4.

ESTIMATING EFFICACY AND SAFETY

4.

ESTIMATING EFFICACY AND SAFETY

Techniques used for estimating efficacy and safety range from the informal methods of individual physicians to randomized clinical trials with complex methodological designs. No technique is universally applicable for every medical technology. In many instances less complex methods may be more appropriate than the more sophisticated approaches. Frequently, combinations of techniques are used. This chapter describes five techniques used in evaluating safety and efficacy: preclinical, informal, epidemiological and statistical, controlled clinical trials, and formal consensus development.

Various laws have been enacted to regulate the efficacy or safety of drugs and medical devices since the passage of the Federal Pure Food and Drugs Act in 1906. Surgical and other procedures that depend primarily on providers' techniques have not been subject to similar controls. Rather, responsibility for assessing the efficacy and safety of these procedures is contained within the profession (125,332,334).

Assessments of efficacy and safety for "products" (drugs and devices) usually differ from assessments of medical and surgical procedures in terms of the source of evaluation and the kinds of techniques applied. The physical nature of products implies a highly consistent formulation that may be unattainable in surgical technique evaluation. Also, investigators can learn much about products before they are tested clinically (394). Many procedures, however, heavily rely on testing for their development.

PRECLINICAL

Many medical technologies are evaluated in biochemical and animal tests prior to human experimentation. These preclinical tests maybe part of the developmental effort, or a requirement for Federal or private approval, or both. The required tests may be of two types: 1) preliminary evidence to gain the right to test with humans (364), and 2) performance standard compliance to establish marketability.

Chemical analyses for purity, quantity, and quality of the active agents are typically undertaken. Other filler and stabilizing substances are evaluated for potential pharmacological activity.

Animal testing provides a guide to potential therapeutic activity as well as capacity to induce toxicity (85). Determining the degree of toxicity, or safety, is the major function of animal studies. A prime factor analyzed in safety tests is the level of median lethal dosage. Toxic effects are evaluated in terms of chemical and physiological analysis. Therapeutic effects may be measured in terms of bioavailability (transport across gastrointestinal membranes) and pharmacokinetics (distribution throughout the body).

The accuracy of animal models in determining the probable effects of drugs on people is a controversial issue. In particular, carcinogenic agent evaluation in animals is a

very complex, multifaceted problem. Questions that arise in these evaluations include short-term high dose versus long-term low dose, animal species selection, population size, and controls (191). Despite some of the inherent problems in utilizing animals, the report by the Office of Technology Assessment (OTA), *Cancer Testing Technology and Saccharin (353),* concludes that they are acceptable models for cancer studies and probably should be regarded as reasonable precursors to clinical studies.

Medical devices are evaluated by chemical and physical laboratory testing in addition to animal studies. Physical testing may seek to determine mechanical strength, material properties, and electrical performance. General manufacturing techniques, such as quality control, precision machining, and sterility, may also be evaluated. Chemical tests using culture or hematologic techniques may determine biocompatibility. Other chemical tests evaluate long-term dissolution in body fluids and the possible presence of toxic residues in the production of plastic materials. Implantable devices also are subjected to complete preclinical animal testing.

INFORMAL

Despite the increasing need to formally estimate the efficacy and safety of medical technologies, the majority of such evaluations are still based on informal approaches. White (426) estimated that 80 to 90 percent of all procedures have been evaluated by informal methods. These informal assessments of medical technologies may take place during medical school and specialty training and through personal peer experience.

Physicians and other health care personnel are constantly exposed to medical technologies throughout medical school, residency, and special courses. Students generally assume that these technologies are efficacious and safe, Technologies recommended to the student have undergone formal statistical studies or professional consensus exercises. However, it is more likely that the suggested uses of technology are based on previous experiences or training received by the instructor.

Personal experience is perhaps the oldest and most common informal method of judging the efficacy and safety of a medical technology. This technique is dominated by qualitative impressions. The control groups are primarily envisioned as experiencing the end result that would occur if there were no clinical intervention (85). Despite its limited statistical value, this technique does have some advantages compared to the more rigorous methods used in certain situations. For example, personal knowledge of the patient may promote beneficial adjustments to the type and level of treatment. Also, many rare side effects are reported in letters to the editor columns by individual physicians (85). Perhaps more importantly, personal experience is the primary method that determines whether or not a medical technology is adopted into widespread practice (79,187).

Peer experience is more explicit than personal experience; information may be exchanged by personal communication, journal articles, pamphlets, and the like. Again, there is little control over the scientific quality of these technical assessments. However, this peer interaction is the core concept of the more formal group consensus discussed later.

It is important to point out that many medical advancements have properly and successfully proceeded without rigorous statistical methodology of evaluation. For example, vitamin B12 treatment for pernicious anemia clearly is justified. Cast application for

forearm fracture (see chapter 3, case 14) is a technique whose efficacy has been established experimentally in medical settings. Alternatives such as bamboo splints exist (170); however, the widespread acceptance and success of casting makes evaluation of other methods unlikely and probably unnecessary. An earlier OTA report, *Development of Medical Technology: Opportunities for Assessment, * (354),* made two points that summarize the utility of informal methods: 1) "despite complexity, and cost, some procedures are so effective in restoring function that few would question their social utility, " and 2) ". . for a disease for which the natural history is fairly well known and the benefits of a new technology are dramatic, alternative methods of evaluation (as compared to controlled clinical trials) may be appropriate. "

Informal techniques are based on the clinical approach of qualitative, artful decisions as compared to the scientific approach of quantitative, mathematical decisions. Ingelfinger, et al. (178) point out the critical issue of statistically significant findings versus clinically significant results. Other sources (24) describe further causes both for separating the informal from the rigorous technique and developing new methodologies to improve medical decisions.

Three concepts summarize the necessity of both the informal and the rigorous techniques for assessing efficacy and safety. First, each extreme may be appropriate in certain situations. Second, many assessments require various combinations of techniques. And third, cooperation between clinicians and statisticians must exist to attain appropriate decisions when more rigorous techniques are used.

EPIDEMIOLOGICAL AND STATISTICAL

Epidemiology is the study of the determinants and the distribution of diseases and injuries in human populations. The term also incorporates the study of the impact of medical interventions on diseases and injuries. Three types of epidemiological methods that are particularly useful in evaluating the efficacy and safety of certain medical technologies are described in this chapter. These three methods are: retrospective, prospective, and controlled clinical trials. The last type of study warrants discussion in a separate section from the other two because of its importance and prevalent use.

Retrospective studies compare groups of people who have a disease with those that do not. These studies are designed to determine whether the two populations differ in terms of percentage exposed to certain critical factors. In addition, attempts may be made to compare standard factors, such as age, sex and race, between the two groups. Data obtained from retrospective studies are summarized as an "odds" ratio** which is defined as the ratio of incidence rate among the exposed group to the incidence rate among those not exposed. Both the relationship between oral contraceptives and thromboembolism* * * and the positive correlations demonstrated between smoking and lung cancer were established by retrospective studies.

Most information used in retrospective studies is derived directly from the patients, their relatives and friends, and individuals' medical and other records. Consequently, the

---

*This report, released in August 1976, described the development and assessment of cardiac pacemakers for heartblock.

* ● The "odds" ratio is a close approximation of the relative risk.

●● *Users of oral contraceptives are four or five times more likely to develop thromboembolic disease than nonusers (81).

uniformity, accuracy, and completeness of information (especially on death certificates) are often in doubt. In addition to incomplete or biased data, the selection of appropriate comparison groups represents another major problem in this type of research.

Despite some inherent problems, general utility of retrospective studies has been frequently substantiated by other experiments in which there is more control (81). Even marketing and manufacturing data may provide critical links to unsafe technologies. Atomizers containing isoproterenol were linked to cardiac arrythmia deaths. Improper usages and overdoses due to poor quality control in manufacture were shown to be probable causes of death. Utility, low cost, and quick results are the major advantages of these studies (237).

Prospective studies follow the histories of persons both exposed and unexposed to a particular factor under study. The incidence of deleterious effect resulting from such exposure is then determined for persons in the two groups. If records of individuals exposed to a particular factor exist, then the study also may utilize past data; however, prototypic prospective studies deal with ongoing events (43). Statistical results from such studies include incidence rates in addition to relative risk.

A major advantage of prospective studies is the relatively clear designation and selection of both the study and the comparison groups by means of matching characteristics with minimum bias before the disease develops. Some of the disadvantages of these studies include their high cost and the possible occurrence of changes in patients and methods over the duration of the test (237).

The Boston Collaborative Drug Surveillance Program (244) is an example of a large study that assesses drug efficacy and safety by utilizing epidemiologic methods. * To date, approximately 12 percent of the drug exposures studied by this program have yielded unsatisfactory results. In addition, statistical techniques were useful in discovering and estimating the frequence of unsuspected adverse drug reactions. The Framington Heart Study, which has been in progress since 1948, has shown a clear correlation between high blood pressure and the occurrence of cardiovascular disease in adults also using epidemiologic methods (81). Currently, some epidemiologic methods are aimed at assessing the efficacy and safety of various antihypertensive treatments.

CONTROLLED CLINICAL TRIALS

All subjects who agree to participate in controlled clinical trials (or simply, randomized clinical trials) are assigned to experimental and control groups. Subjects in these trials are assigned randomly to either the experimental or control group. These trials, and their impartial test and control group establishment, are direct experimental extensions of prospective studies that have no control over the physician's choice of treatment. In a controlled clinical trial intended to assess efficacy and safety, the experimental group would be treated or diagnosed by the technology under examination; usually the control groups would be either treated by an established standard technology or given a placebo. However, in some cases, a standard technology is administered to one of the study groups while a second (control) group receives no treatment. Clinical tests and examina-

---

*The program was initially funded by the Pharmaceutical Manufacturer Association Foundation. Since 1967, it has been supported by a number of other organizations, including FDA and the National Institute of General Medical Sciences of NIH.

tions of the members of each group are used for evaluations of the relative benefits and risks of the technology.

Many controlled clinical trials require a long period of time and large commitments of money, resources, and subjects, The National Institutes of Health (NIH) estimated that the total amount of money* expended for trials underway in FY 1975 (new starts and continuing studies) was $641.8 million for 755 trials. * * Efficacy and safety research often requires money contributions from several sources. For example, it may be appropriate sometimes for the third-party payers to finance part of the evaluation of an established, presently reimbursable technology. In addition, the Food and Drug Administration (FDA) estimates that private drug firms spend $1 million to $4 million to bring a drug to market after it has been developed in the laboratory (406).

Many professionals who conduct research into the efficacy of medical technologies have focused attention on the randomized controlled clinical trial because critical assessments of the efficacy and safety of medical technologies require high-quality research (65). For example, Cochrane (72), Hill (163), and others strongly support the use of the randomized clinical trial in evaluating efficacy or safety. Conversely, others (133) suggest that nonrandom, less well-controlled trials and statistical manipulation of available data can provide results that are as useful as randomized clinical trials.

Randomized controlled trials are the most useful when: 1) the benefit of a new technology is uncertain (e. g., amniocentesis, see chapter 3, case 2), and 2) the relative benefits of existing therapies are disputed (55) (e.g., tonsillectomy, see chapter 3, case 9). There is much statistical theory that supports the scientific utility of such randomization procedures in clinical trials. Byar, et al. (55) discussed three major advantages to randomization. First, and most familiar, bias may be eliminated from the assignment of treatment. Often double-blind techniques are utilized in which neither the patient nor the physician knows the technology used on any specific individual. (However, in comparing drug to surgical treatments, bias may well occur because both the surgeon and the patient know which method is being utilized; and only lower risk patients may be candidates for the surgical operation.) Secondly, randomization prevents bias with respect to variables that exist in the experiment but are not directly considered in the design. This allows comparisons between treatment groups. The third advantage of randomization is the validity of the statistical tests of significance that are used to compare treatments. It should be noted that complete randomization may be inappropriate under certain circumstances; in such cases modifications in the randomizing process may be used (151).

There are many areas of controversy surrounding the use of randomized clinical trials, perhaps the greatest of which is ethical (21). Arguments against randomization and other aspects of these trials are based on a concern for both patient and physician rights and responsibilities. Critiques of randomization include the following statements: physicians must make clinical judgments and act according to their consciences (431); personal physicians must influence whether their patients enter a trial and what treatment is administered; patients must be given the best possible information in consent forms (335); and, patients should be able to choose which treatment is delivered.

Critics of controlled trials or of some of the processes used in trials also point out that certain groups of patients have rights that are easily violated. Appropriate questions regarding the rights of children in particular are raised, For example, when can informed

---

● The total amount here refers to the entire cost of completing trials that were underway in FY 1975.

.● Trials supported by the NIH vary widely in costs. One of the most expensive, the Multiple Risk Factor Intervention Trial (MRFIT), is budgeted at $115.7 million.

consent be given by a child?; at what age?; with what medical conditions or illnesses?; and, who, if not the child, will guard those rights? In addition, the long-term effects of treatments or other medical technology interventions can be especially serious and very long in evidencing themselves in children. Clinical trial protocols must be established with all these and more questions in mind. Similar questions may occur regarding the rights of other groups composed of convicts, the aged, and the mentally retarded, for example.

Many articles defend the ethics of using controlled clinical trials. Byar, et al. (55) state that physicians cannot do just what they "believe" best, their practice must be based upon sound scientific evidence. Similarly, an honest acceptance of the fact that the relative benefits and risks of the best current therapy are not known is the first step in recognizing the need for clinical trials. If each patient is so unique as to be ineligible for statistical randomization, how can the individual physicians use clinical judgments based on past experience as the optimal guideline for determining the treatment of the next patient (55)? Mosteller (249) contends that the rights of patients are protected in their ability to refuse participation in the trial. In addition, proper diagnosis of a patient must precede a decision regarding trial participation. In some cases, patients (or physicians) may also choose to select a treatment but randomize on dosage level. This choice also provides the patient with more control. A final point in favor of randomization is the apparent improvement (although not perfection) of the statistics and planning of recent randomized clinical trials.

There are no unequivocable answers to these concerns. Certain technical improvements in statistical methods allow faster identification of intermediate results, thereby leading to sounder decisions regarding the termination date of certain types of trials. Improved consent mechanisms are being developed and could be applied more widely, Interestingly, many articles note serious complaints about randomization but still recommend cautious use of the technique (335,423).

FORMAL CONSENSUS DEVELOPMENT

The assessment of a specific medical technology may include one or more studies which use any or all of the techniques previously described. If the evidence clearly supports or rejects the relative utility of a treatment, then the analysis of efficacy and safety may be complete (though it may need periodic re-examination). In many cases, however, the evidence does not lead to such an unequivocable decision. Consequently, a consensus group may be formed both to evaluate all pertinent information, which may range from informal to detailed statistical studies, and to recommend its findings to the medical community.

There are two types of consensus groups relevant to this report which are discussed further in the next chapter. Briefly, one type of consensus group evaluates the current state of efficacy and safety knowledge regarding either a particular medical technology or technologies that relate to a specific medical condition. An example of this type of consensus development is the "technical consensus-building" effort of NIH. A second type of group both analyzes a medical technology, particularly devices, and recommends possible standards to be used in the conduct of future efficacy and safety assessments. This type of consensus process is used in the programs of the Association for the Advancement of Medical Instrumentation and the American Society for Testing and Materials.