The New England
Journal of Medicine

Volume 247 JULY 24, 1952 Number 4

THE CLINICAL TRIAL*

A. BRADFORD HILL, C.B.E., D.Sc., PH.D.†

LONDON, ENGLAND

IT IS with mixed feelings that I stand, a layman, before the medical faculty of one of the world's most famous universities. In the past century, or more, statisticians have, it is true, been recognized as having a right of entry to the field of public health, and to its related sciences of epidemiology and preventive medicine. They have been closely associated with medical officers of health. They have worked with all who have been concerned with the promotion and maintenance of the well-being of the community. But it is clear that the health of the community — the group — essentially requires the statistical approach. For we are concerned with the herd rather than with the individual and so we need community measures. I might hope to speak on matters of mutual interest to your School of Public Health. I might, of course, prove boring to it, but I should speak as a bore, not as a stranger. I should promote sleep rather than catcalls; that belief gives me courage. What, however, of your University's Medical School? Dare the statistician now pass from the well tilled (perhaps I ought to say well drained) fields of public health to those more exclusive upland meadows in which are practiced the arts of the clinician — arts that appear to cultivate the individual approach and sometimes even an air of infallibility? It is a bold step. In my own country I have been fortunate enough to be able to take it and can, even at the worst, still exclaim, with the poet Henley, "my head is bloody but unbowed." Over a fairly wide expanse of clinical medicine in Great Britain the statistical approach has been accepted as useful; it is being increasingly applied. But here, like Ruth "amid the alien corn," I stand, if not tearfully, at least a trifle fearfully.

Before setting out I sought support in an anonymous article on Statistics in Medicine that formed part of the *British Medical Journal's* instructive review of the progress of British medicine in the first half of the present century. Between statistician and clinician there had been, it confesses, antagonism: "The medical man charged with responsibility for the patient was contemptuous of the statistician's fundamental approach through the group; and the statistician took a jaundiced view of the conclusions light-heartedly drawn by the practitioner from a handful of cases without allowance for the play of chance."[1] The only comforting thing about that statement is its use — an apparently deliberate use — of the past tense. Has, then, the antagonism gone? I believe it is undoubtedly very much less than it used to be. But I could not lay my hand upon my heart and say it has vanished wholly. There are still noticeable from time to time, in medical journals and at meetings of medical societies, various kinds and degrees of mutual misunderstanding and even scorn. Before launching out upon the philosophy of a clinical trial, I think it would be well to consider them.

MISUNDERSTANDINGS AND MISTAKES

The Statistician

Let me first place in the dock those of my own profession. The statistician, and particularly those not in close contact with clinical medicine, may tend to forget that the physician's first duty is to his patient — to do all in his power to save the patient's life and restore him, as rapidly as possible, to health. That fundamental and ethical duty must never be overlooked — though with the introduction of better, brighter and ever more toxic drugs, and with the wide prevalence of surgical procedures such as tonsillectomy, the onlooker may perhaps with good reason sometimes ask the clinician "are you sure you know where that duty lies?" It seems to me sometimes to be unethical *not* to experiment, not to carry out a controlled clinical trial. But we must never forget the issues. Basically, too, the statistician is concerned with things that can be counted. "In so far as things, persons, are unique or ill-defined, statistics are meaningless . . . ; in so far as things are similar and definite . . . they can be counted and new statistical facts are born . . . Our arithmetic is useless unless we are counting the right things."[2] Clearly, until that happy day ar-

*Read before the Department of Preventive Medicine, Harvard Medical School, May 28, 1952 and based partly on a discussion held at the Medical Society of London, November, 1951.

†Professor of medical statistics, London School of Hygiene and Tropical Medicine; honorary director of the Statistical Research Unit of the Medical Research Council.

rives when every clinician is his own statistician, we need a close collaboration to ensure that the statistician *is* counting the right things — for instance, that he is devising satisfactory groups into which patients can be classified, and that he is setting up the most appropriate and definite measures of the patient's condition before and after treatment. I have seen friction arise through the statistician's inability to do this through a lack of knowledge of the medical problem and of its details.

The art of applied statistics is, needless to say, compounded of two things — a knowledge of statistical methodology and a wide and detailed knowledge of the data to which that methodology is to be applied. I should hesitate to allocate a relative importance to each of these two aspects, but I hold firmly that both are essential and that the use of the methodology, however erudite, without a parallel and exact knowledge of the data under study can be, indeed is likely to be, most dangerous. It follows that the statistician, if he is to play his proper part in a clinical trial must be in it "up to his neck"; he must be in it from its very start — that is, at the initial planning level. If he is a layman he must endeavor to learn and understand its medical as well as its statistical aspects — and should, I believe, be as intensely absorbed and interested in the one as in the other. As a corollary he should almost invariably decline to be brought in at the end of a trial merely to sort out and add up the results or, even worse, to calculate as a kind of super bookmaker, the odds for or against some event of whose pros and cons he is largely ignorant and about which he could not care less. Too often, he makes the mistake of accepting such a task. In short, the statistically designed clinical trial is above all a work of collaboration between clinician and statistician, and that collaboration must prevail from start to finish.

Finally, in my indictment of the statistician, I would argue that he may tend to be a trifle too scornful of the clinical judgment, the clinical impression. Such judgments are, I believe, in essence, statistical. The clinician is attempting to make a comparison between the situation that faces him at the moment and a mentally recorded but otherwise untabulated past experience. There clearly may be gain in the introduction of more objective and more quantitative methods of assessment, which can be recorded, than can be supplied by this clinical instinct. Certainly no harm will be done, as Sir Henry Dale[3] has pointed out, if these newer and more objective methods of observation "are used primarily to supplement, and only by careful stages and with proved advantage to replace, the older and subjective ones. The loss, if any, would arise from haste to discard a coherent though impressionistic picture, and to replace it by a collection of precise but as yet uncoordinated details." At present I would argue that we should, whenever possible, endeavor to harness the clinical impression to our measurements of the results of a therapeutic trial. To that I shall return again.

The Clinician

Turning now to the other side of the picture — the attitude of the clinician — I would, from experience, say that the most frequent and the most foolish criticism of the statistical approach in medicine is that human beings are too variable to allow of the contrasts inherent in a controlled trial of a remedy. In other words, each patient *is* "unique" and so there can be nothing for the statistician to count. But if this is true it has always seemed to me that the bottom falls out of the clinical approach as well as the statistical. If each patient is unique, how can a basis for treatment be found in the past observations of other patients? In fact, of course, physicians do not act like that. They base their "method of choice" upon what they have seen happen before — whether it be in only two or three cases or in a score. But even if human beings are not each unique in their responses to a given treatment they are certainly likely to be variable, sometimes extremely variable. Two or three uncontrolled observations may, therefore, give merely through the customary play of chance, a favorable picture in the hands of one doctor, an unfavorable picture in the hands of a second. And so the medical journals, euphemistically called the "literature," are cluttered up with conflicting claims — each in itself perfectly true of what the doctor saw, and each insufficient to bear the weight of the generalization placed upon it.

Far, therefore, from arguing that the statistical approach is impossible in the face of human variability, we should realize that it is because of that variability that it is often essential. It does not follow, to meet another common criticism of the statistical approach, that it invariably demands large numbers. It may do so; it depends upon the problem. But, it should be recognized, the responses to treatment of a single patient are clearly a statement of fact — so far as the observations were truly made and accurately recorded. And that single case may give, in certain circumstances, evidence of vital importance.

If, for example, we were to use a new drug in a proved case of acute leukemia and the patient made an immediate and indisputable recovery, should we not have a result of the most profound importance? The reason underlying our acceptance of merely one patient as illustrating a remarkable event — not necessarily of cause and effect — is that long and wide experience has shown that in their response to acute leukemia human beings are *not* variable. They one and all fail to make immediate and indisputable recoveries. They one and all die. Therefore, although it would clearly be most unwise upon this one case to pass from the particular to the general, it would be sheer madness

not to accept the evidence presented by it. No statistician (no statistician, let me say, who knew his subject matter) would object to that sample on the grounds that it was too small to be informative.

If, on the other hand, the drug were given to a patient suffering from acute rheumatic fever and the patient made an immediate and indisputable recovery, the statistician might suggest (with customary diffidence) that we have little basis for remark. That recovery may clearly have followed the administration of the drug without the slightest probability of a related cause and effect. With this disease human beings *are* variable in their reactions, — some may die, some may have prolonged illnesses but recover eventually with or without permanent damage, some may make immediate and indisputable recoveries, — whatever treatment we may give them. We must, therefore, have more cases before we can reasonably draw inferences about cause and effect.

I have somewhat labored this point of numbers of observations because the old and fallacious gibe still lingers on, even among persons who should know better: that statisticians would, if given the opportunity, have rejected — or even suppressed, though I am never quite sure how — some of the original and fundamental observations in medicine, on the grounds of their small number. For example, fragilitas ossium, I was once told, was originally described in just two cases; and this number I was also informed, a trifle tartly, "statisticians would regard as useless evidence." Why on earth should they? My retort ran as follows[4]:

> If 'exact descriptions' were given of these two cases and 'beautiful water colours' were supplied to illustrate them, then of course they are scientific evidence, and undeniable evidence of an occurrence — whether it be of one case or two. It would only be in relation to any subsequent appeal from the particular to the general that the statistician — and equally any trained scientific worker — could object. If on the basis of these two cases the clinician (in practice, let us say, near Smithfield Market) should be so unwise as to argue that the condition was specific to butchers, then the statistician might suggest that the experience was both too select and too slender in size to justify any such generalization.

In short there is, and can be, no magic number for either clinician or statistician. Whether we need one, a hundred or a thousand cases turns upon the setting of our problem and the inferences that we wish to draw. Faced with the rapid recovery of a single patient with rheumatic fever under a new drug what, I suggest, the careful clinician would do is not to generalize but to test the treatment on a second case. If it worked well again — or, perhaps, if it did not — he would test it on a third. And, without being accused of undue caution, or even of mathematical leanings, he might go so far as to seek a fourth. Thus, with somewhat halting steps, he unwittingly directs himself up the statistical garden path. I believe he might sometimes fare better if he straightaway walked boldly up the path and without any ado opened the gate to a designed and controlled clinical trial.

The Controlled Trial

The essence of such a trial is comparison. To the dictum of Helmholtz that "all science is measurement," we should add, Sir Henry Dale[3] has pointed out, a further clause that "all true measurement is essentially comparative." Usually, therefore, but again not always, we need not only the group of patients to be submitted to a special treatment, or treatments, under investigation but another group differently treated — by, usually, the older and, at the time, more orthodox methods. The first step in the statistical technic, then, is the *random* allocation of patients to one or other of these groups. (I am, of course, assuming that the ethical situation has first been most carefully considered and the decision reached that such a trial is both proper and possible; if it is not it may yet be possible to devise some other, though probably less informative, type of trial, but with that I am not here concerned.)

There are various ways in which the random allocation of the patients to the treatment groups can be carried out. Most often today and, I think, preferably, it is done by the construction of an order of allocation, unknown in advance to the clinician, and based upon random sampling numbers — a modern substitute for tossing up, and one that is a trifle less embarrassing in the ward or office. Thus using OT to stand for orthodox treatment and NT for new treatment (without necessarily any confession of faith therein) we may have as the order for patients consecutively brought into the trial OT, NT, NT, OT, OT, OT, NT, and so forth. We may easily introduce a device to equalize the number of patients on the new and old treatments at, for example, every dozen — since otherwise by the play of chance we may occasionally be embarrassed by too many of one kind and too few of another. I do not think it necessary to spend time now upon these minutiae. What I would emphasize are the advantages of this random and unpremeditated distribution of patients to groups. Strictly adhered to — and I need hardly say that this is a *sine qua non* — this method ensures three things: it ensures that neither our personal idiosyncrasies (our likes or dislikes consciously or unwittingly applied) nor our lack of balanced judgment has entered into the construction of the different treatment groups — the allocation has been outside our control and the groups are therefore unbiased; it removes the danger, inherent in an allocation based upon personal judgment, that believing we may be biased in our judgments we endeavor to allow for that bias, to exclude it, and that in so doing we may overcompensate and by thus "leaning over backward" introduce a lack of balance from the other direction; and, having used a random allocation, the sternest critic is unable to say when we eventually dash

into print that quite probably the groups were differentially biased through our predilections or through our stupidity. Once it has been decided that a patient is of the right type to be brought into the trial the random method of allocation removes all responsibility from the clinical observer.

It may sometimes, however, be argued — and with truth — that this random allocation gives too much freedom to the play of chance and thus fails to provide groups that are sufficiently alike, or as alike as they could have been if we had used a more deliberate method of distribution. The likelihood of such a failure is clearly dependent upon two things: the size of the groups that we set up and the variability of the patients we bring into them. Often one can effectively meet the difficulty by first dividing the somewhat heterogeneous group of patients who are to be brought into the trial into more homogeneous sub-groups, and then making the random allocation to treatment O, treatment N and so forth within each of the sub-groups. This has been the technic adopted in the joint trials of cortisone and ACTH versus salicylates in rheumatic fever now being carried out to a common plan in a number of centers in the United Kingdom, the United States and Canada. Thus we have six subgroups. Rheumatic fever runs, on the average, a different course in young children and in adolescents. We have, therefore, a subdivision by age into patients under sixteen years old and patients aged sixteen years or over. Apart from age, however, the course of the disease in relation to treatment, particularly with regard to any permanent damage that may occur to the heart, may well be influenced by the speed with which, after the onset of an attack, treatment can be instituted. We have, therefore, within each of the two age groups, three divisions by the duration of time that had elapsed between the recorded onset of attack and the start of treatment: treatment begun within the first two weeks of the illness; treatment begun between two and six weeks after onset; and treatment not begun until more than six weeks had elapsed. Every patient must first satisfy certain criteria of diagnosis, which have been laid down. Being thus eligible for entry to the trial, he is first allocated to whichever of the six subgroups he belongs. He is then allocated to a specific treatment, whether with cortisone, ACTH or salicylates, by means of random orders set up for each of the six subgroups and, separately, for each center taking part in the trial.

It may be, and indeed it is highly likely, that we shall not have enough cases in each of these six more homogeneous subgroups to allow firm conclusions to be drawn. But we shall at least be in the position to amalgamate the six subgroups, if it is justifiable, knowing that the three treatment groups in total will then have the same distribution of patients by age and by duration of illness. By the design of the trial we have ensured equality in the treatment groups in these respects, and by a random allocation applied to the subgroups we have ensured an absence of personal bias.

Such designs, simple though they may be, are of importance. They may often require smaller numbers of patients for the solution of a problem and allow a wider range of questions to be simultaneously tackled. I would usually, however, myself sooner seek a decisive answer to one or two questions than cast my bread upon the waters and be faced finally with seven "not statistically significant" differences. "Not statistically significant," one should always remember, is the equivalent of the "non-proven" of the Scots law rather than the "not guilty" of the English courts. I also emphasize that in spite of the random construction of the treatment groups that I have envisaged, it will in practice be essential to prove that the groups of patients under each treatment have, in fact, turned out to be similar in their features at entry — that is, in such features as may be relevant to a patient's response to treatment. For example, an analysis of the patients at entry to the trial must reveal the level of their temperatures and their blood sedimentation rates, the frequency of their heart murmurs and their bacteriologic companions, and so on. No method of random allocation can absolutely ensure the equality of the groups in all these respects. The goddess of chance may prevent it, particularly when we are concerned with small groups. It is therefore our first and imperative job to show that the groups we have set up are reasonably alike or, if not, to make such allowances as are possible for the differences that have occurred. Such allowances after the event are not always easy to make, and therefore the "stratifying" by subgroups that I have touched upon has considerable advantage in ensuring equalities in major characteristics.

The Treatment

In the great majority of the controlled clinical trials with which I have been associated a specific treatment schedule has been laid down. It may, needless to say, be varied with age or body weight or some other characteristic of the patient; but apart from that it has been rigidly defined in advance and must be adhered to by the clinician (except for the usual over-riding ethical reasons). Clearly, however, in regard to treatment there are an infinite number of questions we can ask of a trial. We can choose one dose out of many; we can vary the interval of administration; we can give it by different routes; we can exhibit it for different lengths of time; and so on. In testing a new form of treatment knowledge at first is necessarily scanty, being usually based upon laboratory work and a few scattered clinical observations. In Medical Research Council trials we have thought it proper, therefore, to choose such a regimen as promised to reveal the potentialities (and often the dangers)

of the drug (or whatever may be concerned). Thus we have a tidy question — for example, if to a defined type of patient 2 gr. of drug X are given daily in four divided doses by intramuscular injection and for three months, what happens?

But perhaps the question is too tidy. We can, of course, extend it after its answer has been reached by experimenting with variations upon the original theme. It may, however, be argued, and sometimes, I think, legitimately, that allowance should have been made during the basic trial for individual idiosyncrasies, that the clinician should have been free to vary the dosage according to his own judgment of the patient's needs as shown by the latter's responses. Statistically, I see no reason why that should not be done so long as two things are observed and remembered. The first is that we have deliberately changed the question asked of the trial; it now runs "if competent clinicians in charge of defined types of patients use drug X in such varying amounts and for such varying durations of time, and so forth, as they think advisable for each patient, what happens?" The moot point is which question, in given circumstances, is the better one to ask. The second point is that at the conclusion of such a trial we can *in no circumstances* compare the effects of the different regimens of treatment that have been used. These regimens have been determined by the conditions and responses of the individual patients; to observe then, at the end of the trial, the patients' differential conditions and responses in relation to their treatments is merely circular reasoning. We cannot possibly measure thus the advantages and disadvantages of differing regimens but only, by an expansion of the controlled trial, include groups randomly allocated to such treatments. The main danger of this free-for-all trial is the apparently almost overpowering attraction to some clinicians of circular motion.

The Answers

These are the opening gambits of the clinical trial — the definition of the patients, their random allocations to the treatment groups, the laying down of the treatment schedule. At the end of play we shall have to add up the score. Clearly the more objective we can make our means of so doing, the safer we shall be from criticism. Naturally, therefore, we turn first to measurable characteristics. Stone-dead has no fellow, and pre-eminent, therefore, stands the number of patients who die. No statistician, so far as I know, has in this respect accused the physician of an over-reliance upon the clinical impression. Fortunately, however (except statistically speaking), many diseases upon which we must test new treatments are not particularly lethal. Other objective characteristics must be sought and are usually found in such features as the duration of fever, the level of the blood sedimentation rate and the presence of infecting organisms. The appropriate measures vary from one disease to another, and I shall not dwell on them now. I would, however, stress that it is essential to decide *before* the trial begins exactly what measurements will be taken, how they will be taken and precisely when. Without those prior decisions chaos will reign. Working under them we shall have measurements by means of which we can see whether the different treatment groups show differences in their average levels of, for example, body temperature or body weight on specified days — and in their ratios, such as the proportion of patients who become afebrile after a given duration of treatment and the proportion who have a raised blood sedimentation rate. Even with prior thought, it is not always a quite simple task to lay down requirements. In another field, that of the law, Lord MacMillan notes that a statute of the reign of King George III enacted that the penalties under the Act were to be given one half to the informer and one half to the poor of the parish. Neither the informer nor the poor would, however, he points out, be likely to be grateful for this benefaction since the only penalty prescribed was fourteen years' transportation to the colonies. If lawyers, with their care for words, can so err, how much more the doctor!

Returning to our answers from the clinical trial, they present, it will be noted, a *group* reaction. They show that one group fared better than another, that given a certain treatment patients, *on the average*, get better more frequently or more rapidly, or both. We cannot necessarily, perhaps very rarely, pass from that to stating exactly what effect the treatment will have on a particular patient. But there is, surely, no way and no method of deciding that. Also it is well to remember that observation of the group does not inhibit the most scrupulous and careful observation of the individual at the same time — if it is believed that more can be learnt that way.

"Whereas I Was Blind, Now I See"

Passing now from such measurable to the less measurable characteristics of disease processes we shall have, in some cases, x-ray evidence and in all of them the clinical assessment to which I referred above. The difficulties inherent in making unbiased and reliable assessments of x-ray films we have endeavored to meet in some Medical Research Council trials, and I think with much success, by having the films read and the radiologic changes assessed by one or more persons, who have in so doing been kept in ignorance of the treatment groups to which the patients belonged. That process resolves to a large extent the problem of bias, and it does not subject the clinician in charge of the case to hardship. The clinical assessment of the patient's progress and of the severity of the illness he has suffered presents considerably greater difficulties. If it is to be used effectively, without fear

and without reproach, the judgments must be made without any possibility of bias, without any overcompensation for a possible bias, and without any possibility of accusation of bias. In other words, the assessment must, whenever possible, be "blind"— the assessing clinician must not know the treatment of the patient. And that, of course, is a difficult procedure to carry out — indeed it is sometimes quite impossible. Thus, in the Medical Research Council's trials of streptomycin in young adult patients with phthisis, in the early days of that antibiotic, the group on streptomycin had to undergo repeated injections — at least twice a day for several months. It would have been out of the question, for very obvious reasons, to give the contrasting group similar injections of an inert substance. There could be no way of disguising the treatment from the clinical observer (in fact, in that particular situation it was not very necessary to do so, for we could judge on the mortality and the radiologic changes that were assessed "blind").

On the other hand, in the Medical Research Council's trial of an antihistaminic drug in the "cure" of the common cold, feelings of enthusiasm or skepticism may well run high in the bosom (or should I say nasal mucosa?) either of the recipient of the drug or of his clinical observer, or indeed of both. If either were allowed to know the treatment that had been given, I believe that few of us would without qualms accept the result that the drug was of value — if such an answer came out of the trial. Here a group given a placebo was essential; here complete ignorance in both patient and doctor of the treatment actually given was essential. It was not very difficult to incorporate both features in the trial (and in passing it is interesting to note that the placebo produced not only the same number of "cures" as the drug but almost as many "side-effects").

Sometimes another possible approach is by means of the administration of a known treatment, such as an injection, by one person and the clinical assessment of the results by another kept in ignorance of that treatment. In short, given a problem of sufficient importance and given sufficient faith in the method, some means of making a "blind" clinical assessment can frequently, if not usually, be devised. It is thus, I suggest, that we may effectively give the "coherent though impressionistic picture" its rightful place in the therapeutic trial.

Responsibilities

Without demanding this "blind approach," or even accurate measurement, it is certainly likely that the mental contrasting of one treatment with another or of a new treatment against the old will give, at least in the hands of able observers, a truthful, if not precise, answer *if* a treatment has a real and considerable effect. Without a strictly controlled trial the merits of a penicillin could not (and, in fact, did not) fail to come to light. With "winners" it is very easy for the critics of the statistical approach to be wise after the event and to say that the general evidence available at the start of a long, and perhaps tedious, trial made it unnecessary or pedantic. They tend — conveniently to themselves — to forget the occasions when the trial has shown that a treatment has little, if any, value — in spite of the general "evidence" that was available. The history of medicine is, surely, sullied by the reigns of such vulgar upstarts whose prestige, lacking a scientific test, lingered on, to the detriment of human welfare.

It is easy to point out to critics, too, that it is difficult to determine through clinical impressions whether a drug is quite useless or of some slight but undoubted value, and that it is even more difficult to determine with uncontrolled and unco-ordinated observations whether, today, one powerful antibiotic is more valuable than another in particular situations. For instance, I do not think the relative values of aureomycin, chloramphenicol and penicillin in the treatment of clinical pneumonias could have been clearly determined without a fairly large-scale and carefully controlled trial.[5] In particular, it is frequently important that negative results be proved and proved without undue delay.

These particular criticisms are not difficult to meet. The real difficulty to my mind lies, or may lie, in the charge that a trial is unethical. For instance in a letter to the *British Medical Journal*[6] a writer protesting — in part, I think, justifiably — against present ways of writing scientific papers referred to "the replacement of humanistic and clinical values by mathematical formulae," the degradation of patients "from human beings to bricks in a column, dots in a field, or tadpoles in a pool" and "the eventual elimination of the responsibility of the doctor to get the individual back to health." This appears to me to be in essence an attack upon the clinical trial, though the writer at the same time admitted that "since science is mensuration, a sound statistical survey of any particular form of treatment is of importance." I am impressed by the editorial comment in the same issue of the *British Medical Journal*, a comment that expresses what I feel and could not myself, I am sure, have put more clearly:

> It is difficult to see how the results of this 'sound statistical survey' can be analyzed other than by a 'sound statistical method' and presented otherwise than in a 'sound statistical table,' with, if necessary, mathematical formulae. By so doing, in what way are we degrading the patients and shirking our responsibility as doctors to get the individual back to health? The whole essence of the survey is to see whether a new treatment will benefit patients. There appear to be two ways of going to work. We may try it on John Smith and Mary Robinson and report the results — that John Smith got well quickly and Mary Robinson, poor soul, died. We have done our best for both. Or we may try it on 30 John Smiths and 20 Mary Robinsons and report the numbers that lived and died. It is likely that the larger numbers will be found more informative than the smaller. But have we done

any less for the 50 than we did for the two? Further, if there be no ethical reasons to forbid — a paramount and overriding proviso — we may contrast the 30 John Smiths and 20 Mary Robinsons with an equal number of men and women not given this new and wholly unproved drug. Wherein have we shirked our duties? In treating patients with unproved remedies we are, whether we like it or not, experimenting on human beings, and a good experiment well reported may be more ethical and entail less shirking of duty than a poor one.

As R. A. McCance,[7] professor of experimental medicine at Cambridge University, has pointed out, "the medical profession has a responsibility not only for the cure of the sick and for the prevention of disease but for the advancement of knowledge upon which both depend. This third responsibility can be met only by investigation and experiment." The statistically guided therapeutic trial is not the only means of investigation and experiment, nor indeed is it invariably the best way of advancing knowledge of therapeutics. I commend it to you as *one* way, and, I believe, a useful way, of discharging that third responsibility to mankind.

References

1. *Fifty Years of Medicine.* London: British Medical Association, 1950.
2. Bartlett, M. S. Some remarks on the theory of statistics. Read before the Manchester Statistical Society, 1951.
3. Dale, H. Measurement in medicine. *Brit. M. Bull.* 7:261-263, 1951.
4. Psychology and science. *Brit. M. J.* 2:1114, 1950.
5. Treatment of clinical pneumonia with antibiotics. *Brit. M. J.* 2: 1361-1365, 1951.
6. Infectious diseases and vital statistics. *Brit. M. J.* 2:1088-1090, 1951.
7. McCance, R. A. Practice of experimental medicine. *Proc. Roy. Soc. Med.* 44:189-194, 1951.

ACUTE PERFORATED PEPTIC ULCER*

Criteria for Operation and Analysis of 500 Cases

Stanley Mikal, M.D.,† and William R. Morrison, M.D.‡

BOSTON

THE purpose of this paper is to present an analysis of 500 consecutive cases of acutely perforated peptic ulcers admitted to the Boston City Hospital from January, 1938, to May, 1950, and to attempt to present some criteria for selecting the perforations that should be treated surgically and those that should be treated nonsurgically.

There were 471 males and 29 females, a ratio of 16:1. Four hundred and seventy-one patients were white, 25 were Negro and 4 were oriental. Most were between the ages of thirty and sixty, but 75 cases occurred after the age of sixty (Fig. 1). The youngest patient was a girl fourteen years of age; the oldest, a woman of eighty-four. A slight increase in the number of perforations occurred in the fall, as compared to other seasons. Perforation occurred in 90 patients while they were at work, 351 during rest and 59 during sleep. The acute catastrophe of perforation appeared to be the first sign of an ulcer in 84 cases. The remaining 406 patients had had ulcer symptoms for from a few weeks to five years. Operation for perforated peptic ulcer had previously been performed in 8.6 per cent of the cases.

Three cardinal symptoms and signs of perforation were: sudden onset of acute abdominal pain, restlessness and abdominal rigidity. The pain often simulated, in order of frequency, that of acute cholecystitis, acute pancreatitis, coronary occlusion, appendicitis, diaphragmatic pleurisy and small-bowel obstruction. Although the abdominal wall was usually tense and spastic, abdominal distention was present in 8 per cent (40 cases). Hematemesis and melena occurred in 12.8 per cent (64 cases) and was associated with multiple peptic ulcers in 15 cases. Any patient with a perforated peptic ulcer associated with bleeding should be suspected of having multiple ulcers.

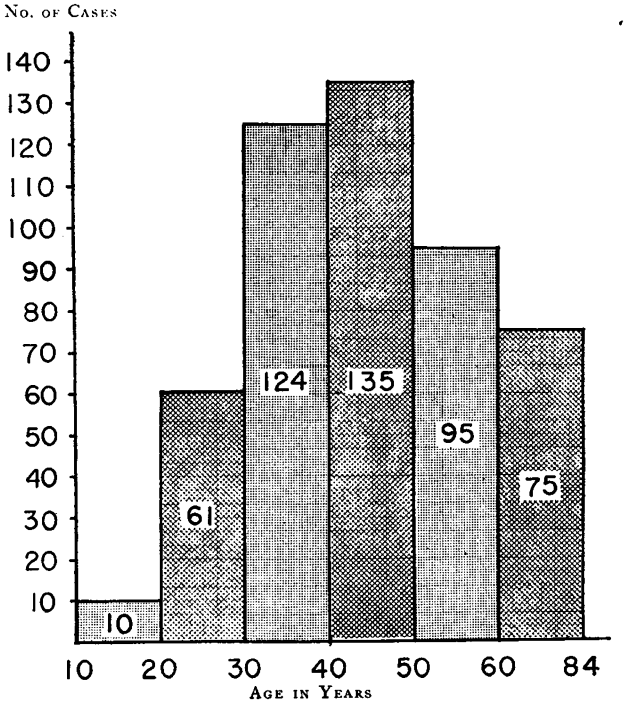

FIGURE 1. *Age Distribution of Patients with Perforated Peptic Ulcers.*

Three hundred and twelve patients had x-ray films for pneumoperitoneum, and in 232, or 74.4 per

*From the Surgical Services, Boston City Hospital.
†Instructor in surgery, Tufts College Medical School.
‡Formerly, surgeon-in-chief, Surgical Services, Boston City Hospital.