972.

after intradermal BCG vaccination may be compared with that after percutaneous BCG). The term treatment is being used in its widest sense as any process whose effect one wants to measure. Each group should be representative of the whole of the population in the trial; furthermore, all possible unknown factors must be equally distributed over the different treatment groups. Two procedures can be adopted.

1. All treatments may be applied to the same individual without fear of interaction, for instance, 2 and 5 units of human PPD, or two types of specific PPDs. However, sites or sides can possibly influence the reaction, therefore random allocation of each injection to every site is required. A simple, satisfactory method, as used in our trials, is explained below.

2. Only one treatment can be applied to each individual, because of possible interaction between the treatments, as would be the case with BCG vaccination.

A treatment must be randomly allocated to every single unit of the population under study. Allocation of two treatments, A and B alternately, such as A, B, A, B, A, B, A, B, is not a randomized allocation, since the chance of getting treatment A is 100% for the third individual and nil for the fourth, once the start and process of the sequence is fixed. A dice, or more convenient, random digit tables, should be used.

Suppression of Bias and Errors

An experiment must be designed in such a way that any bias is suppressed and errors are minimized and/or estimated. The word 'errors' is not synonymous with 'mistakes', but includes all types of extraneous variation which tend to mask the effect of the treatments.

Bias is prevented by randomization of the sample and blindness of the reading. The purpose of blindness, is to eliminate any preference, conscious or unconscious, by the person assessing the result, be it the reading of a tuberculin reaction or a clinical assessment. Consequently, the reading of tuberculin reactions should be carried out without knowledge of the relevant treatment, i.e. type of PPD or dosage of the injection or vaccination status of the person.

Sources of errors are many, and must be minimized or estimated. Sampling errors can be overcome by correct random selection and optimal size of the sample. The interested reader is referred to the excellent work of Sanchez-Crespo.[3] Experimental errors can be minimized in work with tuberculin by observing the following: (*i*) precise quantitation of the antigen, i.e. use of highly standardized PPD only; (*ii*) reduction of glass adsorption by adding Tween 80; (*iii*) use of fresh dilution from powdered substance; (*iv*) precision in diluting; (*v*) precision in injecting (high-quality syringe, well-trained staff).

Most of these requirements are best fulfilled by intradermal injection of a low dose of PPD (Mantoux test). Comparison of multiple puncture and intradermal methods is outside the scope of this article and will be the subject of a separate publication. As mentioned in the introduction, advanced statistical analysis demands quantified measurements. Attempts to quantify the multiple-puncture test (9 points of Collins)[4] have not yet been proved successful. The measurements of the transversal diameter of induration in the Mantoux test will provide satisfactory quantitative data. Reader errors and bias (over- or under-reading) will be minimized by, for instance, adequate training of the reader, two independent readings, and the use of high-precision callipers instead of a ruler.

Examples

The following are the requirements of a single comparative trial using two tuberculins:

1. **Formulating the questions.** In the preparation of a large-scale comparative BCG trial, a few preliminary questions had to be answered. Is the potency of our human PPD exactly the same as the international standard laid down by the World Health Organization? Is the difference between injections performed by two nurses negligible? If not, can we estimate this source of variability? What is the accuracy in reading, comparing two unskilled readers with a highly qualified professional?

2. **Planning an investigation.** As the trial is not intended to give an estimate of a parameter in the population, for instance the infection rate in a group, there is no need to draw a representative sample from a large population. Fifty-five patients in a tuberculosis hospital near Pretoria were taken as the population or universe for this study. Four treatments were considered: (*a*) injected by nurse A, South African PPD, (*b*) injected by nurse A, international standard tuberculin, (*c*) injected by nurse B, international standard tuberculin, (*d*) injected by nurse B, South African PPD. All treatments could be applied at the same time to the same patient. Random allocation and blindness of the reading was achieved by a random allocation list. Every patient was allocated a sequential number (his bed number). The four intradermal injections were given on the flexor surface of the proximal part of each forearm, with a minimum of 3 cm between two injections. Every one of the four treatments (injections) were randomly allocated to one of the four sites, i.e. upper left forearm, lower left forearm, upper right forearm, lower right forearm. Sites were used in a constant order, i.e. upper left to lower right. Every ampoule of PPD and the corresponding syringe, was labelled with the number of the treatment only, i.e. 1, 2, 3 or 4. The injector was kept ignorant of the corresponding injection, i.e. blindness in injection. A random permutation of the 4 digits was matched with the sequential list of numbers. A 'mother list' of 180 random permutations was easily obtained, using a random digit table and juxtaposed to the sequential list of numbers. For instance, patient 35 had the permutation 4132, treatment 4 (bottle and syringe 4) was first used (left forearm, upper site), then treatments 1, 3 and finally 2 (right forearm, lower site). Every treatment was given equal chance to be applied to any site in any patient, and every patient had the same chance to be given the four treatments in any specific order. Precision and accuracy in injecting was achieved by

using the standard WHO procedure for intradermal injections. Permanent supervision was maintained by an independent person as to the correctness of the treatment applied, as prescribed by the random allocation list. For the purpose of assessing the reliability of untrained readers, three independent readings were planned.

3. **Collecting and recording data.** Precise measurement of the reaction was limited to a single aspect, viz. the induration. It was done with a caliper, after gentle palpation of the site. A soft, ill-defined swelling can easily escape notice unless the site is palpated. Blind reading was accomplished by each reader, reading independently, first the upper site on the left arm of all patients, then going back to the first patient in the ward to read all the lower sites of the left arm, and so on. A clerk or assistant recorded the results and checked the sequential number of the patient. The purpose of this apparently complicated procedure, is to exclude any influence the reading of the first reaction may have on the reading of the other reactions. Naturally, the random allocation list was inaccessible, and the readers had no way of knowing what they were reading. Care was taken to keep each reader uninformed of the results of others.

4. **Organizing and presenting the data in tables and graphs.** The results and the random permutation of 4 digits were processed by computer IBM 360/65. Within a day, correct listing with heading of all individual results, histograms and graphs were available as printout.

5. **Analysing the data and recognition of new facts or relationships,** and

6. **Interpreting the results.** This small trial was not designed to study factors correlated to tuberculin hypersensitivity in a given population, but merely to answer the questions as stated under 1.

The mean induration for each treatment was computed for each reader. Any mean induration is no more than an estimate of the hypersensitivity of the given population to a given PPD, as is any single measurement in any single patient. Obviously, both estimates do not bear the same accuracy and some statistical analysis is required to give significance to the results. Presenting data in tables or graphs, is by no means an analysis. Probability statements on the confidence or reliability of the estimates and tests of significance must be made as part of a scientific investigation. A hypothesis on the mechanism underlying the significant difference observed can be put forward and eventually confirmed by further experiment.

7. **Reporting on the project.** Most of the rules for a good publication are applicable. Complete description of the methods and analysis is of fundamental importance, to enable the reader to make his own judgement.

CONCLUSION

Great care must be exercised in drawing a representative sample for an epidemiologic survey, and in the random allocation of the treatments in a comparative trial. Careful and detailed planning, use of reliable randomization techniques and control of strict blindness, are basic prerequisites for scientifically valid research work, especially in tuberculin investigation.

The method used in our tuberculin trials is not perfect, but it can be considered as a correct approach. In previous experiments a few difficulties arose. Bed numbers were used as the only record. Within the period of 3 days between injection and reading of reactions, changes occurred in the wards and the errors due to reshuffle of patients were irreparable, because of the blindness of the recording. Currently, the hospital number and the name of the patient are recorded and checked again.

Additional work is created by the strict rules. The readers of the test may be frustrated because they are left in ignorance. The blindness in reading a tuberculin test must not be taken as lack of confidence in the readers' competence or honesty! Because man is able to reason and everybody's judgement is influenced by unconscious thoughts, objectivity must be safeguarded by blindness.

Too many so-called 'random' surveys are reported all over the world. Premature conclusions are often made from tables or graphs, which must be considered as a way of illustrating the results, and therefore do not contribute additional information. Advice of a statistician is needed, both for the design of the survey and for the interpretation of the results.

ANNEXURE

A random allocation list is easily constructed by using a random digits table, available in any statistical textbook. In a perfect table of random digits, 10% of the digits are zeros, 10% are ones, and so on through to the nines. Any single digit has a probability of 0,1 of appearing at any place. Any pair of digits (00 to 99) has a probability of 0,01.

Allocation of Four Treatments to a Single Individual

A random permutation of 4 digits = 1, 2, 3, 4 is obtained first by writing down all the possible permutations, i.e. 24 permutations. Each is allocated a number: 01 to 24. The reader is now referred to the extract of the random digits of Fisher and Yates[5] (Table I) and to look for pairs of digits between 01 and 24 in the top horizontal line. Every single pair has the same probability of appearing, therefore any of the 24 permutations has the same chance of being selected, and any specific treatment 1, 2, 3 or 4 has the same chance of being applied at any of the 4 sites. Using the top row, pairs 14, 19, 11, 11, 17, 17, 20, 18 would be selected. In our example it happens that the same pair is twice selected by chance. No attempt to 'correct' the choice must be made.

To save time, a 'mother' list of 180 random permutations was constructed and roneotyped, and strips of

TABLE I. RANDOM DIGITS OF FISHER AND YATES[5]

72	84	71	14	35	19	11	58	49	26	50	11	17	17	76	86	31	57	20	18
88	78	28	16	84	13	52	53	94	53	75	45	69	30	96	73	89	65	70	31
45	17	75	65	57	28	40	19	72	12	25	12	74	75	67	60	40	60	81	19
96	76	28	12	54	22	01	11	94	25	71	96	16	16	88	68	64	36	74	45
43	31	67	72	30	24	03	94	08	63	38	32	36	66	02	69	36	38	25	39
50	44	66	44	31	66	06	58	05	62	68	15	54	35	02	42	35	48	96	32
22	66	22	15	86	26	63	75	41	99	58	42	36	72	24	58	37	52	18	51
96	24	40	14	51	23	22	30	88	57	95	67	47	29	83	94	69	40	06	07
31	73	91	61	19	60	20	72	93	48	98	57	07	23	69	65	95	39	69	58
78	60	73	99	84	43	89	94	36	45	56	69	47	07	41	90	22	91	07	12
84	37	90	61	56	70	10	23	98	05	85	11	34	76	60	76	48	45	34	60
36	67	10	08	33	98	93	35	08	86	99	29	76	29	81	33	34	91	58	93
07	28	59	07	48	89	64	58	89	75	83	85	62	27	89	30	14	78	56	27
10	15	83	87	60	79	24	31	66	56	21	48	24	06	93	91	98	94	05	49
55	19	68	97	65	03	73	52	16	56	00	53	55	90	27	33	42	29	38	87
53	81	29	13	39	35	01	20	71	34	62	33	74	82	14	53	73	19	09	03
51	86	32	68	92	33	98	74	66	99	40	14	71	94	58	45	94	19	38	81
35	91	70	29	13	80	03	54	07	27	96	94	78	32	66	50	95	52	74	33
37	71	67	95	13	20	02	44	95	94	64	85	04	05	72	01	32	90	76	14
93	66	13	83	27	92	79	64	64	72	28	54	96	53	84	48	14	52	98	94

suitable length were cut at any level from the main list and juxtaposed onto the list of sequential numbers allotted to patients, for instance bed numbers.

Tables regarding permutations exist in textbooks, and they are needed whenever large numbers of digits are involved.

Allocation of Only One Treatment to Each Individual

When the treatments are mutually exclusive, as in comparative BCG trials, the problem consists of allocating an individual to one group out of five, for instance.

For expediency, the digits 0, 6, 7, 8, 9 may be disregarded. The digits 1, 2, 3, 4, 5 are then easier to correlate to the groups or treatments 1, 2, 3, 4 and 5. One uses, say the left-hand column of digits of the random digits table, looking for 1, 2, 3, 4 and 5 only. In this case the individuals will be allocated to one of the five groups as follows: 4, 4, 5, 2, 3, 3, 1, 5, 5, 5, 3, 3 . . . . . .

I wish to thank the staff of the Cullinan Hospital, the Tschepong Settlement and the Knights Chest Hospital for their participation; Drs I. F. P. Willers and P. E. Pistorius for their active co-operation; and Dr H. H. Kleeberg's advice and criticism in preparing this manuscript.

REFERENCES

1. Guld, J. (1957): Bull. Wld Hlth Org., **17**, 225.
2. O'Toole, A. L. (1964): *Elementary Practical Statistics*, p. 26. London: Collier-MacMillan.
3. Sanchez-Crespo, J. L. (1967): Bull. Wld Hlth Org., **36**, 821.
4. Collins, T. F. B. (1968): S. Afr. Med. J., **42**, 1285.
5. Fisher, R. A. and Yates, F. (1963): *Statistical Tables for Biological, Agricultural and Medical Research*. Edinburgh: Oliver & Boyd.