36

REGULATORY ISSUES

PAUL LEBER

For over 60 years, the United States has relied primarily on a federal system of premarket drug product clearance to ensure the quality of the nation's drug supply. When the premarket clearance system was first introduced in 1938 in the aftermath of the Elixir of Sulfanilamide tragedy in which over a hundred patients needlessly died because of a drug manufacturer's carelessness (1), federal law required only that new drugs be tested and shown, prior to marketing, to be "safe for use." Since 1962, however, the law requires that new drug products also be shown to be "effective in use" under the conditions of use recommended in their proposed labeling.

This chapter considers how these two fundamental requirements of the Federal Food, Drug and Cosmetic Act (FFDCA), our national drug regulatory law, are currently interpreted and applied by the Food and Drug Administration (FDA) in its evaluation of new drug products. Issues that are singularly important to the evaluation of products intended for use in the management of psychiatric conditions are identified and explicated.

THE US FEDERAL DRUG REGULATORY SYSTEM: LEGAL BASIS, STRUCTURE, AND MODE OF OPERATION

The basic structure and operation of the US federal drug regulatory system are established under the provisions of the FFDCA.

The FFDCA makes it unlawful " . . . to introduce or deliver for introduction into interstate commerce any new drug, unless an approval of an application . . . is effective with respect to such drug."

The application to which the Act refers is a New Drug Application (NDA). By law, authority to approve NDAs resides with the Secretary of the Department of Health and Human Services, but the Secretary delegates the actual authority to review and approve NDAs to the Food and Drug Administration (FDA, the agency).

Paul Leber: Neuro-Pharm Group, LLC, Potomac, Maryland.

The Act instructs anyone (i.e., a sponsor) seeking to market a new drug product to submit and gain FDA's approval of an NDA for the product prior to marketing it. Importantly, although it is not widely appreciated, NDAs are not approved for drug substances (i.e., chemical entities), *per se,* but for one or more specific "claimed" uses of a specific drug product (i.e., a specific formulation of the drug substance) under a specific set of conditions of use recommended (i.e., described) in the product's proposed labeling.

The Act describes, albeit in rather general terms, the information and reports that each NDA must contain. The details need not concern us, however. Suffice it to say that a sponsor's NDA is required to provide all the information necessary to allow the FDA to determine whether or not the drug product that is the subject of the application meets the standards set out in the Act for a lawfully marketed drug product. These standards address not only matters bearing on the product's safety and effectiveness in clinical use, but on its method of manufacture; chemical identity, purity, and strength; pharmaceutical performance; bioavailability; and proposed labeling.

The Act requires the FDA, in turn, to "file" (i.e., accept for review) any NDA that appears on initial inspection to provide, in a reasonably organized and coherent format, full reports of all the tests necessary to evaluate whether or not the drug product meets the requirements just cited.

Finally, the Act instructs the FDA to review and to approve a sponsor's NDA within 180 days of its submission, *unless,* on review, the agency determines that the reports it contains *fail* to establish that the drug product identified in the application fully complies with the Act's requirements.

THE INVESTIGATIONAL NEW DRUG APPLICATION

Because the Act forbids the introduction into interstate commerce of new drugs unless they are the subjects of an approved NDA, the lawful clinical testing of unapproved new drug products would be a practical impossibility if the Act did not provide for an exemption to this ban.

The original FFDCA (1938), accordingly, provided for

precisely such an exemption, known then as a "Notice of Claimed Investigational Exemption for a New Drug." The exemption is still available, but it is now officially known as an Investigational New Drug (IND) application.

Initially, an investigational exemption could be obtained largely for the asking. Between 1938 and 1963, the sponsor of an IND had only to agree to keep records and clearly label its new drug as to its status as an unapproved investigational new drug, but little else.

With the passage of the Kefauver and Harris amendments of 1962, however, the IND requirements were extensively revised and expanded. Congress was led to alter the requirements for investigational use because of yet another public health disaster involving a drug product. In this case, fortunately, the drug thalidomide, although widely marketed in Europe, was not marketed in the United States. However, the potent teratogen thalidomide was widely distributed under INDs in the United States; worse, when its teratogenicity was recognized and efforts were undertaken to recall the supplies of it that had been distributed, the extent of domestic distribution was not easily determined.

Although very few American women who had received thalidomide under an IND bore children with limb reduction defects, the episode raised substantial concerns about the safety of human research subjects (2,3). Thus, Congress amended the Act so as to give the FDA the authority to monitor and control the conduct of clinical drug research within the United States.

Under the 1962 amendments, the agency gained explicit authority not only to establish mandatory prerequisites for the granting of INDs, but also the power to prevent the initiation and/or suspend the conduct of a clinical investigation (i.e., impose a "clinical hold") being carried out under an IND if and when the agency concludes that an investigation poses an unreasonable or unnecessary risk to human subjects. In 1997, with the passage of the Food and Drug Modernization Act (FDAMA), FDA's authority under the Act's IND provisions was clarified and explicated in more detail, but not substantively modified.

In sum, since 1962, the IND serves not only as a license sponsors must obtain to allow them lawfully to ship unapproved new drugs in interstate commerce, but also the device through which the agency monitors and maintains control over the way in which clinical research with new drugs is conducted within the United States.

WHAT AN APPROVABLE NDA MUST DEMONSTRATE

The text of the Act, at least insofar as safety and effectiveness standards are concerned, speaks almost entirely to broad goals and generic principles. Responsibility for interpreting the Act and developing, revising, and promulgating the regulations and policies necessary to secure the aims Congress had in mind in drafting the Act are delegated to the FDA.

Safety

Insofar as safety is concerned, the Act demands that a sponsor provide full reports of all tests necessary to establish that its product will be safe for use. The Act instructs the agency to reject a sponsor's application, if, on review, it determines that the drug has been inadequately tested, or, if tested adequately, the findings of the tests conducted are inadequate to show that the drug, as recommended for use, is safe for use, or show that the drug, again as recommended for use, is unsafe for use.

Efficacy

The Act instructs the FDA to approve an NDA *unless,* on review of the reports submitted, it concludes there is a lack of "substantial evidence" that the drug is effective as claimed in its proposed labeling.

HOW THE AGENCY INTERPRETS THE ACT'S GENERIC REQUIREMENTS

Safety for Use

No pharmacologically active drug substance is ever likely to be entirely free of risk. Accordingly, the agency maintains that a regulatory determination that a drug is "safe for use" is, in actuality, a favorable "risk-benefit" determination (i.e., a conclusion, based on the information evaluated, that the risks imposed by the use of the drug are acceptable in light of the benefits it provides).

Risk-benefit assessments, however, are hardly straightforward undertakings. To begin, their reliability is in large part a function of the extent and quality of the information on which they are based. Unfortunately, the information ordinarily available to inform a regulatory risk-benefit assessment is limited in scope; a typical NDA, for example, is approved based on experience gained with a drug product in perhaps 1,000 to 2,000 human subjects *in toto.*

The information that is available, moreover, is in many respects marginal in regard to its aptness; the individuals who participate in drug development programs, although they are reasonably representative of the patient population to which the drug will be administered when marketed, are not fully representative of it. Accordingly, risks uniquely, or more likely to be, associated with the use of a drug in various subgroups of the population, particularly when those subgroups are rare and/or under-represented in the samples studied during the drug's development, are almost never appreciated, let alone factored into the regulatory risk-benefit determination.

Data bearing on the risks of a drug are collected during premarket testing under conditions of use (e.g., dose, regimen of administration, duration of use, restricted use of concomitant medications, etc.) that vary substantively from those under which a marketed drug is likely to be used. This is of especial concern where duration of use is concerned.

Ordinarily, the bulk of premarketing data are obtained in relatively short-term clinical trials (weeks or months) although the product under development typically will be used, once marketed, over much longer intervals (months to years). As a consequence, a typical drug development program has little, if any, chance of detecting untoward effects of a drug that emerge only after an extended period of exposure.

Regulatory officials are well aware of these limitations, but for practical (i.e., a politic word for economic and political) reasons, must tolerate them. The International Conference on Harmonization (ICH) in which the US participates, for example, has issued a guideline (4) that states that it is ordinarily sufficient to evaluate a new drug, prior to marketing, in no more than 300 to 600 patients for 6 months and 100 patients for a year.

Anyone familiar with the arithmetic of risk estimation will recognize that an experience of ''safe passage'' on a drug gained in such limited numbers of patients is not very reassuring. The failure to see even one catastrophic or fatal event in a sample of 300 patients only reduces to 5% the chance that the drug investigated causes such unobserved events at a rate no greater than in one of every 100 patients exposed to it (5).

(The notion of ''safe passage'' in drug safety assessment is the author's invention; it likens regulatory premarket clinical trials to journeys undertaken on uncharted waters. Much as early seafarers determined which of several routes between two points was safer by comparing the risks of one with another, society determines whether or not a new drug is safe for use from the proportion of patients exposed to it who enjoy safe passage.)

Whatever one's personal sense of the size of such a risk, it is truly enormous from a public health perspective. Just imagine the horror if a new antidepressant drug product caused a fatality in one in every thousand patients exposed to it. Yet, our society, like those of other developed nations, is seemingly content to market drugs without being able to confidently exclude a 10-fold greater risk.

In recent years, highly publicized withdrawals of new drugs shortly after their approval for marketing because of previously unrecognized risks of use have focused public attention on the limitations of premarket drug safety assessment. Both the agency and regulated industry have, consequently, been urged to develop new and better methods and paradigms to predict the likelihood of new drugs to cause injury (6).

No doubt, both the agency and regulated industry would be delighted to do so if they could. It is somewhat difficult to fathom, however, what kinds of methods will make it possible to identify drug-induced injuries in advance of their occurrence caused by pathogenetic mechanisms that have yet to be recognized, let alone characterized.

On the other hand, it is difficult to deny that the systematic study of a drug product's capacity to cause effects that have in the past been associated with an increased risk of untoward effects should sometimes be useful. The risk of pharmacokinetic interactions, for example, should be predictable if the major metabolic pathways involved in the elimination of a new drug, its metabolites, and the pathways of elimination of other drug products likely to be coadministered with the new drug are identified and adequately characterized. Knowledge of a drug product's metabolism also makes it possible to identify individuals in the population that might be at unique risk of suffering injury because of their diminished capacity to metabolize the drug (e.g., 6% to 8% of the white population are ''poor'' metabolizers of drugs that are CYP 450 2D6 substrates). Presumably, as our knowledge of the human genome expands, our ability to predict drug-induced risks on such grounds will grow.

Efforts to screen drugs prior to marketing for specific properties that predict drug-associated harms are still largely in their infancy, however. Moreover, such approaches have inherent limitations. Their utility is typically predicated on the assumption that the indicator of risk employed (e.g., a capacity to prolong the QTc interval on the surface ECG) reliably and consistently predicts a drug's capacity to cause harm. As is the case with almost all surrogate indicators, however, there is always a possibility that the association between the surrogate and harm found in one set of circumstances will not hold in another.

In contemplating the development of new approaches to premarket safety assessment, it is important to be mindful that many of our expectations may be unrealistic, even magical. Congress, for example, did not intend that premarket testing would successfully identify every unsafe or unfit drug product. If it had, it would not have authorized the agency (Section 505(e) of the FFDCA), to withdraw approval of any NDA for a drug if new information, not available at the time of approval, becomes available which shows the drug is unsafe for use as labeled. Moreover, it is evident that Congress not only anticipated that new risks would be recognized after a drug's approval for marketing, but expected that postmarketing surveillance would detect them. Specifically, to ensure that adverse information and reports bearing on the safety of marketed drug products would be collected and made available for evaluation, the Act (Section 505(k)) requires the sponsors of marketed drugs " . . . to establish and maintain . . . records, and make . . . reports to the Secretary, of data relating to clinical experience and other data or information, received or otherwise obtained by [the sponsor] . . . with respect to . . . [its drug] . . . to enable the Secretary to determine, or facilitate a determina-

tion, whether there is or may be ground for . . . [withdrawing approval of the NDA].''

In actuality, however, only a relatively small proportion of marketed drug products are withdrawn from the market on grounds of being unsafe. This fact is frequently overlooked in the midst of the sensationalist publicity and second-guessing that so often accompanies product withdrawals. In the vast majority of instances, in fact, after their evaluation, adverse reports received on a drug from postmarketing surveillance sources lead, at most, to revisions being made in the product's approved labeling.

The nature of the labeling changes made, and the publicity given to them, is a function of the severity, estimated frequency, potential reversibility, and likelihood of mitigation or avoidance of each newly appreciated risk. From a regulatory perspective, a labeling change is a sufficient legal remedy, even for relatively serious newly appreciated risks, provided it remains possible for the agency to sustain its earlier conclusion, albeit under the newly revised labeling, that the drug product is ''safe for use'' *as labeled.*

The foregoing discussion reveals just how subjective and tenuous drug safety assessments actually are. Indeed, even if the risks associated with the use of a drug product were known exhaustively and in detail, its risk-benefit assessment would likely remain arguable. Paradoxically, the crux of disputes about risk-benefit determinations less often concerns the seriousness of the harms a drug causes, than the value of the benefits its use provides.

There is good reason for this. Except in those few instances in medical therapeutics were the use of a drug can be shown to reduce the absolute incidence of death or a serious, otherwise irreversible, injury, a substantively meaningful numerical estimate of the value of the benefit provided by a drug lies beyond reach. This occurs because there is a substantive distinction between a drug with an effect and an effective drug treatment.

Given the proper choice of experimental design, an appropriate patient sample, and the use of validated outcome measures, a competently executed clinical experiment can, of course, generate a numerical estimate of the magnitude of a drug's effect on some assessment instrument or scale of measurement. What such a trial cannot do, at least in a way that can be understood in a public sense, is to provide a meaningful measure of the value of the drug's clinical benefit. Unfortunately, it is the clinical benefit provided, and *not* the numerical estimate of the magnitude of a drug's effect as measured on some rating instrument that must be considered in a substantively meaningful risk-benefit assessment.

Thus, in areas of medicine, like psychiatry, where the beneficial effects of a treatment are *not* ordinarily measured by counts of cures or numbers of actual deaths prevented, but in the degree of symptomatic relief afforded by treatment, it is ordinarily impossible to obtain a publicly meaningful, let alone quantitative, estimate of a drug's value.

It is to be acknowledged that there are statistical estimates of ''effect size'' available (7), but these are intended only to gauge the relative magnitudes of a measured effect and the natural variation of that measured effect among individuals in the population being investigated. Although such ''effect size'' estimates are necessary to make informed ''guesses'' about the numbers of patients that must be admitted to a controlled clinical trial to obtain a statistically significant result, they say nothing whatsoever about the value of the effect being measured.

In sum, although a regulatory risk-benefit determination is widely represented to be a reasonable and responsibly considered judgment that derives from a disinterested weighing of the gains and harms known to be associated with the use of a drug, it is much more accurately depicted as a gestalt informed by an inchoate process that mixes, in undetermined proportions, evidence, sentiment, and personal values.

Nonetheless, because the Act requires that marketed drugs be shown to be safe for use prior to marketing, the FDA must take the information and reports presented in a sponsor's NDA, and with the assistance of its consultants and advisors, determine whether a reasonably qualified and informed expert, in possession of the data that are available, could conclude, fairly and responsibly, that the drug is ''safe for use'' as labeled for use within the meaning of the Act.

Effectiveness in Use

Effectiveness determinations under the Act are, at least in comparison to those involving safety, relatively straightforward, provided, however, that a modicum of agreement exists within the community of qualified medical experts as to: (a) what constitutes a beneficial treatment effect in the area of therapeutics involved, and (b) how that therapeutic effect is to be measured.

Given agreement on these two matters, it is a relatively simple, although not always an easy or quick, task for a sponsor to procure the evidence necessary to meet the Act's substantial evidence standard, provided, of course, that the sponsor's product is truly an effective one. If controversies are extant within the community concerning the condition that is the target of treatment (e.g., its features, diagnostic boundaries, cardinal manifestations, etc.), as they often are in the field of psychiatry, the task of demonstrating a drug's effectiveness in use becomes considerably more complicated but still possible.

The key question that remains to be addressed in this section then, is what constitutes substantial evidence within the meaning of the Act.

Between 1962 and 1997, Section 505[d] of the Act offered the following definition of substantial evidence:

> . . . the term ''substantial evidence'' means evidence consisting of adequate and well-controlled investigations, including clini-

> cal investigations, by experts qualified by scientific training and experience to evaluate the effectiveness the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof.

A number of points about the agency's traditional interpretation of the substantial evidence standard are noteworthy.

The agency has long held that "substantial" evidence must derive, at least in part, from the findings of valid *clinical* experiments. This presumably reflects the view that the value of drugs intended for the treatment of a human disease or condition can only be evaluated meaningfully in tests conducted in human subjects actually afflicted with or at risk of developing that disease or condition.

Agency regulations make clear that for evidence to be deemed substantial, it must, in part, be adduced in scientifically valid experiments (i.e., adequate and well-controlled investigations). Inferences based on scientific theory alone, or on the findings of uncontrolled clinical studies or observations (e.g., case reports, case series, etc.) will not suffice (e.g., see 21 CFR 314.126 (e)).

Neither can clinical judgment or professional opinion, per se, contribute to the body of evidence required to meet the substantial evidence burden. In light of the importance seemingly given to expert opinion in the statutory definition of substantial evidence, this assertion may seem at odds with the Act's requirements. A careful reading of the statutory definition reveals, however, that substantial evidence does not include what experts believe. To the contrary, the reference to experts in the definition of substantial evidence serves only to describe the character of the evidence that can be deemed substantial. (The evidence must be of a kind, quality, and quantity that would allow a disinterested and informed *expert* to conclude, "fairly and responsibly," from the evidence that the drug will have the effect its sponsors claims it has.)

It is also important to be mindful that "substantial" is an arcane legal term. It is frequently, and incorrectly, taken to mean definitive, even compelling, but substantial does not have that connotation within the meaning of the Act. Authorities on the legislative history of the Act (8) note that Congress considered, but decided not to employ, a "preponderant" standard of evidence, choosing, instead, the legally less demanding "substantial" evidence standard. Congress concluded, evidently, that the public health would be better served if the Act allowed the marketing of a drug even if only a minority of qualified experts agreed that it had been shown to be effective. It may help, therefore, to think of substantial evidence as the quanta of evidence sufficient to persuade at least a sizable number, but by no means a majority, of disinterested experts that a drug has been shown to be effective as claimed.

Among all the agency's interpretations of the substantial evidence standard, none, perhaps, has caused more concern and enduring complaint than its determination that positive findings from *more than one* adequate and well-controlled clinical investigation are ordinarily required to establish a drug product's effectiveness in use.

Although the requirement that experimental findings be independently substantiated prior to their formal acceptance is fully consonant with common scientific practice and epistemological principle, the agency's interpretation of the statute in this fashion proved so controversial and so politically vexatious that it eventually led Congress (Food and Drug Administration Modernization Act of 1997 [FDAMA]) to add the following sentence to the Act's definition of substantial evidence.

> If the Secretary determines, based on relevant science, that data from one adequate and well-controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness, the Secretary may consider such data and evidence to constitute substantial evidence for purposes of the preceding sentence.

FDA authorities contend (circa 1999 to 2000) that the revised definition does no more than confirm FDA's long standing authority (its regulatory discretion) to interpret the Act's efficacy provision flexibly and responsibly so as to secure the "aims of Congress" and advance the interests of the public health. In fact, before the Act was amended in 1997, the agency had, on occasion, taken exception to its traditional interpretation of the effectiveness standard, approving an NDA on the basis of findings from but a single adequate and well-controlled clinical study (9).

Agency spokespersons assert, further, that in all but highly unusual circumstances corroborating positive findings from more than one adequate and well-controlled clinical investigation will continue to represent the minimal quanta of evidence sufficient to satisfy the Act's substantial evidence requirement.

Evidently, because Congress failed to explain clearly what it intended by the terms "relevant science" and "confirmatory evidence" in drafting FDAMA, agency officials seemingly enjoy a degree of latitude as to whether and when to rely on the findings of a single controlled clinical investigation. Whether FDA officials will elect to or be able to retain this flexibility in the long run is far from certain, however.

THE EVALUATION OF DRUG PRODUCTS INTENDED FOR USE IN THE MANAGEMENT OF PSYCHIATRIC CONDITIONS

The Division of Neuropharmacological Drug Products (DNDP) is the organizational unit within the FDA's Center for Drug Evaluation and Research (CDER) that is responsible for the monitoring of INDs and the evaluation of NDAs for psychiatric drug products.

It is the Division's obligation to review an application, determine whether or not it meets the requirements of the FFDCA as interpreted under FDA's prevailing regulations and policies, and, based on that review, take appropriate regulatory action.

For IND issues, signatory authority on most matters is at the level of the Division Director. Signatory authority is divided where NDAs are concerned. NDAs for new chemical entities (NCEs) are approved or disapproved at the Office level (Office of Drug Evaluation 1); supplemental NDAs (those involving new claimed uses of already marketed drug products) are approved at the Division level.

It is important to be mindful that Divisional policies evolve over time. A reader who wishes to understand whether, and if so, how a particular policy applies to a specific problem or drug product would be prudent, therefore, to seek fresh guidance on the subject directly from an appropriate Division representative.

From a regulatory perspective, the fact that a drug product is intended for the management of a psychiatric, sign, symptom, condition, or disease presents no unique problem.

A drug claim can be advanced for virtually any effect on the structure or function of the body of humans, for the cure or management of a disease or condition, or relief of a sign or symptom. The Act makes no distinction between the value of symptomatic treatments and those that are advanced as cures for a disease. Almost any claim can be made, provided that it can be presented in product labeling in a way that does not make the product's labeling, "false or misleading" in any particular.

In psychiatry, as in most other therapeutic areas, the effectiveness of a new chemical entity will almost always have to be demonstrated in *more than one* adequate and well controlled clinical investigation. It is possible, however, when an application is submitted for a claim closely related to one for which the drug product is already marketed, that a single controlled clinical study with robust and internally consistent findings might suffice. Because reliance on a single study is an exception to ordinary practice, however, the decision whether or not to take this approach will invariably be made on an ad hoc basis.

The agency's regulations (21 CFR 314.126) enumerate five control conditions that may be suitable for the evaluation of the effectiveness of new drug products. In light of the variability in course and outcome among samples of patients assigned the same psychiatric diagnosis, it is highly unlikely that either the historic control or the no control designs would ever be deemed acceptable for the evaluation of a drug for a psychiatric indication.

A randomized controlled and blinded trial employing any one or combination of the three other enumerated control conditions (placebo, graded dose, and standard active drug), provided it produces a statistically significant ($P \leq 0.05$, two-tail) difference favoring the investigational drug over a control condition, will invariably serve as one source of evidence contributing toward the quanta required to establish substantial evidence of a psychiatric drug product's effectiveness in use.

Among all the designs conforming to the requirements enumerated in 21 CFR 314.126, however, one that includes both a placebo and a standard drug control is especially attractive. Not only does such a design allow an estimate of what might have been had no treatment been administered (i.e., the response among placebo-assigned subjects) (10), but it provides a test of the capacity of the sample of patients participating in the experiment to respond to a treatment of established effectiveness (i.e., the response among patients randomized to the active control)

The three- (or more) arm design just described provides an internal means to assess what Modell and Houde (11) describe as an experiment's "assay sensitivity," its capacity to discriminate an inert substance from placebo and one level of an active drug substance from another.

Knowledge of an experiment's "assay sensitivity," or lack thereof, is especially important in the evaluation of psychiatric drug products because a sizable proportion of psychiatric drug trials (e.g., close to 50% or so of antidepressant trials) (12 –14) fail to discriminate between active and inert treatments. Obviously, the trial's failure can be discounted if the failure of a study to find a drug placebo difference is owing to the inability of the sample of patients randomized in the study to respond to drug. On the other hand, if the sample of patients enrolled can respond to standard treatment, but not the investigational drug, the trial must be viewed as a source of evidence that speaks against the effectiveness of the investigational drug.

Why the agency encourages the use of an active control arm in a placebo-controlled study, incidentally, is often misunderstood. It bears emphasis that the active control is not included to obtain an estimate of the standard drug's performance relative to that of the investigational drug, but solely to gauge whether or not the experiment has "assay sensitivity."

When an investigational drug appears to have a relatively low therapeutic ratio, the use of fixed-dose graded designs is advantageous because it provides the surest means to identify the dose or dose range most likely to be acceptable for a typical patient. The agency is mindful, of course, that a single dose is unlikely to be the best choice for all patients; nonetheless, for dose evaluation purposes, a fixed-dose design is more likely to be interpretable than a dose titration design. Clinicians often find this assertion counterintuitive, but designs allowing up-titration for therapeutic nonresponse commonly produce an inverted dose-response relationship (i.e., treatment resistant subjects who show poor response are given the highest doses of drug).

Another vexing issue regularly confronting regulators is how best to extrapolate the results of clinical investigations of new psychiatric drug products to labeling claims.

Although the agency's regulations require that the sample of patients evaluated in controlled effectiveness trials be "reasonably representative" of the population of patients for which a drug will be recommended for use, the patients recruited in typical commercially sponsored drug trials are never truly representative of the population, at least not in any formal statistical sampling sense. To the contrary, the choice of patient subject is almost always based on the sponsor's convenience and its desire to maximize the statistical efficiency of its study (15). This sampling strategy is not objectionable from a regulatory perspective when the primary goal of a clinical study is to establish that the investigational drug product has the effect claimed for it in at least some patients with the condition for which the treatment will be marketed.

What claimed uses should be granted to a sponsor of a drug based on the results of such studies is yet another arguable matter. During the author's tenure as DNDP's director, for example, clinical trials conducted in acutely exacerbated schizophrenic patients were deemed sufficient to support a claim for the use of a neuroleptic drug product in the "management of the manifestations of psychotic disorders." In a good faith attempt to adhere to the Act's requirement that product labeling not be false or misleading in any particular, the text of the Indications and Usage section also briefly described the clinical investigations that supported the "antipsychotic" claim, including the nature of the patient samples that had been employed in them and whether or not long-term maintenance trials had been conducted. The strategy employed was intended to reserve for practitioners the right to determine the extent to which the results of a sponsor's effectiveness trials applied to psychotic conditions other than schizophrenia. DNDP's current leadership takes a different view. At DNDP's annual morning session at the NIMH's annual NCDEU meeting (June 2, 2000) held in Boca Raton, Florida, Tom Laughren (group leader for DNDP's Psychopharmacology Unit) announced that henceforth claimed uses for psychiatric drug products would be more narrowly defined (i.e., ordinarily limited to the population of patients actually studied). Thus, drugs shown to be effective in studies enrolling schizophrenic patients will get claims for use in schizophrenia, not psychosis. Incidentally, the newly announced approach to product labeling is perfectly reasonable and certainly consistent with the requirements of law, although it is obviously not the one that the author prefers.

Difficulties arise in the extrapolation of study results in many other areas as well. Drugs are administered to individuals, not diagnoses. There is advantage in knowing, therefore, whether, and if so how and to what extent, various individual patient characteristics (sex, age, race, severity of illness, etc.) and the interactions among them affect response to a drug. Offsetting the interest in obtaining the data necessary to address these issues are the difficulties and costs (both time and money) encountered in obtaining them; recruiting representative patients from even the more important of the patient subpopulations (e.g., children, the elderly, the very ill, etc.) is often exceedingly difficult, and sometimes, just not feasible.

For decades, the FDA took a relatively passive stance in regard to the demands it made on the regulated industry for data that might better inform the use of prescription drug products in children. Groups interested in the welfare of children have lobbied long and hard for making the study of investigational drugs in children a premarket obligation. FDAMA attempted to encourage sponsors to conduct remedial pediatric studies by offering the incentive of a 6-month extension on the patent life of certain drug products. In 1998, however, the agency issued new regulations (21 CFR 314.55 and 21 CFR 201.23), asserting its authority to require sponsors to evaluate new drugs in clinical trials enrolling subjects of pediatric age.

It seems unlikely to the author that the agency would actually refuse to approve an otherwise safe and effective new psychiatric drug product on the grounds that the effects of the drug had not been adequately studied in children, unless, of course, the primary use of the product was likely to be in children (i.e., the drug is intended as a treatment for ADHD).

The fact that the official psychiatric nosology is continuously evolving further complicates the extrapolation of study results to product labeling claims. For example, prior to the promulgation of DSM-III in 1980, sedatives, as they were then known, were granted broad and nonspecific claims for anxiety, anxiety neurosis, etc. As a result of DSM-III, a distinction had to be made between generalized anxiety disorder and panic disorder. In recent years, claims have been further expanded to include not only long-established entities such as obsessive–compulsive disorder (OCD), but also new entities such as social anxiety.

On each occasion that the psychiatric nosology expands or changes, the agency is confronted with a set of new problems including not only how to evaluate claims for newly created entities, but how to conform existing drug product claims to fit with the revised nosology. In dealing with these issues, the agency has to consider whether or not already approved older claims subsume the new entities. (Is a new claim simply a re-expression of a previous one, a claim for a subset of the patients covered by the previously approved claimed use, or an entirely new claim for a previously nonexistent entity?)

The agency also must decide how broad a claim or set of related claims may be. For example, sponsors often seek to define a claimed use in a way that will allow the unique promotion of their drug product (i.e., distinguish their drug from others within the same therapeutic class). When a claim links some diagnostic entity or subtype of entity to a drug's effect and that linkage is irrelevant to the expression of the effect, the claim is considered, "pseudospecific." (The author early in his tenure as DNDP director coined the

term ''pseudospecificity''; it was first applied in connection with claims advanced for the use of benzodiazepines in anxious patients suffering from specific medical conditions [e.g., the anxiety of heart disease, cancer, etc.].) Such claims, moreover, are misleading because they seek to promote a distinction without meaning; consequently, they can be rejected by the agency because they can be held to be a violation of the Act's requirement that a product's labeling not be false or misleading in any particular.

A claim that a marketed antibiotic is effective for the pneumonia of dementia, for example, even if based on empirical evidence that the drug is effective in curing pneumonias in patients with dementia, is pseudospecific because the linkage between the pneumonia and the diagnosis of the patients treated is of no pharmacologic or biological importance, existing solely because of the sponsor's decision to select demented patients with pneumonia as subjects for study. A legitimate (i.e., nonpseudospecific) disease-related claim requires a demonstration that the effect of the drug is in some way conditioned on the presence of the diagnosis (i.e., the diagnosis of the disease controls to what extent, if any, the effect of the drug is expressed).

The distinction made by the Division between pseudospecific and legitimate disease-related claims has often proved to be both unpopular and a source of continuing controversy. Claims advanced by sponsors for the use of marketed antipsychotic drug products in the management of psychotic demented patients are a case in point. During the author's tenure as Director of DNDP, these claims were regularly deemed to be pseudospecific, in the absence of evidence to prove the contrary.

Based on testimony and discussion at a Psychiatric Drug Products Advisory Committee (PDAC) held in March, 2000 on treatments for the management of behavior in dementia, however, it appears that the Division is now inclined to accept ''the psychosis of Alzheimer's disease'' as a bona fide entity for which drug product claims may be made.

The seeming reversal in the Division's prior position, as far as the author can determine, came about because the current membership of the PDAC endorsed the psychosis of Alzheimer's disease as an entity *sui generis.* Support for the existence of this entity derived primarily from the testimony of psychiatrists who treat demented patients as to what they believe is the true state of nature. The fact that a diagnostic algorithm for the capture of patients with the ''psychosis of Alzheimer's disease'' had been recently proposed, seemingly gave further support to the reality of the putative entity. Although there is little doubt that the algorithm endorsed by the PDAC does capture demented individuals who exhibit behaviors that can be deemed psychotic, the fact that it does so in no way establishes that there is, in nature, a unique psychosis of Alzheimer's disease. Indeed, what the discourse at March meeting of the PDAC revealed is that when a diagnostic nosology is based on a taxonomic system controlled by authoritative figures (i.e., ''opinion leaders'' in the terminology of marketing departments within the regulated industry), the existence of a diagnostic entity may owe more to politics than biology.

Incidentally, the taxonomic nature of our official psychiatric nosology not only complicates the task of drafting of accurate product labeling, but also contributes to the high failure rate of studies intended to document the effectiveness of psychiatric drug products.

It bears note that DSM-III was developed, at least in part, in the hope that phenotypic similarities among patients would reduce within diagnostic category genotypic variability and, thereby, make psychiatric diagnoses more predictive of course, outcome, and treatment response than were the diagnostic categories based on the pseudoexplanatory, dynamic systems employed in DSM-II (16).

Unfortunately, this goal of DSM-III has yet to be achieved, in part, perhaps, because the effort to improve psychiatric nosology has been confounded by issues and interests that have little to do directly with biologic classification (e.g., a focus on new entities that expand the size of a drug's potential market, or that advance a professional career, or a political interest, etc.).

Whatever the explanation may ultimately be, however, sample-to-sample variation in both drug and placebo response rates among samples of patients assigned identical psychiatric diagnoses documents that our current psychiatric nosology has poor predictive power, at least insofar as drug responsiveness is concerned. As a consequence, to demonstrate that their products are effective in use, sponsors have had little option but to conduct trials using relatively large sample sizes.

Some critics have complained that the FDA, by allowing sponsors to conduct these ''overpowered'' trials, has turned the drug regulatory system into an institution that certifies as valuable drug products that provide little, if any, more benefit than placebo (17). Perhaps there is some truth in these allegations, but, then again, who is to say what drug effect is so small as to be dismissible? As discussed, from a strictly regulatory perspective such criticisms are irrelevant. It is important to recall that Congress could have set out other and/or more demanding standards for establishing drug effectiveness, but it did not.

The agency does have an obligation under the law, however, to ensure that its effectiveness determinations are based on fair and reliable estimates of a drug's measured effects. Accordingly, a large part of the effort expended in the review of an NDA is devoted to excluding the possibility that bias, rather than a true drug effect, accounts for the positive study reports submitted in the application.

Great pains are thus taken to ensure that a sponsor's claims are supported by valid statistical analyses of outcome measures prospectively identified in each study's protocol. Each study report is scrutinized for lapses in the execution of a study that might have introduced bias into the estimates

of drug effect it provides. Reviewers search for evidence that the randomization process failed, the treatment mask was penetrated, subjects failed to comply with protocol requirements, etc. The extent, pattern, and timing of premature discontinuations are carefully analyzed in an effort to determine whether the censoring process has created a biased estimate of the drug's effect.

The immediate goal of these efforts is to ensure that the comparison between the investigational and control treatments can be made "ceteris paribus" (all other things being equal).

The overarching aim is to determine whether or not the evidence submitted meets the Act's substantial standard fairly and responsibly.

THE FUTURE OF DRUG REGULATION

These are, at once, the best of times and the worst of times for medical therapeutics. On one hand, we have access to more information about the function and body of humans than we have ever had; therefore, we have good reason to hold sanguine expectations for the future in regard to the discovery and development of effective new treatments. On the other hand, the armamentarium stands at increasing risk of being polluted by worthless, even unsafe drugs. Signs of this potential risk abound.

The public is gullible and uncritical where therapeutic claims are concerned, exhibiting not only unrealistic expectations for the countless potential treatments touted as promising by the medical establishment and the pharmaceutical industry, but a willingness to spend billions of dollars on unproved, perhaps unsafe, remedies and nostrums.

Although the FDA's substantive powers under the FFDCA to regulate the drug supply, even after the passage of FDAMA, have been constrained to only a relatively minor extent, FDA's mission has noticeably changed.

Today's agency is expected to work as a "partner" of the regulated industry in a joint effort to expedite the development of new and promising treatments, not merely to keep unfit drugs off the market. It is arguable whether such dual expectations truly advance the interests of the public health, however. As the former head of the FAA noted when she resigned in the wake the ValueJet crash in Florida's Everglades, it is difficult for a federal regulatory agency to serve simultaneously and well the interests of both the consumer and the industry it regulates.

Congress has limited FDA's authority in less obvious ways as well. An egregious and illuminating example of its behavior is the Dietary Supplement Health and Education Act (DSHEA) of 1994. DSHEA has, by fiat, simply removed whole classes of drugs from FDA's jurisdiction and oversight.

By what is as close to an Orwellian *1984*-like maneuver as one is likely to find, the Congress simply declared whole classes of drugs to be something else: nutriceuticals, botanicals, food supplements, etc. These substances escape the premarket drug clearance requirements of the FDDCA because they are deemed by DSHEA *not* to be drugs.

Some may consider the legerdemain of DSHEA to constitute a relatively minor threat to the public health. Vitamins, for example, have long escaped the premarket clearance requirements of the Act, and have not proved dangerous, except perhaps when some are taken at excessive doses. Subsumed within the ill-described mix of botanicals, nutriceuticals, food supplements, and other materials deemed to be "nondrugs" under DSHEA, however, are large numbers of incompletely characterized, pharmacologically active drug substances that have yet to be fully and reliably evaluated for their safety and efficacy.

The lack of data bearing on the safety and efficacy of these drug substances did not deter the Congressional Leadership, however; to the contrary, when Senator Orin Hatch, a sponsor of DSHEA, introduced the bill he asserted that

> . . . dietary supplements can help promote health and prevent certain diseases, a fact substantiated by an ever-growing body of scientific studies and other evidence. In my own state of Utah, healthy life-styles, coupled with common use of dietary supplements have made a real difference. Our state is one of the healthiest in the nation and we enjoy one of the lowest incidence rates for cancer and heart disease. In Utah, the use of herbs is a well-accepted practice that has passed from generation to generation (18).

What is chilling here is not Senator Hatch's espousal of his state's virtues and the benefits of herbs, but his seeming willingness to approach an important public health issue in his nonscientific manner. Even more alarming is the inference congressional passage of DSHEA allows about the general state of our society's respect for science and the scientific method. Evidently, a majority of our elected representatives believes the benefits of scientific progress are attainable without the use of sound scientific methods.

If the armamentarium is to be kept reasonably free of unsafe and ineffective drugs, the antidrug regulatory drift of our national politics must be reversed. The FDA has demonstrated that it can, with the limited authority it has under the FFDCA, serve as an effective guardian of the drug supply. It will continue to serve effectively in this role, however, only if it obtains the necessary political support.

Given the current antiregulatory political environment, and the unrealistic expectations of the body politic for near magical breakthroughs in therapeutics, that support will not be forthcoming without the full support of the medical, scientific, and health care communities.

REFERENCES

1. Jackson CO. Doctor Massengill's elixir. In: *Food and drug legislation in the new deal.* Princeton, NJ: Princeton University Press, 1970:151–174.
2. Silverman M, Lee PR. Drug efficacy: the case of the emperor's

old drugs. In: *Pills, profits and politics.* Berkeley: University of California Press, 1974:107–137.
3. Silverman M, Lee PR. The impossible dream: search for the absolutely safe drug. In: *Pills, profits and politics.* Berkeley: University of California Press, 1974:81–106.
4. Anonymous. The extent of population exposure to assess clinical safety: for drugs intended for long-term treatment of non-life-threatening conditions. *Fed Reg* 1995;60:11270.
5. Leber P. Postmarketing surveillance of adverse drug effects. In: Lieberman J, Kane J, eds. *Adverse effects of psychotropic drugs.* New York: Guilford Press, 1992:3–12.
6. Wood AAJ. Thrombotic thrombocytopenic purpura and Clopidogrel—a need for new approaches to drug safety. *NEJM* 2000; 342:1824–1826.
7. Cohen J. *Statistical power analysis for the behavioral sciences,* rev ed. Orlando, FL: Academic Press, 1977.
8. Hoffman JE. Administrative procedures of the FDA. In: Adams DG, Cooper FM, Kahan JS, eds. *Fundamentals of law and regulations,* vol. 2. Washington, DC: Food and Drug Law Institute, 1977:13–53.
9. Anonymous. *Guidance for industry: providing clinical evidence of effectiveness for human drugs and biological products.* Rockville, MD: Food and Drug Administration, 1998.
10. Leber P. The use of placebo control groups in the assessment of psychiatric drugs: an historical context. *Biol Psychol* 2000;47: 699–706.
11. Modell W, Houde RW. Factors influencing clinical evaluation of drugs with special reference to the double-blind technique. *JAMA* 1958;167:2190–2199.
12. Leber PD. Hazards of inference: the active control investigation. *Epilepsia* 1989;30:S57–S63.
13. Leber PD. Is there an alternative to the randomized controlled trial? *Psychopharmacol Bull* 1991;27:3–8.
14. Khan A, Warner H, Brown W. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the FDA database. *Arch Gen Psychiatry* 2000; 57:311–317.
15. Leber PD, Davis CS. Threats to the validity of clinical trials employing enrichment strategies for sample selection. *Controlled Clin Trials* 1998;19:178–187.
16. Wilson M. DSM-III and the transformation of American psychiatry: a history. *Am J Psychiatry* 1993;150:399–410.
17. Fisher S, Greenberg RP, eds. *From placebo to panacea: putting psychiatric drugs to the test.* New York: John Wiley, 1997.
18. Statement of Sen. Hatch on Introduction of S. 784/April 7, 1993, Appendix 1: Section H. In: Bass IS, Young AL. *Dietary Supplement Health and Education Act: a legislative history and analysis.* Washington, DC: Food and Drug Law Institute, 1996.

Neuropsychopharmacology: The Fifth Generation of Progress. Edited by Kenneth L. Davis, Dennis Charney, Joseph T. Coyle, and Charles Nemeroff. American College of Neuropsychopharmacology © 2002.