4
Factors Affecting the Impact of RCTS on Medical Practice

4. Factors Affecting the Impact of RCTS on Medical Practice

The decision to conduct a randomized clinical trial (RCT) creates a potential impact on medical practice. The act of participating in a trial may have a limited impact, on the practice of at least those physicians directly involved. Once an RCT is complete, both its own characteristics and those of the technology it is used to evaluate determine the trial's impact. This chapter first outlines the objections and alternatives to RCTS that may bear on the decision to carry them out. The latter part of the chapter describes those characteristics of RCTS that appear to most influence their impact, which include:

1. the timing of the trial with regard to the technology's degree of development and diffusion;
2. the constituency supporting the technology prior to the trial;
3, the quality of the trial, both in statistical and other design features;
4. the fact of whether the trial is conducted through one or more centers;
5. the form of disseminating trial results; and
6. other important characteristics.

OBJECTIONS TO RCTS

Objections are rarely if ever raised to the principles of controlled experimentation on which RCTS are based. RCTS themselves, however, are not universally accepted. Two objections are commonly raised against them:

1. that they are too difficult to conduct; and
2. that they may violate the ethical principles that apply to all experimental research involving human beings.

Practical Problems in Conducting RCTS

Objections to RCTS because of their practical problems focus on the use of resources. RCTS are expensive compared with other study designs, can require long periods of followup, and can be administratively complex. If other study designs could answer the questions asked as RCTS can, these objections would be compelling. This is not the case, however, as a later part of this chapter explains ("Alternatives to RCTS").

With regard to cost, it is easier to put a price tag on an RCT than on the expense of not doing one. The widespread adoption and use of ineffective technologies can waste scarce resources. For instance, before a great deal of diffusion, RCTS checked the use of hyperbaric oxygen treatment for cognitive deficits in the elderly, a practice that could have become widespread (see box F in ch. 5). The balance sheet for RCTS might look different if their "credits" could be shown as easily as their "debits. " This is not to claim that every RCT saves money in the long run.

RCTS, especially multicenter RCTS, can be complex administratively. Like all other good research, they require careful planning, execution, and data handling and analysis. These do not appear to be valid reasons for not undertaking RCTS. To some extent, the pratical problems have been lessened by the widespread availability of computers for data handling.

Ethical Issues in conducting RCTS

The most frequent objections to RCTS are on ethical grounds. These objections center on the rights of patients to get the best treatment available and the responsibility of physicians to pro-

98-825 0 - 83 - 4

vide it. Clearly, certain kinds of experimentation on human beings are not acceptable. When the evidence is overwhelming that a newly developed therapy is efficacious—penicillin for pneumococcal pneumonia, for instance—it would be unethical to withhold this therapy from a control group, although RCTS might be appropriate to determine its optimal regimen. The choice between competing technologies or the superiority of an innovation is not always clear. Ethical issues are most difficult in the middle ground where uncertainty is greatest.

The decision to fund an RCT, or any human research, involves at least an implicit decision that the trial is ethical and that it addresses an important question about which uncertainty exists, After this point, the mechanism protecting the individual's rights are procedures of "informed consent ." While informed consent may appear a simple idea, universally acceptable methods of seeking and obtaining informed consent still elude us, though progress has been made.

International bodies have developed ethical codes addressing the particular problems of research. Such codes include the Nuremburg Code and the World Medical Association's Declaration of Helsinki (13). In this country, the Department of Health and Human Services (the Department of Health, Education, and Welfare prior to 1979) has conducted a number of studies on human research that offer guidance on ethical issues (132).

A number of other measures have been proposed to minimize the subtle coercion of patients in obtaining their consent to participate in trials. For example, the World Medical Association suggests that a physician who is not part of the investigation discuss informed consent with the patient. The Department of Health, Education, and Welfare's National Commission recommends giving patients adequate time to decide whether to participate and reducing other potentially coercive environmental conditions (132),

The Cancer Research Campaign Working Party in Breast Conservation recently recommended the following points to improve methods of seeking in formed consent in breast cancer trials (33):

(1) Eligible patients should be given the option to take time to consider giving their consent, perhaps along the lines described by Simpson at the Wellington Hospital in Australia [77]. Here the patient is fully informed about the trial by her physician or surgeon but an informal consent in principle only is obtained. At a later date the procedures are again explained and only then is formal consent obtained by asking her to sign a consent form.

(2) The consent form should be fairly non-specific but it must be backed up by as much verbal explanation as possible. Signature to such a form in the presence of a witness might have legal validity if it included the phrase "the effect and nature of such treatment have been explained to me, " but only if it could be proved that the explanation had been given **[68].**

(3) Ideally, a trained nurse counselor or other suitably qualified person should help to obtain informed consent, and the patient should be made aware that she may resume this continuing dialogue at any time.

(4) Ethical committees should view the issue of informed consent as a top priority, bearing in mind its various applications—in the ordinary clinical situation, in therapeutic trials, and in experimental research. They should reconsider the type of guidelines to propose to doctors, with reference to the Declaration of Helsinki and other national and international codes and regulations; they should consider practical ways of improving consent procedures in their hospitals; and they should monitor these procedures, perhaps by requesting reports at stated intervals.

(5) Those doctors who treat patients with cancer but do not participate in randomized clinical trials should realize that they too have an obligation to discuss alternative forms of treatment with their patients. In our view the fact that they are not formally randomizing their patients does not reduce their obligation in this respect.

While RCTS in this country today require the "informed consent" of the participants, the procedures used to obtain consent vary considerably. Critics point out that the rights of certain classes of patients, e.g., children, the aged, the mentally retarded, and prisoners, are easily violated. The steps taken to protect patients' rights are generally reviewed by at least the funding organization and any institutional review board with jurisdiction over the investigators. While patients' rights are a major concern, mechanisms have been established to protect those rights.

A unique issue arises in seeking patients' informed consent to participate in an RCT. Seeking consent for a particular procedure is more easily accomplished than seeking consent to be randomized and to undergo the uncertainties of such assignment. The role of and need for randomization must be communicated, as well as the risks and benefits of all possible treatment assignments.

The difficulties of seeking informed consent for RCTS may be daunting, but they are not reasons to abandon RCTS. If the same standards of informed consent were applied to experiments with control treatments, the difficulties might appear less (33):

> This argument may be taken to its logical conclusion: that clinicians treating patients outside any protocol in any area of controversy also have the obligation to inform their patients of the alternative treatments that are being offered in different parts of the country at the same time.

A common contention is that control groups are deprived of the benefit of therapy by participating in an experiment. It is a common misconception that control groups are administered only a placebo or no treatment at all. In any case where a standard or accepted technology is challenged by a new technology, the ethical comparison is usually between the standard and the new. In some cases, it may be ethical to use a placebo even if some treatment is available, for instance, in a trial of a new headache remedy, a placebo might be used instead of aspirin.

A frequent objection to randomizing is that some patients will be denied access to the innovative intervention. On the other side of the coin, objections may be raised because participants are subjected to new technologies with unknown risks (21). A general conviction that research subjects are exploited or manipulated regardless of the benefits they might receive contributes to the ethical objections (2 I).

The responsibility of a physician is to give patients optimal treatment. Ethical arguments against randomizing state that physicians should act on the best information available and choose the intervention they believe is superior. When uncertainties about new or existing interventions allow no clear distinction, "a physician makes the intellectually honest admission that best therapy is not known, and than an ethical course of action is to undertake a randomized clinical trial to find out" (32). In fact, the ethical failure of relying on uncontrolled experiments is that lack of effectiveness and side effects are recognized much later than they would be if tested in RCTS (33).

Ethical issues may confound attempts to evaluate practices that are questionable but so entrenched in medical practice to make an RCT all but impossible. Hiatt (110) cites as examples coronary care units (CCUS) in hospitals in this country, and cytologic screening for cervical cancer. Treatment in a CCU indisputably adds greatly to the cost of care but is of unknown value in lowering mortality from myocardial infarction. The development and subsequent widespread use of cytological screening for cervical cancer (the Papanicolau or Pap test) followed a decline in the incidence of that cancer. The value of this screening and the optimal interval for its use are unknown. Both these interventions use a great deal of health care resources: the first mainly because each episode of its use is costly, the second because it is applied to almost half the adult population, and up to 40 or more times during the course of each woman's life. In the case of CCUS, two RCTS in Great Britain found no advantage of CCUS over home care. Nonetheless, RCTS in this country would be extremely difficult to do, and if results were contrary to current practice, they would probably be received unfavorably.

Ethical concerns do not disappear once a trial starts. As data are continually gathered and endpoints recorded, answers about safety and efficacy may emerge more quickly than anticipated. In the case of detecting unsuspected adverse effects, as occurred in the Coronary Drug Program (ch. 5, "RCTS in Cardiovascular Disease") and the University Group Diabetes Project, a decision must be made about when to discontinue treatment. In such cases, however, there are no rules to rely on. Differences of opinion arise about questions of safety as well as of efficacy. Some investigators will be convinced earlier than others that one therapy is better than another. Decisions to stop large-scale trials are generally made by an oversight committee of some sort, and are reached by con-

sensus. Klimt (124) discusses the major issues involved in terminating a long-term trial.

Whether enrolled in a clinical trial or not, a patient deserves the best possible treatment from his or her physician. Particularly in long-term chronic disease trials, patients' conditions may change during the course of the trial so that different treatments are indicated. In whatever way a trial is organized, a physician retains and must exercise the responsibility to withdraw the patient at any time, or to offer the competing treatment or a different one, whenever any such change is in the patient's interest. In recent trials of coronary artery bypass surgery that assigned individuals to surgical or medical treatment, a large number of those assigned to medical treatment have subsequently undergone surgery for intractable angina. These necessary changes of treatment have changed the research question from "Which is more effective, medical or surgical treatment?"to "Which is more effective, immediate surgical treatment or immediate medical treatment, followed by surgery only in those patients for whom medical treatment is insufficient?" The second question conforms more closely to actual practice than the original one.

Another ethical concern is how long researchers and funding agencies should follow those patients who participate in clinical trials (255). Perhaps a lifetime followup is desirable for some classes of participants. The potential long-term effects of some chemotherapeutic agents are worrisome, especially those of anticancer drugs. At present, funding agencies do not routinely provide for long-term followup.

ALTERNATIVES TO RCTS

The money and time that RCTS require have led to a continued search for alternative means to determine the safety and efficacy of medical technologies. It is generally argued that any acceptable method must compare a group that undergoes the new treatment (or other intervention except in the rare case that the experimental treatment is an obvious major breakthrough) with a group that does not. The arguments center on the ways in which these groups are assembled. The major rival to RCTS has been the type of study that uses "external controls, " most frequently "historical controls. " External controls are those drawn from populations that may differ, in unknown ways, from the study population. Historical control trials (HCTS) compare a group of patients treated by the new intervention with a group treated sometime in the past in another way. Another type of external controls are patients treated during the same time period at the same or different institutions from the experimental group, but who are not assigned to treatment according to the experimental plan ("concurrent controls").

The data on historical controls ranges from dim personal remembrances to that gathered carefully and in detail by investigators (24). Historical controls may have been treated at the same or a different institution as the experimental group. They are generally chosen from the literature, from the immediately preceding trial in a sequence of trials, or matched from a previous study (88). Successful matching assumes knowledge of important prognostic factors, which is often not a valid assumption.

The attractions of historical controls are several. HCTS sidestep the question of whether it is ethical to randomize patients. Studies with historical controls require the active cooperation of fewer participants since data need be newly collected only for the experimental patients. Requiring fewer participants makes studies proportionately cheaper. Recruitment into the study is improved to the extent that patients need not consent to randomization and are sure of the treatment they will receive beforehand.

Gehan and Freireich (88) argue that clinical trials in cancer research should sometimes use a selected rather than a randomized control group. They cite the following kinds of cases: 1) when the study attempts to determine the absolute rather than the relative effectiveness of the treatment, 2) when large differences in response rate

between treatment groups are based on preliminary trials, and 3) when a therapy can be compared to a standard therapy evaluated in a recent trial. Addressing at least the second kind of case, Chalmers, Block, and Lee (45) argue for randomized controls on the grounds that most drugs tried to date in cancer therapy have been relatively ineffective.

In the past, data on external controls have usually been gathered from patient records by abstracting the relevant information. Because the primary purpose of such records is for patient care rather than research, the requisite information is often not recorded. Data banks are a relatively new development that may improve the quality of external controls, but this is yet unproven. Medical data banks are usually created by establishing a common vocabulary to describe clinical histories, and then observations on patients are entered as events occur (234). The uniform information available about patients can be used to improve the comparability of an experimental group and a group of controls (who are chosen from a data bank). Nevertheless, data banks do not solve the problem of treatment changes over time that may render groups incomparable, particularly because not all medically significant variables can be identified. While data banks may be useful in discovering some important prognostic factors, they are not good enough to compare treatments (99). In this regard, Byar observes (31):

> The great danger seems to me to be that data banks will be seen as a replacement for randomized trials, whereas in fact the most useful data which could be stored in data banks would be those obtained from randomized studies.

When a technology is so widespread or well established that use of untreated controls would be questionable, investigators must then rely on historical data. When random assignment to groups is possible, however, the available evidence suggests it is superior. Wortman and Saxe (252) compare the validity of RCTS with that of HCTS (and other epidemiologic study designs). The major advantage of RCTS is their internal validity; i.e., high probability that the effects they reveal result from using the technology and not from some other factor. HCTS, in contrast, often lack internal validity. Whether identifiable or not, changes over time in medical practice or the patient population are often equally likely explanations of effects detected in HCTS. This is illustrated by changes in the treatment of osteogenic sarcoma. The history of this treatment points to the hazards of comparing aggregate survival rates from time periods before and after a procedure is introduced (252):

> Following the development of this treatment in the early 1970's, researchers began to experiment with ways to treat patients with the drugs before their cancer metastasized. Historical controls drawn from patients' records dating from the 1960's were used in this research, and the results provocative. Nearly half the patients treated lived 2 years without a recurrence of the disease, compared to only 20 percent of the patients in 1960.
>
> Unfortunately, the change in therapy from 1960 to 1970 was also accompanied by other changes in diagnosis, treatment, and patients. The use of the computed axial tomography (CAT) scanner in the 1970's provided a much more sensitive test for detecting patients who did not have metastasis. At the same time, surgeons began removing metastasis in the lungs. At the Mayo Clinic, where both of these techniques were employed without chemotherapy, the survival rates equaled those of patients treated with the drugs.
>
> In addition, the patient mix probably changed over time so that those with the worst prognosis no longer constituted the majority of those treated. These criticisms of the research design and findings of a small controlled trial have convinced the National Cancer Institute to support a multicenter RCT to assess the efficacy of adjuvant chemotherapy for osteogenic sarcoma.

Sacks, Chalmers, and Smith (197) compared the outcomes of RCTS for six therapies that each had been tested by at least two RCTS and two HCTS. In every case, HCTS indicated these therapies were more beneficial than did RCTS, the difference lying mainly in the outcomes of the control groups. In HCTS, control groups fared considerably worse than controls in RCTS, while the treatment groups fared about the same. To provide a better comparison, the results of some HCTS were adjusted to account for differences in prognostic factors between HCT and RCT groups. Sacks and colleagues found that this had little effect on the analysis and concludes that little can be done to improve the accuracy of HCTS. The problem of

using historical control is not the existence of bias per se, but the impossibility of detecting, measuring, or removing it.

HCTS are more likely to favor a new treatment because of the nature of historical controls. RCTS are more likely to find no difference between treatments even if a difference exists. Although other factors may contribute to not detecting an effect when it actually exists, the main culprit is an inadequate sample size, and not an inherent weakness of RCTS. The problem could partly be solved by greater emphasis on power considerations in experimental design, with planning for sample sizes large enough to ensure finding any important difference in treatment groups.

Sacks and colleagues (197) suggest in addition that the "nearly automatic" use of a p value of 0.05 as a measure of statistical significance may not always be appropriate. Such an association means that the prespecified result is expected to occur by chance alone 5 times out of 100, given the sample size of the trial. They suggest that positive results of RCTS might be accepted as true positives even assuming **a** greater possibility that the results may be due to chance. On the other hand, given the bias in favor of new interventions in HCTS, a more stringent significance level might be required of them for the same level of proof.

Wortman and Yeaton (253) synthesized the results of studies of coronary artery bypass graft (CABG) surgery. They looked at both RCTS and nonrandomized studies with concurrent controls reported between 1970 and 1981. They conclude that both kinds of trials favor surgical treatment, but that nonrandomized studies tend to overestimate its benefit. They combined data on survival and mortality from 9 RCTS and 16 nonrandomized studies by means of two different synthesis techniques. In both cases they found that the average benefit to the surgical patients as computed from nonrandomized studies is four to eight times greater than that computed from RCTS.

Studies to date comparing RCTS and other types of studies indicate that RCTS are and should be the favored method for evaluating major clinical recommendations and should be abandoned only when special conditions preclude them.

CHARACTERISTICS OF RCTS THAT AFFECT THEIR IMPACT

Timing **of** RCTS

At what point in the life of a medical intervention should it be tested in an RCT? The law and regulations answer this question for new prescription drugs and vaccines, requiring RCTS of nearly all. The safety and efficacy of pharmaceuticals must be demonstrated before they can be widely used. To other kinds of interventions, e.g., surgical and radiological ones, no such law applies. RCTS have typically been initiated when a critical amount of skepticism has developed about the effectiveness of an intervention. By then it may have attained widespread popularity, with its attendant consequences—e.g., major investments in learning skills, such as surgical techniques, or in equipment, Many people have been subject to an intervention of unknown efficacy, including ineffective ones, such as gastric freezing for duodenal ulcer (see box D) and some that are actually harmful. These problems may be confounded by the usual delay inherent in changing even a bad technology, and the increased grounds for malpractice suits for an abrupt public admission of error.

One approach to the timing of trials is to "randomize the first patient." Chalmers is one of the main proponents of randomizing patients to treatments with the first use of a new intervention. He cites several times this has occurred, including trials of prophylactic use of portacaval shunt surgery (a procedure to allow blood flow to bypass the liver) for portal hypertension (abnormally high blood pressure in the veins of the liver, a frequent complication of liver cirrhosis) and colon bypass for chronic encephalopath, (a degenerative disease of the brain) in patients with cirrhosis (41). Randomizing from the very first is possible in some cases, but there are convincing arguments to delay the start of RCTS (though not to delay establishing formal systems to collect data),

Box D.—Gastric Freezing

The rise and fall (1962 to 1969) of "gastric freezing" **in treating duodenal ulcer is a classic story. The procedure consists of a patient swallowing an uninflated balloon to which tubes are attached. Once in the stomach, the balloon is filled with a coolant, maintained at** -10° **C for about an hour, after which the balloon is deflated and removed. Claimed by its originator, Owen Wangensteen, a leading academic surgeon, to decrease gastric secretions, to relieve pain, and to be safe, simple, and relatively inexpensive (245), gastric freezing quickly gained popularity. The only rival treatment to gastric freezing was palliative medical treatment with antacids, sedatives, and changes in living habits, or in severe cases, surgery with a mortality rate of 5 to** 10 **percent (160).**

Despite enthusiastic adoption of gastric freezing, enough doubts about it remained to spur the planning of a multicenter RCT in 1963. When **the** results **appeared in 1969 showing no difference in outcome between the group that had received gastric freezing and the group given a sham procedure, 2,500 gastric freezing machines were in use. According to Miao, the convincing results of the trial led to rapid abandonment of the procedure (160).**

In a somewhat different interpretation of the events, Fineberg suggests that even before publication of these results, gastric freezing was on its way out. The negative result of the RCT, he claims was "of little practical consequence, as if a marble tombstone were erected over the grave of a patient already several years **deceased (71)."**

Arguments in favor of early RCTS are supported by the use of untested interventions later proved either ineffective (e. g., bed rest for hepatitis, the Sippy milk diet for gastric ulcer [40]) or harmful (e. g., prophylactic portacaval shunt surgery for portal hypertension, which was both ineffective and caused a type of brain damage in some patients **[27]).**

Doubts have been raised about the efficacy and safety of some technologies, yet years pass before they are tested in RCTS. Radical mastectomy was introduced around the turn of this century. In 1948, the simple mastectomy was proposed as an alternative. RCTS, which demonstrated the equality of the two procedures in patient survival rates, waited until **1969 and 1973.** RCTS of bed rest for hepatitis, a bland diet for peptic ulcer, and diethylstilbestrol to prevent spontaneous abortion were delayed for similar periods of time (40).

Three facts argue against very early RCTS of surgical procedures, First, as surgeons' skills in performing a procedure improve, the results of performing it may improve, as measured in mortality or morbidity rates. Secoond, as experience accumulates, improvements to the procedure itself will be made, not only by clinicians involved in trials but by other practitioners. If the procedure evolves to a somewhat different and improved form, the ethical and methodological question arises whether a trial in progress should continue. The Veterans Administration's (VA) RCT of CABG surgery was a well-designed trial, but had minimal impact, in part because changes in techniques made the results irrelevant to practice by the time the trial had ended **(20).** In this trial, the procedure initiaI1y used, the Vineberg implant, was replaced with the newer CABG surgery. Data analysis was further complicated by a higher rate of operative mortality in the earlier CABG patients compared with the later ones. Third, when an innovation is better known, it may be applied to a changing set of patients. In particular, a promising but risky therapy may be applied to patients in earlier stages of disease, patients who may in fact benefit more from the procedures because they may have not yet begun to suffer some permanent late effects of the disease.

Bonchek (20) cites two well-designed RCTS in which problems arose because of the trials' delay in relation to the diffusion of the technology. The Coronary Artery Surgery Study began in 1974 after much experience with the procedure had accumulated. Excluded from the study were some high-risk patients of great interest (e.g., those with unstable angina). By the time the study began,

their physicians presumably preferred them to have surgical treatment. Recruitment into the study was slower than expected, so the enrollment period was extended. Such delay in recruitment creates its own problems owing to evolutionary changes that take place, as was discussed above. A similar problem in recruiting patients occurred in a single-center study of unstable angina at the University of Oregon. Recruitment declined as physicians diverted their patients from the university hospital, not wanting them to be randomized.

Problems with the timing of trials are difficult, and there are advantages and disadvantages to carrying out trials at specific points in the diffusion process. In general, however, the arguments for earlier trials are stronger. The earlier RCTS occur, the sooner sound information is available for medical decisionmaking. The examples mentioned of "late" RCTS, and of no RCTS at all (for most current procedures) are more typical than those of RCTS conducted too early.

The Constituency Behind the Intervention

A strong interest group obviously supports the trials of new drugs. Those with a financial stake in these trials see that the results of positive ones are translated into practice as widely as possible. There is a general consensus that the results of positive drug trials are disseminated widely, and that physicians rapidly adopt new drugs. If there is any problem in adopting new drugs, it is their overuse. Although drug companies cannot label their products for indications other than those for which they have been given FDA approval, physicians are not bound by any law to prescribe according to RCT results.

When RCTS of already marketed drugs have negative results, the situation can be quite different. Beginning in 1961, the University Group Diabetes Program (UGDP) tested a popular hypoglycemic drug, tolbutamide, used in treating adult-onset diabetics to control their blood glucose. Early results of this trial indicated that the drug was unsafe (see box E), and the corresponding part of the trial was discontinued. This finding on tolbutamide set off a heated debate, which is now 13 years old and still alive.

Procedures also have their constituencies. The developers of new procedures and techniques have a professional stake in having them accepted and widely used. Financial interests may also be present when capital equipment is involved, e.g., imaging equipment or devices like heart valves, and joint implants, and when procedures are regarded as high reimbursement items by third-party payers. Positive results seem to have a greater impact in these cases than negative results. A potentially beneficial new procedure is welcomed by practitioners, particularly when the condition it treats is life-threatening and there is no alternative treatment. Rather than abandon a procedure for no treatment, even if an RCT shows little or no benefit, physicians may prefer to continue what they see as the only hope.

The Quality of RCTS

"Quality" in research cannot be precisely and categorically defined but criteria can be established to measure some of its features. Bailar (6) suggests two methods to judge quality: 1) evaluating the quality of the published research report, and 2) evaluating the quality of the work itself. Publications concerned with the quality of RCTS have taken both approaches. Regardless of whether better quality RCTS will have greater impact than those of poor quality, on general principle it is worthwhile to ensure that they are of the highest quality possible.

Most writers who focus on the quality of RCTS use the published literature as their source of data. Some have reviewed published RCTS to determine what features of the trials are reported, with the aim of judging the quality of the published reports. Others have taken data from these publications, i.e., the number of participants and other quantitative items, to judge the quality of the research. These two types of evaluations are discussed below.

The Quality of RCT Reports

Chalmers and colleagues propose a method to evaluate the quality of published RCTS, and a quality index based on this evaluation (47). They give heavy weight to the form of blinding, including blinding during randomization, that of physi-

Box E.— The University Group Diabetes Program

In 1961, the University Group Diabetes Program (UGDP) began an RCT "unique in the amount of rancor it has aroused and the length of time it has lasted" (142). **The trial was sponsored by the National Institute of Arthritis, Digestive, and Metabolic Diseases, to settle longstanding questions about the treatment of "adult-onset" diabetes. The disease is characterized by the impaired ability to metabolize carbohydrates, stemming from the inefficient use of endogenously produced insulin. Traditionally, treatment consisted of controlling blood sugar (glucose) levels by injections of exogenous insulin, dietary management, or taking oral hypoglycemic drugs (agents that act to lower the level of glucose in the blood). The actual value of controlling blood sugar, however, was unknown. Two schools of thought were prevalent at the time: one holding that strict control was warranted, the other that the discomfort, inconvenience, and anxiety of strict control were not worth its benefits (142).**

One **aim of the UGDP RCT was to evaluate the control of blood glucose on the development of major complications of diabetes, particularly atherosclerotic heart disease, the most common cause of death among diabetics. The trial also set out to study the natural history of complications of the disease and to improve methods in clinical trials.**

About 1,000 patients in 12 centers were instructed in dietary control and randomized to one of four treatments: 1) insulin in variable dosages to keep blood glucose at specified levels, 2) insulin in fixed dosages, 3) tolbutamide (an oral hypoglycemic agent widely used at that time), and 4) placebos in the same form and scheduling as tolbutamide. A fifth group, receiving a new oral hypoglycemic agent, was added after the study had begun.

The trial employed rigorous techniques of data collection and patient evaluation, relying whenever possible on objective measures of pathology and functional impairment. Many of these quality assurance and control measures had never before been employed in a large-scale trial. The followup was scheduled to last 10 **years.**

By **the end of the eighth year, higher rate of cardiovascular mortality, one significantly higher than** in **any** other group, **had occurred in the group taking tolbutamide. The investigators discontinued its use and announced the results, touching off a controversy still unresolved. Their further conclusion, that insulin was no more effective than dietary control alone in preventing fatal vascular complications, added fuel to the fire.**

A hue and cry arose from diabetologists, drug manufacturers, and publishers who carried advertisements for the drugs. **The study was scrutinized and attacked on two major counts:** 1) **that treatment of the participants in the trial did not measure up to standards of clinical practice at the time; and 2) that a failure of randomization placed more high risk individuals in the tolbutamide group than in the others, rendering the results invalid.**

In response, the National Institutes of Health (NIH) reviewed the trial and found it valid. The Biometric Society undertook a 2-year review of all the statistical aspects of the trial and came to the same conclusion. The Food and Drug Administration (FDA) conducted a 2-year audit, visiting the treatment sites and checking the data. They found no error (43). The data were finally reviewed by the courts during 10 years of legal action against the principal investigator. The UGDp trial is surely one of the few whose data **have been found satisfactory by the Supreme Court of the United States.**

The UGDP results were published in 1970. Not until 6 to 8 years later did sales of hypoglycemic agents begin to decline (43). In this case it may take the emergence of a new generation of physicians and patients for the practice to change entirely. One effect of the trial may be the policy decision of drug companies not to develop new hypoglycemic agents; none have attempted to seek approval for **such agents since the controversy started.**

Aside from its medical conclusions, the UGDP led to great debate about the value of RCTS in general, and revived the old issue of the relative value of inference and clinical judgment.

cians and the patients with regard to the therapy given, and that of physicians with regard to ongoing results. Analytic techniques, control of bias, description of patient population and treatments, and various aspects of quality control are counted as well. Adherence to the standards set down by these authors might raise the quality of RCTS, and might also facilitate comparing and synthesizing the results of small trials, particularly those with conflicting results.

DerSimonian and colleagues (62) studied the quality of reports of RCTS in 67 articles published in the New *England Journal of Medicine,* the *Journal of the American Medical Association,* the *British Medical Journal,* and the *Lancet* during specified time periods in 1979 and 1980. They chose 11 items of methodological importance and determined how often each was reported. A low score might indicate a poorly conducted study, a poorly reported one, or both. Information about statistical analyses, the names of statistical tests used, and the fact of random allocation to treatment were relatively well reported—at least 80 percent of the articles mentioned these items. Only 19 percent reported the method of randomization, 37 percent the eligibility criteria for admission to the trial; 57 percent whether patients were blinded, and 30 percent whether those assessing outcomes were blinded. The least frequently reported item was the statistical power of the trial to detect differences in outcomes, which was reported in only 12 percent of the articles. There were substantial differences among the journals, but they were as great as within-journal variation among articles. DerSimonian and colleagues conclude that journal editors could influence the quality of published trials by setting standards for reporting. The items of information they identify as important should be available to all authors, and could theoretically be reported 100 percent of the time.

Mosteller, Gilbert, and McPeek (165) came to similar conclusions in their reveiw of RCTS in cancer research. They looked at the frequency of reporting of five statistical and two procedural aspects of trials: randomization, statistical method, blinding, statistical power, sample size, patient survival rate, and informed consent. Each item was in **O** to 50 percent of the articles, with "24 percent as a reasonable overall single-number summary, " of the frequency any item was reported. (The authors' recommendations based on this study are discussed in ch. 6.)

Haines (103) notes a number of deficiencies in reports of RCTS in neurosurgery, in addition to low statistical power. He found inadequate descriptions of blinding, of interventions tested, and of the eligibility criteria used. Haines did note a trend, though weak, toward improved quality over time, as determined by the scoring system of Chalmers and his colleagues (47). The participation of a biostatistician in the study, as evidenced by authorship or acknowledgment, was the most important correlate of whether a study was judged of good quality.

The Quality of RCT Research

Hemminki (107) cites 29 reviews on the quality of clinical trials published between 1950 and 1977. Hemminki's work was prompted by her previous review of clinical trials submitted to the drug licensing authorities of Sweden and Finland, which showed many trials to be both poorly reported and poorly done. Her conclusion echoes that of the authors of the original reviews, namely, that the majority of published trials were inadequately controlled or otherwise methodologically inadequate. Among the common deficiencies she cites, e.g., lack of statistical power, and lack of information about randomization and blinding techniques, Hemminki includes the unsatisfactory cojoining of information about adverse effects and beneficial effects. Adverse effects, which are generally rare, are usually analyzed separately from indications of effectiveness in comparing therapies. Hemminki suggests expressing both adverse and beneficial effects using the same scale, as in cost-effectiveness analyses. The most frequent criticism of many RCTS is that their sample sizes have been inadequate. Combined with other factors, small sample sizes lead to trials that have little power to detect moderate differences between groups. Statistical power and statistical significance in RCTS are discussed after reviewing other issues of quality in their design, execution, and analysis.

The use of appropriate statistical tests, and the analysis of "crossovers" and withdrawals from trials sometimes have important implications. In the trials of CABG, a large proportion of patients randomized to medical treatment eventually undergo surgery. These "crossovers" are so numerous that these trials do not compare surgical with medical treatments, but rather immediate surgery with initial medical treatment followed by surgery if indicated. That is to say, the trial tests a question of medical management rather than one of clinical efficacy. Data analysis in CABG trials is by "intention to treat. " In some cases data are analyzed according to actual treatment, or the analysis may include both options.

Counting Events

Important methodological issues have been raised by a recent multicenter double-blind RCT, the Anturane Reinfarction Trial (ART). This RCT compared a placebo with Anturane (sulfinpyrazone), a platelet-active drug (one that inhibits blood clotting), in preventing cardiac mortality after myocardial infarction. A publication of the trial's results appeared in *The New England Journa] of Medicine* in 1980 (4), reporting a reduction in cardiac mortality as a result of the drug. The difference was attributable to a decrease in sudden deaths (those deaths occurring within the first 6 months after myocardial infarction) in the experimental group. FDA later criticized the study on two grounds (220): 1) that the criteria used in classifying causes of death were ambiguous and illogical, and 2) that the criteria were not applied consistently. FDA also questioned the exclusion of certain participants and deaths in the analysis. Reanalysis of the data, including a reclassification of deaths by an independent group and by the ART Policy Committee, showed different results, though the same trend that was originally reported. The observed difference in overall mortality was no longer significant, though there were still fewer sudden deaths in the Anturane group compared to the group taking the placebo (3).

The disagreement over the ART in part concerns the way events are counted and attributed in RCTS (196). Decisions about which participants and events should and should not be counted in the analysis to some degree rest on whether the trial is considered one of medical management or clinical efficacy, though there is debate even on this point. In medical management trials all randomized patients are included in analysis, and all events during followup are counted. In trials of clinical efficacy, designed to test the biological effects of interventions, only those patients actually taking the treatments as prescribed are included in analysis, and only those prespecified events likely to be influenced by the treatment are counted. ART was a trial of clinical efficacy using debatable rules for counting, as well as some faulty applications of these rules.

Methods of Randomizing

Randomization does not ensure the equal distribution of characteristics, but it does ensure the valid use of statistical significance tests. Improper randomization, which has occurred many times, ensures neither. Various allocation schemes, more and less successful at randomization, have been based on date of birth, date of visit to the physician or hospital, alternating assignments as patients enter a trial, and other plans. Mosteller, Gilbert, and McPeek (165) review the biases of faulty allocation schemes. For example, in using the flip of a coin or the draw of a playing card, investigators might be tempted to even out groups if they begin to look unbalanced. Alternating assignments can be biased when two patients arrive simultaneously and a decision must be made about who gets which treatment. Physicians may know what the next treatment is and schedule patients accordingly, or they may selectively enter patients only when they approve the next "random" assignment.

In spite of such practical problems, random numbers can be reliably obtained from tables and from computer programs, and there are methods to ensure that investigators do not know which treatment a participant will be assigned. For example, in many multicenter trials treatments are assigned by telephone after patient eligibility y has been established. The person enrolling a patient, therefore, has *no* control over group assignments.

Deviations From Treatments and Protocol

In the course of an RCT, events may not take place according to plan. In one well-known case,

high-oxygen environments were evaluated as a possible cause of retrolental fibroplasia (a condition leading to blindness) in premature newborns. Some attending nurses in one of the studies were so strongly convinced that low-oxygen environments were harmful to the infants that they increased the levels of oxygen. Recognizing this practical problem in carrying out the trial, in another study the oxygen concentration was only partly reduced until the harmful effect of high oxygen concentration was firmly established **(252).** Not adhering to a protocol, as in the first study above, may invalidate the findings of an RCT if the deviation is widespread or unknown in extent. An investigator's lack of adherence to study protocol is probably the most serious type of deviation.

Patients may also deviate from the study protocol. In general, however, their lack of compliance, unlike that of investigators, can be planned for as another aspect of the RCT itself. Protocols can be designed to allow some patient noncompliance without compromising the results. RCT designers usually want to know about clinical efficacy in both experimental and ordinary conditions, making a certain amount of compliance necessary, on the one hand, and the quantifying of compliance necessary on the other. In some cases the percentage of compliant patients may be as important as the biological effect of the intervention, and compliance itself may be designated as an experimental endpoint. If a drug, for exampie, is known to be effective but patients will not take it, it has little value.

Blinding

"Blinding" is keeping secret the treatment assignments (experimental or control) of trial participants (see ch. 1 for more discussion of blinding). Blinding compensates for the expectations of patients and physicians which, whether positive or negative, can affect the experiment's outcome. A patient's sense of well-being maybe enhanced by belief in a treatment, and a physician's assessment of the patient's condition may be strongly affected by the physician's expectations about the treatment.

Blinding in drug trials is accomplished commonly by the use of a placebo, usually an inert substance resembling the experimental drug. Blinding can fail even using a placebo, if, for example, the experimental drug has unmistakable side effects. A failure of blinding can raise doubts about an experiment's conclusions.

Blinding is not possible in some trials, notably those comparing surgical and medical treatments or other markedly different interventions. For example, in the Multiple Risk Factor Intervention Trial the experimental group received intensive counseling while controls went their normal route of care (166). The question arises in such a case whether the effects observed in experimental subjects are attributable to the treatment itself or to the attention they received. If all such trials are considered purely medical management trials, the importance of that distinction is diminished.

Other Issues Concerning the Quality of RCT Research

One criticism of most RCTS, which probably applies to much clinical research, is the information they fail to obtain on how interventions affect "quality of life." McPeek, Gilbert, and Mosteller (152) focused some attention on this issue based on a review of research evaluating new surgical procedures. Many RCTS show that as far as they can be measured, the interventions compared cannot be distinguished in efficacy or safety, Such is often the case in RCTS of cancer treatments. Thus, an important factor in deciding between therapies is the way they affect the patient's quality of life. Research in this area requires developing methods to define and appraise quality of life and developing administrative methods for the long-term followup of pertinent questions without great inconvenience to physicians and patients. Greater cooperation between social and clinical scientists has been recommended to develop RCTS (152).

Little is taught about clinical trials in medical schools, and from this might result poor quality of design and participation in RCTS. Improving physicians knowledge of the value of RCTS and

of their conduct, both in medical schools and in continuing medical education, could motivate their better participation in RCTS.

Statistical **Power and Statistical Significance***

A frequent criticism of RCTS is that they have lacked sufficient statistical power to detect important effects. In practical terms, this means that the number of cases studied is so small that even if the experimental technology is superior (or inferior) to the control treatment, the difference will likely not be detected in the RCT. Failure to detect such an effect is called a "Type II error, " and is analogous to a "false negative. " The probability of this type of error is expressed as "beta. " "Power" is equal to l-beta. Commonly sought power levels are 0.80 and 0.90.

Another type of error, less frequent in RCTS but closely related to lack of power, is concluding that there is an effect when, in fact, there is none. This can and does occur purely by chance because of sampling error. It can lead to adopting or rejecting a treatment that is neither more nor less effective than the tested alternative. This is known as "Type I error" and *is* analogous to a "false positive. " The probability of this kind of error is expressed as "alpha, " which is commonly called the level of statistical significance. Common alpha levels are 0.05 and 0.01.

The power of a trial is the probability of detecting an effect of at Ieast a specified magnitude at a specified level of statistical significance. For example, a trial might have a power of 0.80 to detect a 50 percent better outcome in the experimental than in the control treatment at the 0.05 level of statistical significance.

As power is a function of sample size, it is essential in designing an RCT to determine the sample size needed for an effect of a specified magnitude to be judged statistically significant. Specifying the magnitude of effect depends in turn on the investigator's judgment of how large an effect would be practically significant and at the same time, how large an effect can be realistically expected. The larger the sample size, the higher the probability the test has of detecting an effect of a given magnitude, or, alternatively, the smaller the effect the test can detect as statistically significant. As sample size increases, however, so does the cost of the study. It would be wasteful to choose a sample size so large that it would detect a difference that has no practical significance. The investigator must make a judgment weighing cost and statistical power. Investigators frequently overestimate the effectiveness of the treatment under study and therefore underestimate the size of sample needed to detect a statistically significant effect. For example, the sample size may be chosen on the premise that the experimental treatment is 50 percent better than the control treatment, whereas in reality it is only 20 percent better. Statistical analysis is likely to lead to the erroneous conclusion that the experimental treatment is "not statistically significantly better" than the control even though the investigators might have considered the improvement of 20-percent important. Had the investigators chosen the larger sample size needed to detect a 20-percent improvement as statistically significant, they would have avoided this Type II error.

Small studies do have a place in the greater scheme of research, as pilot and feasibility tests, and, should a real breakthrough occur, they can detect such a big effect. Small studies in themselves are not the problem. Too often, though, they are treated as definitive, and not evaluated in light of their probability of finding a true difference.

Small study sizes and concomitant lack of statistical power are well illustrated by reviews of published cancer RCTS. Mosteller, Gilbert, and McPeek (165) surveyed the sample sizes in over 400 trials referred to in the volume *Randomized Trials in Cancer: A Critical Review by Sites (211;* discussed in ch. 5) as well as 54 RCTS from the journal *Cancer* that Zelen and colleagues review in an earlier paper (258). Zelen concluded that the median sample size was about 50 per treatment group. Mosteller and colleagues (165) found this calculation to be "a bit optimistic. "

A "typical trial, " conducted on 50 patients, has a probability of less than 0.40 to detect a difference from 20 percent of patients responding in the

*This section benefitted considerably from reference 14.5.

group, to 40 percent in the other (at the **0.05 level of significance) (258). Referring to these same data, Mike (161) noted that the studies could provide reasonable power only for differences in outcome so large as to be highly unlikely.**

Zelen (257) has addressed the problem of false positives in cancer research. Given the small sample sizes and the low probability of success of most trials of cancer therapies, Zelen calculates that of every five such trials with positive results, only two are true positives (see ch. 5 for a fuller discussion). Most positive results are published without being confirmed in a second trial, and oncologists are prone to accept them uncritically into their practices (258).

Haines evaluated the statistical power of published RCTS in neurosurgery. Of the 51 trials published since 1945, half had less than a **50/sO chance of finding a difference in outcome as large as 50 percent between the experimental and control groups (103).**

Sometimes the sample sizes chosen for studies are based on unrealistic estimates of a treatment effect, making the studies too small to detect lesser but still important effects. Clinicians dream of spectacular new therapies, but in fact most progress occurs in small, incremental steps. Statisticians should be conservative in determining necessary sample sizes, and should aim for significance levels higher than are seemingly needed (25).

Greater cooperation between statisticians and clinicians is a way to improve the quality of trials. Haines showed that the best sign of a well-designed trial in neurosurgery was the participation of a statistician. This is probably true in a wide range of research. About 10 years ago, the National Center for Health Services Research studied the factors that affect approval of research grant applications. They found the most important single factor was the presence of a biostatistician on the proposed staff. Presumably this finding reflects the work of a biostatistician in preparing the proposals and accordingly, the proposal's substantive merit, rather than the mere presence of a biostatistician's name (145).

Recruitment

Many studies are never completed or not adequately completed because of poor patient recruitment (87). This stems, at least in part, from the tendency of clinicians to overestimate the number of patients available for study. Brown (25) states that clinicians overestimate the number of patients that can be recruited by at least twice, and sometimes as much as 10 times.

The problems of recruitment were graphically illustrated in the National Heart, Lung, and Blood Institute's (NHLBI) Lipid Research Clinics Coronary Primary Prevention Trial. The protocol called for 3,550 men to be recruited from physician referrals, advertisements in the media, clinical laboratories, blood banks, occupational screening, and other sources. The number of likely subjects was seriously overestimated, causing the project to fall behind schedule. While 46.5 percent of those referred from physicians and laboratories were recruited, only 2.5 percent of those from the other sources were. This experience was not unique. Tallying the numbers from four large-scale studies—this study, the National Diet Heart Study, part of the Hypertension Detection and Follow-Up Program, and VA's Mild Hypertension Study—almost 1 million contacts were screened to yield about 11,000 entrants (129).

Recruitment should take place as quickly as possible to avoid time-dependent trends that may complicate comparisons between patients recruited early and those recruited later.

The need to recruit many patients quickly has led to greater numbers of multicenter trials, an arrangement that appears to improve the quality of trials for reasons other than reliance on sheer numbers (see "Multicenter v. Single Center Trials," below). A related development, especially in RCTS of cancer treatments, is including community hospitals along with major research and teaching hospitals in multicenter RCTS. This reflects the trend of treating cancer patients in the community setting.

Multicenter v. Single Center Trials

More than half the RCTS in the 1979 NIH Inventory of Clinical Trials involved the participation of more than one institution. Such trials have a number of advantages.

Regardless of the experiment's protocol, recruiting at a number of institutions shortens the time necessary to enroll the participants. Such trials may take longer in planning, but prolonged recruitment can cause difficulties for RCTS (see "Recruitment"). In studying rare diseases, the cooperation of a number of centers is necessary to enroll even a modest number of patients. Permanently constituted "cooperative oncology groups" have been a mainstay of cancer therapy RCTS, especially in allowing clinical trials of therapies for rarer cancers (see ch. 5, "Impact of the Cooperative Oncology Groups"). The use of multiple centers has made possible the large-scale prevention trials in heart disease. Because of their larger sample sizes, multicenter studies generally have greater statistical power than single-center trials,

A second advantage of multicenter trials is that they often have more highly refined protocols and organization. In well-run trials, all investigators participate, both in planning and throughout the triaI. Problems are likely to be worked out early. The effects observed in the trial are not likely to result from one investigator's personal style. Multicenter trials generally have better arrangements for data analysis and data monitoring, and more often employ statisticians in planning the collection and coordination of data.

A third advantage of multicenter RCTS is that they can enroll a more heterogeneous patient group. One criticism of RCTS, and a reason sometimes offered for the irrelevance of their findings, is that RCT participants represent only a small proportion of the total patient population. The results lack external validity, that is, they can't be generalized to real treatment decisions. Multicenter studies do not entirely eliminate this problem, but insofar as they are geographically distributed, the heterogeneity of the patient population is increased.

Traditionally, most institutions participating in multicenter RCTS have been large university research hospitals. (One exception is VA Cooperative Studies Program trials, carried out in VA hospitals.) More recent trials have sought to include community hospitals and small group practices, with varying degrees of success. One investigator claims that the data submitted by smaller institutions are inferior to those of the larger institutions (215). This claim has been questioned by multicenter research groups that include smaller institutions. They argue that in well-organized trials with strong central administration and sufficient training and orientation provided for the smaller institutions, no such difference can be seen (14). Thomas and colleagues (221) comment that "more clearly written protocols, orientation sessions for physicians, and more effective monitoring of satellite performance would go a long way toward keeping protocol studies open to a broader array of institutions, physicians and patients. This is particularly desirable if the knowledge gained from protocols is ever to be incorporated into standard treatment. "

There are also arguments made against multicenter trials. For example, some argue that the complex administrative arrangements these trials require, if there is no established cooperative system, are too great an impediment. Multicenter trials are generally more expensive than single-center trials, mainly because of the number of participants. In fact, they are not necessarily more expensive per patient. Meinert calculated that the cost per patient in a multicenter RCT (based on the 1979 NIH Clinical Trials Inventory) is $523, while that for single center trials is $587 (158).

Even when multicenter trials are preferred in resolving clinical questions, there is a role for single-center investigations. First, there is a legitimate need for small-scale preliminary studies in the early stages of evaluation. Almost everyone would agree that RCTS should not be undertaken without some evidence from smaller studies on which to base the trial. In some cases these preliminary trials might be HCTS rather than RCTS. There are technical limitations to multicenter tri& in that they require special skills or equipment. Unfortunately, multicenter trials may be foregone simply because the details of their design and execution are not the sufficiently known. In some

poorly planned studies, data collection is expected to be part of regular patient care, and is not seen as research requiring extra time, an incorrect assumption.

Multicenter trials are often viewed as overly complex and not worth the effort. They are difficult to begin without some funding, and the initial stages of planning usually require more money than is available. As a result, the planning of large-scale trials in some fields falls more often to the Federal Government and not to other researchers in the field. This has been the case with NHLBI-funded trials, while the impetus for developing trials funded by the National Cancer Institute (NCI) is largely in the hands of the extramural community. Incentives to participate are less when investigators have little or no say in the design of trials.

Another problem of conducting multicenter trials is the lack of written material about the methods of large-scale RCTS, although this is changing. In the past few years, a number of articles have addressed such questions, including a number of articles in the journal Controlled *Clinical Trials.*

Investigators have little incentive to participate in multicenter RCTS because participating investigators are given little recognition. The "author" of publications reporting the trial is often given **as** the name of a group, e.g., the Multiple Risk Factor Intervention Trial Research Group (see ref. 166), and institutions award little status for participation. Academic promotions are rarely based on participation in large trials (159). (Related recommendations to encourage multicenter RCTS are discussed in ch. 6.)

Dissemination of RCT Results

As the number of trials conducted, including large-scale trials, has increased, their results are not so effectively disseminated by simply publishing them, even in distinguished journals.

The trials drug companies sponsor for FDA approval of New Drug Applications are often reported in obscure journals. Because drug companies have their livelihoods at stake, they take other steps toward disseminating their results. The two main avenues they use to reach the practicing physician are advertising (both in major journals and, perhaps more importantly, in "throwaway" publications) and the use of representatives who visit physicians' offices. The throwaway publications are distributed free of charge to most practicing physicians in the country. Advertising in major medical journals also receives widespread attention.

Drug companies' representatives, their salespeople, personally visit private physicians and medical institutions to distribute literature on their products, to dispense samples to physicians, and to encourage the physicians to prescribe their products. In general, neither advertising nor drug companies' representatives stress the design and conduct of the trials, but rather the uses of the drugs.

In a study of physicians' prescribing practices, Avorn found that "pharmaceutical advertising has become the major source of continuing education for the American physician" (156). This study indicated that both advertising and drug company representatives have a marked influence on prescribing habits, yet that most physicians believe both have only minimal influence.

The research community could profitably borrow from the practices of the drug industry in disseminating their results. It is very likely that research results would be better disseminated if increased resources were devoted to the effort. Funding bodies should recognize this more fully. At NIH, NHLBI, for instance, has a well-developed strategy toward disseminating research results (described in more detail in ch. 5). "Analysis and Dissemination" is a separate phase of all NHLBI's large-scale trials, and the Institute requires a plan for dissemination of trial results. The vehicles of communication it recommends are conferences, activities of professional societies, workshops, and articles in less specialized medical publications and the popular press. NHLBI's methods of dissemination are still evolving, but its progress is apparent. Its recently completed Multiple Risk Factor Intervention Trial received attention in all of the major medical journals, in newspapers and magazines, and on radio and television. NHLBI followed up the publications with a workshop (February 1983) to discuss the results

of the trial with practitioners and policy makers. Other NIH institutes, such as the National Eye Institute and NCI, have also developed mechanisms to disseminate research results. While every trial cannot expect to become famous, efforts to publicize results, should be greater, including important negative results.

Effective dissemination of results depends on knowing how physicians get information. Medical journals and textbooks, continuing medical education courses, and discussions with colleagues appear to be the most influential sources aside from drug advertising (214). Depending on the subject, multiple sources of information may be important. Experimental programs have effectively used physician tutorials in hospitals for selected problems in the management of their patients (122). Nevertheless, not enough is known about how best to translate clinical research findings into practice.

At present, much dissemination of information is left to chance. Kessner has suggested a few measures to improve the situation:

1. identify the primary audience the results should reach,
2. communicate early with selected journal editors, and
3. allocate a small percentage of research funds to dissemination (122).

Other Factors Affecting the Impact of RCTS

Other characteristics of RCTS influence their impact. For example, investigators and their institutions, especially those of repute, can influence the acceptance of results.

Whether an RCT'S results are negative or positive can affect its impact. Positive results are generally more enthusiastically embraced than negative ones. Positive results are also more likely to be published than negative results, and thus may have a greater impact.

The risk associated with a technology affects the way practitioners use information about its efficacy. Technologies perceived to be of low risk, such as many diagnostic tests, may still be used despite evidence questioning their efficacy. Some time-honored treatments, such as bed rest for hepatitis, persist despite the evidence, typifying the "it can't hurt" philosophy (40).

93-825 0 - 83 - 5