

5

The Impacts of Clinical Trials on Medical Practice

The Impacts of Clinical Trials on Medical Practice

The use of randomized clinical trials (RCTS) grew enormously in the late 1960's and 1970's. By the mid-1970's, literature began to appear about their impact on medical practice. The interest in RCTS has continued to grow, but the body of literature evaluating their impacts is still small.

RCT results can have several effects. They can encourage the adoption or abandonment of technologies through treatment decisions by individual physicians and by institutions (e. g., those resulting in the purchase of equipment or in establishing screening programs), or through changes in policy, for example, Federal guidelines (e.g., for immunization practices). All these effects, insofar as they are actually supported by RCT results, are positive.

On the negative side, an RCT favoring the use of a therapeutic agent may encourage the agent's extensive but unjustified use. The drug cimetidine (Tagamet®), for example, was found in an RCT to be effective for treating duodenal ulcer. It then became widely used for conditions and indications for which it had never been tested by RCT (51).

RCTS are only one kind of research that can be done on a promising medical intervention, however. Because they are not the sole source of evidence, it is difficult to separate their impacts from those of the other factors.

The literature about the impact of RCTS is of two general types. The first begins with the results of specific RCTS or the results of RCTS in a specific area (e.g., RCTS of treatments for hypertension), and then examines whether physicians are aware of the results, or what their treatment practice is compared with the recommendations that arise from the RCTS. The second type starts with medical practice, either through literature reviews or by questionnaires, and determines how well practice agrees with the results of appropriate RCTS. An important element of some papers is their quantification of the delay between publica-

tion of RCT results and changes in practice. Many papers that describe RCTS and their results also make claims about their impact, but without citing supporting data. These papers are difficult to interpret.

An increasing number of papers review the results of a number of RCTS in a field and make recommendations for practice in light of those results. These range from qualitative reviews of the literature to formal statistical "meta analyses" synthesizing data from more than one study into a single set of statistics.

Most authors conclude that the impact of RCTS on medical practice has been less than optimal or that their impact is exceedingly slow to develop. The literature as a whole demonstrates great variation in the use of RCTS and in their influence in different medical areas. These studies of RCTS' effects have evolved in method. Earlier papers concentrated on showing the lack of influence of RCTS. More recent articles, going beyond simply showing this fact, have identified some of its possible explanations (discussed in detail in ch. 4). Information from all these studies has contributed to researchers' and funding agencies' greater awareness that the dissemination of research results plays a major role in determining their impact. The National Heart, Lung, and Blood institute (NHLBI) is now taking more rigorous measures to disseminate the results of RCTS, and to make followup studies of how profoundly these results have affected practice. NHLBI has just completed a followup of two recent large-scale RCTS, the Coronary Drug Project (CDP) and the Aspirin Myocardial Infarction Study (AMIS), and plans similar followup of the recently completed Multiple Risk Factor Intervention Trial (MRFIT) and the ongoing Lipid Research Clinics. The National Cancer Institute (NCI) also has instituted a major new program for disseminating information about ongoing studies. Protocol Data Query System (PDQ) is an international computerized

data base currently including information about treatment protocols for about 700 research programs.

The published literature on the impact of RCTS by no means covers all medical practice. More attention has been given to the impact of RCTS in cancer research, though there is now increasing interest in RCTS related to cardiovascular disease. These two medical areas have inspired the majority of clinical trials and the greatest expenditures for such trials. A 1981 conference on the recent history of RCTS, concerned at least in part with their impact, focused on cancer and heart disease. (The proceedings were published as the September 1982 issue of controlled *Clinical*

Trials.) A good deal has been written about RCTS in surgery. The main complaint for surgical RCTS is that too few are done, and that when they are done, they are late. Some authors have focused on controversial trials that illuminate particular issues, for instance, the University Group Diabetes Program (see box E in ch. 4), perhaps the most controversial trial of all time. The remaining published articles about the impact of RCTS are about diverse topics from nursing practices to pediatrics.

Because of the extent of related literature, the influence of RCTS on treatment of cardiovascular disease and cancer and on surgery are specifically discussed in later sections of this chapter.

RCTS AND CONCORDANCE WITH MEDICAL PRACTICE

In one of the earliest articles on the topic, Chalmers concluded that physicians' practice in the 1950's and 1960's was often at odds with data from RCTS (39). McGrady came to the same conclusion in a 1982 survey of family practitioners. Asked about their treatment of a variety of common problems, there was little concordance between their practice and the results of controlled trials (149).

Christensen, Juhl, and Tygstrup reviewed 65 RCTS on treatment of duodenal ulcer and compared the results to recommendations in medical textbooks. They found that RCTS had little influence on these recommendations (49). Tygstrup, Lachin, and Juhl (224) concluded that the results of RCTS have had little effect on gastroenterological therapy.

In a discussion of various types of research studies in ambulatory pediatrics, Hoekelman concluded that the results of RCTS had little influence on physicians' behavior (114).

Moskowitz, Sacks, and Chalmers reviewed RCTS of alcohol withdrawal treatment. They concluded that such treatment using drugs had been established as superior to that using only a placebo. They then polled physicians about their practices and examined review articles on alcohol withdrawal treatments. In this case, the authors found that practicing physicians were using the

treatment that RCTS had shown to be effective before it had been recommended in review articles (163).

Baum and colleagues focused on RCTS' effects on later research, instead of their effects on practice. After surveying clinical trials of antibiotic prophylaxis in colon surgery, they concluded that the results published showing antibiotics superior to a placebo apparently had little effect on the design of later studies (12).

In a preliminary report, Boissel and colleagues conclude that the results of RCTS had no influence on the prescribing habits of French physicians for four classes of drugs—beta blockers, long-acting nitrates, clofibrate, and platelet antiaggregants (19).

Stress and Harlan found that only 28 percent of family physicians and 46 percent of internists were aware of the results of a major multicenter study using photocoagulation to treat diabetic retinopathy (Diabetic Retinopathy Study [DRS]), a year and a half after the study had been published (213). Their study shows that even the results of well-conducted large-scale studies must be brought explicitly to physicians' attention or these results will not affect practice. The DRS was reported in an ophthalmologic journal, not inappropriately, but leaving uninformed the general practice physicians who usually treat diabetics.

Medical practice might have benefited more from DRS had it been given greater coverage initially, e.g., as a report of a clinical advance, rather than one of the study itself, in a general medical journal with wide circulation.

Stress and Harlan also found that many who knew about DRS had learned about it from ophthalmologists or other colleagues, not from the medical literature. This argues for encouraging communication among physicians in local areas. Continuing medical education could also give greater emphasis to new findings in clinical research.

The National Institute of Mental Health (NIMH) of the Alcohol, Drug Abuse, and Mental Health Administration played a key role in evaluating hyperbaric oxygen treatment for cerebral dysfunction in the elderly and also in seeing that the evaluation had appropriate impact (see box F).

NIMH has continued to fund RCTS when promising but controversial treatments appear. As of 1980, in response to reports that schizo-

phrenics can be treated with hemodialysis (244), NIMH funded three double-blind RCTS, two still under way. Carpenter and colleagues (36) have reported their finding from the study that is complete, a small study of 15 patients. They used a “cross-over design” for the study. They randomized patients to one treatment or the other initially, and switched to the other treatment midway through the trial. The experimental treatment was dialysis and the control treatment, sham dialysis. Carpenter and his colleagues found no difference between the effects of real and sham dialysis on the symptoms and behavior of schizophrenia. The results of this trial (along with the other two) may have a direct impact on practice, depending on coverage decisions for the procedure by Medicare. In response to a request for evaluation from the Health Care Financing Administration, the National Center for Health Care Technology found that the evidence for the procedure’s safety and efficacy was inconclusive and recommended that it not be covered under Medicare (235). With evidence from the other RCTS, this initial decision may be either affirmed or overturned.

RCTS IN CANCER RESEARCH

Characteristics of Cancer RCTS

RCTS are employed in developing cancer drugs in “phase III” clinical testing. Preclinical tests identify potential anticancer agents, and then test them in rodents and larger mammals. Phase I clinical studies establish the tolerated dosages of the drugs and their toxicities and measure any therapeutic effects they have. Phase II trials evaluate drugs in treating specific kinds of tumors. In phase III trials, RCTS are used to compare a new treatment with whatever the standard treatment is at that time.

Anticancer drugs are generally very active compounds with marked toxicities, and the patient populations on which they are tested reflect their risks. In testing most other kinds of drugs, phase I studies are carried out on relatively healthy subjects, and only later studies on those with the conditions for which the drug is intended. In contrast, the first clinical studies of cancer drugs are car-

ried out on those with very advanced cancers, who have not improved through any other treatment, and for whom there is little other hope. These clinical studies then progress, if the drug shows promise, to testing the drug on patients with early cancers who are more likely to benefit from therapy.

The earlier the stage of a cancer, and the greater the survival rate for that kind of cancer, the less acceptable is treating that cancer using a drug with known and unknown risks, and unknown value. This fact has affected the use of RCTS in cancer research. More RCTS have tested treatments of acute leukemias, for example, than of chronic leukemias, in part because the acute forms were rapidly fatal, and at least in acute lymphocytic leukemia (ALL), most victims were children. People with chronic leukemias can live for years, and those affected are usually older.

Clinical trials of cancer therapies can be somewhat more complex than clinical trials of therapies

Box F. Hyperbaric Oxygen Treatment for Cognitive Deficits in the Elderly*

Considerable excitement arose in both scientific and lay communities over a 1969 article in the *New England Journal of Medicine*, reporting that repeated exposure to pure pressurized oxygen in a hyperbaric chamber enhanced the cognitive functioning of elderly male patients with organic brain syndrome (117). No effective treatment had been available before for the memory loss associated with brain changes due to arteriosclerotic disease or Alzheimer's disease. This finding by Jacobs and her associates was especially compelling because five of their control subjects exposed to an air mixture failed to show improvement initially, but did improve later when they were "crossed over" to the oxygen treatment. Perhaps 10 percent of those over 65 years of age are affected by cerebral dysfunction, and so the potential impact of this therapy was enormous.

Five other published studies confirmed Jacobs' observation (16,22,66,115,116), but only one used a control group. Two additional studies failed to replicate Jacobs' findings (95,222). One using 21 experimental subjects and 4 control subjects failed to note any significant differences between the experimental and control subjects (222).

One of the major problems in evaluating the efficacy of hyperbaric oxygen as a treatment was the paucity of studies that employed control subjects and the small number of control subjects in those that did. One reason for the investigators' reluctance to include control subjects was that the control condition was more dangerous than the experimental one. Experimental subjects breathed pure oxygen, control subjects an air mixture containing nitrogen, presenting some danger of the "bends" if care was not taken in timing decompression.

Because of the importance of Jacobs' results and the obvious need for their confirmation using a sufficient number of control subjects, in 1973 the Psychopharmacology Research Branch of NIMH and the New York Medical Center undertook a collaborative RCT of the treatment.

This study failed to confirm that oxygen administered under pressure improves cognitive functioning in the elderly. The study had also investigated whether some subgroups of patients might be especially aided by the treatment. Again, there was no evidence of differential treatment effects as a function of initial severity of illness, sex, or presumed evidence of cerebrovascular disease. Subjects in the

study had well-documented evidence of memory problems but were still able to reside in the community and to respond meaningfully to intelligence, psychological, and psychometric tests. On the basis of the findings of Jacobs and others (117), many of these patients should have shown a favorable response to hyperbaric oxygen treatment, but this was not the case.

Jacobs' findings had been picked up early on by the news media, especially the more sensational press, and hyperbaric oxygen was widely touted as a cure for a variety of the infirmities of old age as well as for memory loss. A number of special centers in this country were already offering hyperbaric oxygen to treat memory loss in the elderly at substantial fees. At one, the fee was \$5,000 for 15 days of treatment. The problem of the established use of this treatment was not easy to resolve. Scientific findings are generally not disseminated widely prior to their publication in a respected scientific journal, where the lag time between receipt of a manuscript and publication may run a year or more. To offset this delay, researchers decided to present the new findings at a meeting of the American Geriatric Society and to release a statement to the press once word was received that the paper had been accepted for publication (186).

Although publication of the study findings and dissemination of the results through the press and television did not completely eliminate the practice, the coverage did appear to dampen enthusiasm significantly. The study findings also had an effect on the policy of health insurance carriers and that of the Medicare program, which at one time had considered paying for the treatment. The insurance carriers and Medicare have since ruled that use of hyperbaric oxygen is not a medically accepted or effective treatment for cognitive deficits in the elderly, and they will not pay for it.

By identifying the need for an RCT, and acting quickly, NIMH halted the spread of an ineffective treatment. This case points out the importance of appropriately disseminating scientific findings. Information that promises relief to suffering individuals may be disseminated quickly and extensively—perhaps exceedingly so—when testing has been inadequate. In such cases, later valid findings must be given the widest and most rapid dissemination possible.

* Adapted from *Assessing the Efficacy and Safety of Medical Technologies* (225).

for other diseases. Four major types of treatment are now given to those with cancer: 1) surgery, 2) chemotherapy (treatment with drugs), 3) radiotherapy (treatment with ionizing radiation), and 4) biological response modification. The best therapies now available for most solid tumors* combine several of these treatments. Most RCTS have tested chemotherapies and more recently, types of biological response modification. Chemotherapy itself is not a simple treatment. Combinations of three or more drugs often provide the best results. The possible variations in chemotherapy, including dosages, timing of drug administration, and types of drugs, are almost limitless. The greatest limiting factor for such possible variations is probably the number of active anticancer drugs available; there are now about 20.

Most RCTS in cancer research are of chemotherapeutic agents. Surgery and radiotherapy have been tested far less often, in part because the first has been a mainstay of cancer treatment since the last century, and the second, since early in this century. The major developments in these therapies occurred before RCTS were in common use.

At least two volumes and a number of papers have addressed specifically the impact of RCTS on cancer therapies. *Randomized Trials in Cancer: A Critical Review by Sites* contains a number of papers by experts on all major anatomical sites of cancer and groups of these sites. These papers review the bases for treatment and the contribution of RCTS to current recommendations (211). *Methods and Impacts of Controlled Therapeutic Trials in Cancer (5,37)*, published as part of a project of the International Union Against Cancer, reports on RCTS from their initiation to their conclusion, and determines the extent to which the results have altered therapeutic methods in subsequent years. A second part lists treatments available for specific cancers, including colorectal-bronchogenic, breast, melanoma, and osteosar-

*There are three main classes of neoplasms or cancers. Cancers of the epithelia, including the external epitheliums (the skin and the lining of intestinal and respirator tracts) and internal epithelia (the lining of various glands) are called carcinomas. Cancers of supportive tissues (e. g., bone, muscle, tendon, and cartilage) are called sarcomas. Carcinomas and sarcomas together are termed "solid tumors." Cancers of blood are called leukemias and those of the lymph tissues lymphomas.

coma, and attempts to identify the roles of randomized and nonrandomized clinical trials in establishing their treatments.

Impact of the Cooperative Oncology Groups on RCTS

The mid-1950's saw the development of NCI "cooperative groups," to carry out multicenter studies in cancer treatment. These groups conducted the first RCTS in cancer research, studying treatment for childhood acute leukemia and for a variety of solid tumors. Fourteen groups are now active: five include multidisease, multiprotocol studies; six specialize by disease (e.g., National Wilms' Tumor Study Group and National Surgical Adjuvant Breast and Bowel Group); and three are "related resource groups" (Lymphoma Pathology Reference Center, Radiologic Physics Center and Cancer Clinical Investigations Coordinating Center) (59). Each group consists of 30 to 50 institutions (59), with more than 1,000 institutions participating altogether, including affiliates from 41 countries outside the United States. While these foreign affiliates are rarely funded, they find it important to participate (35). The cooperative groups are active in phase II as well as in phase III clinical trials (RCTS).

One of the main advantages of the cooperative groups is that they can recruit relatively large numbers of patients for trials in far shorter time than can single institutions. As is discussed below, small studies abound in the cancer treatment literature, more noticeably than any other field,

From the administrative necessities of large cooperative efforts the groups have developed well-formed organizations. Each has an elected chairman, an elected or appointed statistician, and several other elected and appointed positions and committees. The scientific sections of the groups vary, but include committees representing treatment modalities and specific diseases. Another important feature of the cooperative groups is that each has a statistical coordinating center. As in other areas, the presence of statistical expertise is a key factor in ensuring the high quality of RCTS.

The Cooperative Groups ensure a high quality of research by stringent internal review mech-

anisms, in addition to the usual external reviews of Government-supported research. Group members are evaluated at regular intervals on specific criteria related to the quality and productivity of trials (35). These evaluations can include auditing original clinical documents for accuracy of reporting (255).

The Cooperative Group members have traditionally been university hospitals or major treatment centers. Cancer patients are increasingly treated in community hospitals, however, as more oncologists are trained and enter the medical work force. The Cooperative Groups have thus recently arranged for community hospitals to participate in clinical trials. This should improve the efficiency of trials by extending the population from which patients are recruited, and improve the impact of trials by involving a greater number of oncologists and institutions. The Eastern Cooperative Oncology Group (ECOG) published their first evaluation of community hospital participation in their clinical trials. It indicated that the contribution of 112 community hospitals is equal in quality to that of the larger member institutions. Quality was measured by relative enrollment rates in trials, compliance with the protocol, and submission of data, as well as measures of outcome—e.g., survival and positive and toxic responses to treatment. (Community hospitals have shown similar performance in multicenter trials of heart disease (83).)

ECOG has found in addition, through a survey of affiliated hospitals, that while 16 percent of cancer patients were enrolled in a trial, a further 35 percent were treated in accordance with an experimental protocol.

Impact of RCTS on Cancer Treatment

RCTS have contributed to developing successful treatments for a number of cancers, e.g., those for ALL, Hodgkins' disease, and Wilm's tumor, and adjuvant chemotherapy for breast cancer. The clinical trials for these therapies have been part of larger targeted research programs, which were prompted by the discovery of significant drugs. The therapeutic regimens now actually employed were then developed gradually by trying the different drugs and their combinations in

RCTS and building new trials on the results of previous ones. The sustained support of these programs and rational process through which they developed treatments appear to be the reasons for their success. Had uncoordinated trials been conducted in many places after the initial discoveries were made, it is doubtful that this progress could have been made as quickly and efficiently. It can be argued, on the other side, that new approaches and ideas may have been sacrificed by concentrating the effort.

RCTS have also had a major impact, though one difficult to document or quantify, in preventing costly but ineffective and debilitating cancer therapies from becoming part of medical practice (208).

Gamier, Flamant, and Fohanno (86) have shown that RCTS in cancer research are not conducted in proportion to the incidence or importance of the disease, but are heavily influenced by whether or not worthwhile treatments are available to be tested (table 6). While the highest incidence of cancer is at sites in the gastrointestinal tract, only 10.8 percent of RCTS are on treatments for cancers at those sites. The leukemias and hematosarcomas (circulatory cell neoplasms) account for 26.7 percent of RCTS, while the incidence of these cancers is less than one-third that of gastrointestinal cancers. The RCTS referred to here are those registered with the International Union Against Cancer between 1968 and 1978, nearly 1,000 RCTS.

A series of therapeutic advances, such as in treating ALL, depends on an initial breakthrough. For most cancers, particularly the solid tumors, such breakthroughs are rare. Most clinical trials in treatments of these tumors consist of testing drugs that have shown anticancer activity against a number of tumor types in phase I and phase II trials. These trials are usually small and conducted at single centers, with too few participants to showing a significant effect of the drug, if it has one. In part this is because a "significant" effect of an anticancer drug may be smaller than such an effect in treating less serious and more treatable diseases.

Thousands of cancer therapy RCTS have been generated by combining chemotherapeutic, often

Table 6.— Distribution by Site of the 945 Trials Registered at the International Union Against Cancer Information Office, and Related Incidence Rates

Site	Incidence ^a	Percent of trials by site
Gastrointestinal tract	76.0	10.8
Genito-urinary sites	46.9	5.9
Breast	40.4	15.9
Lung	40.0	12.5
Gynecological sites	30.4	6.9
Leukemias and haematosarcomas	23.0	26.7
Miscellaneous	15.3	3.3
Head and neck	15.0	5.4
Brain and nervous system	4.9	4.3
Skin (including melanoma)	4.2	5.4
Bone and soft tissue	2.8	3.0

^aAverage annual age-adjusted incidence rate per 100,000 Population, United States

SOURCE H Garnier, R Flamant, and C. Fohanno, 'Assessment of the Role of Randomized Clinical Trials in Establishing Treatment Policies,' *Contr Clin Tr* 3(3)227.234, September 1982

two to four in one regimen, along with radiotherapy and surgery. Though drug combinations are based on some prior information, there is no satisfactory scientific basis for designing combinations. Given that the prior probability of success—the expectation that the trial will have positive results—is low in cancer research (judging from the history of cancer therapy RCTS), and that most of these RCTS employ few patients (a median of 25 per treatment), a large proportion of the positive results obtained must be false positives. The consequence is that many ineffective treatments may be applied in the clinic because clinicians do not have adequate information to distinguish effective from ineffective ones.

Many of the contributors to Staquet's book identified areas in which ongoing trials would provide some answers in the next few years and areas in which studies were needed (211). The contributors to the International Union Against Cancer's two-part publication concluded that RCTS have in most cases been more useful than nonrandomized studies in developing cancer treatments (5,37).

Garnier and colleagues looked at the treatment policies for head and neck cancers at the Gustave-Roussy Institute during two periods: from 1960 to 1967 and after 1967. They then examined the possible reasons for policy changes **between the** two periods. They set out to answer three questions about treatments for each main site of cancer: 1) whether there was a consensus about treatment, 2) the reasons for the choice of a specific

treatment, and 3) the correlation between the treatment problems yet unsolved and the trials being conducted by the international cooperative groups (86). These authors did not complete the task they set for themselves. To have done so might have been a monumental undertaking. In fact, their attempt raises the larger question of how, whether, and to what end the impact of RCTS can be correctly and completely determined.

The authors did conclude, however, that there is consensus mainly about treatments that have not been tested in RCTS, namely those of surgery and radiotherapy.

Breast Cancer

The treatment of breast cancer has given rise to more RCTS than any other cancer site (37), and the impact of those trials has gradually been felt. In 1977, McPherson and Fox reviewed the reports of selected RCTS published since 1965, when the first RCT report demonstrated that radical mastectomy had no survival advantage over a more conservative operation (119). McPherson and Fox concluded that the RCTS had little impact: the radical procedure was still the treatment of choice based on surgery rates in 1970 (153).

A **more recent** paper on breast cancer (190) presents the view of the National Surgical Adjuvant Project for Breast and Bowel Cancers (NSABP), which is more optimistic about the impact of RCTS. Initial NSABP RCTS of breast cancer ther-

apy focused on the treatment of local-regional disease (not metastatic), comparing radical mastectomy, total (simple) mastectomy with radiation, and total mastectomy alone with removal of axillary nodes only when they become affected. Underlying the study were competing hypotheses about the nature of breast cancer. The traditional belief, on which the rationale for the radical mastectomy is based, is that breast cancer follows an orderly progression from local-regional to systemic disease. The competing hypothesis is that the disease is often systemic very early on, so that even considerable improvements in local-regional treatment alone will not substantially affect the outcome of the disease. The trial results support the second hypothesis, with little difference in long-term survival observed among the treatment groups. The more extensive surgery involved in a radical mastectomy is not better than less extensive surgery in this regard.

The remaining NSABP trials have studied the effects of chemotherapy. Like the trials that developed a treatment regimen for ALL, these trials developed breast cancer treatments by increments. New trials are now being conducted in this area with a wide range of patients.

Continuing progress has been made through NSABP over the past 10 years, particularly in the use of adjuvant chemotherapy. The advances would have been difficult to document without the use of clinical trials in a structured program. In an overview of NSABP, the principal investigators of the program come to some generalizations about clinical trials of cancer treatments (190):

1. There is a need for larger sample sizes than are generally used in adjuvant phase III clinical studies. The heterogeneity of the patient population along a number of important prognostic lines, both known and unknown, make this particularly important.
2. Because of the relatively good prognosis for breast cancer patients, long followup is necessary, and overall survival, not necessarily disease-free survival, may be the appropriate measure.
3. The need for large numbers necessitates the need for multicenter participation. The de-

velopment of straightforward, clear aims and reasonable data collection requirements is essential for success. In addition, particularly with long-term studies, constant refamiliarization of staff at participating institutions, where turnover may be high, is necessary.

4. Finally, the authors point to the need for clinical trials to be integrated into a general program aimed at the disease, which is predicated on an understanding of the natural history of the disease, and seeks to gain biological information about the disease.

The authors conclude that RCTS have contributed substantially to treating primary breast cancer in its early stages, and that NSABP trials have had a "strong impact in changing the clinical management of breast cancer over the past decade." Their conclusion is supported to some extent by trends in surgery for breast cancer between 1972 and 1981 (2). While the number of patients with breast cancer given radical mastectomies has dramatically declined (from about 50 percent in 1972 to about 3 percent in 1981), the shift has not been so much to simple (total) mastectomy or lesser surgery, but to a compromise between the radical and simple mastectomies, the modified radical mastectomy. In 1972, less than 30 percent of those with breast cancer had modified radical mastectomies; in 1981, over 70 percent. Between 1976 and 1981, there was a modest increase in women given a "wedge excision" (lumpectomy), from about 3 to 8 percent of those with breast cancer.

Early Detection in Cancer

The best secondary prevention for cancer is breast cancer screening. Miller and Bulbrook reviewed all major studies, randomized and nonrandomized, of all methods of breast cancer detection: self-examination, physical examination by medical personnel, thermography, mammography, and combinations of techniques. The combination of mammography and physical examination has proven most valuable (162).

The first trial of breast cancer screening, conducted by the Health Insurance Plan of New York, studied 62,000 women who were randomized either to mammography and clinical examination

or to their regular pattern of care. The results showed a benefit of screening for women over 50 (204), though there is still some controversy over this study. Current studies in Canada and Sweden are designed to determine whether screening younger women is worthwhile (162).

Based on the available evidence, Miller and Bulbrook conclude that there is value in screening asymptomatic women over 50 by physical examination and mammography, but that the desirability of introducing screening on a larger scale requires answers to some outstanding questions. Studies in progress should provide the necessary information within the next decade. Regarding the potential impact of these studies on practice, "it should be noted that results from experimental studies cannot necessarily be directly translated into practice." This transition requires information in several areas: the training of personnel, the factors affecting participation in screening programs outside experimental settings, and the quality control of screening.

There has been relatively little improvement in survival for most common forms of cancer during the past three decades. Because survival is better for many cancers treated in earlier stages, early detection may hold the greatest current potential for lowering overall cancer mortality (226). of such early detection techniques, breast cancer screening has received the most attention. There are also now three RCTS of lung cancer screening in progress, each testing both sputum cytology and X-rays. A preliminary finding in two of those

is that sputum cytology is relatively ineffective. In addition, they have found that the benefits of screening, if proven, will be in detecting non-small-cell cancers* (85), which comprise the majority of lung cancers.

RCTS could also make the use of existing screening techniques more effective. The Pap smear, an examination of cells from the cervix, was introduced in 1943, to detect cervical cancer in asymptomatic women. The technique has been widely promoted and accepted, even though its efficacy has never been demonstrated in an RCT. In 1973, 75 percent of U.S. women over 17 had had at least one Pap smear. In recent years a controversy has developed about the efficacy of this screening, focusing on four issues: the natural course of cervical cancer, the accuracy of the test, the appropriate interval between screening tests, and the efficacy of screening while the incidence of death from cervical cancer is declining. OTA concluded (225):

Once the Pap smear was in widespread use, the very extent of use and professional consensus of its efficacy argued against carrying out a controlled trial. As the risks to women whose tests were found falsely positive by the Pap smear have never been seriously documented, it is possible that a controlled trial to examine that question may be of value.

*Non-small-cell lung cancers include adenocarcinomas, squamous cell carcinomas, and large-cell carcinomas.

CARDIOVASCULAR DISEASE

The major problems in the treatment and prevention of cardiovascular disease have been well-studied in the United States, Canada, Europe, and Australia. RCTS are the primary instruments for resolving issues of therapy and prevention. NHLBI and the Veterans Administration (VA) have been key players in this field in the United States. Their large-scale multicenter RCTS, many with thousands of participants, have had a major impact on the treatment of heart disease.

These trials are mostly of two types: prevention trials based on evidence from epidemiology and physiology, and trials of therapeutic surgery and drugs. In the first category, the most intensively studied interventions for cardiovascular disease are those for lowering blood pressure, those for lowering levels of blood lipids and those for preventing thrombosis (blood clots), each of which has spawned large-scale primary and secondary prevention trials. Therapeutic trials have

focused on surgical procedures (most importantly coronary artery bypass surgery), on beta-blocking drugs, and on antithrombotic agents. In general, trials for cardiovascular disease have not been undertaken without strong hypotheses to test and unless the intervention they test has a reasonably good chance of success.

RCTS of treatments for cardiovascular disease have progressed along a number of lines. One important trend in this field has been toward large multicenter trials. A second trend, illustrated by RCTS in hypertension, is a progression from those of treatments toward those of secondary, and more recently, primary prevention. The first major trials in hypertension studied severe hypertensive, and then later those with moderate and mild hypertension. A new NHLBI trial is testing interventions to prevent hypertension in those who are likely to develop it.

A third trend in research on cardiovascular disease results from knowing that it may have many causes. Early trials in the area concentrated on interventions related to single risk factors. More recent trials have studied several risk factors at once, notably MRFIT, which focused simultaneously on the risks of hypertension, high blood lipid levels, and cigarette smoking.

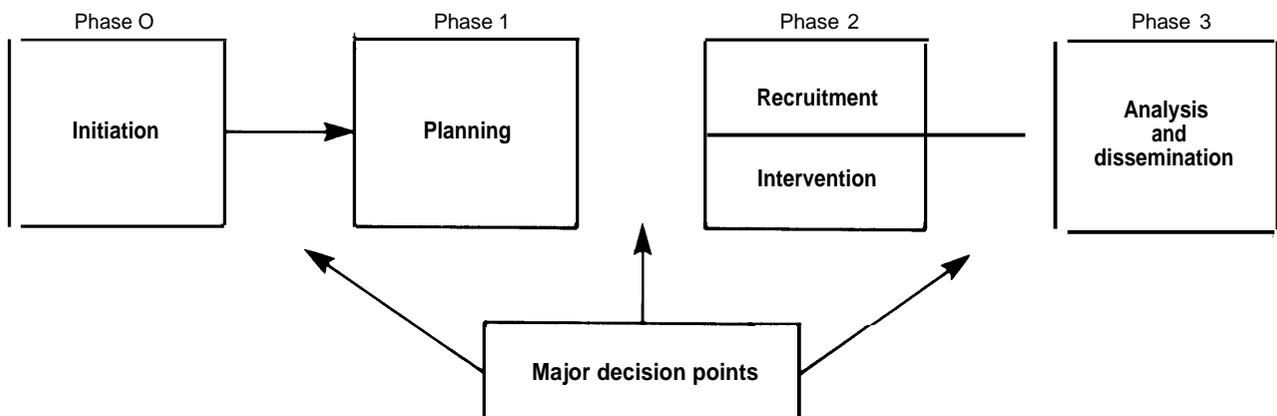
NHLBI and RCTS

NHLBI bases its decisionmaking about RCTS on an idealized view of the progression from basic research to health practice (fig. 1). The philosophy underlying NHLBI's use of clinical trials is well articulated by Levy and Sondik (134):

Advances in knowledge at the basic research level result in hypotheses on potentially effective approaches for the prevention, management and control of disease in man. One objective of clinical research involves the testing of these hypotheses in controlled settings. Clinical trials serve to bridge clinical research and demonstration, prevention, education, and control activities. The clinical trial tests and validates the effectiveness of therapies before their introduction into the health care system. In some cases, however, trials are used to determine which of several alternative treatments already in use is most effective.

NHLBI's model could serve in other circumstances as one for decisions about clinical trials (fig. 1). Of particular relevance to this paper is NHLBI's phase 3, "Analysis and Dissemination." The success of preceding phases is, of course, required for that of phase 3: the initial concept must address an important question that can be answered in a clinical trial, planning must be ade-

Figure 1.—The National Heart, Lung, and Blood Institute's Clinical Trial Decision Process



SOURCE: R. Levy and E. Sondik, "Decision-making in Planning Large-Scale Comparative Studies," *N. Y. Acad. Sci.* 304:441-457, 1978

quate to ensure answering the question, the trial must be carried out in accordance with the protocol and its progress well monitored. The dissemination of results depends on a well-designed, well-executed trial if the results are to have a positive impact on health care.

Data analysis is an ongoing activity in clinical trials, and interim results are sometimes published. The major effort to disseminate results follows the final data analysis, and begins with their publication in the medical literature. This is also their final resting place in many cases. NHLBI stresses that every institute-supported clinical trial must employ all available avenues of dissemination to be useful, including conferences, professional societies, workshops, and articles in less specialized medical publications and the popular press. A few months after the MRFIT results were published, for example, NIH held a 2-day workshop to discuss the results and their implications.

In addition to publicizing trials, it may be useful to find out how effective dissemination has been. NHLBI has completed a followup of CDP and AMIS (described below), and has a similar contract for the MRFIT and the Lipid Research Clinics.

The Coronary Drug Project and Aspirin Myocardial Infarction Study Followups

The fact that trials are well designed and well run does not guarantee that their results will influence practice. Given its heavy investments in clinical trials, NHLBI has an equal interest in knowing how influential they are. A few years ago NHLBI began an effort to find out the impact of two major RCTS, the CDP, which began in 1974 and AMIS, which began in 1980. It interviewed about 1,800 physicians nationwide about their knowledge of the studies and the studies' results, and about their treatment practices. Of all groups, cardiologists were the best informed, though probably not from having read the original reports of the trials. Internists and general practitioners were less well informed.

The results of the followups have not yet been published except in abstract form, and NHLBI has

made no formal changes in policy for disseminating results, but the study suggests certain improvements. The dissemination of information must be local to reach most physicians. The national meetings of specialty societies already disseminate study results and treatment recommendations, but they could increase these efforts. Greater coverage of study results in the throw-away journals with wide circulations would reach physicians who don't read technical journals regularly.

RCTS and their impact on those areas of cardiovascular disease most actively investigated are described briefly.

Hypertension

High blood pressure, or hypertension, is one of the principal conditions leading to heart disease and stroke. The main strategies for controlling hypertension include diet modification, weight loss, behavior modification to reduce stress, and drug treatment. RCTS have tested several interventions in these areas, especially drug treatments.

Drugs to control hypertension first became available in the early 1960's following a search beginning after World War II. Their availability set the stage for large-scale RCTS. The VA Cooperative Studies Program (CSP) carried out the first large-scale RCT of drug treatment of severe hypertension (diastolic blood pressure [DBP] defined as above 115mmHg). The report of the study's results in 1967 showed convincingly that drug treatment helped to prevent death and disability from stroke, congestive heart failure, and kidney disease. A second study, published in 1970, extended the population studied to include men with DBP of 105 and above. Since that time, further studies in this country, under the auspices of VA and NHLBI, and in Europe and in Australia, have attempted to determine whether treatment of mild hypertension (usually defined as DBP between 90 or 95 and 104 or 109) also reduces morbidity and mortality.

Whether mild hypertensives should be treated with drugs is a question of more than passing interest. Perhaps 15 percent of the U.S. population has a DBP reading into the range of 90 to 104

DBP. McAlister describes this question as one . . . with awesome social and economic implications” (146). Freis estimates that if the 40 million people in this country with blood pressures of 90 to 99 DBP all were given drug therapy, the annual cost of treatment might be as high as \$20 billion (81).

In considering whether mild hypertensives should be treated, another important point should be weighed. There are qualitative as well as quantitative differences in the medical characteristics and of mild and severe hypertension. These affect the design of RCTS as well as the hopes for these patients’ treatment. Severe hypertension has its own symptoms, in addition to its association with complicating disease. The treatment of severe hypertension both relieves these symptoms and reduces the risk of complicating disease. In contrast, mild hypertension is a symptomless condition. The major complication of mild hypertension is coronary heart disease. The major complications of moderate and severe hypertension are hemorrhagic stroke, renal failure, congestive heart failure, and aortic dissection (81).

With the trend toward treating milder hypertension in RCTS came the need for larger trials and proportional increases in cost. These trials illustrate a general point. The statistical power of trials (ch. 4, “Statistical Power and Statistical Significance”) depend much more on the number of endpoints counted in each group than on the number of *participants* in a trial. Endpoints of importance in hypertension trials—stroke, heart failure, or death from some cardiovascular cause—occur much less frequently among those with mild than those with severe hypertension. Far more participants have been required for the later trials than those required for trials that tested treatments for severe hypertension. The first VA trial, whose participants were men with DBP over 115, provided convincing support for treatment with only 143 participants. The more recent Hypertension Detection and Followup Program (HDFP) required nearly 11,000 participants (about 8,000 with mild hypertension), and MRFIT, nearly 13,000 (about 8,000 with mild hypertension) for what was considered sufficient power.

The HDFP and MRFIT, along with a large Australian study (of about 3,400 with mild hyperten-

sion) and at least three smaller RCTS, have increased the debate over drug treatment of mild hypertension. All have provided information, but none an answer. The controversy focuses on the benefits of treatment and especially on the risks, known and unknown, of possible lifetime administration of antihypertensive drugs.

The HDFP showed that treatment reduced mortality by 20 percent in mild hypertensives (see box G). Pickering (183) puts this figure in a different light by expressing the 20-percent reduction in other terms, i.e., the reduction in the mortality rate from 7.7 percent in the control group to 6.4 percent in the treated group. In other words, of every 100 untreated patients, 7.7 died, while of every 100 treated patients, 6.4 died. Only 1.3 treated patients per 100 enjoyed a benefit. Pharmaceutical companies have used this information to claim that “HDFP findings justify early and aggressive management of mild hypertension,” while some researchers have concluded that the studies provide no such basis for treatment (121).

The MRFIT study participants all had a high risk of cardiovascular disease, as defined by a rating included two other risk factors as well as hypertension: smoking and high blood lipid levels. A disturbing and unexpected finding in the MRFIT was a higher rate of death from coronary heart disease in the experimental than in the control group, in those hypertensive men who had abnormal baseline resting electrocardiograms. Subgroup analyses must be viewed cautiously, however, especially when they are not based on prior hypotheses. Nevertheless, in an editorial accompanying the MRFIT report, Lundberg commented that this result was “so major as to demand caution, since the results fly in the face of current medical dogma and practice” (138). His prediction that the observation would “no doubt foster substantial debate” was certainly correct. Only a few months after publishing the initial MRFIT results, the *journal of the American Medical Association* carried two related articles and an editorial about the treatment of mild hypertension (121,146,183). Another related article, “Mild Hypertension: The Gray Zone Gets More Confusing” appeared in *Medical World News* during that interval (144). MRFIT results and resulting controversy have been publicized widely in both

Box G.—The Hypertension Detection and Followup (HDFP) Program*

The HDFP was a community-based RCT that studied 10,940 people with high blood pressure. The trial compared the effects on 5-year mortality of a systematic antihypertensive treatment program (stepped care, or SC) and referral to community medical care (referred care, or RC). SC patients were offered therapy in special centers, and therapy was increased stepwise to achieve and monitor reduction of blood pressure to specified levels. RC patients were sent to their usual sources of care, with special referrals for those with more severe hypertension or organ system damage. Patients were first grouped by age, sex, and race, and then further by the value of their DBP: 90 to 104; 105 to 114; and 115 or greater.

The study was designed to answer questions unresolved by previous studies conducted in VA's medical care system:

1. Is a systematic approach to antihypertensive therapy (SC) more effective in reducing risk of 5-year mortality for all hypertensive adults in the community compared to community care (RC)?
2. Can a substantial proportion of all hypertensives, detected in general populations, be pharmacologically managed to maintain blood pressure at normotensive levels?
3. Do the benefits of therapy exceed its toxicity in those with mild hypertension as well as in those with more severe hypertension?
4. Is antihypertensive therapy effective in young adults and in women and equally effective in blacks and whites?
5. Can morbidity and mortality from coronary artery disease be decreased by antihypertensive therapy?

The results of this large clinical trial, which cost nearly \$70 million, showed that more intensive care with available therapies could lead to a significant decrease in mortality and morbidity from hypertension and that these benefits were found in treating "mild" hypertensives as well.

The results of HDFP were first published in the *Journal of the American Medical Association* in December 1979. A survey of physicians revealed that 40 percent of family physicians knew of the study within 2 months of publication, and 63 percent of internists within 6 months. Of the family physicians who knew of the study, 98 percent were able to correctly answer questions about the observed reduction in mortality and the benefits of treating mild hypertension. Eighty percent of the family physicians and 50 percent of the internists learned of the study from medical journals, and 40 percent of the internists learned of it from continuing medical education courses (the remainder learned of the study from colleagues or the lay press).

In sum, as a result of these RCTS and related educational activities, the public is much more aware that hypertension is a disease with serious but preventable consequences. The new information developed in HDFP disseminated rapidly to the medical community.

● Based largely on *Technology Transfer at the National Institutes of Health* (235).

the medical and the popular press. Each of the major trials has contributed to knowledge of hypertension, but at such expense that some find the results disappointing.

The newest NHLBI supported trial in this area is one of primary prevention of hypertension through dietary interventions in those aged 18 to 40. These interventions include altering the intake of sodium and potassium and helping patients to lose weight. This represents a logical step in the

progression of related drug and diet trials that have been completed. Medical researchers would like to reduce the need for drugs in treating hypertension. The drugs carry some risk and are expensive. In treating a younger population, these RCTS also move toward the goal of primary prevention.

From the beginning, the trials of hypertension treatments have had a major effect on medical practice and on the design of subsequent trials.

Part of the NHLBI strategy has been the National High Blood Pressure Education Program, begun in 1972 to educate the medical community and the public about hypertension. Surveys of public knowledge about high blood pressure conducted in 1973 and in 1979 showed the following changes. First, those believing that hypertension is a serious condition increased from 63 percent in the 1973 survey to 73 percent in 1979. Second, 83 percent of those surveyed in 1979 had had their blood pressure measured within the past year, compared with 73 percent in the 1973 survey. Third, about twice as many people knew in 1979 what constituted normal blood pressure. Fourth, 40 percent more people understood in 1979 that hypertension did not have reliable symptoms. And fifth, in the 1979 survey, more people knew that effective treatment was available, and more were also following their prescribed therapies.

The early VA studies provided the first clear evidence of the benefit of drug treatment for severe and moderately severe hypertension. The first evidence from RCTS on the treatment of mild hypertension came in 1979 with publication of the HDFP (see box G). Even before that time, 92 percent of New York State physicians who responded to a questionnaire were treating patients with DBP in the range 90 to 104 (121). Since the publication of HDFP and the results of a large Australian trial, the use of drugs in treating hypertension has probably increased (121). MRFIT results pointed out the need to reexamine treatment policies, which, as described above, are now being debated in the literature.

The progression of hypertension trials has been orderly. New trials have built on the results of previous ones, not only those carried out in this country by VA and NHLBI, but also on those of trials in other countries. The available data allow some conclusions to be drawn and the reshaping of questions that remain for this field of research. Pickering makes three summary statements about treating mild hypertension (183):

1. Cardiovascular risk factors other than BP [blood pressure] should be taken into consideration. Therapeutic benefit is less likely to be seen in patients who have a low overall level of risk than in high-risk groups. Thus, two groups who have so far shown no benefit (in

both the HDFP and Australian trial) are white women and men younger than 50 years. There is, therefore, no sound justification to treat all such patients.

2. For those who are at relatively high risk, treatment is more likely to confer protection against cerebrovascular events than coronary heart disease.
3. In doubtful cases, there is nothing to be lost by delaying the start of drug treatment. In both the HDFP and Australian trial, there was a substantial decline of BP in the control groups during the period of observation.

Freis makes similar recommendations based on RCT results: "By such a discriminative approach, many millions of people could be spared needless lifelong exposure to drugs" (81).

The evidence from RCTS in this field "does not support dogmatic guidelines" (146), but they do provide physicians useful information in considering each patient individually. Rather than supplanting clinical judgment in treating hypertension, the results of RCTS would appear to enhance it.

Hyperlipidemia

Known from epidemiologic studies, the strong relationship between high blood lipid levels (cholesterol and other fats) and the increased risk of atherosclerosis, has led to many large RCTS aimed at lowering blood lipid levels in the hope of reducing death rates. One of the first of these trials was conducted in Norway from 1956 to 1963. Since that time, trials have been under way continuously, each building on the results of earlier trials. (Buchwald, Fitch, and Moore discuss the major trials in this field (26).)

A notable evolution has occurred in trials that study the lowering of blood lipid levels. Early trials tested dietary interventions. These were mainly secondary prevention trials, and included only individuals with proven atherosclerotic disease. Lowering saturated fat was accomplished either by controlling total fat intake, or by substituting unsaturated (e.g., corn or soybean oil) for saturated fat (e.g., animal fat and butter).

Around the mid-1960's, more emphasis was placed on lowering lipid levels with drugs, while

dietary recommendations were often provided to both experimental and control groups. A number of large trials in the United States and Europe tested the most promising drug at that time, clofibrate. Early results of these trials were also promising (26). In later trials, however, notably CDP funded by NHLBI, the benefits of clofibrate were small, particularly in light of some serious side effects. A European primary prevention trial confirmed the risks of the drug. The use of clofibrate has declined since the results of these studies were published (82).

Clofibrate was one of five treatments tested in CDP. Of the remaining four treatments, three were discontinued before completion of the trial because of adverse, at times lethal, effects. The discontinued drugs were estrogen (given in two dosage regimens) and dextrothyroxine. The last drug, niacin, also appeared to cause unwanted effects. It was, perhaps, effective in preventing recurrent nonfatal myocardial infarction, but not in altering mortality rates.

The Lipid Research Clinics, a primary prevention trial, is using a cholesterol-lowering diet for all participants and the drug cholestyramine for the experimental group. Results from this study are expected by the end of 1983.

One RCT still under way has been relatively successful in lowering blood lipids, the Program on the Surgical Control of the Hyperlipidemias (POSCH). POSCH also uses the most drastic intervention for such control: partial ileal bypass to reduce circulating blood cholesterol levels. Survivors of one myocardial infarction with high serum cholesterol levels, but with no other major risk factors, are eligible for the trial. Not surprisingly, recruitment for this trial has been slow. Complete recruitment of the 500 subjects required for each group may not be achieved. Early results show a 31-percent reduction in serum cholesterol in the surgical group over the first 3 years. Even if successful, because this procedure is radical, and has significant though not yet fully known side effects, it is unlikely to become a model for secondary prevention of cardiovascular disease,

A recent generation of trials, notably MRFIT in this country and the Oslo Heart Study in Norway, are primary prevention trials that use mod-

ifications in diet as the intervention to lower blood lipid levels. Both trials include interventions for more than one factor related to cardiovascular disease.

For the most part, the results from lipid-lowering trials have been less than promising (26):

All completed randomized clinical trials of lipid intervention for atherosclerotic cardiovascular disease have shown no convincing evidence for disease retardation, arrest, or reversal associated with plasma cholesterol reduction; albeit in no trial has cholesterol reduction been marked and in many it has been minuscule,

These trials have served important purposes, in spite of their disappointing results. First, they have provided evidence against a number of drugs that might have been widely used without the trials. In addition, all the major diet intervention trials have shown some therapeutic benefit, if not as much as hoped. The trials, especially CDP, have generated a great deal of information about the natural history of cardiovascular disease. One finding is that serum cholesterol does not appear to be as prognostically important after myocardial infarction as before. This finding has important implications for treatments following myocardial infarction and for RCTS conducted of those treatments.

Coronary Artery Disease

Early surgical RCTS for coronary artery disease tested a procedure called internal mammary artery ligation. The procedure was based on the hypothesis that if the mammary arteries were tied off, blood flow to the heart would increase. The technique, though never widespread, gained brief popularity in the 1950's. At that time, two RCTS were conducted, comparing this surgery with a sham surgery. (These are the only RCTS that have used a sham surgical procedure (251).) The studies showed the sham procedure to be "at least as effective as internal mammary artery ligation" in treating angina pectoris. The procedure was rapidly abandoned after publication of the RCT'S results. Fisher and Kennedy attribute this rapid change to the RCTS themselves (74).

The surgery in this field now under study is coronary artery bypass graft (CABG) surgery. Over

100,000 of these operations are now performed yearly in the United States (74), having rapidly increased from their first use in 1968. CABG surgery clearly relieves the pain of angina pectoris, and this is the reason for its widespread acceptance. However, the use of the procedure appears to have gone beyond its accepted indications (235).

The debate over CABG, which has inspired both U.S. and international RCTS, is over whether the procedure prolongs life, and if so, in which subset of patients. Controversy arose when the initial results of the full CABG study were released in 1977 showing no difference in survival between medically and surgically treated patients. The *New England Journal of Medicine* ran an editorial by Hiatt decrying the haphazardness of assessing surgical procedures, and suggesting that more orderly tests were called for (109). The trial was scrutinized from all angles and criticized on a number of points, especially the high rate of mortality in the surgery group early in the study.

Fisher and Kennedy conclude that in spite of this controversy the VA study convinced some that, while CABG prolonged the survival of those with left main artery disease, its effect on the survival of other patients was equivocal (74). More recent data from the study have also shown significantly increased survival in patients with three-vessel disease (without left main disease).

Wortman and Yeaton have identified nine RCTS of CABG surgery since 1974 (253). The first RCT of CABG surgery to have a major impact was the VA Cooperative Study. Fisher and Kennedy claim that this study "has had the most impact among the randomized studies published" (74). The trial began as one of a different operation, the Vineberg Implant, in 1968. This procedure was changed to CABG when it became evident that CABG was a superior operation. The early results on CABG showed it was better than medical therapy in prolonging life for those patients with left main artery disease. These results were readily accepted.

After 5 to 8 years of followup, a European RCT of CABG surgery found significantly increased survival in patients with three-vessel disease, those with stenosis in the proximal third of the left

anterior descending artery, and insignificantly decreased survival in patients with left main artery disease (69). This trial has not elicited the reaction that the initial VA results did, probably in part because it justifies practices already current.

An NHLBI trial scheduled to end in 1983, the Coronary Artery Surgery Study, has suffered from entering the game rather late. A number of centers would not randomize patients because the evidence from other studies favored surgical treatment. A large registry is being kept as part of the study, including patients at one of those centers not randomizing.

Fisher and Kennedy drew several conclusions from their review of surgical trials for coronary artery disease (74). First, they found that these RCTS, especially the large, multicenter trials, have had a significant impact on clinical practice. The influence has not been uniform, however, nor has it been associated only with the quality of studies. Results that agree with current practice are readily accepted, as was VA'S first report that patients with left main disease benefit from surgery. Results at odds with practice, on the other hand, are carefully scrutinized and criticized (see ch. 4, "Constituency Behind the Intervention").

Wortman and Yeaton compared the results of randomized and nonrandomized studies of CABG surgery, and synthesized the RCTS' results (253). They point out the value of RCTS by showing that nonrandomized studies consistently overestimate the benefit of surgery compared with randomized studies. This conclusion held regardless of whether the endpoint measured was mortality, survival, or size of effect. The discrepancy could not be explained by differences in distribution of patients' risk categories, crossover rates, or the timing of the trials. The different results between the two types of studies occur primarily because nonrandomized studies find that the medically treated group fares considerably worse than RCTS find. The surgically treated groups were not so different in outcome, though their results were slightly better in RCTS.

Antithrombosis Trials

Blood platelet aggregation is an important factor in thrombosis and in atherogenesis. A number

of agents have been tested to prevent this aggregation. Aspirin, a well-known inhibitor of platelet aggregation, has been tested on heart attack survivors in at least six RCTS. The NHLBI AMIS, the largest RCT in this field with over 4,500 patients, showed that aspirin had no effect on survival.

Soon after publication of this trial's results, the Society for Clinical Trials reviewed it along with five other studies (including two other newly published trials). Together these trials studied over 10,000 myocardial infarction patients randomized between aspirin and double-blind placebo controls. During the studies, 1,000 of the patients died. Each study individually provided no clear evidence of aspirin's benefit. Taken together, however, they indicated that aspirin did reduce the risk of death, though at a lower rate than the individual tests could reliably detect. It was estimated that the overall reduction in the odds of reinfarction in all six trials was 21 percent (standard error + 5 percent) and that about 70 deaths had been prevented (126a).

Reviewing the evidence from the six aspirin trials, an editorial in *The Lancet* concluded:

It may be that the small benefit indicated thus far by both the antiplatelet and the anticoagulant randomized trials realistically represents all that can be achieved by any form of interference with haemostasis in the months or years after MI [myocardial infarction].

Other antiplatelet agents have been evaluated in RCTs—e.g., Persantine (dipyndamole) and Anturane (sulfipyrazone) (see ch. 4 “the Anturane Reinfarction Trial”).

NHLBI is now funding jointly with NCI a primary prevention trial to test the hypothesis that aspirin may help prevent initial MI. More than 20,000 healthy male U.S. physicians have been enrolled as participants in a double-blind placebo-controlled trial of aspirin to prevent cardiovascular disease in addition to testing beta carotene (a precursor of vitamin A) for cancer prevention.

Beta Blockers

In 1965, a nonrandomized study showed a reduction in mortality in those given propranolol, a beta-blocking drug (106), after a myocardial infarction. Though beta-blockers clearly have anti-hypertensive, antiarrhythmic, and antiplatelet properties, the mechanism through which they reduce mortality after MI unclear. Nonetheless, since then at least 41 placebo-controlled RCTS have tested at least 7 beta blockers in varying regimens (128).

Completed trials have most reliably evaluated the effect of “moderately prolonged beta-blockade in the period after discharge from hospital” (128). While most of these trials were too small to demonstrate a statistically significant benefit (using $p = 0.05$), in nearly all the trials mortality was reduced in those who took beta blockers. When the trials are pooled, a strongly significant result emerges. Based on the joint results, the total number of deaths was reduced by about 25 percent in those who took beta blockers over the course of the trials. “This effect will be widely regarded as sufficient to justify routine use of long-term beta-blockade in many patients for perhaps the first year or so after discharge from hospital” (128).

It is gratifying that RCTS have produced reliable information in this field, but questionable whether so many trials were necessary. Rose comments that given limited resources, “this sort of uncoordinated proliferation has been extremely wasteful” (193).

Two big questions remain about treatment regimens for beta blockers: 1) whether treatment should begin “early” (between a few hours and about 3 days after the infarct) or “late” (3 days later or more), and 2) how long the treatment should last. A number of studies of early beta-blockade are in progress, and answers to these questions may be available within the next few years. It is generally thought that beta blockers are used extensively for treating heart attack patients, and that their widespread use preceded convincing evidence from RCTS.

SURGERY

The impact of RCTS on surgery has been minimal, largely because RCTS in surgery are the exception rather than the rule. When RCTS are done, they are often criticized for coming too early or too late in the life of the innovation (see ch. 4, “Timing of RCTS”).

It is instructive to consider the origins of surgery. Most current surgical practice has its beginnings before RCTS were available as a tool—i.e., before the middle of this century. Historically, much of the practice of surgery was in setting bones or suturing wounds. These procedures are clearly effective. As in treating acute diseases, a surgeon would know quite quickly whether the treatment worked. In many cases, the treatment could be repeated (e.g., a bone reset) if it failed the first time.

The removal of diseased or cancerous organs also seems to make such good sense intuitively that the value of such procedures was rarely questioned. If the patient died, it was not necessarily a failure of the operation, but a sign that the patient was beyond help. The theory behind much cancer surgery, which has been available since the last century, is that survival depends on removing all diseased tissue. (This assumes that all disease is visible, and that no spread of cancerous cells in the bloodstream occurs until late in the disease. The treatment of breast cancer has shown this not to be the case.) Successful surgery, meaning an aseptic operation that the patient survives, was considered successful treatment, and for many operations this is a good rule. Long-term outcomes have generally not been considered.

The nature of surgical procedures contributes to the difficulty of testing them through RCTS. Bonchek compares RCTS for surgery to those for drugs (20). Unlike drugs, which are fixed compounds, surgical procedures evolve. The efficacy of a drug is in many ways unrelated to the skill of the physician administering it. In surgery, the skill of the surgeon is vital, and this skill itself changes over time. Love observes (137):

Drugs come as packaged preparations to be given by dosage. Operations are conceptual plans that require execution, and the details of a given

operation change with time among surgeons and from patient to patient. It should be abundantly clear that techniques for evaluating the one cannot be used to evaluate the other.

Bunker and colleagues attribute the limited use of RCTS in surgery to the “very real conceptual, practical, ethical, and economic difficulties of carrying out in adequate numbers and sizes experiments involving complex surgical procedures in human beings” (30). They also conclude that not conducting such trials can cost more in dollars and lives than a trial adequate to answer the question.

Surgical RCTS in cancer treatment follow much the same pattern as those in other fields. Trials of chemotherapy by far outnumber those in surgery or radiotherapy. Many surgical oncologists resist participation in such trials, and trials that have been done have come long after a procedure is introduced. The history of surgical techniques used in treating breast cancer illustrates this. The proposal that a lesser operation be used in place of a radical (Halsted) mastectomy was published in 1948. Not until 1967 was a trial carried out. Even today, though the practice has gradually declined, many women undergo radical mastectomy when a modified procedure would be equally effective and less disfiguring (see the section “Breast Cancer” above and ref. 226).

The literature on the impact of RCTS in surgery is limited, considering the size of the field. One volume, *Costs, Risks, and Benefits Surgery*, covers a wide range of topics in surgical innovation and evaluation, including RCTS (28). The editors conclude with a series of recommendations, including those for improving the study of surgical procedures (see ch. 6).

Bunker and colleagues (29) studied the introduction and evaluation of four modern surgical procedures, three that were eventually assessed by RCTS. They note the particular problem of carrying out RCTS of new therapies for conditions that previously had no effective therapy of any kind. Withholding treatment in these cases can pose difficult ethical questions. The use of shunt surgery for portal hypertension is one example. After decades of use, the procedure was subjected

to evaluation by RCT only because of two developments: the recognition that the surgery had a serious side effect (encephalopathy), and the extension of the use of the operation beyond its original indications. The uncertainty about the use of the surgery for new indications, using it prophylactically rather than just therapeutically, led to RCTS with the newly indicated group of patients. After these trials showed shunt surgery to be ineffective prophylactically, further trials demonstrated its lack of efficacy for its original therapeutic uses.

Three case studies in *Assessing the Efficacy and Safety of Medical Technologies* discuss surgical procedures that require evaluation, largely because RCTS of them have been inadequate or simply not done (225). These three case studies are summarized below.

Tonsillectomy, the third most common surgical procedure in U.S. hospitals, is thought by many physicians to be overused. Reports of tonsillectomy reach back as far as 600 B. C., yet the first RCT of the procedure in this country began in 1973. Tonsillectomy differs from some other procedures with long histories, such as cast application for bone fractures, in that its efficacy is not obvious and the indications for use not well understood. The National Institutes of Health (NIH) sponsored a workshop in 1973 on Tonsillectomy and Adenoidectomy that recommended a nationwide multicenter RCT. That idea was later endorsed by another NIH-convened group, the NIH Ad Hoc Advisory Panel on Tonsillectomy and Adenoidectomy. In 1978, a third group did not agree to go ahead with the trial.

Appendectomy is another frequently performed surgical procedure that has not been evaluated by an RCT in this country. The different rates of appendectomy in different regions of the country (from 100 to 620 per 100,000 for 1965-73) and evidence from other parts of the world provide strong support for the need to understand the appropriate use of this procedure. The OTA report concluded that an RCT might be warranted in view of "strong evidence suggesting that appendicitis may be treated with substantially fewer appendectomies without increased loss of life."

Hysterectomies are performed for a wide variety of conditions, including the traditional indications of premalignant states, localized cancers, descent and prolapse of the uterus, and obstetric catastrophes (e. g., functional problems). Performed in over 600 per 100,000 women each year, this major operation is more frequently performed than any other. In assessing the costs, risks, and benefits of elective hysterectomy, Korenbrot and colleagues reviewed studies indicating that at least 30 percent of hysterectomies performed were not justified by medical indications alone (126). The implication, though unprovable, is that most were performed for sterilization or cancer prophylaxis. Lack of clarity about the procedure's appropriate indications and the substantial risks and poorly known aftereffects of the surgery itself emphasize the need for controlled trials. In 1978, OTA was unable to identify any clinical trial of hysterectomy in this country.

Neurosurgery

Haines has recently examined RCTS in neurosurgery based on an exhaustive search of the English language literature (103). In an earlier paper, he reviewed 4,685 scientific articles appearing between 1944 and 1977 in the *Journal of Neurosurgery*, finding that only 18 could be classified as controlled clinical trials, and of those, 10 used random allocation procedures (104). One of the ten used blinding procedures. His later, more extensive review (103) identified a total of 51 RCTS of neurosurgical procedures, adjuncts to neurosurgical procedures or medical treatment of neurosurgical diseases. Half these studies were published after 1977. Most of the studies (61 percent) were of adjuncts to surgical therapy (e.g., radiation and chemotherapy for malignant primary brain tumors), 15 directly tested a neurosurgical procedure, and 5 nonsurgical therapy, such as antibiotic treatment of shunt infection.

The increased use of RCTS in neurosurgery is encouraging, but Haines asks: "Have any important questions been resolved by such *studies*?" He answers with a qualified "no." A large percentage of the trials were methodologically inadequate and permitted no conclusions. The well-conducted

studies, however, though they failed to put important questions to rest, did gather important information about the natural history of diseases. The case has been made for more definitive trials in this field, some of which are under way. In neurosurgery, and probably in other surgical areas, the quality of trials is a serious problem. Statisticians have not been routinely involved in design, which proves to be a major determinant of trial quality (105). Progress has been relatively slow, and will come only with surgeons' greater appreciation of the value of RCTS.

Haines reports a case of negative results in small RCTS with low statistical power, that encouraged an unwarranted decline in a neurosurgical practice (105). A standard practice in the late 1970's was the use of antifibrinolytic agents in treating patients with subarachnoid hemorrhage from ruptured intracranial aneurysm. The purpose of the treatment is to prevent recurrent hemorrhage dur-

ing the waiting period between first hemorrhage and surgery. Haines reports that three recent reviewers have seriously questioned the efficacy of this therapy, based on the evidence from RCTS, and have suggested that antifibrinolytic agents may aggravate another problem, vasospasm. Haines' reassessment of the RCTS yields a different conclusion. The four trials that showed the treatment was ineffective all had a less than one chance in three of finding a 50 percent better outcome in the treated group, if such a difference existed. The three studies with the greatest statistical power showed some benefit from the therapy, and little evidence for its aggravation of vasospasm. Haines concludes that discarding antifibrinolytic therapy is premature. He recommends further clinical trials to study both its efficacy and safety, in studies that are well designed and large enough to produce significant answers.

RCTS IN OTHER FIELDS

Chalmers and colleagues (40) have been engaged over about the last 5 years in the development of a computerized data base of RCTS. As of 1982, about 2,700 RCTS were entered, indexed by groupings of the International Classification of Diseases (WHO, 1977). From their data base, Chalmers and colleagues have identified common disease states for which a relatively large number of RCTS are available, and have evaluated the quality of the trials according to an index they have developed (see ch. 4, "Quality of RCTS"). Where possible, they have synthesized the results of studies to draw conclusions about therapies tested. Topics addressed have been: surgical therapy of duodenal ulcer, early mobilization and discharge of acute myocardial infarction patients, antithrombotic agents in acute myocardial infarction, cost and efficacy of the substitution of ambulatory for inpatient care, treatment of acute

alcohol withdrawal, treatment of acute infections and alcoholic hepatitis, nephrology, tropical diseases, effects of steroids in the gastrointestinal tract, and emergency diagnosis and treatment of gastrointestinal hemorrhage.

The degree to which RCTS are used in different fields of medicine varies greatly, hence the impact of RCTS must vary. Certain areas have not been mentioned specifically in this chapter, for instance pediatrics, and obstetrics and gynecology. In these areas too few RCTS have been conducted to allow much impact. While it is easy to focus on deficiencies of studies that are done, it is more important though more difficult to identify medical fields which lack RCTS altogether. Very little has appeared in the literature in this regard, except in the case of surgery, which was reviewed in this chapter.