A Multiattribute-utility-function Approach to Weighing the Risks and Benefits of Pharmaceutical Agents

Stuart Eriksen and L. Robin Keller
Med Decis Making 1993 13: 118
DOI: 10.1177/0272989X9301300205

A Multiattribute-utility-function Approach to Weighing the Risks and Benefits of Pharmaceutical Agents

STUART ERIKSEN, PhD, L. ROBIN KELLER, PhD

Both the selection of doses of pharmaceutical agents and comparisons between pharmaceutical agents have long been based on the nonquantified concept of the risk–benefit ratio. Though useful, this concept implies a data comparison that is difficult to make: the toxicity *versus* the efficacy of a drug compound. This research demonstrates an approach for weighing risks and benefits by combining utility functions for human efficacy and toxicity with animal and laboratory toxicity information to develop an overall multiattribute utility function for an ophthalmic pharmaceutical agent, l-bunolol, intended for the treatment of glaucoma. With this multiattribute function and a small portion of the published data available for this drug, the expected utilities for six doses (including a control) could be compared and the value of this approach in drug-dosage selection demonstrated. *Key words*: decision analysis; expected utility; dosage selection; compound comparison; pharmaceutical decision; multiattribute utility; risk–benefit ratio. **(Med Decis Making 1993;13:118–125)**

The decision by a pharmaceutical firm to ask the United States Food and Drug Administration (FDA) for marketing approval for a pharmaceutical agent is made on the basis of an evaluation of the efficacy and toxicity information available at the time of the decision. That information is obtained both from human studies and from animal or laboratory studies.[1] Human studies produce both efficacy and toxicity information, while animal and laboratory studies generally produce only data related to toxicity. The pharmaceutical decision maker uses data from all three information sources when making the decision concerning a request for marketing approval.

This paper demonstrates the development of a multiattribute utility function for the combination of the utilities for efficacy and toxicity data from human and animal sources. With this approach, the relative merits of a particular dosage and/or a particular drug compound can be evaluated so that comparisons and selections among compounds and dosages can be made on the basis of both efficacy and toxicity. This paper focuses on the decision of the company intending to market a drug, and does not address the equally important decision of a physician to prescribe a drug for an individual patient.

DOSE AND DRUG COMPARISONS

The most important of the decisions made during the development of a drug involves whether or not the compound and/or the dose being considered is likely to have sufficient efficacy in humans to warrant the toxicity demonstrated. Such decisions use similar types of information and are concerned with two basic questions: 1) *which of the drugs* being tested should be pursued further with human studies (a comparison between drugs); 2) *which doses* of the drug should be recommended for these subsequent human studies or for eventual marketing (a comparison between dosages).

Comparisons between drugs are necessary because a pharmaceutical firm often conducts simultaneous research on several different compounds, all of which are designed to have essentially the same (or closely related) therapeutic effects. This occurs when a pharmaceutical research program is structured to produce more than one active modification of related compounds.

Compounding the problem is the fact that laboratory efficacy and toxicity screening methods are usually insufficiently selective to allow identification of *the one* compound that will have the balance of benefits and risks desired for the eventual human user. It is necessary, therefore, to make decisions using tradeoffs between efficacy and toxicity with information obtained from a variety of human, laboratory, and animal sources. These decisions are customarily made using heuristic judgment methods derived from past successes and failures. A decision method using some quantifiable summary characteristic, such as our proposal to use the expected utility of the overall efficacy and toxicity, would permit use of all the efficacy and toxicity information available and would be of significant value.

In a related comparison problem, the compound

Received November 13, 1990, from the Graduate School of Management, University of California, Irvine. Revision accepted for publication July 24, 1992.

Address correspondence and reprint requests to Dr. Eriksen: 13341 Eton Place, Santa Ana, CA 92705.

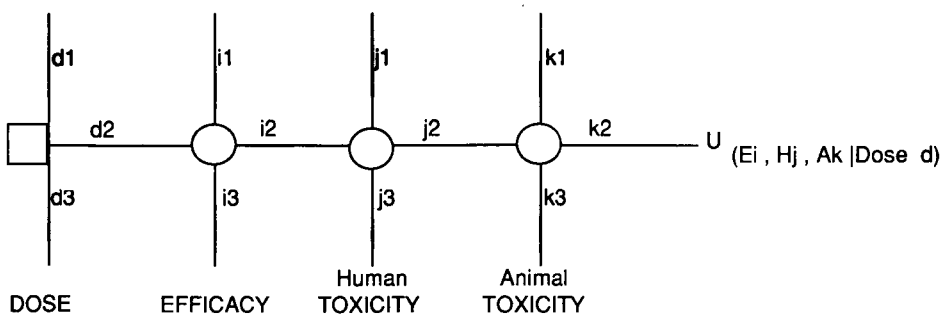

FIGURE 1. Overall schematic decision tree.

being considered by the firm may have one or more competitors already in the marketplace. The decision to request approval from the FDA to sell the drug on the U.S. market must be based on some efficacy or toxicity advantage of the proposed compound over the competitors. The expected utility of the overall efficacy and toxicity for the new compound should quantify this advantage better than the heuristic judgments currently being used.

As important as the drug comparison problem discussed above is the selection of the dose (or doses) to be recommended. When preliminary clinical studies have been completed to provide a body of data describing the action of the formulation, the most desirable dosage (e.g., concentration, amount and/or frequency of administration) must be selected for the final clinical studies. Only one dosage or, at most, two or three dosages are desired for this last and most expensive aspect of clinical testing. Expected-utility comparisons of several dosages using all the available information on efficacy and toxicity, will permit the ones providing the best balances of efficacy and toxicity to be selected for these last tests. Finally, at the termination of the last studies, the *dose* having the highest utility for the potential patients to be treated must be selected. The expected-utility evaluation of the doses studied in the earlier phases of the research should permit this decision to be made for the final studies.

THE DECISION TO BE MADE

The schematic decision tree for a sample drug having an efficacy attribute E, a "grouped" human toxicity attribute H, and a "grouped" human-related animal attribute A for all the doses considered is shown in figure 1.

In this schematic decision tree, $U_{(Ei, Ak|Dd)}$ represents the overall utility of the consequence: efficacy at level i, grouped human toxicity utility at level j, and grouped animal toxicity utility at level k, all conditioned on the selection of dose d of the drug in question. The human and animal toxicity attributes are noted as "grouped" because each of these involves several subattributes; this is discussed in more detail below.

Following custom, in figure 1 the dosage decision is depicted as a square or "decision node," and the probabilistic events leading to different outcome levels of an attribute are depicted as circles or "chance nodes."

DOSE EFFECTS

The utility assigned by the decision maker to any particular *efficacy* outcome level (such as the utility assigned to an intraocular pressure of 30 mm Hg is independent of the dose of the drug actually given. Only the probability of obtaining that specific outcome depends on the dose. Similarly, the utility for a specific human *toxicity* outcome level (e.g., moderate yet prolonged irritation, or pulse rate decrease of 5 beats/min) is also independent of the dosage given. Again, it is the probability of the occurrence of the specific toxicity outcome that is dose-dependent.

The utility assigned by the clinical investigator to a specific human toxicity manifestation that is related to an animal or laboratory test is also not dependent on the dose to be administered to humans. The probabilities of human outcomes, as estimated from the animal results, do depend on the "closeness" of the animal dose tested to the anticipated human dose.

So, a firm can have decision makers determine utility functions for toxicity and efficacy before any actual laboratory or clinical trials are done, so long as the potential health outcomes and toxicity measures are known. Then, for each drug and dose under consideration, the *probabilities* associated with each of the human efficacy and toxicity attribute nodes are determined from clinical studies at the doses under consideration. Those probabilities for the chance nodes associated with the animal toxicity are determined from the animal test data and the proposed dosages by the toxicologist involved in the decision.

THE EXPECTED UTILITY OF DOSES

As in all evaluations of decision trees, the expected utility (EU) of each chance node (a circle in figure 1) can be calculated (for example for animal toxicity) to be:

$$EU = \sum_{k=1}^{m} p_k \cdot U_{Ak}$$

where m is the number of levels of animal toxicity (A) being considered. For a sequence of chance nodes (efficacy, human toxicity, animal toxicity) as shown, the utility of a branch (consisting of one outcome for each chance node along the branch) is $U_{(Ei,Hj,Ak)}$ and the probability of that branch is the product of all the probabilities of the individual outcomes.

Given one dose, D_d, the overall expected utility on all attributes is:

$$E(U_{EHA}|D_d) = \sum_{i=1}^{q} \sum_{j=1}^{r} \sum_{k=1}^{s} p_{ijk} \cdot U_{(Ei,Hj,Ak|Dd)}$$

The term p_{ijk} is the product of all the probabilities related to the branch being evaluated, and the utility term is the multiattribute utility for that branch. The terms i, j, and k represent the levels of human efficacy, human toxicity, and animal toxicity being used, which are assumed to have q, r, and s levels, respectively. Using the customary decision criterion, selecting the dosage branch with the highest overall expected utility, $E(U_{EHA})$, the best dosage for the drug d could be selected.

THE UTILITY OF HUMAN EFFICACY FOR AN ANTI-GLAUCOMA DRUG

The utility function for the efficacy of an agent involved in a pharmaceutical risk–benefit decision will normally contain only one efficacy attribute. Most pharmaceutical agents do have several therapeutic actions, and more than one of these may have potential benefits. For example, while the pharmaceutical example used in this research (l-bunolol) has anti-glaucoma effects,[2,3] it is also active in the control of high blood pressure.[4] However, because of the costs involved in the clinical studies required for approval, it is customary to select only the efficacy attribute with the most significance for the sponsoring company and the marketplace it serves. If the other clinical effects are also deemed to be worthwhile from a marketing standpoint, subsequent marketing-approval decisions may be made, but commonly only one efficacy attribute is considered at a time.

In the example presented, standard forms of questions and scenarios[5] were used to determine the utility function U_E for intraocular pressure (IOP) treatment efficacy from decision makers in the clinical department of a pharmaceutical firm involved in glaucoma drug research. All utility functions were scaled from 0 to 1, with an IOP level of 12–15 mm Hg as the "best" with a utility $U_E = 1$. An IOP level of 50 mm Hg was scaled "worst," with a utility $U_E = 0$.

THE UTILITY OF HUMAN TOXICITY FOR AN ANTI-GLAUCOMA DRUG

Unlike the situation for efficacy, there are nearly always several human toxicity attributes in a pharmaceutical decision. In our example, the utilities of two commonly studied human toxicity attributes were assessed. Pulse rate decrease (a continuous variable) and ocular irritation (a discrete variable) were selected as representative and appropriate for topically applied anti-glaucoma β-blockers such as l-bunolol.[6–8]

The multiattribute utility function for the overall human toxicity utility (U_H) for the drug was determined to be a multiplicative function combining the utility for pulse rate, U_P, and the utility for irritation, U_I. The coefficients for each decision maker were calculated. The specific form of this multiattribute utility function is given later in the section describing the overall multiattribute utility function.

THE UTILITY OF ANIMAL TOXICITY FOR AN ANTI-GLAUCOMA DRUG

The problem of how to use animal toxicity for decisions about drug use in humans is the subject of considerable discussion and research by toxicologists at this time. For the purpose of this work, a procedure similar to that proposed by Keeney, Sarin, and Winkler[9] was employed. The utility functions ascribed to those human toxicities that were related to these animal toxicities were determined. The actual animal test results themselves and the proposed dose were then used by the toxicologist to estimate the probabilities of achieving related human utility levels.

Two commonly used animal tests for toxicity, the LD_{50} and the SGOT, were used in the example presented here. The former of these is a test of "overall" toxicity that measures only the dose causing death (from whatever cause) in 50% of the test animals. The latter is a test of the effects of the drug on serum glutaminase oxaloacetic transaminase (SGOT), an enzyme important in the livers of both animals and humans. The first is a proxy measure of the relative general toxicity of the drug for a human, and the second a similar proxy measure of human liver damage caused by the drug.

As an example of the process for the SGOT testing, the utility (U_S) of a human SGOT of 8 standardized spectrophotometric units (spu), indicating no liver damage, was scaled as equal to 1, and that of an SGOT of 1,000 spu (significant damage) equal to 0. Using the SGOT values obtained from the animal tests (beagle-dog–feeding studies), the toxicologist was asked to estimate the probabilities of observing a particular human SGOT result, e.g., "considering only this information, and a proposed dosage of xx mg/kg/day, what is the probability that a human user will have an SGOT of 8 spu? Of 100 spu? Of 1,000 spu?"

Because the LD_{50} is a measure of overall toxicity in an animal species, and cannot be directly compared with an analogous human effect as can the SGOT, a slightly different procedure was used. The LD_{50} is rather a suggestion of the overall toxicity one might expect in humans, most often resulting from misuse of the drug. The human effect being measured is therefore all-or-none, toxic or not ($U_L = 0$ or $= 1$). The animal LD_{50} value may be used as an indicator of the probability of danger for a user when the dose and the amount of drug proposed to be prescribed are considered. This request was framed, "considering only the LD_{50} data, what is the probability for the appearance of toxicity in humans if the dosage is 0.4 mg/day?"

It should also be clear that although only two toxicity attributes were studied, the process described need not be limited to two tests or to animal study results. Laboratory test results regarding the potential toxicity would seem to be equally amenable to this procedure. One might just as easily include all the attributes associated with the toxicity tests commonly used by the FDA or a particular firm in its decision making.

A Multiattribute Utility Function for Efficacy and Toxicity

The multiattribute utility function describing the overall utility of one test outcome for a pharmaceutical agent may have one of several forms. But only the additive and multiplicative cases were considered here, the equations having the following forms[5]:

Additive

$$U_{(\text{outcome})} = k'_E U_E + k'_H U_H + k'_A U_A \quad (1)$$

Multiplicative

$$U_{(\text{outcome})} = [1/K_o][(K_o k_E U_E + 1)(K_o k_H U_H + 1)(K_o k_A U_A + 1) - 1] \quad (2)$$

In the additive equation, k'_E and U_E are the weighting coefficient and utility of one efficacy outcome for the drug; k'_H and U_H are the weighting coefficient and utility of one human toxicity outcome for the drug; and k'_A and U_A are the weighting coefficient and utility of one animal toxicity outcome for the drug. In the multiplicative equation, the terms k_E, k_H, and k_A have the same weighting coefficient definitions, but defined for the multiplicative form. K_o is the scaling factor for the overall multiplicative equation, to ensure that the overall utility, $U_{(\text{outcome})}$, equals 1 when all attribute utilities equal 1.

While studying human and animal toxicity utilities, it was found that U_H and U_A were themselves multiattribute utility functions and were described best with multiplicative forms. These equations may be written:

$$U_{(\text{pulse rate P, irritation I})} = [1/K_h][(K_h k_P U_P + 1)(K_h k_I U_I + 1) - 1] \quad (3)$$

$$U_{(LD_{50},\text{SGOT})} = [1/K_a][(K_a k_L U_L + 1)(K_a k_S U_S + 1) - 1] \quad (4)$$

In equation 3, k_P and U_P are the weighting coefficient and utility value for the human pulse rate outcome; k_I and U_I are similar values for human irritation. In equation 4, k_L and U_L are the coefficient and utility for the human toxicity outcome related to the animal LD_{50} data and k_S and U_S are the similar values for the animal SGOT data. K_h and K_a are the multiplicative scaling constants for the human and animal equations.

The selection between additive and multiplicative forms of the multiattribute equations was made with questions of the form[9]:

Consider two treatments:

> Treatment 1 with a 50/50 chance of producing *either* an IOP of 50 mm Hg along with all human and animal toxicity values at their worst values, *or* an IOP of 10 mm Hg and no detectable toxicity, human or animal;
>
> OR
>
> Treatment 2 with a 50/50 chance of producing *either* an IOP of 50 mm Hg, along with no human toxicity manifestation, but with the worst animal toxicity values, *or* an IOP of 10 mm Hg and no detectable animal, but the worst human, toxicity.

The decision maker was asked to select between the two treatments.

Notice that with both treatments there is a 50/50 chance of getting 1) the best or the worst IOP levels, and 2) the best or the worst human and animal toxicity outcomes.

The combinations of efficacy and toxicity levels that would occur at one time, however, vary between the two treatments. If a person is indifferent between treatments 1 and 2, then an additive multiattribute utility function may be appropriate, since the decision maker is indifferent between the combinations of efficacy and toxicity levels that occur at one time, so long as the probabilities of different efficacy and toxicity levels are the same with both treatments.

For the clinical decision makers used in this study, indifference was not found between the two treatments at several IOP levels. Thus, the additive form (equation 1) was not suitable for the multiattribute utility function, and the multiplicative form (equation

Table 1 • Scaling Constants for Human Toxicity*

	Decision Maker 1	Decision Maker 2
K	−0.96	9.33
k_P	0.72	0.250
k_I	0.90	0.225

*Responses from decision maker 2 were used to design the questions in the human toxicity tests; his responses were therefore not used either for utility or for scaling values.

Table 2 • Weighting and Scaling Constants for Animal Toxicity*

	Decision Maker 2	Decision Maker 3
K	9.83	0
k_L	0.217	0.5
k_s	0.25	0.5

*Questioning decision maker 3 indicated nonadditive independence, but the direct assessment indicated additivity ($\Sigma k = 1$). Decision maker 1 felt unable to respond to questions regarding animal toxicity.

Table 3 • Weighting and Scaling Constants for the Overall Multiplicative Equation*

	Decision Maker 1	Decision Maker 2	Decision Maker 3
K_o	0.56	11.97	4.68
k_E	0.40	0.20	0.25
k_H	0.23	0.071	0.13
k_A	0.23	0.089	0.13

*The constant k_A is the coefficient for grouped animal data. The k_H term is the coefficient for the grouped human toxicity data, and k_E is the coefficient for the one efficacy attribute.

2) was used in its place.*

Scaling and weighting constants for human and animal toxicity. Each toxicity group (human and animal) was used as a single attribute in the overall multiattribute function. The usual additivity and weighting questions produced direct assessment of the multiplicative equation constants for human and animal toxicity shown in tables 1 and 2 (for details of these studies, see Eriksen[11]).

Constants for overall efficacy and toxicity. Following the procedures recommended by Keeney and Raiffa,[5] the coefficients for the overall multiattribute utility function were determined by the direct-assessment procedure. The multiplicative coefficients for these clinician decision makers are shown in table 3.

CLINICAL TRIAL INFORMATION

The raw results of clinical studies are normally expressed in terms of the percentage of the test population that achieves a particular efficacy or toxicity outcome.[12] It is assumed that such a percentage represents the best estimate of the probability of the occurrence of that outcome in the general population of patients to be treated.†

DOSE AND SUBSTANCE DIFFERENCES

By evaluating the decision tree (fig. 1), the overall expected utility with efficacy, human toxicity, and animal toxicity [$E(U_{EHA|Dd})$ as defined below] can be calculated for a dose of the drug in question. In the example used, with one efficacy, two human toxicity, and two animal test toxicity attributes, the equation for this expected utility is:

$$E(U_{EHA}|D_d) = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} p_{ijk}\{[1/K_o][(K_o k_E U_E(i) + 1)(K_o k_H U_H(j) + 1)(K_o k_A U_A(k) + 1) - 1]\} \quad (5)$$

The term P_{ijk} is the product of all the probabilities related to the branch being evaluated, and the term in braces ({..}) is the multiattribute utility for that branch. The remaining terms have the meanings discussed previously, expanded somewhat to indicate that there are q levels of the efficacy measure used, r levels of human toxicity outcome groups, and s levels of animal toxicity outcome groups. Using the customary decision criterion, selecting the branch with the highest overall expected utility, $E(U_{EHA})$, the best dosage for the drug d could be selected.

A Sample Calculation for the Expected Utility of a Series of Dosages

The data from one decision maker are used in this section to illustrate the calculation of the expected utilities of a series of doses. This manager decision maker could answer all the questions, since he had both clinical and toxicological experience. The equations elicited from this decision maker are‡:

*The implications of the nonadditive question responses were interesting, "I'm *not* indifferent to these two treatments because in the case of #1 at least some of the patients treated would have the best outcome."

†This direct association between the tested group and the eventual group of patients for whom the drug is intended has not, to the authors' knowledge, been tested. With careful sample selection it is generally assumed to be true. That assumption is also made here.

‡In one case the U_A *lottery* choices made indicated that an additive function described that manager's function well, $\Sigma k = 1$.

$$U_A = 0.5 \cdot U_S + 0.5 \cdot U_L \quad (6)$$

$$U_H = [1/9.33][(9.33 \cdot 0.25 \cdot U_P + 1)(9.33 \cdot 0.225 \cdot U_I + 1) - 1] \quad (7)$$

$$U_{EHA} = [1/4.68][(4.68 \cdot 0.25 \cdot U_E + 1)(4.68 \cdot 0.132 \cdot U_H + 1)(4.68 \cdot 0.132 \cdot U_A + 1) - 1] \quad (8)$$

The animal and human coefficients for equations 6 and 7 were taken from tables 1 and 2, respectively. The coefficients for equation 8 were taken from table 3.

BRANCH PROBABILITIES

Side data. L-Bunolol is relatively nontoxic, as evidenced by its LD_{50} values, some of which are shown in table 4. Using an appropriate scenario and questions along with the data in table 4, the probabilities for the occurrence of toxicity or potential danger in humans (as estimated by the toxicologist decision maker based on the LD_{50} data) were estimated, given the indicated human doses (these probabilities appear in table 5).

The liver function effects of l-bunolol, as measured by the SGOT during a one-year dog-feeding study in beagle dogs (controls and animals treated orally at 2, 6, and 24 mg/kg/day) are shown in table 6. The toxicologist was confident that no significant level of SGOT would result in humans, based on the beagle-dog-feeding studies. Therefore, the estimate of the probability of abnormal SGOT values (≥ 40 spu/ml) was assumed to be = 0.01 for all dosages.

Human data. A large number of clinical studies were done to evaluate the efficacy of l-bunolol[6,8] and were presented in the New Drug Application sent to the FDA. For the purpose of illustrating the power of the process developed here, however, only one multidose human clinical study was selected. It involved one investigator (external to the firm making the decision), 49 patients treated with five concentrations of the drug, and one drug-free control.[8] Many more patients would be required for accurate estimates of the probabilities involved, but even this small number adequately illustrates the use of the multiattribute-utility-function approach proposed here.

The clinical study was performed with solutions containing 0% (placebo), 0.03%, 0.3%, 0.6%, 1%, and 2% concentrations. The data (efficacy and toxicity) for the individual patients were grouped into subdivisions and the percentage of the group at each dosage producing that result calculated.[§] The center point of the group was used to calculate the utility for that group.

Overall utility estimation. Using the utility functions, along with the scaling and weighting coefficients determined above, the overall expected utilities were plotted in figure 2 for the doses/concentration used.

Although the small number of patients in each dos-

Table 4 • Acute Oral and Intravenous Toxicities of l-Bunolol

Species	Strain	Route*	Sex	LD_{50} (mg/kg)
Rat	CFN	po	M	700
			F	800
		iv	M	25
			F	28
Mouse	MF_1	po	M	1,530
			F	1,220
		iv	M	78
				84†

*Lethal doses administered by intravenous injection (iv) are usually significantly smaller than orally administered (po) lethal doses. Therefore, LD_{50} values cannot be compared across routes of administration.

†For comparison purposes, two drops of a 1% solution administered four times a day (a common dosage regimen) to a 150-pound patient represents 0.059 mg/kg.

Table 5 • Probability Estimates for Abnormal Human Outcomes from l-Bunolol Based on LD_{50} Data

Dose (Concentration) (%)	Probability Values* p_L
0.03	0.0001
0.3	0.0002
0.6	0.0005
1.0	0.001
2.0	0.009

*These probability values for human toxicity were estimated by the toxicologist, based on the potential for toxicity conditioned on "ingesting one 20-ml container at one time." The probability of *that* occurrence is small (see Eriksen[11]), estimated at not more than 1/1,000; thus, the overall probability of toxicity based on the LD_{50} would be $(p_L) \leq (p \times 0.001)$.

Table 6 • SGOT Test Results for Dogs Fed Several Concentrations of l-Bunolol for One Year

		Treatment		
Week	Controls	2 mg/kg/day	6 mg/kg/day	24 mg/kg/day
0	31.6 (2.0)*	25.9 (1.3)	22.0 (0.9)	36.4 (1.6)
13	26.6 (1.0)	25.4 (1.3)	28.7 (1.3)	26.1 (0.5)
26	30.4 (1.6)	25.3 (0.3)	30.0 (1.1)	28.7 (0.8)
52	36.2 (2.2)	31.3 (2.8)	33.3 (1.9)	33.8 (1.1)

*Mean value in spu/ml and (standard error) for the determination. The 95% confidence limits for the historical controls for this test are reported to be 10.28–39.90 spu/ml.

[§]No patient discontinued treatment during the study, but clinical observations of interfering conditions dictated that pressure data from four subjects not be used. These were eliminated for that part of the evaluation.

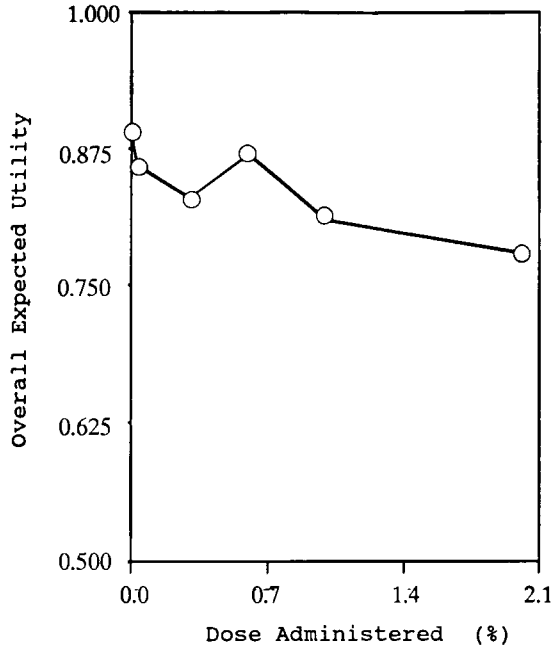

FIGURE 2. Overall expected utility for l-bunolol (decision maker 3).

age group prevents selection of a single best dosage from this study alone, the fall-off in utility from increasing toxicity is apparent. The small number of patients in the 0% (placebo) group and the chance occurrence that the average intraocular pressure in this group was rather low made that sample, which should have had a low utility, have a higher than anticipated utility, and obscured the lower value one would anticipate for the overall expected utility, $E(U_{EHA})$, at the 0% dose.

The curve is generally concave but is relatively flat over the lower dosages, suggesting that no significant change in utility should be anticipated from any of the dosages until the 1% or 2% concentration is reached. At this point the toxicity increase begins to affect the utility without a concomitant increase in utility due to efficacy. After nearly 2,000 clinical cases had been evaluated, the sponsoring firm came to a similar conclusion, and it has received FDA approval to market 1/4%, 1/2%, and 1% concentrations of l-bunolol.

Discussion

This paper presents a prototype multiattribute-utility-function method for the application of decision analysis to aid pharmaceutical firms in making product-development decisions. Decision-analytic approaches to medical and pharmaceutical questions are common in the medical decision literature. They most commonly address either the sociopolitical aspects of the medical questions[13–15] or the choice of medical care.[7,12,16,17] Business decisions involved in pharmaceutical development, such as those described here, have not been addressed in this manner.

The research presented here has used the multiattribute-utility-function approach to render the scales by which *efficacy* (relative benefit) and *toxicity* (relative risk) outcomes are customarily measured into easily relatable units (utilities). A multiplicative multiattribute function was used to combine these utilities into one equation in order to evaluate each of the potential overall outcomes. This approach has shown itself to be capable of determining an overall expected utility value for a drug and dosage, a value one might propose to be related to the well-known risk–benefit ratio of a drug and a dosage. This overall utility for a drug, used here to suggest the best dosage of a pharmaceutical agent, should also be related to the market advantage of one drug over another.

We found that experienced clinical decision makers were able to provide information needed to assess the utilities of several of the usual types of efficacy and toxicity test results (both animal and human) obtained for an ophthalmic anti-glaucoma drug. With these responses, the multiattribute utility functions, the branch probabilities associated with the results, and consequently the overall expected utilities for several dosages of the drug could be determined. A graph from one study of the overall expected utility for a drug versus the doses given suggests a concave, downward-trending curve, implying that the lower dosages tested have a higher utility for patients. Thus, the balance between the benefits of the drug's efficacy and the risks of its toxicity was best at lower dosages.

This study used only a small fraction of the data produced to support the new-drug-approval application for l-bunolol. It seems likely that if the same procedure has been used on *all* of the data, more effective dosage comparisons would have been possible. Because there do not appear to be any special factors involved in assessing ophthalmic drugs that are not present in assessing other pharmaceuticals, it is reasonable to assume that other experienced clinical people would be able to answer similar questions for other, nonophthalmic drug outcomes. The prototypical process suggested for ophthalmic determinations should thus be generalizable to any pharmaceutical agent.

The apparent disparity that we found between the utilities and coefficients assigned to these efficacy and toxicity data points by different decision makers within one firm should be of concern to a firm using such decision makers to evaluate drugs. It would not seem to be in the firm's best interest to have decisions about the same drug dependent on the person to whom the evaluation was assigned. It was clear during the discussions, however, that the differences observed were not errors in assessment. These differences were defended by the decision makers and supported by their experience and opinions. A similar, but unrelated, problem involving differences in utilities, where the

problem involved investment personnel within one firm, has been discussed by Spetzler.[18] The suggestion in that work was to devise a fitted group utility function pooling all decision makers' judgments for the firm to use. A similar solution to the problem might be suitable in this case.

The pooling of these different assessments to obtain one best value for utility and weighting coefficients is beyond the scope of this study and is not considered here. It is apparent, however, that there is significant merit in a firm's having only *one* consistent set of utilities and weighting factors for routine decisions. The firm could hold discussions to eliminate any differences and form a consensus on the "corporate utility function."

At least three future research areas are strongly suggested by the preliminary results presented here. First, a full study of the complete set of data for one drug would permit the complete development of the overall expected utilities for many doses. Full evaluation of at least two agents competing in the same marketplace would have additional and significant value. Success in the marketplace should be at least partly related to the overall utility of the product at the dose selected. While financial success depends on many factors (only one of which is the utility of the drug), the clinical experience of relatively unbiased academic clinicians might offer a measure of the drug's eventual therapeutic value and acceptability by the profession.

Second, a comparison should be made of the relationship of the clinical study group versus the general population to be treated. The availability of postmarketing information (data that are collected by the sponsoring firm after the approval of the drug and its subsequent marketing) offers the possibility of studying the relationship between the results obtained in clinical study patients and in the actual population for which the drug was developed.

Third, deeper investigation is needed to elucidate the distinctions used by clinical and toxicological decision makers in evaluating animal and laboratory data as they apply to humans. The determinations of the utilities and probabilities for some of the laboratory and animal data during the development of this research suggested that toxicological decision makers and clinical decision makers feel quite differently about their meanings (e.g., LD_{50} data). The importance placed on these pieces of toxicity information by both the FDA and clinical decision makers suggests that this difference and similar uncertainties regarding other laboratory and animal studies should be further studied to make their use clearer and more amenable to this type of evaluation.

The authors thank the clinicians who kindly consented to respond to the questioning required for this research and also the Allergan Pharmaceutical Company, both for permission to refer freely to the New Drug Application for l-bunolol and for financial support of Dr. Eriksen.

References

1. New Drug Law, 21 USC Section 102(c) of the Drug Amendments of 1962. Amending Section 505(d) (21USC 355(d)) of the 1938 Act.
2. Kolker AE, Hetherington J. Medical techniques: diagnosis and therapy of the glaucomas, 5th ed. St. Louis: C. V. Mosby, 1983.
3. Lesar T. Drug review: comparison of topical glaucoma therapy. Clin Pharm. 1987;6:451–63.
4. Novack GD. On cardiovascular effects of topical glaucoma therapy. J Toxicol; Cutaneous and Ocular Toxicol. 1984;3:243–50.
5. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: John Wiley and Sons, 1976.
6. Long D, Zimmerman T, Spaeth G, et al. Minimum concentration of levobunolol required to control intraocular pressure in patients with primary open-angle glaucoma or ocular hypertension, Am J Ophthalmol. 1985;99:18–22.
7. Moskowitz AJ, Pauker SG. A patient with new Q waves: methods for decision making in the individual patient. J Am Coll Cardiol. 1989;14(3):29A–37A.
8. Partamian LG, Kass MA, Gordon M. Dose + response study of the effect of levobunolol on ocular hypertension. Am J Ophthalmol. 1986;95:229–32.
9. Keeney RL, Sarin RK, Winkler RL. Analysis of alternative national ambient carbon monoxide standards. Manage Sci. 1984;30:518–28.
10. Fishburn PC. Methods of estimating additive utilities. Manag Sci. 1967;13:435–53.
11. Eriksen S. Decision making with information from apparently non-comparable studies. Dissertation, University Microfilms International. 300 N. Zeeb Road, Ann Arbor, MI 48106.
12. Kassirer JP, Moskowitz AJ, Lau J, et al. Decision analysis: a progress report. Ann Intern Med. 1987;106:275–91.
13. Jacoby I, Pauker SG. Technology assessment in health care: group process and decision theory. Israel J Med Sci. 1986;22(3-4):183–90.
14. Patrick DL, Bush JW, Chen MM. Toward an operational definition of health. J Health Soc Behav. 1973;14:6–23.
15. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health-status index. Health Ser Res. 1973;8:228–45.
16. Pauker SG. Coronary artery surgery: the use of decision analysis. Ann Intern Med. 1976;85:8–18.
17. Young MJ, Fried LS, Eisenberg JM, et al. The single-cutoff trap: implications for Bayesian analysis of stress electrocardiograms. Med Decis Making. 1989;9:176–80.
18. Spetzler CS. The development of a corporate risk policy for capital investment decisions. IEEE Trans Sys Cybern. 1968;SSC-4:279–300.