4
Methods of Drug Evaluation;

4
Methods of Drug Evaluation;

TYPES OF STUDIES

The primary objective of postmarketing studies is to develop inforrnation about drug effects under customary conditions of drug use.

The initial clues about a drug's potential effects come from the experimental studies carried out with both animals and humans in the premarketing period. Spontaneous or voluntary reporting (e.g., in letters to the editors of medical journals) is the oldest, and to date, the most productive source of new information about a drug's possible effects once a drug is marketed. Other types of studies are used to examine in more detail the possible effects of a drug. In general, these other types of studies use either cohort or case-control methods. Thus, four types of studies are generally used to identify drug effects: 1) controlled clinical trials, 2) spontaneous or voluntary reporting, 3) cohort studies, and 4) case-control studies (23, 50,61,77).

Controlled clinical trials match treatment and control groups as closely as possible, minimize bias through such methods as randomization and "double-blinding," and directly monitor patients for the duration of the study. For example, patients can be randomly assigned to either the control or treatment group. The control group receives a placebo or an active comparison drug that looks exactly like the drug being tested, and both the investigators and patients do not know who is receiving the real drug. (Personnel not directly involved in the tests would of course know what substance each patient was receiving). In this method, possible drug effects, both therapeutic and adverse, are closely monitored, so that they are discovered as they occur.

The controlled clinical trial is considered the most definitive method for evaluating a drug's efficacy and safety, but the use of rigorous criteria for patient selection usually means that the patients tested represent only a special class of the anticipated users of the drug(s), and the carefully controlled conditions allow for study of fewer patients than in the other methods, Thus, for example, to observe drug effects that are rare or that appear only after long-term use, controlled clinical trials might be impractical or too expensive.

Voluntary reporting may be spontaneous, such as in a "letter to the editor" of a medical journal about an unusual condition observed in a patient on a particular drug, or it may be more organized, as with the "yellow card" system in Great Britain. Most of the reporting to the Food and Drug Administration (FDA) is by pharmaceutical manufacturers, who are required by law to report adverse reactions. In practice, most of the information obtained by the manufacturers originates from physicians and other health professionals. Such observations serve as warnings of possible adverse drug effects, so that the inference of an association between a drug and an observed health condition may be further studied by cumulative, careful reporting, and confirmed or disconfirmed by more vigorous methods. Underreporting is a serious deficiency of this voluntary method, and a drug may also be wrongly associated with an adverse effect until the suspected association fails to show up in repeated, statistically validated studies.

Cohort studies follow a defined group of patients (the cohort) for a period of time. In this method, patients are not randomly assigned to groups, and there is no blinding. Cohort studies are usually prospective, and observe the cohort from the beginning of drug *use.* A group of patients taking the drug of interest is assembled and followed to see, for example, if any adverse reactions occur. A second group of patients (the controls) with the same medical condition, who are not taking the drug and who may be receiving alternative treatment, but who are otherwise matched as closely as possible with the cohort, may be studied in parallel. The control group is used to identify the frequency of occurrence of

any condition observed in the drug-exposed group, but which must be due to causes other than the drug (the "background incidence" of the condition). In this method, patients can be directly monitored to ensure they take the drug appropriately and to observe the drug's effects; or monitoring can be less systematic. With less monitoring, a larger cohort can be followed, but bias is thus increased.

Although uncommon, a retrospective cohort study may also be conducted when purported drug-induced effects have already been observed at the time the study is started. A retrospective cohort study must accurately ascertain patients' past drug use. In principle, this could be done by a "closed" pharmacy record system, in which users have been restricted to a single supplier or payment source. In practice, the use data on individuals needs to be connected to data on their outcome, as through computer files. Medicaid, Medicare, the military, and some prepaid health maintenance organizations could be employed for this purpose (76).

Case-control studies identify patients with the adverse effects to be studied (the cases), and compare them with a sample (the controls), drawn from the same cohort that gave rise to the cases. Controls are matched as closely as possible with the cases, except with regard to the drug's suspected adverse effect, to examine whether exposure to the drug is the cause. Patients with conditions suspected of being associated with a certain drug would have their medical records reviewed or be interviewed concerning the use of that drug. The histories of the controls would also be studied for information about drug use in the general population. By comparing the proportion of drug users among the cases with the proportion of drug users in the general population, it is possible to infer the relative frequency with which adverse reactions occur in users of certain drugs as compared with nonusers. A sufficient number of appropriate cases must be identified and accurate histories of exposure to drugs must be obtained.

Among the advantages of retrospective case-control studies compared with prospective cohort studies are the smaller number of patients required in retrospective case-control studies, the relative ease of carrying out the study, the lower cost, and the shorter time needed. A disadvantage of the retrospective case-control method is that a condition must have been already identified and suspected as the effect of drug use. It is also harder to reduce bias in a retrospective study than to do so in a prospective one (65).

Bias is equally possible in cohort and case-control studies, though each kind of study is liable to a different kind of bias. For example, bias in the observations can arise with respect to identifying the effects of the drug in cohort studies and with respect to identifying the exposure in case-control studies.

Controlled clinical trials and prospective cohort studies can be used to determine a drug's beneficial as well as adverse effects. Case-control studies are usually used to trace adverse effects back to prior drug use. Voluntary reporting can uncover additional uses of drugs as well as their adverse effects, but reporting of adverse effects is much more common.

DETECTION AND ASSOCIATION

The ability of a particular surveillance method to detect a drug's effect depends on two factors: 1) the time that transpires between use of that drug and the occurrence of the drug's effect (the latency period), and 2) how often the effect occurs (its frequency). There are many other determining factors, such as accuracy of observation, and accuracy and completeness of medical records, but these factors are more a problem in the design of a study's details,

The latency of some drug effects presents a serious problem for their study. Some effects occur immediately, or within days or weeks after drug use, or with continued use of a drug. But other effects may occur long after a drug has been dis-

continued, or only when another drug is taken simultaneously, *or* only in patients with certain predisposing conditions. Other effects may not be manifest in the patients themselves, but rather in their children. The use of DES (diethylstilbestrol), in pregnant women, for example, has been associated with vaginal cancer in their daughters, but only after the daughters have reached adolescence.

Controlled clinical trials, because of their relatively short duration, will detect only acute or subacute effects. Long-term cohort studies can detect delayed effects, but the data bases necessary for such long-term, large studies are still sparse. Voluntary reporting is usually the way in which long-term effects are first identified. Long-term effects are usually confirmed through retrospective case-control studies, but their reliance on historical data such as medical records can limit the accuracy of these studies.

The chance that a particular study will discover a drug effect also depends on the study's sample size and the frequency of the drug effect. For example, in a cohort study, if a drug causes blindness in 1 out of every 100 users (1/100), how many users must be observed to find one case of blindness? If there were 1 million users of the drug, there would be 10,000 users blinded. But in a small sample of only 100 users, the probability of finding one or more cases of blindness in the sample would only be **63** percent. If the sample were 200 users, the probability would increase to 86 percent. With a sample of 500, the probability would be 99 percent that at least one case of blindness would be found in the observed users.

To state it another way, what number of users would have to be observed to be 95 percent sure of finding one or more cases of blindness when they occur at a frequency of 1 in 100 users? The answer is 300 users, and the general rule is that the number of users in the sample must be three times the reciprocal of the frequency; e.g., for a frequency of 1 in 1,000, the sample would have to be 3,000 to be 95 percent sure of observing at least one case. Table 5 summarizes: 1) the probabilities of observing an adverse drug reaction (ADR) for different sample sizes and frequencies, and 2) the sample size required for various frequencies of an adverse reaction to be 95 percent sure of observing that reaction. The numbers in the bottom row are the sample sizes needed to be 95 percent sure of observing at least one ADR. In table 6 are shown the sample sizes needed to be 95 percent sure of observing one, two, or three adverse reactions for the frequencies shown.

In a sample of 3,000, effects occurring at any rate higher than 1/1,000 (e.g., 1/10, 1/50, 1/200, etc.) should be observed, and there will be a 95 percent probability of observing at least one effect that occurs at a frequency of 1/1,000. How-

Table 5. —Likelihood of Observing an ADR (95% **likelihood)**

	Threshold for an ADR					
	1/100	1/500[a]	1/1,000[b]	1/5,000[c]	1/10,000[d]	1/50,000[e]
Number of patients in ADR study:						
100	*0,63*	*0.18*	0.10	0,02	0.01	0.002
200	0.86	0.33	0.18	0.04	0.02	0.004
500	0.99	0.63	0.39	0.10	0.05	0.01
1,000	0.99	0.86	0.63	0.18	0.10	0.02
2,000	0,99	0.98	0.86	0.33	0.18	0,04
5,000	0.99	0.99	0.99	0.63	0.39	0.10
10,000	0.99	0.99	0,99	0.86	0.63	0.18
Number of patients required to be 95°/0 likely to observe an ADR	300	1,500	3,000	15,000	30,000	150,000

[a] For example, lymphoma from azothioprine
[b] For example, eye damage from practolol
[c] For example, heart attack in older women frp, oral contraceptives
[d] For example, anaphylaxis from penicillan
[e] For example aplastic anemia from chloramphenicol

SOURCE D L Sackett, et al 'Compliance,"' in *Monitoring for Drug Safety* W H W In man (ed) (Phi l adel phia J B Lippincott Co 1980)

Table 6.—Number of Patients Required To Detect One, Two, or Three ADRs With No Background incidence of Adverse Reaction (950/0 likelihood)

Expected incidence of adverse reaction	Required number of adverse reactions		
	1 ADR	2 ADRs	3 ADRs
1 in 100	**300**	**480**	**650**
1 in *200*	***600***	**960**	**1,300**
1 in 1,000	3,000	4,800	6,500
1 in 2,000	6,000	9,600	13,000
1 in 10,000	30,000	48,000	65,000

SOURCE J A. Lewis, "Post-Marketing Surveillance How Many," *Trends in Pharmacological Sciences* 2.93, 1981

ever, except for some effects that are unique to a drug, many drug effects (e. g., stroke, heart attack) are indistinguishable from conditions due to other causes. This "background incidence" of a condition must be known before purported drug effects observed in a study can rightly be attributed to a drug.

The sample size needed to be reasonably sure that an observed condition is the drug's effect depends on the following factors: 1) the background incidence of the condition, 2) the additional incidence due to the drug, and 3) the size of the control group, if the background incidence is unknown (47). The quantitative relationships among those factors are summarized in table 7, which presents the sample sizes required to be 95 percent sure that an observed condition is due to the drug and not to some other cause.

Larger sample sizes are needed to determine a drug's effect as the background incidence of a condition increases and as the frequency of the drug's added contribution to a condition decreases. This is best explained graphically, as in figure 2. In this example, the background incidence in every case is 1/100. As the incidence of the added effect decreases from 1 in 100 to 1 in 10,000, the sample size has to increase from *1,600* to 11 million to remain 95 percent sure of observing the added effect (see the rows in table 7 representing a known background incidence of 1 in 100).

The relationship between background and added incidence is also revealed in considering sample sizes at the extremes. For a known background incidence of 1 in 1,000 and an added incidence of 1 in 100, the sample size needed to observe at least one case of the added effect is only **500.** But when the background incidence is 1 in 10 and the added incidence is only 1 in 10,000, the sample size must be 98 million (table 7).

Table 7 also illustrates that another factor could increase the sample size that a study requires. If the background incidence is not known and has to be estimated through observation of control groups, the smaller the size of the control group, the larger the sample size of drug users must be for the same degree of confidence in the results (compare the sample sizes required for control groups equal to the treated group and for those five times its size).

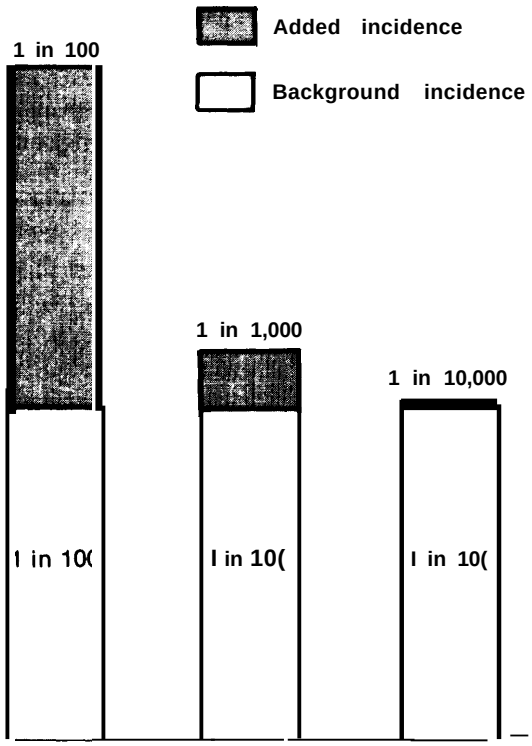
If there is a background incidence, the sample size needed to observe a drug-induced effect rises dramatically. Recall that a sample size of 3,000 was needed to be 95 percent sure of observing at least one drug-induced effect with a frequency of

Table 7.–Number of Patients Required in Drug-Treated Group To Detect One ADR With Background Incidence of Adverse Reaction (95% likelihood)

Size of control group	Background incidence of adverse reaction	Number of patients required with additional incidence of adverse reaction to drug		
		1 in 100	1 in 1,000	1 **in 10,000**
"Infinite" (background incidence known)	**1 in 10**	**10,000**	**980,000**	**98,000,000**
	1 in 100	**1,600**	**110,000**	**11,000,000**
	1 in 1,000	**500**	**16,000**	**1,100,000**
Five times as big as treated group	1 in 10	12,000	1,200,000	120,000,000
	1 **in 100**	**1,900**	**130,000**	**13,000,000**
	1 in 1,000	700	19,000	1,400,000
Equal to treated group	1 in 10	**20,000**	**2,000,000**	**200,000,000**
	1 in 100	**3,200**	**220,000**	**22,000,000**
	1 in 1,000	**1,300**	**32,000**	**2,300,000**

SOURCE J A Lewis, "PostMarketing Surveillance How Many, " *Trends in Pharmacological Sciences 2"93, 1981.*

Figure 2.—Comparison of Additional Drug-Induced Effects of Decreasing Incidence



SOURCE Office of Technology Assessment

1 in 1,000 and with no background incidence (table 6). If the background incidence is also 1 in 1,000, the sample size required rises to 16,000. A background incidence higher than the drug's added incidence increases the sample size required even more dramatically (to 110,000 and 980,000 for background incidence of 1 in 100 and 1 in 10 in table 7).

Many studies monitor for several effects, not just one, and some apparent drug effects may have occurred by chance. Minimizing these chance relationships also increases the sample size required. Table 8 illustrates the hypothetical case where 100 effects will be examined in one study, and the sample size reflects a 95 percent chance that the observed effects will be related to use of the drug.

These sample size illustrations reflect, at the 95 percent confidence level, that an effect will be observed and whether that effect can be attributed to drug use. They do not provide a good estimate of the added incidence of the effect from the drug (recall from table 6 that a sample size of 3,000 is needed just to detect at least one effect that has an incidence of 1 in 1,000).

Controlled clinical trials are used primarily for evaluating drug efficacy, not safety, because they are carried out on hundreds, or, at the most, a few thousand drug users. Their use for evaluating drugs already on-the market is also limited by their high cost and logistical problems, In fact, the use of controlled clinical trials for determining efficacy alone is already constrained by these two factors (9,46).

Table 8.—Number of Patients Required in Drug-Treated Group To Allow for Examination of 100 Adverse Reactions (950/' likelihood)

Size of control group	Background incidence of adverse reaction	Number of patients required with additional incidence of adverse reaction to drug		
		1 in 100	1 in 1,000	1 in 10,000
"Infinite" (background incidence known)	1 in 10	23,000	2,200,000	220,000,0%
	1 in 100	3,100	250,000	24,000,000
	1 in 1,000	800	32,000	2,500,000
Five times as big as treated group	1 in 10	27,000	2,600,000	260,000,000
	1 in 100	4,000	300,000	29,000,000
	1 in 1,000	1,300	40,000	3,000,000
Equal to treated group	1 in 10	46,000	4,400,000	440,000,000
	1 in 100	7,200	510,000	48,000,000
	1 in 1,000	2,900	73,000	5,100,000

SOURCE" J A Lewis, "Post-Marketing Surveillance How Many, "*Trends* m *Pharmacological Sciences 2:93, 1981.*

These limitations of controlled clinical trials in evaluating the safety of marketed drugs have led to relying on cohort and case-control methods for postmarketing studies. While these latter methods can only indicate an association between a drug and observed conditions, but not that the relationship is causal (49,77), the cumulative experience of multiple cohort and case-control studies that show consistent associations between a drug and a suspected effect can lead to a high degree of confidence that the relationship is causal.

The following three case studies illustrate the use of the four methods of drug evaluation: controlled clinical trials, voluntary reporting, cohort studies, and case-control studies.

CASE STUDIES IN DRUG EVALUATION

Oral contraceptives, the most investigated drugs in use, were first associated with cardiovascular disease in 1966 by British researchers. Earlier that year, they had noticed that Britain's yellow card system had revealed that relatively more reports of thromboembolism (heart attacks, strokes, deep vein thrombosis of the legs with or without emboli to the lungs) were associated with one type of oral contraceptive than with another. A case-control study later that year established that there was a significant association between oral contraceptive use and thromboembolism, but it found no significant difference between the two contraceptive types (38). Publicity over these adverse effects increased voluntary reporting, so that within 3 years more than 1,300 reports of thromboembolism had been received. By 1969, researchers established that it was the total dose of estrogens, not the different types of estrogen, that was responsible for any differences (39). As a result, high-dose estrogen oral contraceptives were removed from the market in **1970.** The association between oral contraceptives and thromboembolism has been confirmed in other studies (15,51, **52,73,74).**

It is now known that the risk of thromboembolic disease from oral contraceptives increases with age. Young women have a chance of 1 in 20,000 of dying from the cardiovascular effects of oral contraceptives (62); but a woman taking oral contraceptives from age 35 to 45 has a risk of death of about 1 in 1,000 over the 10-year period (32). The risk of heart attack from oral contraceptives has also been found to increase for patients who smoke cigarettes (52), so that older women who smoke have the greatest risk.

Other conditions associated with oral contraceptives include gall bladder disease (6, 7), liver tumors (l), yeast infections of the vagina (27), and increased bacteria in the urine (78). Oral contraceptives have also been associated with jaundice, but only in patients with a rare genetic condition in which there is impaired biliary excretion of bilirubin, the Dubin-Johnson syndrome (14). Decreased effectiveness of oral contraceptives has been associated with the use of anticonvulsants and other drugs (54), but oral contraceptives have also been associated with reduced incidence of nonmalignant breast disease **(8)** and ovarian cysts (58). Most, if not all, of these other conditions associated with oral contraceptive use have been observed in only a few studies, and the role of contraceptives in producing them is not so clear as it is in the case of the occurrence of cardiovascular conditions.

To illustrate that an association between a drug and an observed condition may be too casually inferred, it should be noted that in 1970, a Swedish physician reported Down's syndrome (or trisomy 21, the most obvious symptoms of which are mental deficiency and mongoloid features) in two children whose mothers had taken oral contraceptives before pregnancy. At the same time, one hospital reported increased numbers of children born with the syndrome. Sweden's monitoring system therefore compared the incidence of the syndrome during a period (1968-70) when oral contraceptives were widely used with earlier periods when oral contraceptives were not available. The incidence of the syndrome turned out to be lower in the period of oral contraceptive use (48). The public did not know of the investiga-

tion until it was completed. "One can easily imagine the worldwide alarm such a suspicion could have caused" (3).

Streptokinase, a drug derived from group C beta-hemolytic streptococci for use in dissolving blood clots, was released for marketing in the United States in November 1977, following its use in Europe for many years. Bleeding, allergic reactions, and fever had been found in the premarketing clinical trials carried out on 535 patients. Hoechst-Roussel Pharmaceuticals, manufacturers of the drug, implemented a postmarketing surveillance study at the request of FDA. The study lasted until May 1980, at which time FDA lifted the postmarketing study requirement.

Hoechst-Roussel included a reporting form in each package of the drug vials sent to hospital pharmacies, and requested that the pharmacist have treating physicians complete the forms. Based on sales, the company estimated that reports were returned for about 20 percent of treated patients, a total of 306 patients of 260 physicians in 44 States and the District of Columbia,

Physicians were asked to rate treatment outcomes as "successful, " "partially successful, " or "unsuccessful. " Thirty-nine percent were reported as "successful, " 32 percent as "partially successful, " and 30 percent as "unsuccessful" (percentages rounded to nearest point).

Fifty percent of the patients (153/306) had a total of 208 adverse reactions, with 49 patients reported to have 2 or 3. The reactions reported were the same as those seen in the clinical trials—allergic reactions, fever, and bleeding. Fifteen patients suffered severe bleeding, and 48 less severe bleeding. Three patients with severe bleeding died, and one had nerve damage secondary to bleeding (femoral hematoma). Most patients recovered from the drug reaction, but a total of 25 of the 306 patients died during or within a few weeks of therapy. The investigators concluded that no adverse reactions were discovered that were not already known from the premarketing clinical trials and that their incidence and severity were similar (71).

This postmarketing study is instructive for a number of reasons. First, it was necessary to select the sample from marketing sources and to rely on voluntary reporting, so there was a bias in patient selection, there were no controls, and the criteria for effectiveness were vague. The low reporting rate of 20 percent is also typical of voluntary methods. Second, the purpose of the study was "to determine whether the incidence and severity of adverse reactions with widespread use would differ from that seen during clinical trials. " The investigators concluded that incidence, severity, and type of reaction were similar, but noted that the final sample size was only 306, smaller than that of the 535 patients in the premarketing clinical trials. Adverse reactions with incidence less than those observed in the clinical trials were not likely to be discovered in the smaller postmarketing study. The strongest conclusion that can be reached on the basis of the later study is that the types of drug reactions in an uncontrolled patient population were similar to those of the more carefully selected patients in the clinical trials.

Finally, this study illustrates a problem that is often encountered in evaluating drugs. In contrast to patients taking oral contraceptives, these patients were being treated for underlying diseases, which can complicate the interpretation of what is observed. For example, 25 of the 306 patients died, 15 from pulmonary emboli and 4 from deep vein thrombosis, both conditions that are related to the diseases being treated, The investigators concluded that these deaths occurred about as frequently as they did during the premarketing trials.

Cimetidine, an anti-ulcer drug that blocks the release of stomach acid, was approved for marketing in August 1977. Because it was a new class of therapeutic agent and was expected to be widely used—it was an alternative to surgery in reducing stomach acidity—a postmarketing study was requested of the drug's sponsor—Smith, Kline & French Laboratories. The drug had become available a few years earlier in some other countries, including Great Britain.

The sponsor established a cohort of about 10,000 patients by using its sales representatives to enroll over 1,000 physicians who were asked to complete case report forms on at least **10** patients for whom they would prescribe cimetidine during a 3-month period starting in March 1978.

A second phase (not reported by early 1982) was to cover the following 6-month period. The expected 10,000 patient cohort would provide a 90-percent probability of observing a drug effect with an incidence of 1/4,348 and an 80-percent probability of observing one with an incidence of 1/6,250. The physician response rate was 85 percent, with 9,907 case reports used in the study. A control group was not included because of design and cost problems.

A **total** of 577 adverse reactions were reported in 442/9,907 (4.4 percent) cases. The investigators concluded that these adverse reactions did not differ in type or incidence from those observed in the premarketing controlled clinical trials (31).

The most common drug reactions observed in the patients were diarrhea and nausea or vomiting, Previous reports in the literature (i. e., spontaneous reporting) from Great Britain had postulated an association between cimetidine use and mental confusion, blood dyscrasias, and endocrine effects. Some patients did develop such symptoms, but the investigators concluded that only the 18 cases of gynecomastia (breast swelling in males) and one case of blood abnormality were related to cimetidine use.

There were 65 patient deaths during the 3-month review period, none considered to be cimetidine-related. Previous spontaneous reports had also led to postulating an association between cimetidine use and stomach cancer (19,34,60,70). However, the hypothesis was viewed with skepticism, as people who have ulcers also have a higher chance of developing stomach cancer. Though the investigators did not address this possible effect of cimetidine, they reported that the 19 patients with tumor-related deaths were diagnosed for corresponding conditions before cimetidine therapy began.

The year after this study was published, British researchers reported experiments in humans on the basis of which they speculated about how cimetidine could produce stomach cancer (59). The investigators pointed out that the toxicological data from long-term studies in rats and dogs did not uncover any evidence of the drug's carcinogenicity, but they contrasted these findings with their experimental findings and with the spontaneous reports associating cimetidine use and stomach cancer in humans.

In contrast to the streptokinase study's enrollment of a smaller cohort than the number of patients evaluated in the premarketing tests, the cimetidine postmarketing study was able to enroll its goal of a cohort of nearly 10,000 patients. The cimetidine postmarketing study, however, is of interest for other reasons.

The published report covering only the first 3 months of cimetidine use found no difference from the premarketing controlled clinical trials in the types or incidence of adverse reactions. One problem with postmarketing studies, however, is to detect adverse reactions without prejudice about what will be found. The findings of the premarketing clinical trials can provide guides for what reactions to anticipate, but such guides might cause the observer to overlook other adverse reactions associated with the drug under study.

In contrast to the published preliminary findings, FDA's review of this study, not yet complete as of early 1982, suggests a lower rate of adverse reactions than that observed in the premarketing trials, although the types of adverse reactions and their relative rates in the two are similar (10). This last observation is in agreement with at least one pharmaceutical company's experience (66):

> Some people might think postmarketing surveillance can sharpen up estimates of incidence of adverse reactions seen in phase III trials. Phase III trials, however, use trained investigators and they ordinarily are treating the more severe patients. Both of these facts contribute to higher adverse reaction rates in phase III studies than one expects to get and actually gets in postmarketing surveillance studies.

Thus, in designing postmarketing studies, the investigator must both be guided by the results of the premarketing clinical trials and not be overly influenced by them.