THE DESIGN OF EXPERIMENTS IN MEDICAL RESEARCH

702330

REPOSITORY ARGONNE - CHR
COLLECTION A. BRUES
BOX No. A. BRUES #9, REPRINTS H-HI
FOLDER H-HI (JOURNAL ARTICLE AUTHORS)

by

LEE HERRERA

Instructor in Medical Statistics, New York University Bellevue Medical Center, New York University College of Medicine, 550 First Avenue, New York 16.

Reprinted from Special Section on Medical Statistics appearing in the

Journal of the Indian Medical Profession

Vol. 6, No. 12, pp. 3029 to 3032 and 3040, March, 1960

Published by

A. MacRae & Co. (Private) Limited,

United India Building, Sir P. Mehta Road,
Bombay-1, India.

0017608

THE DESIGN OF EXPERIMENTS IN MEDICAL RESEARCH

As an introduction to the subject of design of experiments, it may be useful to consider two very common first questions which medical researchers ask of medical statisticians. These questions may be phrased as follows.

1. I have taken before and after treatment measurements on a certain number of subjects; can you tell me if the treatment has produced a statistically significant difference before and after treatment?

2. I wish to set up an experiment to compare two forms of treatment on a group of subjects of a certain kind; can you tell me how many subjects I will need to obtain a statistically significant difference between the treatments?

Unfortunately, the statistician must generally answer "No" to the first question and "I don't know" to the second, although in the second case a thorough study of the problem may eventually lead to a positive answer. The reasons why the answers are so disappointing are in both cases connected with problems of the design of the experiment; in the one case, the difficulty lies primarily in the probable inadequacy of the design for statistical analysis, and in the other, it lies in the lack of sufficient specification of the design as well as of other necessary information.

The Meaning of Design of Experiment

The design of an experiment, according to Finney (1955), may be thought of as consisting of the following.

1. The set of treatments selected for comparison.
2. The specification of the experimental units to which the treatments are applied.
3. The rules by which the treatments are allocated to the experimental units.
4. The specification of the measurements or other records to be made on each unit.

However, these questions regarding what one might call the mechanics of a specific experiment must be preceded by the consideration of broader questions which may also be considered to lie within the compass of the design of experiments. A list of those should include these additional considerations.

a. The questions which the experiment is planned to answer.

b. The population to which the results shall be applicable.

c. The choice of subjects for the experiment.

Test of Significance

Before trying to show the relationship between the medical researchers' questions and the problem of experimental design, it is helpful to review briefly the general nature of tests of statistical significance. A statistical test consists of finding the probability of the occurrence of certain events by chance alone. When this probability is sufficiently low (e.g., P, the chance probability, less than 0.01) for a given observed event, we decide to reject the hypothesis of chance (often called null hypothesis, meaning the no-real-difference hypothesis) as an explanation of that event. Basically, chance probabilities can be considered as relative frequencies of the occurrences of events under closely specified conditions. The term relative frequency implies that there is some sort of aggregate of events, collected under specified conditions, in which one can count how many times each (type of) event occurs. Thus the chance probability which a statistical test attaches to the outcome of a given experiment is an expression of the relative frequency of such outcomes or events by chance alone, if many experiments were performed under the same specified conditions (i.e., if the same experiment were performed many times).

Population to which Results are Applicable

We are now ready to begin to examine the first of the medical investigator's problems with regard to the adequacy of the experimental design. Starting with more general considerations, suppose the experimenter will tell us the questions which the experiment was planned to answer and describe the subjects included in the experiment. Clarification is needed regarding the population to which the results shall be applicable. In the technical sense, this is a question about the way in which the experiment would be repeated, if it were possible to run the experiment once again. As we saw above, this kind of information forms the basis for the application and interpretation of a test of statistical significance.

There are two ways in which one may visualize a repetition of an experiment.

1. The repetition of the experiment may consist of the collection of more observations on the very same subjects who participated in the first experiment. In that case, the population to which the results shall be applicable consists only of the subjects actually participating in the experiment. The implication here is that the objective of the experiment was to determine whether or not the particular subjects included would respond to the treatments in a certain manner.

2. The experiment may be visualized as repeated on a group of subjects taken from the same population as the first group of subjects. In that case, the population to which the results shall be applicable is larger than the group of subjects participating in the experiment, and the objective of the study would be to determine how *subjects of the type represented* respond to the treatments in question. This second kind of picture is the more common in medical research. In that situation, the subjects included in the experiment should, ideally, according to the statistical theory, be chosen at random, i.e., by pure chance. However, in practice, our samples of subjects are not random samples—at best, they may be chosen at random from some clinic population. All we can usually do is to insure that the subjects are similar with respect to certain characteristics considered important to the experiment, i.e., that they are all members of a certain specified subpopulation, and hope that other special characteristics of the subjects will not seriously affect the outcome of the experiment. Clearly, it is quite possible that a given sample of subjects contains unusual features of a non-random kind and that, in spite of our best efforts, the sample may in the end be representative only of itself.

Suppose now that, as we questioned the experimenter regarding the kind of population to which he wished to apply his results, he told us that he stopped his experiments after eleven subjects because he then felt satisfied that there was a clinically important difference, say, in the mean values before and after treatment. What sort of repetition of the experiment is implied in that case? Obviously, his experiment was not at the outset restricted to a specific group of subjects, since he would have been willing to use more subjects, if the experiment had not seemed conclusive after the first eleven subjects. The population to which he would wish to apply his results would be the one of which his subjects are representative. Now if he repeated his experiment, he could not tell at the outset how many subjects were to be included, because, in his scheme, the number depended on the outcome. The usual tests of significance are based on sample sizes fixed in advance of the experiment and

0017609

therefore would not be applicable to the problem presented by the experimenter. While there are statistical tests based on sample size determined by outcome of experiment, the so-called sequential tests, these require advance planning of another kind and therefore could not be applied to his problem either.

Let us now suppose that the experimenter had decided on the number of subjects without regard to the outcomes on those subjects and that the subjects were representative of the kind of population to which he wished to apply his results. We are then ready to proceed with the remaining considerations involved in the design of an experiment.

Obviously the experimenter has made his decision regarding the set of treatments selected for comparison. Regarding the specification of the measurements or other records to be made (point 4), decision must also be made in advance of the conduct of the experiment for it must be uninfluenced by the outcome to insure avoidance of bias due to selection of a criterion favorable to one's particular hypothesis or opinion.

The Problem of the Control Group

Some clarification is needed regarding points 2 and 3, the specification of the experimental units and the rules by which the treatments are allocated to them. In order to be clear about what is implied here, it is useful to think in terms of a simple example.

Suppose we wish to conduct an experiment to evaluate a new type of gastric test meal with respect to its effect on the secretion of free hydrochloric acid. We choose a number of suitable subjects, administer the test meal on an empty stomach and, say, 30 minutes later we obtain a specimen of gastric juice and determine its free hydrochloric acid concentration. We then find the mean hydrochloric acid concentration for the group of subjects as well as their standard deviation which is a measure of the variability in the concentrations among the subjects. (The variability could probably be decreased or increased by changes in the definition of the population of subjects to be included in the study).

We now may wish to apply a test of significance in order to answer the question: what is the chance probability of the occurrence of a mean concentration of hydrochloric acid as high as the one found in this study? The trouble with this question is that it is meaningless, because statistical tests of significance can only be applied to comparative experiments, that is, to observed differences between treatments, not to absolute values. Given an observed difference between treatments, it may be meaningful to find the probability of obtaining a difference of that size (or larger) by chance alone. Whether or not it is meaningful depends on the way in which the difference was obtained. If we had used a published figure on the average hydrochloric acid concentration after a standard test meal for comparison with our observed value on the new test meal, a small chance probability would prove very little regarding differences in response to the two different test meals. The reason is that a difference between the two hydrochloric acid concentrations may be due to so many other factors besides the difference in the test meals and chance. The published figure on hydrochloric acid concentration after the standard test meal may simply not apply to our group of subjects— in any case, we have no way of knowing whether it does or not. The answer to this problem obviously is the use of a control group, that is, a group of subjects who are given the standard test meal for comparison with the subjects given the new test meal. We then have a set of treatments consisting of the standard test meal and the new test meal. The experimental units are, say, human subjects, with one form of treatment per subject. With respect to the rules by which the treatments are allocated to the subjects, the treatments are allocated to the subjects at random (with the help of a table of random numbers). If the treatments are allocated at random and if there is no real difference in response to the two test meals, then any observed difference in the averages between the two groups of subjects can be accounted for by the variability in the response observed within each group. That is to say, we then have an appropriate yardstick against which to measure the observed difference between the groups. If our test of significance gives us a very small chance probability for the observed difference, taking into account the within group variability, we can feel confident that there is a real difference in response to the two test meals.

Before and After Treatment Observations.—The study originally under discussion was somewhat more complicated in that each subject in a sense served as his own control by having both a pre-treatment and a post-treatment observation. The criterion of response was the difference between the pre-treatment and post-treatment observations. As an example of such experiment consider a study of the effect of breathing against positive pressure on the flow of urine. The subjects were asked to empty their bladders every fifteen minutes. The control specimen was the specimen voided immediately before the institution of pressure breathing; the treatment specimen was the one voided at the termination of pressure breathing. An estimate of variability between subjects is afforded by the variability in the before-after treatment differences. However, for a meaningful test of significance regarding the effect of pressure breathing, it is necessary that differences before and after pressure breathing can be accounted for either by chance alone or the effect of pressure breathing plus chance.

In all experiments in which the control value is really a "before treatment" value we cannot rule out the possibility that the circumstances of the experiment (in addition to the effect of the experimental treatment itself) have influenced the after treatment value in a consistent manner. The before treatment "control" value may no longer be representative of control conditions after the experiment is completed; therefore, it may not be an appropriate basis for estimating the effect of treatment. Time has elapsed, and conditions may have changed. This is by no means saying, however, that the change in urine flow in an individual before and after pressure breathing is not the appropriate criterion by which to evaluate the effect of positive pressure breathing. It does mean that the estimate of the effect of treatment may be subject to bias and that, therefore, a test of significance to determine whether the observed change in urine flow is different from zero has no unequivocal interpretation. Basically, this example is no different from the previous one in which it was a question of determining whether an observed hydrochloric acid concentration was different from some probably inappropriate published control value. The solution to the problem is therefore also the same: the use of a control group of subjects, this time consisting of a group which is exposed to all the circumstances of the experiment with the only exception that of pressure breathing itself. That is to say, the control group is a group of subjects receiving a dummy treatment (the placebo of clinical trials). In the pressure breathing experiment, it would mean that "before treatment" specimens are collected on the control subjects, that the subjects are connected to the apparatus for the same length of time as those exposed to positive pressure, etc. It would also be highly desirable that the subjects not be informed regarding whether they are in the positive pressure group or in the control group, to avoid introducing a psychological bias. Other devices may be used to give the control subjects the impression that they are subjected to a real treatment. Subjects must be allocated to the control and pressure breathing groups at random and should also be admitted to the experiment in random order, so that small systematic changes in technique will not bias the final comparison. When these precautions are taken, the difference between the before and after treatment changes in the two treatment groups will be an unbiased estimate of the change in urine flow due to positive pressure breathing. If positive pressure breathing has no effect on urine flow, any difference between the changes before and after "treatment" in the two groups can be attributed to chance alone. Under those conditions, it is appropriate to apply a test of significance to the changes observed in the two groups in order to determine the probability of meeting the observed difference between the changes by chance. If chance alone rarely brought about such difference, we would run little risk in attributing the difference to the positive pressure.

To sum up, the question '.....can you tell me if the treatment has produced a statistically significant difference before and after treatment?' must be answered with "No" primarily because we do not have a comparison with a control group in which the treatment was not given. The design recommended for such experiment was to allocate the real treatment and the dummy treatment to a group of subjects strictly at random and then observe the change in before and after "treatment" values in each subject. That is to say, the experimental units to which the treatments are applied are individual persons, the treatments are allocated to the experimental units at random, and the measurement to be made on each experimental unit is the difference before and after "treatment". This completes the description of the experimental design according to Finney's list of specifications.

The Completely Randomized Design

The design suggested above is an example of what has been called the completely randomized design which consists of completely

0017610

random allocation of subjects to all treatments to be compared, with one treatment per subject. One of the treatments may be a dummy or placebo, of course. There must be several subjects per treatment in order to obtain an estimate of variability in subjects treated alike. (An experiment which comprises all treatments to be compared and provides an estimate of variability in subjects treated alike — sometimes misleadingly referred to as "error"—is called a self-contained experiment.)

The completely randomized design is by far the most common and useful in clinical medicine and the one to be recommended to the majority of medical researchers for its simplicity both in the conduct of the experiment and in the analysis of results. The fault with statistical analysis in the medical literature often is that methods which are appropriate for a completely randomized design are applied to experiments in which no such design was employed. This is true for many t-tests and chi-square tests etc. employed in studies without appropriate control groups, without random allocation or without predetermined end point of the experiment (e.g., predetermined sample size). Sometimes one sees t-tests etc. applied in designs which are not in themselves faulty, such as the ones just mentioned, but which require a different form of analysis. A reader of the medical literature should be warned to hesitate accepting the verdict of a statistical test of significance without a careful study of the design of the experiment. Unfortunately, the information given about design is often inadequac, possibly because its importance is not yet fully appreciated by investigators and editors who wish to conserve space. However, if no mention is made of the method by which subjects are assigned to treatments, it would be fairly safe to assume at the present time that no strictly random methods, e.g., random numbers, was employed.

The Randomized Blocks Design

Before going on to consider the question regarding required sample size, it may be useful to mention some other experimental designs which are not quite so simple as the completely randomized design in that they have an internal structure within which random allocation takes place. The most basic of these designs is the so-called randomized blocks design from which developed other more complicated designs which will not be discussed here. The term "block" has been taken over from agriculture where it represents a block of plots of ground with similar inherent fertility. The randomized blocks designs are popular in agriculture and industry, but also have their application in medicine. Their essence is that the large experimental units, the blocks, are subdivided into smaller units, the plots; within every block, the various treatments are assigned to different plots, the particular plot for a given treatment being chosen at random. Literally speaking, therefore, it is not the blocks which are randomized, but the location of the treatments within the blocks. When all the treatments can be applied within the same block, the differences between the treatments will be less affected by differences in fertility than if each block receives only one form of treatment; between block differences in fertility (over and above within block differences) will be averaged out in the differences between treatments, which will therefore be estimated with greater precision.

In medicine it is easy to see that it is sometimes possible to apply two or more treatments to the same subject, with the treatments given one after the other in a random order. In that case, each subject represents a block. If such scheme is feasible, it is advantageous because average treatment differences can be estimated without the variability introduced by differences in the levels of response of the various subjects. These differences cancel out, if each subject is given all treatments.*

Examples of Randomized Blocks Designs

One example of randomized blocks in medicine with the subject as a block might be the application of various doses of radiation to small areas on the patient's back in order to determine the threshold dose producing erythema. The location of the various doses on the patient's back must be allocated strictly at random, for every patient anew, in order to avoid the introduction of bias due to possible gradients of sensitivity to the radiation. It might be mentioned here too that the observer noting the erythema should be unaware of the particular random allocation used to rule out the effect of personal bias.

Another example of the randomized blocks design is in the field of evaluation of analgesics used for episodic pain, such as headache or menstrual pain, in which the patient may be given a different analgesic for every one of a specified number of attacks. The analgesics are given to the patient in a random order which is allowed to vary with each patient. However, even if the patient as well as the questioner are kept unaware of the nature of the medications, this design may be invalidated by psychologicl biases introduced somewhere along the sequence of treatments being given. Such a bias is a form of carry-over effect of treatments, or interference between units of treatment, against which special precautions must be taken. Because of such carry-over effects, the randomized blocks design frequently is unsuitable in clinical work.

In clinical trials in the field of chronic disease, the randomized blocks design may fail because changes in the patient's condition with time may affect his reaction to treatment. The difference in the patient's condition between two (or more) "plots" of time may be so great that comparing two treatments in the same patient may hold no advantage over comparing the treatments on different patients. It may even be a disadvantage. In addition, there is the problem of having to follow the patient for twice the length of time required in the completely randomized design.

If the treatments to be compared are repeated in a random order within the block, we have a randomized blocks design with replication (strictly speaking, replication within blocks). An example of such design could be a sequence of daily experiments in the laboratory. Here the block would consist of an experiment carried out in one day. On each day, we have a certain number of animals (or test tubes, etc.) to which the treatments to be compared are assigned strictly at random; that is, within each block we have a completely randomized experiment. Here one often sees that the experiments are simply pooled and the data analysed as if there were only one experiment. This may be throwing away valuable information besides possibly invalidating a test of significance. The fact that more than one experiment was performed enables one to check on the consistency of materials and methods and to test whether the treatment effects themselves differ between experiments. If they do, the interpretation of results may be considerably affected.

If real differences exist between experiments in terms of the level of the measurements, pooling of the data as if they all came from one experiment will invalidate the test of significance for treatment effects. The reason is that the estimate of variability used as the yardstick in the test of significance will be increased by the difference between the experiments, while, as we have seen, in a randomized blocks design (with equal numbers on each treatment), the differences between treatments are not affected by differences in levels of measurements between the blocks. Clearly, the appropriate estimate of variability must also be unaffected by differences in levels of measurements between the blocks, i.e., it must consist of the variability observed between units treated alike with respect to treatment *and* experiment (an intrablock estimate of variability). With the inflated estimate of variability, the probability of falsely proclaiming a real difference between treatments will be reduced below the stated level of, say, 5 per cent, and we will be less likely to detect a real difference when it exists.

Perhaps another more interesting example of the randomized blocks design with replication is the clinical trial with stratification. Here the total group of subjects is divided into a number of subgroups or strata consisting of subjects that are similar with respect to one or more characteristics, such as age, sex, severity of disease, etc. The subgroups of similar subjects form the blocks in this kind of experiment. Within the subgroups, the treatments are, as usual, allocated completely at random. When equal numbers of patients are assigned to each treatment within each subgroup (difficulties arise in the case of losses), differences between treatments will not be affected by differences in general level of response between the various subgroups. Subgroups should generally be sufficiently large, however, to make it possible to determine whether or not treatment differences vary between subgroups.

---

* *There is quite a popular design for two treatments in which one random half of the subjects receives the two treatments in one time order and the other half receives them in the reverse time order; this design has been called a reversal-pair or cross-over design. Because of the restraint on the variation in time order, this design no longer is a true randomized blocks design.*

0017611

Required Sample Size

We are now ready to return to the second original question, the question regarding the required sample size. In order to try to answer, the statistician must pose a number of counter questions regarding the various features of the experiment, including, of course, its design. If we assume that the basic questions regarding the purpose of the experiment, the selection of subjects for the experiment and the nature of the treatments have been settled, there remain the following six questions.

1. In what manner will the subjects be assigned to the treatments (Finney's points 2 and 3)? Will this be a completely randomized experiment or will there be some form of block design due to the distinction of subgroups or the application of both forms of treatment and/or the repetition of treatments on the same subject?

2. What is the criterion to be used for evaluation of the treatments? Is the criterion in the form of a measurement or simply a category? If the criterion is in the form of a category, the required sample size is generally larger than if it is a measurement. Moreover, in the case of a block design, category (frequency) data are more difficult to analyse statistically than are measurement data. Frequently it is desired to study more than one criterion for evaluation, with each theoretically requiring a different sample size for successful results; in such case it would probably be necessary to fix on the most important criterion for the determination of sample size.

3. What is the variability of the criterion under like treatment conditions in the group to be studied? What constitutes this variability (the "yardstick") will depend on the experimental design. For instance, in the case of a completely randomized design, it means the variability of the observations chosen as a criterion among all subjects given the same treatment. In a randomized blocks experiment without replication in which two treatments are compared on the same subject, the quantity required is an estimate of the variability among all the subjects of the differences between two repeat observations on the same subject.

The estimate of variability under like treatment conditions is usually (particularly with measurement data) expressed in terms of a standard deviation and must come from previous, similar work (There is no special virtue in the use of the standard deviation as a measure of average variation, unless one deals with the Gaussian curve.) The trouble is, of course, that there frequently is no such previous work; even experiments which look similar in the literature often were performed on somewhat different groups of subjects with sufficiently different materials and methods to render estimates obtained from them highly questionable. Because of the difficulty regarding the estimate of variability, we are frequently unable to determine the required sample size with any degree of exactitude. Often it is useful to conduct essentially two experiments instead of one and estimate the sample size required in the second experiment from the variability observed in the first. A decision must of course still be made regarding the size of the first experiment.

In the case of a completely randomized design in which the criterion for evaluation can assume only two values, e.g., improved, not improved, the problem of required sample size is somewhat simplified. Here the question regarding variability (on the standard or control treatment) is equivalent to the question: what percentage of improved cases do you expect with the control treatment? The reason is that, for any population percentage, we know (experimentally as well as mathematically) the variability in percentages that can be expected between random samples of any specified size.

4. What is the size of the difference between the two treatments in the units of measurement of the criterion which it is important to detect? This question frequently requires a great deal of soul searching on the part of the investigator. Often the answer is found only by means of a compromise between the difference which it would be nice to detect and the difference it is feasible to detect with the resources at hand — because the smaller the real difference which we wish to detect, the larger will be the sample size required to detect it with any assurance of success.

5. What is the size of the type I error which you are willing to commit? The type I error is the error of concluding that a real difference exists, when in fact there is none. Every test of significance is subject to this error which can never be completely eliminated. It should always be fixed in advance of the experiment. Usually, it is set at 5 per cent or 1 per cent (of experiments), but other values may be chosen depending on the particular nature of the experiment.

6. What is the size of the type II error which you are willing to commit? The type II error is the error of concluding that a real difference of a size considered important does not exist, when in fact it does. This error also can never be completely eliminated, but, again, its size can be controlled. The choice of the type II error is an individual matter; however, it would seem that once one has decided that a certain (minimum) difference between the treatments is important to detect, one should aim for a high degree of assurance of being able to detect it. Therefore it is suggested that the type II error be set at no greater than 5 per cent for a difference of the given size. Smaller differences will then be missed more often than 5 per cent of experiments, while larger differences will be detected with still greater certainty (greater than 95 per cent of experiments).

When the criterion of evaluation consists of the percentage of subjects responding in a specified way and we have a completely randomized experiment with two treatments to be compared, it is possible to draw up simple tables which show the sample sizes required for the detection of differences between various population percentages (of improved, cured or whatever the criterion). We are therefore fortunate to be able to conclude on a positive note by giving as an example of required samples sizes of 100 or less a small excerpt from such a table (Mainland, Herrera and Sutcliffe, 1956), with a type I error no greater than 5 per cent and specified type II error of about 5 per cent or less.

Table I

SAMPLE SIZES REQUIRED FOR DETECTION OF DIFFERENCE BETWEEN TWO SPECIFIED POPULATION PERCENTAGES

Two-sided type I error of 0.05, and specified type II error about 0.05 or smaller.

Population Percentages Group A (smaller %)	Group B (larger %)	Sample Size Required (each group)	Type II Error
1%	15%	100	2.7%
1%	25%	50	3.5%
5%	33%	50	5.2%
5%	67%	15	5.2%
10%	67%	20	3.5%
10%	75%	15	5.6%
15%	50%	50	4.2%
15%	75%	20	3.4%
25%	50%	100	5.5%
25%	75%	30	3.0%
33%	67%	70	2.2%
33%	90%	20	3.5%
50%	75%	100*	5.5%
50%	85%	50*	4.2%

* *An experiment with 50% in group A and 75% in group B, where the percentage stands for percentage improved, say, is equivalent to an experiment with 25% in group A and 50% in group B if the percentage stands for percentage not improved; similarly, 50% in group A and 85% in group B is equivalent to 15% in group A and 50% in group B. Thus each entry in this table can be used in two ways.*

ACKNOWLEDGEMENT

This paper was written as part of a project entitled "Promotion of Biometrical Methods in Medical Research" and supported by a grant RG — 6100 from the United States Public Health Service.

0017612

REFERENCES

1. *Cox, D. R. (1958). Planning of Experiments. Wiley, New York.*

2. *Finney, D. (1955). Experimental Design and its Statistical Basis. The University of Chicago Press, Chicago.*

3. *Herrera, L. (1955). "Bias in the Allocation of Treatments by Random Numbers"*, Science, *122, 828.*

4. *Mainland, D. (1952). Elementary Medical Statistics: The Principles of Quantitative Medicine. Saunders, Philadelphia.*

5. *Mainland, D. (1954). "Statistics in Medical Research"*, in *Methods in Medical Research. Year Book Publ., Chicago.*

6. *Mainland, D., Herrera, L. and Sutcliffe, M. I. (1956). Statistica Tables for Use with Binomial Samples — Contingency Tests, Confidence Limits, and Sample Size Estimates. Department of Medical Statistics, New York University College of Medicine, New York.*

● ● ● ● ● ● ●

The series of articles on Medical Statistics starts in the Journal of the Indian Medical Profession with effect from the June 1959 issue. The articles published so far are as follows. They are being continued each month.

Ramachandran K. V.: History and Development of Statistics.
Siegel Morris: Aim and Basis of Statistical Methods.
Chakrabarti M. C.: Statistics in Medical Science.
Deshpande V. A.: Collection of Data.
Rele J. R.: Classification, Tabulation and Digrammatic Representation of Data.
Jain S. P.: Vital Statistics.
Ramachandran K. V.: Measures of Location or Central Tendency.
Seetharam B. G.: Measures of Dispersion and Skewness.
Vashi D. G.: The Normal, Binomial and Poisson Distribution.
Gorrill R. H.: Statistics in Relation to Medical and other Sciences.
Bhanot Mrs. Indira: Collection of Data.
Stewart Alice M.: Social Medicine Surveys.
Gorrill R. H.: Therapeutic Trials.
Heredia A. F.: Assessment of Therapeutic Trials.
Herdan G.: Therapeutic Trials (Clinical Trial vs. Animal Experiment).
Kshirsagar A. M.: Regression and Correlation.
Duncan Ethel H. L.: Some Points to be Considered When Planning A Survey.
Chakrabarti M. C: Problems of Sampling in Public Health Work.
Linder Forrest E. and others: The U.S. National Health Survey.
Sprowls R. Clay.: Tests of Significance and Chance Differences.
Carpenter R. G.: Generalizing from the Results of Investigations, and Designing. Surveys that give Reliable General Conclusions.
Cooksey Eddie V.: The Medical Record Library and Statistics.
Gourlay R. John: Community and Special Group Surveys.
Siegel Sidney: Considerations in the Choice Between Parametric and Nonparametric Tests.
Clelland Richard C.: The Role of Statistics.
Mukshkin J. Selma and d'A Collings F.: Problems in Defining and Measuring Indirect Costs of Disease.
Chandrasekhar S.: Demographic Statistics in India.
Clelland C. Richard: Uses of the Chi-Square Distribution.
George P. A. and Mukerji B.: Statistical Approach to Therapeutic Research.
Seal K. C: Importance of Medical and Health Statistics in Health Planning.
Perry N. C.: An Introduction to the Technique of Probit Analysis.
Sen N., George P. A. and Mukerji B.: Some Uses of Non-Parametric Tests in Medical Research.
Herrera Lee: The Design of Experiments in Medical Research.
Rao P. S. R. S.: Design of Experiments and Presentation of Results.
Jain S. P.: Life Tables.
White Carl L.: The Statistical Approach to Research.
Curey, D. V.: Importance of Medical and Health Statistics in Health Planning.
Clelland C. Richard: Nonparametric Methods.
Grizzle E. James: Choosing A Sample Size in Clinical Trials.
Behm Hugo: Medical Profession and Health Statistics.
Burdick Douglass E.: Common Errors in Statistical Applications.
Tate Merle W.: Quick Inferences From proportions, Means and Correlation Coefficients.
Syme Leonard S.: The Epidemiology of Noninfectious Disease in Sociological Perspective.
Enowitch B. I.: Genetic aspects of the Epidemiology of Chronic Disease.
Fletcher E. and Blum H. L.: The Pittsburg (California) Study: A Report of Information Available in a Local Health Department on a Census Tract Bassis.

Burdick E. Douglass: The Statistical Approach to Research.
Varchaver Catherine: Vital Statistics.
Askovitz S. I.: The Mean Deviation Determined Graphically.

Printed at the Onlooker Press, Colaba, Bombay-5.

0017613