Title
Drug Evaluations: Type I vs. Type II Errors

Author
Intriligator, Michael D.

Publication Date
1996-05-01

Drug Evaluations: Type I vs. Type II Errors

by

Michael D. Intriligator

Professor of Economics, Political Science, and Policy Studies

University of California, Los Angeles

A Presentation to the UCLA Pharmaceutical Economics and Policy Seminar, May, 1996

ABSTRACT

Drug testing in the United States is currently biased toward the minimization of "Type I" error, that is, toward minimizing the chance of approving drugs that are unsafe or ineffective. This regulatory focus of the Food and Drug Administration (FDA) ignores the potential for committing the alternative "Type II" error, that is, the error of not approving drugs that are, in fact, safe and effective. Such Type II errors can result in the loss of significant benefits to society when the sale of drugs that are safe and effective is prohibited. The present drug approval system puts enormous stress on Type I errors and largely ignores Type II errors, thereby raising the cost of drug testing and delaying the availability of safe and effective drugs. A more balanced set of FDA drug approval standards, accounting for the consequences of both Type I and Type II errors, could result in better outcomes, as compared to the present system.

1. Introduction

Drug testing in the United States, as supervised by the Food and Drug Administration (FDA), has a definite focus and bias, stressing one type of error over another. This bias should be reconsidered so as to improve outcomes in terms of drug approvals.

FDA drug testing concentrates on "Type I error" as it is referred to in the statistics literature, the error of approving a drug that is unsafe or ineffective. At the same time, this testing largely ignores the implications of "Type II error," the opposite error of not approving a drug that is safe and effective. This paper considers the reasons for this bias, outlines the basic statistical methodology involved, and then proposes that FDA drug testing should consider both types of errors. The FDA would then be taking explicit account of both the tradeoff between safety and efficacy of drugs and the lost benefits of safe and effective drugs that were not introduced because of the stringent testing requirements it has imposed.

The general decision-theoretic framework presented below implies that current FDA practice is a very special case in terms of its implicit loss function, one that focuses on one type of error and ignores the other. It would appear better for society in terms of potential social benefits if the FDA were to use a more general framework, one involving a loss function that incorporates the benefits and losses of both types of errors. Instead of arbitrarily focusing on one type of loss, excluding consideration of the other, this more informed approach to decision making at the FDA would consider the expected losses from both types of errors. Such an approach to drug testing would provide an improvement in societal benefits. It would strike a balance between the two possible errors, taking explicit account of the consequences of not approving safe and effective drugs in terms of delays in their availability. In particular, an informed reexamination of both types of errors could reduce the social cost that excessively stringent standards with

regard to only one type of error have had on both increasing the cost of bringing a new drug to market and imposing social costs on those unable to obtain safe and effective drugs. (Peltzman, 1973a, 1973b, 1974 and McGuire, Nelson, and Spavins, 1975) This new approach would also reduce the high costs and long delays in the introduction of new drugs that have resulted in many fewer new drugs becoming available and at much higher cost. It would also have major implications in lowering the costs of clinical testing, which would be socially beneficial since some of these costs involve what constitutes a deadweight loss for society, with little social value. In particular, there is a type of diminishing returns to additional drug testing in that the last tens of millions of dollars spent have very little impact on decisions being made by the FDA, yet they impose additional costs on society.

2. The FDA Approach and its History

An important reason why the FDA has this focus and bias is concern over certain past cases in which a drug was approved that turned out to be unsafe, resulting in fears of a similar situation recurring. A leading case of this type was the Thalidomide tragedy of 1962. (Sherman and Strauss, 1986) The tranquilizer Thalidomide had been approved in Germany for use by pregnant women to counteract morning sickness, but the FDA staff member who was responsible expressed doubts about the safety of the drug because of reported side effects. As a result, she limited its use in the United States to (uncontrolled) testing in 3897 women of childbearing age. This limitation was indeed fortunate since the drug was later found to lead to fetal malformations and to the birth of severely deformed

children in the case of a small fraction of the women using it. Since it was limited to testing, its use in the U.S. led to only nine women out of the 3897 giving birth to malformed babies, as opposed to some 8,000 malformed babies born in Europe. Even though a potentially even more severe crisis had been averted, the public outcry after the Thalidomide tragedy led to demands for much greater caution in the approval of new drugs. The result was the Kefauver-Harris bill of 1962 that amended the Food, Drug, and Cosmetic act in a way that substantially enhanced the power of the FDA over new drug testing and marketing. The Kefauver-Harris bill required that manufacturers demonstrate not only safety but also therapeutic efficacy for a new drug to gain FDA approval. Partly as a result of the Thalidomide experience and partly as a result of this legislation, the FDA since 1962 has been trying most strenuously to avoid approval of a drug that would later be found to be unsafe. Fear of such an outcome has led to extreme risk aversion and very conservative behavior, including the requirement of massive clinical trials before the FDA will approve a new drug. Before this change, clinical drug testing was not tightly controlled, with sample design not subject to strict standards and with heavy reliance on physicians' testimonials as to the efficacy of a drug. By contrast, since 1962, the approval of new drugs has been subject to extremely stringent regulations that have contributed to substantial delays in drug approvals and to enormously higher drug testing costs. (Sarett, 1974)

The process of clinical testing of a new drug for FDA approval currently involves three phases. (Spriet and Simon, 1985) In Phase I

the new drug is administered to a small number of healthy volunteers (often prisoners) to test for absorption, metabolism, and possible toxicity in various dosages. This phase took a median of 15.5 months for a sample of 93 clinical trials initiated between 1970 and 1992, and it had an attrition rate of 25 percent. (DiMasi et. al., 1991) In Phase II, which sometimes overlaps with Phase I testing, the new drug is administered to a few and then dozens of persons with the disease to be treated. This phase took a median of 24.3 months between 1970 and 1992, and it had an attrition rate (conditional on passing through the prior stage) of 52 percent. Finally, in Phase III the new drug is administered in double-blind tests (the true drug not being revealed to either the patient or the person administering the treatment) to at least two large samples, sometimes numbering in the thousands, of persons with the disease to be treated. This phase took a median of 36.0 months between 1970 and 1992, and it had an attrition rate (again conditional on passing through the prior stage) of 36 percent. On average for this sample of new drugs, only 23 percent of the substances that entered Phase I testing between 1970 and 1992 ultimately gained FDA approval.

One important implication of these more stringent testing requirements is that there have been substantial increases in both the costs and the delays involved in bringing a new drug to market. The result has been a steep drop in the number of new drugs approved for use. A comparison of the number of new drug approvals prior to the 1962 law with those granted in its aftermath demonstrates the effects of the stringency of the new FDA

standards. Table 1 shows that the average number of new drug approvals each year fell significantly from 50 - 70 new chemical entities (NCEs) in the 1950s to the early 1960s to some 15 - 25 NCEs per year, starting in the late 1960s and continuing up to the recent period. (Grabowski, 1976; Grabowski and Vernon, 1977, 1978; Temin, 1980; and DiMasi et. al., 1991) Part of this decline can perhaps be attributed to other factors, such as the exhaustion of obvious opportunities for research, but some must be attributed to the tighter regulations on drug evaluations imposed by the FDA, as confirmed by international comparisons of drug approvals. The result has been a significant decline in the supply of new drugs and increases in the costs and the length of time for a new drug application to obtain approval. (DiMasi et. al., 1994)

3. The Basic Statistical Methodology

The problem of drug evaluations facing the FDA can be considered a special case of the general statistical problem of hypothesis testing. Here the FDA is testing the hypothesis that a new drug is safe and effective. In general, any statistical problem of hypothesis testing involves four basic elements: a set of data, a probability model that provides the mechanism that generates the observed data, a hypothesis of interest, and a decision to be made concerning this hypothesis on the basis of the data and the model. The problem is one of making a decision on the basis of the available data, and some degree of uncertainty must be tolerated about the accuracy of the decision. Given a null hypothesis of particular interest, which is that the assumption to be tested is correct, the

decision is usually referred to as "accepting" or, alternatively, "rejecting" the null hypothesis.

The concept of Type I and Type II errors was developed as part of the Neyman-Pearson theory of statistical inference concerning hypothesis testing. There is some hypothesis concerning the matter under investigation, the null hypothesis, which can be either true or false. The result of the statistical investigation is a conclusion on the basis of statistical data or other evidence that the hypothesis is true or false. Thus, there are four possible outcomes. Two of them involve correct decisions, namely those of accepting a true hypothesis and rejecting a false hypothesis. Considering the other two possible outcomes, one type of error is that of accepting a false hypothesis, referred to as a "Type I error." The final possible outcome is that of rejecting a true hypothesis, referred to as a "Type II error."

The Neyman-Pearson approach to statistical inference usually sets a "significance level," which is the probability of committing a Type II error, at some arbitrary value, such as .05 or .10, meaning a 5 or 10 percent chance of rejecting a true hypothesis. The probability of committing a Type I error is then minimized, reducing to as small a level as possible the likelihood of accepting a false hypothesis. There is, in general, an interaction between the two types of errors in that reducing the probability of one usually results in an increase in the likelihood of the other. Thus, there is a tradeoff between these two types of errors. Minimizing the chance of one type of error, given the chance of the other error, leads to an "optimal surface," where the decision maker cannot reduce the

probability of one error other than by increasing the probability of the other. This nonlinear optimal surface of "admissible outcomes" is shown in Figure 1. It is the analog of in statistics of Pareto optimality of outcomes in economics, for which it is impossible to improve one person's situation without making another person worse off. In this case of statistical inference, along the optimal surface it is impossible to reduce the probability of one type of error without increasing the probability of the other error. The problem facing the decision maker is then that of choosing a particular point on this optimal surface, where the slope of the optimal surface at any point is a measure of the tradeoff between the two types of errors. In terms of this nonlinear optimal surface, FDA practice since 1962 has moved up this curve in the direction of reducing to infinitesimal the likelihood of a Type I error. As a result, given the tradeoffs shown in this curve, FDA practice has greatly increased the likelihood of a Type II error.

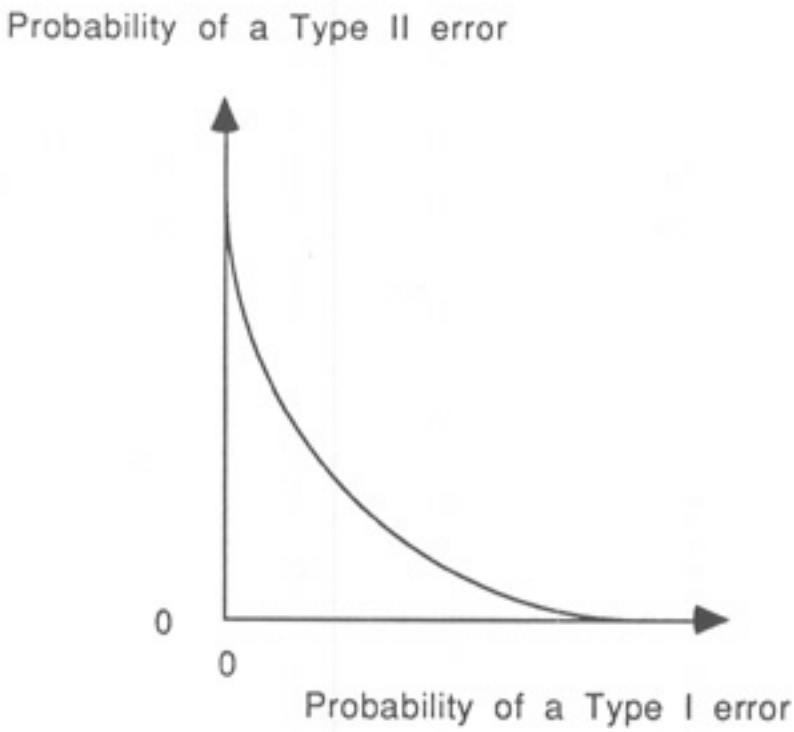

Figure 1: The Optimal Surface of "Admissible Outcomes"

This optimal surface holds cost constant, and it depends on the particular methodology used in testing. Thus, the optimal surface is endogenous and would shift if there were a change in the FDA drug testing methodology. Thus, more stringent FDA requirements, entailing a higher cost of testing, could result in a shift of this frontier toward the origin, reducing the probability of both types of error. Given a particular FDA drug testing methodology, however, there would exist such a surface, and there is then the issue of what point the FDA chooses on this surface. By contrast, as the methodology changes the entire frontier can shift. The difference between the two is analogous to the difference between a movement along a demand curve, as opposed to the shift in this curve.

The Neyman-Pearson approach deals with the interaction between the two types of errors on the optimal surface in a straightforward but arbitrary way. It sets one probability, that for a Type II error, at an arbitrary level and then minimizes the probability of a Type I error. In general there is a nonlinear tradeoff between the two types of error along the optimal surface and the problem facing a decision maker is that of choosing a particular point on this surface. The optimum point on this surface is where the marginal rate of substitution between the two errors equals their marginal rate of transformation, where the latter is measured as the slope of the optimal surface. As in the theory of the consumer, the optimal point is where an indifference curve is tangent to the budget set. In this case the slope of the indifference curve is the marginal rate of substitution and the budget set is the optimal surface itself.

4. Statistical Decision Theory and Its Application to Drug Testing

Statistical decision theory provides a useful generalization of the classical problem of hypothesis testing to allow explicitly for a loss function that takes account of both types of error, without arbitrarily specifying in advance the probability of one type of error as in the Neyman-Pearson approach. According to this decision-theoretic framework, the decision to accept or reject a particular hypothesis should depend on the losses due to both types of error, and it should be derived from the minimization of expected loss or, more generally, the maximization of expected net benefit. Current

FDA practice appears to be consistent with one special form of the loss function.

These concepts of statistical decision theory can be applied to the problem of drug testing. (See also Scherer, 1996) When the FDA faces the problem of approving a proposed new drug for use, it essentially is testing a type of null hypothesis that the drug is safe and effective. The alternative actions are to certify or not to certify the drug for use (or to continue testing), while the outcomes are that the drug is safe and effective or that it is not. Thus, the result must be one of four possible outcomes: accepting a true hypothesis, rejecting a false hypothesis, accepting a false hypothesis (Type I error) and rejecting a true hypothesis (Type II error). Approving a drug that is safe and effective and not approving a drug that is not are both correct decisions. The Type I error here is one of approving a drug that is unsafe or ineffective, while the Type II error here is one of not approving a drug that is safe and effective, thus denying or delaying its use. Thalidomide provides an example of a Type I error. There are several examples of drugs that were not approved by the FDA but that have been used successfully in other countries, providing examples of a Type II error. When drugs are introduced elsewhere, typically in Europe, before they are available in the U.S., the result is a "drug lag." An example from the 1970s was the drug propranolol, widely used abroad for management of angina pectoris and hypertension but not approved by the FDA for this application. Thus, some doctors who would have prescribed it were not able to do so. (Wardell, 1974, Peltzman, 1975) As another example, it has been claimed that if the beta blocker practolol had been available in the

1970s, it would have been able to save at least 10,000 lives per year. (Wardell, 1979, Temin, 1980) There are many cases of drugs that were available in Europe years before their approval in the U.S. (Wardell, 1978, Wardell and Lasagna, 1975) In recent years, according to the Pharmaceutical Research and Manufacturers of America, some 61 percent of new drugs and vaccines approved by the FDA between 1990 and 1994 were approved first in other countries, even though most originated in the U.S. According to the same source, medicines are approved, on average, two years earlier in Britain than in the U.S., and the average time for drug development and approval has increased from about 8 years in the 1960s to over 14 years in 1995.

The FDA sees the penalties from these two different types of errors as strongly asymmetric, given the strong expected criticism if it were to approve a Thalidomide-type drug that is unsafe, but given little or no criticism or even knowledge of the opposite error of not approving a drug that is safe and effective. As a result of this perceived asymmetry, current FDA practice since 1962 tends to focus on minimizing Type I error. It does so by carrying out enough stringent clinical trials to reduce the probability of approving a drug that is unsafe or ineffective to extremely small levels. Judging on the basis of the different phases of clinical trials and the few drugs that pass them, it appears that the FDA takes this responsibility to an extreme by trying to ensure that the probability of this error is minuscule. Its aversion to Type I errors is most likely due to the perception that the anticipated consequences of approving drugs that are later revealed to be harmful are much more severe than those for

not approving drugs that may benefit certain patients. Thus, the implicit FDA loss function has become essentially the likelihood of a Type I error, largely ignoring the Type II error of not approving a drug that is safe and effective.

As a result of the FDA practice of trying to ensure that the probability of a Type I error is essentially zero, clinical trials are extremely expensive and very time consuming, resulting in long delays in approvals and high costs. Certainly the FDA should continue to try to make Type I errors as small as possible, but it should recognize that setting an extremely stringent standard leads to long delays and very high costs for drug approval. One estimate of the cost per approved new NCE, where out-of-pocket costs are capitalized to the point of marketing approval, was $231 million in 1987 dollars (DiMasi et. al., 1991)

As a result of the FDA practice of virtually ignoring Type II error, many potential users of a safe and effective drug are forced to obtain it through other sources, including through foreign and illegal channels. Others are not able to obtain the drug at all, which means that this error imposes high costs on those who are not able to acquire those safe and effective drugs. The excessive concentration of the FDA on one type of error to the virtual exclusion of any consideration of the other stems perhaps from a type of failure of vision, focusing on identifiable individuals who might potentially be injured by drugs that are unsafe and ignoring hypothetical ones who might potentially be helped by drugs that are not approved. Type I errors harm identifiable persons, while Type II errors largely refer to hypothetical people in that they refer to the

likelihood that a drug not yet introduced will cure or help someone. As noted in other contexts, there is frequently a significant asymmetry in terms of actions when it involves a specific reduction in risk to a specific individual, such as preventing a Thalidomide baby, as opposed to a statistical one, such as providing a drug to some hypothetical person. Such an asymmetry calls for correction, however. Thus, as another example of this asymmetry, medical expenditures tend to concentrate on therapy to identifiable individuals as opposed to prevention for a group whose members may not be known and thus is largely hypothetical. (Viscusi, 1992; Pratt and Zeckhauser, 1996)

As one way of appreciating the nature of Type I vs. Type II errors, consider the reductio ad absurdum cases (where the logical consequences of a proposition lead to absurd conclusions). One clear way that the FDA could always reduce Type I error to zero is simply never to approve any drug for use. The result would be a zero probability of a Type I error, but the certainty of a Type II error. At the other extreme, if the FDA were to approve all drugs for use then there would be a zero probability of a Type II error, but the certainty of a Type I error. The FDA currently chooses a point on the optimal surface in Figure 1 that is close to the former extreme, where the probability of a Type I error is reduced to an extremely small value, the result being a high probability of a Type II error. Thus, the FDA approval process reduces to a level of insignificance the chance of Type I error and, at the same time, virtually ignores the chance of Type II error. While many would applaud the FDA striving to reduce the chance of authorizing the use of an unsafe drug, some

consideration should be given to those who could not obtain access to drugs due to the stringent FDA testing. Again, considering the reductio ad absurdum, few would applaud the FDA if they simply did not authorize any drugs for use, which would clearly avoid Type I error altogether but would represent an extreme policy that would prevent the use of all new drugs. The FDA policy today is, in fact, not very far from this extreme in that it sets such a very high standard as a result of its hypersensitivity to Type I error. The outcome is considerable Type II error and, in addition, huge costs and long delays in new drug approvals. A more balanced approach would take account of both types of errors and establish in the approval process a reasonable tradeoff between them.

5. A Decision-Theoretic Approach to Drug Testing

A decision-theoretic approach to drug testing could improve the overall performance of the FDA with regard to approving new drugs. This approach would allow explicitly for the benefits and losses that stem from avoiding or making both types of error and then maximize the expected benefit.

Before considering the more general framework of benefits and losses, consider the simpler case of only losses from the two types of errors. In the simplest case of four possible outcomes (where the action of continued testing is treated as not approving the new drug) the losses are shown in the loss matrix for drug approvals in Figure 2.

		FDA Action	
		Approve New Drug	Do Not Approve New Drug
State of Nature	Drug Safe	0	L_n
	Drug Not Safe	L_a	0

Figure 2: Loss Matrix for Drug Approvals

In this figure, the row "Drug Safe" refers to the state of nature where the drug is both safe and effective, while "Drug Not Safe" refers to the state of nature where it is not safe. The column "Approve New Drug" means the FDA action of approving the drug and "Do Not Approve New Drug" means the FDA action of not approving it. Figure 2 focuses on the losses, and it shows no loss for approving a safe drug and conversely, no loss for not approving a drug that is not safe. The loss stemming from approving a drug that is not safe, that is, stemming from a Type I error, is given in Figure 2 as L_a, while the loss stemming from not approving a drug that is safe, that stemming from a Type II error, is L_n.

Given this loss matrix, the expected loss of approving the drug is given as

(1) $EL_a = p_{a|n} L_a$,

where $p_{a|n}$ is the (conditional) probability of the Type I error of approving a drug given that it is not safe. Conversely, the expected loss of not approving the drug is given as

(2) $EL_n = p_{n|s} L_n$,

where $p_{n|s}$ is the probability of the Type II error of not approving a drug given that it is safe.

From the vantage point of the decision maker, the FDA should approve the drug if the expected loss from approving the drug is less than the expected loss from not approving the drug, i.e., if $EL_a < EL_n$, or if

(3) $p_{a|n} L_a < p_{n|s} L_n$.

This decision depends not only on the losses but also on the probabilities of the errors, which are related to one another as shown in the nonlinear optimal response surface of Figure 1.

Given this decision rule, if the FDA sets a very high value for L_a, the loss stemming from a Type I error, the minimizing expected loss requires the FDA to choose a very conservative course. It approves new drugs only if it believes that it has been clearly proven as a result of clinical trials that $p_{a|n}$ is essentially zero. Thus it gives approval only after it has convincing evidence that the probability of the Type I error of approving a drug given that it is unsafe is nil. Otherwise, given any positive value of the right-hand side of (3), no matter how small, the inequality would not be satisfied given the very high value the FDA attaches to the loss L_a. The FDA does not need to quantify the values of the losses, since, given a very high value for Thalidomide-type error L_a, it must drive the probability of such Type I error down to virtually zero.

A more general decision-theoretic framework would take explicit account of the benefits of not making an error as well as the losses stemming from the two possible errors. Benefits and

losses in the same case of four possible outcomes are shown in the payoff matrix for drug approvals in Figure 3:

		FDA Action	
		Approve New Drug	Do Not Approve New Drug
State of Nature	Drug Safe	B_a	L_n
	Drug Not Safe	L_a	B_n

Figure 3: Payoff Matrix for Drug Approvals

Here, considering the act of approving a new drug, the benefit stemming from approving a drug that is safe (and effective) is given as B_a while the benefit stemming from not approving a drug that is not safe is given as B_n. The losses are as in Figure 2.

Given the more general framework of Figure 3, the net expected benefit of approving the drug is given as

(4) $EB_a = p_{a|s} B_a - p_{a|n} L_a$,

where $p_{a|s}$ is the probability of approving a drug given that it is safe and $p_{a|n}$ is the probability of the Type I error of approving a drug that is unsafe. Conversely, the net expected benefit of not approving the drug is given as

(5) $EB_n = p_{n|n} B_n - p_{n|s} L_n$,

where $p_{n|s}$ is the probability of the Type II error of not approving a drug given that it is safe and $p_{n|n}$ is the probability of not approving a drug given that it is not safe.

From the vantage point of the decision maker, the FDA should approve the drug if the net expected benefit of approving the drug

exceeds the net expected benefit of not approving the drug, i.e., if $EB_a > EB_n$, or

(6) $p_{a|s} B_a - p_{a|n} L_a > p_{n|n} B_n - p_{n|s} L_n$.

This general formulation accounts for all possibilities in this simple decision-theoretic framework, explicitly treating not only the losses from both Type I error and Type II error as in Figure 2, but also the benefits from correct actions and the probabilities of being in each of the four possible cases. This formulation allows for differential weighting of Type I and Type II errors, as indicated in the values attached to the loss functions. Similarly, it can allow for differential weights attached to the two benefit outcomes stemming from decisions that are correct. Current FDA practice can be viewed as a special asymmetric case of this more general framework in which both L_n and B_n are ignored. In this special case, both the loss from not approving a drug that is safe, one stemming from a Type II error, and the benefit stemming from not approving a drug that is unsafe are treated as negligible and therefore ignored. The decision rule in (6) then becomes

(7) $p_{a|s} B_a > p_{a|n} L_a$,

which includes only terms involving the approval decision. With these simplifications of the general decision rule and, implicitly, a very high value for L_a, the loss stemming from a Type I error, the FDA then chooses a very conservative course. It approves new drugs only after it believes it has been proven beyond any doubt as a result of clinical trials that $p_{a|n}$, the probability of the Type I error of approving a drug given that it is unsafe, is reduced to essentially zero. Then the right-hand side of (7) becomes zero, so any efficacy

of the drug, as measured by $p_{a|s}\, B_a$, would justify the decision to approve the new drug.

Current clinical testing of a new drug for FDA approval tends to focus, as the FDA does, on potential dangers of approving the drug and the resulting Type I error. The FDA might be able to obtain some information on the potential dangers of not approving the drug by considering control populations of persons with the disease to be treated who are not given access to the drug, leading to Type II error. It would then have some information from this expanded clinical testing that it could use in the more general decision-theoretic framework, as expressed in (6). It is also possible to use a higher weighting for Type I errors than for Type II errors while giving at least some consideration to the latter.

The inequality in (6) provides a useful framework in which to study drug evaluations. Since all the expressions in this inequality are themselves subject to some uncertainty, a useful further step would be to estimate their distributions and to express the inequality as a probability statement. Thus, the benefits and losses would be stated in probabilistic terms, rather than discrete terms. Using this framework, if the probability that the inequality is met is high with a high degree of confidence then the drug should be approved, whereas if it is low with a high degree of confidence then the drug should be rejected. In the middle range of this probability or where the degree of confidence is not high then it would be reasonable to conclude that testing should continue. This approach can also be used to take account of different possible time paths involved in the terms in (6), accounting for the possible future

approval of a drug not currently approved through the present discounted value of the future streams of benefits and losses.

The overall implication of this decision-theoretic approach to drug testing is that the FDA should be using a broader loss function that takes account of both types of errors in order to make more informed decisions on drug approvals.

6. Conclusion and Policy Implications

This paper gives an interpretation for current FDA practice in drug evaluations in terms of statistical decision theory. According to this interpretation, the FDA has chosen a very conservative course, approving a new drug for use only after it has clearly proven that the chance of approving a drug that is unsafe has been reduced to essentially zero. The results of this practice of setting very high hurdles for demonstrating safety and efficacy have been long delays in the process of drug testing; very high costs in new drug approvals; the failure to approve many drugs that are both safe and effective, with consequent suffering on the part of those who are unable to obtain such drugs; and high costs of research as drug manufacturers must establish the safety and efficacy of new drugs at an unnecessarily high level of confidence.

A policy conclusion of this interpretation is that the FDA should reevaluate its policies so as to take explicit account of both Type I and Type II errors and their resulting losses as well as taking account of potential benefits of approvals. It should reexamine Type II errors and consider explicitly the consequences of not approving safe and effective drugs. It could expand its current clinical testing to provide the information needed to make more informed decisions

on drug approvals, taking account of both types of errors. With this information, and while still strictly controlling the testing process, it should maximize the expected net benefits stemming from its decisions, taking explicit account of all relevant losses and benefits. Such a decision rule would be a better one for society in terms of expected net benefits than the short-sighted one currently employed by the FDA that ignores certain relevant losses and benefits.

References

DiMasi, Joseph A., Ronald W. Hansen, Henry G. Grabowski, and Louis Lasagna, "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics*, 10: 107-142, 1991.

DiMasi, Joseph A., Mark Seibring, and Louis Lasagna, "New Drug Development in the United States from 1963 to 1992," *Clinical Pharmacology & Therapeutics,* 55:, June 1994.

Grabowski, Henry G., *Drug Regulation and Innovation: Empirical Evidence and Policy Options,* Washington, DC: American Enterprise Institute, 1976.

Grabowski, Henry G. and John M. Vernon, "Consumer Protection Regulation in Ethical Drugs," *American Economic Review*, 67: 359-364, February 1977.

Grabowski, Henry G. and John M. Vernon, "Consumer Product Safety Regulation," *American Economic Review*, 68: 284-289, May 1978.

Grabowski, Henry G. and John M. Vernon, *The Regulation of Pharmaceuticals: Balancing the Benefits and Risks,* Washington, DC: The American Enterprise Institute for Public Policy Research, 1983.

McGuire, Thomas, Richard Nelson, and Thomas Spavins, "'An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments': Comment," *Journal of Political Economy,* 83: 655-662, 1975.

Office of Technology Assessment, *Pharmaceutical R and D: Costs, Risks, and Rewards*, Washington, DC, 1993.

Peltzman, Sam, "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments," *Journal of Political Economy,* 81: 1049-1091, 1973a.

Peltzman, Sam, "The Benefits and Costs of New Drug Regulations," in Richard Landau, Ed., *Regulating New Drugs,* Chicago: University of Chicago Center Policy Study, 1973b.

Peltzman, Sam, *Regulation of Pharmaceutical Innovation: The 1962 Amendments,* Washington, DC: The American Enterprise Institute for Public Policy Research, 1974.

Peltzman, Sam, "'An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments': A Reply," *Journal of Political Economy,*, 83: 663-667, 1975.

Pratt, John W. and Richard J. Zeckhauser, " Willingness to Pay and the Distribution of Risk and Wealth," *Journal of Political Economy,*, 104: 747-763, 1996.

Sarett, L. H., "FDA Regulations and their Influence on Future R and D," *Research Management,* 27: 18-20, March 1974.

Scherer, F. M., "Pharmaceuticals," Chapter 9 in *Industry Structure, Strategy, and Public Policy,* New York: HarperCollins, 1996.

Sherman, Max and Steven Strauss, "Thalidomide: A Twenty-Five Year Perspective," *Food Drug Cosmetic Law Journal,* 41: 458-466, 1986.

Spriet, Alain and Pierre Simon, *Methodology of Clinical Drug Trials,* Basel: Karger, 1985.

Temin, Peter, *Taking Your Medicine: Drug Regulation in the United States,* Cambridge: Harvard University Press, 1980.

Viscusi, W. Kip, *Fatal Tradeoffs: Public and Private Responsibilities for Risk,* New York: Oxford University Press, 1992.

Wardell, W., "Therapeutic Implications of the Drug Lag," *Clinical Pharmacology and Therapeutics,* 15: 73-96, January 1974.

Wardell, W., *Controlling the Use of Therapeutic Drugs; An International Comparison,* Washington, DC: The American Enterprise Institute for Public Policy Research, 1978.

Wardell, W., "Rx: More Regulation or Better Therapies?," *Regulation,* 3: 25-33, September-October 1979.

Wardell, W. and Louis Lasagna, *Regulation and Drug Development,* Washington, DC: The American Enterprise Institute for Public Policy Research, 1975.

Note: I would like to acknowledge the valuable suggestions of William Comanor and the research assistance of Tony Kim and Umar Latif.

Table 1: Number of New Chemical Entities (NCEs) Approved in the United States, 1950-1989

Year	Number	Year	Number
1950	33	1970	16
1951	39	1971	14
1952	40	1972	11
1953	54	1973	19
1954	43	1974	18
1955	40	1975	16
1956	47	1976	15
1957	54	1977	18
1958	49	1978	23
1959	66	1979	13
1960	51	1980	11
1961	48	1981	23
1962	29	1982	22
1963	16	1983	12
1964	18	1984	21
1965	23	1985	26
1966	12	1986	20
1967	25	1987	18
1968	11	1988	16
1969	9	1989	21

Sources: Temin, 1980, p 6 and DiMasi et. al., 1991, p. 139