Thomas Jefferson University
Jefferson Digital Commons

Department of Pharmacology and Experimental Therapeutics Faculty Papers

Department of Pharmacology and Experimental Therapeutics

April 1979

A scale for assessing the severity of diseases and adverse drug reactions: Application to drug benefit and risk

Ronald J. Tallarida
Temple University

Rodney B. Murray
Thomas Jefferson University, r.murray@usciences.edu

Carl Eiben
Temple University

Let us know how access to this document benefits you

Follow this and additional works at: http://jdc.jefferson.edu/petfp

Part of the Medical Pharmacology Commons

A Scale for Assessing the Severity of Diseases and Adverse Drug Reactions: Application to Drug Benefit and Risk

Ronald J. Tallarida, Ph.D., Rodney B. Murray, B.A. and Carl Eiben, M.D.
Department of Pharmacology
Temple University School of Medicine
Philadelphia, PA

Physicians were interviewed to assess their willingness to risk adverse drug reactions among patients. These untoward reactions were ranked according to severity and weighted against the primary illness being treated. A specially designed questionnaire in the form of a matrix was used. Severity was divided into seven classes denoted by progressively increasing numerical scores, W_1 to W_7, whose values could be calculated from analysis of the completed questionnaires. The questionnaires presented several cases, in each of which an illness of specified severity was to be treated with a drug whose untoward reactions differ in severity from that of the primary illness. Each case involved a different permutation of the severities. Analysis of the completed questionnaires yielded the mean values of the scores which were found to range from $W_1 = 1.00$ (the mildest case) to $W_7 = 817$ (the most serious case). It is our opinion that this type of scale is preferable to nonnumerical descriptions of severity such as 'mild" or "serious," since, when combined with data on frequency of occurrence, a numerical scale permits a determination of expectation of both benefit and risk.

Adverse drug reactions (ADRs) are an important problem in therapeutics. Most important is harm to the patient. Also important, however, is the impact that ADRs have on the medical profession, the pharmaceutical industry, and the FDA, since these groups share the responsibility for the development and the use of drugs needed to treat illnesses and save lives. The probably beneficial effects of new treatments and new drugs must not be overlooked nor de-emphasized, and these must be properly weighed against probably untoward effects.

The determination of a proper balance between patient benefit and ADRs is, however, a formidable problem. First, there is the difficulty in evaluating an ADR: "Is there a clear cause-and-effect relationship between a specific drug and a clinical event?" Much controversy surrounds this point since ADRs often present no unique clinical or laboratory findings which distinguish them from the patient's underlying illness.[6,7] Many effects frequently attributed to drugs occur commonly in presumably healthy individuals taking no medication[9] and in patients taking placebos.[3] The awareness of this problem has resulted in the setting of defined tolerance limits in linking a drug with a particular clinical event.[6] Second, there is the problem of data gathering which is necessary to determine the frequency of occurrence. At present there are virtually no reliable quantitative data on the frequency of clinical response or adverse reactions to specific drugs as they are used in general clinical practice. Programs such as the Boston Collaborative Drug Surveillance Program[5] have been started for this purpose.

In addition to the problems of recognizing ADRs and of determining the incidence of such reactions, there is at least one other consideration in balancing risk with benefit. This is the formation of a numerical scale which measures severity of both the illness being treated and of the ADR resulting from the treatment. Without such a scale, we cannot properly assess probable benefit and probable risk, leaving us with only imprecise judgments, many inconsistencies, and substantial differences in the attitudes of regulatory agencies in different countries.[11] It is the broad evaluation of a drug's benefit and risk with which we are concerned – the process of evaluation that regulators, manufacturers, and the general public employ. We can gain important information regarding severity by gathering information about views on drug efficacy and safety.

A drug of demonstrated efficacy is considered sufficiently safe if its risks are judged to be acceptable; all drugs have potentials for both. Thus there is a judgmental component in the definition of safety. Measuring risk is an objective but probabilistic pursuit. Judging the acceptability of that risk is a matter of personal and social value judgment. Such decisions range from the very personal to the broadly social. Hence, a severity scale will, of necessity, reflect this wide range of judgments and will not have precise numerical boundaries for each category of illness. Nevertheless, a numerical classification of severity, even through approximate, is as necessary in determinations of drug benefit and risk as are the equally imprecise data on the frequency and identification of untoward reactions.

In this study we interviewed 53 physicians with the use of a questionnaire designed to assess the physicians' willingness to risk ADRs ranked according to severity and frequency, these untoward responses being weighted against the primary illness under treatment. The objective was to test a method as a starting point for the development of numerical scales of severity of ADRs and diseases. Such scales, combined with data on frequency of benefit and ADRs, should permit quantitative determination of drug benefit and risk.

Theory

The theory underlying the analysis of physicians' prescribing practices in regard to benefit and risk is the same as that in games of chance. In the analysis of such games, if an event occurs with a probability P, and if its occurrence yields a prize of value W, then the expectation E of winning is defined to be the product of P and W[1]:

$$E = P \times W$$

The expectation may be thought of as the fair price for the privilege of playing the game.* In connection with drug treatments, W represents the severity of the illness being treated (removal of that illness is the prize value sought) and P is the probability of

_____

*As an example, suppose a state lottery yields a cash prize of $500 and that the probability of winning is 1/1,000. The expectation is 50 cents and is the "fair price" in the sense that in a long series of plays at 50 cents/play, you will break even. If you were charged less than 50 cents, say 45 cents, you would be making an average profit of 5 cents per play.

its achievement. We may apply this reasoning to drug treatments. In such an application the expectation for winning is the efficacy of the treatments, and the cost is risk in the form of adverse reactions to the treatment. Following the notation used in the game model, we denote by W the severity of the illness being treated and by P the probability that the treatment will be efficacious. The expectation for winning is then the patient benefit E:

$$E = P \times W$$

If there is an ADR, we determine its expectation in the same manner, i.e., by taking the product of its probability and severity. We shall use primes on the symbols which stand for quantities related to the ADR, W′ denoting its severity, P′ its probability, and E' its expectation which is the risk of the treatments:

$$E' = P' \times W'$$

Often more than one ADR accompanies treatment. A second independent ADR will add to the total risk E'. If we denote the probability and severity of this second ADR by P″ and W″, respectively, then we add its contribution, P″ * W″, to the risk, giving

$$E' = P' \times W' + P'' \times W''$$

The "profit" of the treatment is the difference between benefit E and risk E′. Table I illustrates the theory.

Table I. Benefit-risk model for drug treatments (analogy to games of chance)

Lottery	Drug Treatment
Value of prize = W	Severity of illness treated = W
Probability of winning = P	Probability of efficacy = P
Expectation, E = W × P	Benefit, E = W × P
Risk, E' = Cost of playing	Risk: Adverse reactions of severity W′ and W″ with respective probabilities P′ and P″
	Risk, E′ = W′P′ + W″P″

Determination of severity

Unlike games of chance, in which the prize has some numerical value, in the benefit-risk analysis of drugs we do not have numerical values for the severity of the illness being treated or of ADRs. We do have, however, some feeling for ranking. For example,

cancer is more severe than arthritis, which, in turn, is more serious than a simple headache. This ranking of severity is implicit in reasonable treatment programs. Thus, we tolerate headache as an ADR when treating cancer, regardless of how often the headache might occur, but we would not tolerate cancer as a side effect in our choice of a drug to treat headache, even if the incidence of cancer were very low. Severity scores can be computed if information on incidence is known.

To illustrate the computation, we consider the following theoretical situation involving a pair of clinical situations. (Since this is only an illustration, any pair of illnesses can be used.) It is desired to find the relative severity, expressed as numerical scores, of two situations: (1) a cardiac arrhythmia consisting of occasional premature atrial beats and (2) a low-grade infection. We can find the numerical severities if we consider two cases. In the first, the illness is the arrhythmia and the side effect of the drug therapy is the infection. In the second case, the illness is the infection, and the treatment might induce the arrhythmia as the side effect. By thus interchanging the illness and the side effect and by ascertaining tolerated frequencies, we can arrive at numerical values for the severity of each condition. Table II illustrates the cases. In each case, we assign probabilities that the treatment will be successful, and ask what probability would be tolerated for the productions of the side effect.

Table II. *Illustration of a 2 × 2 matrix-questionnaire for determining severity scores*

Case No.		Infection		Arrhythmia
1		( )		0.2
2		0.9		( )

The first case asks the question: "Given that a drug has a probability of 0.2 for success in treating the arrhythmia but that it might also result in the infection, what is the highest incidence that should be tolerated for the production of this ADR?" A typical answer is 1; i.e., we would use this drug even if the probability of infection is 1 because the arrhythmia is judged to be more serious. The value "1" is entered in the top row, representing Case 1.

Table III. *Completed matrix.*

Case No.		Infection		Arrhythmia
1		(1)		0.2
2		0.9		(0.01)

Case 2 presents a different, though related, question. Here one is treating the infection, this time with a drug whose probability for patient benefit is 0.9, but this drug can induce arrhythmia as a side effect. What is the highest incidence that should be tolerated for this untoward reaction? A typical answer might be 0.01. The completed table is shown in Table III. (A rectangular array of numbers such as that of Table III will be called a "matrix.") From the entries in Table III, we can derive for each case the benefit and risk. For Case 1, the benefit is E, the product of probability and severity: $E = (0.2) \times$ (severity of the arrhythmia) and the risk for Case 1 is $E':E' = (1.0) \times$ (severity of the infection). The profit is the difference of E and E':

$$(E-E') = (0.2) \times \text{(severity of arrhythmia)} - (1.0) \times \text{(severity of infection)}$$
$$\text{(Case 1; Equation 1)}$$

Applying similar reasoning to case 2, we get from the second row elements of the completed matrix of Table III, the profit in this case:

$$(E-E') = (0.9) \times \text{(severity of infection)} - (0.01) \times \text{(severity of arrhythmia)}$$
$$\text{(Case 2; Equation 2)}$$

If we assume that in each case the same profit is desired,* then the left-hand sides of Equations 1 and 2 are equal, and we may solve these equations simultaneously, yielding: (severity of arrhythmia) = (9.05) x (severity of infection).

Thus, based on the tolerated probabilities entered in the matrix for each case, we may conclude that the respondent's answers are consistent with the judgment that this particular cardiac arrhythmia is about nine times as severe as a low-grade infection. In each of the theoretical cases presented in this example, only a single side effect was assumed, leading to a 2 x 2 matrix. If, say, two independent side effects accompanied each case, we would have a 3-row, 3-column matrix, and 3 cases would have to be considered in order to compute the severities.

Methods

Disease states were divided into seven classes, characterized by unknown numerical severity scores $W_1, \ldots, W_7$, indicating a progression of severity as described in Table IV. Each of the states 1 to 7 in Table IV is briefly described in what we emphasized as a progression of situations of increasing severity, along with an example of a clinical situation which can have the described severity. The table was held in view throughout the interview. The physicians were told that the example was merely a guide** (since the

_____

*Other criteria may be used, in which case the computation would lead to different values. In the absence of any obvious choice of other criteria, we have used equal profit for both cases.

**The use of male gonorrhea may not have been a good example for, although it can fit into the description of class 3, it is not generally a side effect of treatment. We mention again that the description, not the example, was emphasized.

_____

named disease could have degrees of severity) and that the description should be emphasized in the assessment of severity, keeping in mind the progression of severities. The descriptions were purposely kept brief, with just enough detail to discriminate among them. Such descriptions allow the respondent some latitude in classifying. After the respondent studied Table IV, we presented him with a questionnaire which consisted of a matrix which was to be completed.

Table IV. *Severity of disease states**

W_1:	Mild disease or condition with symptoms which are not progressive and which cause only a mild degree of discomfort or incapacity. *Examples:* Mild headache, hay fever
W_2:	A more severe disease or condition whose symptoms differ in degree from the W_1 classification; the symptoms do not occur with regularity *Example:* Severe headache
W_3:	A chronic disease or condition more serious than W_2 which may interfere with normal activity and which is incapacitating for intermittent periods of time *Examples:* Bronchial asthma, grand mal epilepsy, male gonorrhea
W_4:	A chronic disease or condition which is completely incapacitating (confining a patient to bed) but which is usually not considered life-threatening or life-shortening *Example:* Rheumatoid arthritis
W_5:	A disease or condition which is serious enough to shorten life expectancy but which is not considered life-threatening *Examples:* Hypertension, chronic pulmonary disease
W_6:	A disease or condition which is considered life-threatening (death in 1 to 2 yr) but which is not a medical emergency *Examples:* Congestive heart failure, tertiary syphilis
W_7:	A disease or condition which is considered a medical emergency and which will terminate life within 1 yr or less *Examples:* Acute lymphatic leukemia, severe cardiac arrhythmia

   *Seven degrees of severity characterized by increasing numerical weights, $W_1, \ldots, W_7$. Examples are only guides to diseases that can have the degree of severity described.

Table V. *Sample questionnaire*†*

Case No.	W_1	W_2	W_3
1	( )	( )	0.2
2	( )	0.8	( )
3	0.9	( )	( )

   *This matrix deals with state W_1, W_2, and W_3. Two others were used, with column headings W_3, W_4, W_5, and W_5, W_6, W_7.
   †In Case 1, an illness of severity W_3 is treated with a drug which is efficacious with probability 0.2, but may produce adverse reactions of severity W_2 and W_1. Respondent indicates the highest probabilities tolerated for each and enters these values in the appropriate columns. Similarly, in each of the other cases, the probability of efficacy is given and tolerated probabilities for side effects of different severities are sought.

A matrix of 3 rows and 3 columns was used, each row of this represented a theoretical case in which an illness of specified severity was to be treated with a drug which produces two independent adverse reactions with severities different from that of the primary illness. The column headings, such as W_1, W_2, and W_3 in Table V, represented the numerical scores of the severities to be considered in each case. For Case 1 in Table V, the illness being treated is one of severity W_3. The subject was told that there was a single drug known to be beneficial in 2 of 10 cases (probability, 0.2) when used properly. The entry "0.2" was thus given as an element of the first row under the column heading W_3. This treatment, however, may induce side effects of severity W_1 and W_2 which, by definition, are less serious than the illness being treated. These side effects are treatable. The question was, "What is the highest probability (or frequency) that you would tolerate for the production of side effects of severity W_1 and W_2?" The respondent's replies, the tolerated probabilities, were entered in the appropriate blanks of the first row. In most cases the frequencies selected were from the set (1, 0.1, . . . , 0.000001). The subject was then asked to consider Case 2. In this case, the primary disease was one of severity W_2 and the drug available had a beneficial effect in 8 of 10 cases; hence, the value "0.8" was given in the second row under W_2. The drug produced side effects that were more severe (W_3) and less severe (W_1). Again, "What is the highest incidence of each that you would tolerate?" was asked. The subject was subsequently asked to consider Case 3, a case in which the least severe disease (W_1) was being treated with a drug known to be beneficial in 9 of 10 cases; hence, "0.9" was given in row 3 under heading W_1. In this case, however, the two side effects, W_2 and W_3, each more serious, could occur, and the entries represent the subject's judgment of the highest probability that he or she would tolerate for the production of each of these.

In addition to the column headings "$W_1 - W_2 - W_3$" used on the matrix described, matrices with two other headings were used, "$W_3 - W_4 - W_5$" and "$W_5 - W_6 - W_7$" but with the same given diagonal elements. Thus there were three matrix questionnaires in all, which together spanned the scale W_1 to W_7. This division of $W_1, . . ., W_7$ into the three subsets was done in order to keep the time devoted to answering down to a reasonable length and yet to allow each respondent sufficient time to study the description and to reflect on his or her answer so as to minimize errors associated with fatigue, impatience, or interruptions.

The subjects were 53 practicing physicians, either senior staff or residents, in the Philadelphia area. Most were from Temple University Hospital and from Lankenau Hospital and represented the following specialties: internal medicine, 11; cardiology, 10, pulmonary, 6; anesthesiology, 5; rheumatology, 4; surgery, 4; oncology, 3; gastrointestinal, 2; pediatrics, 2; nephrology, 2; neurology, 1; psychiatry, 1; endocrinology, 1; and radiology, 1.

All of the interviews were conducted by the authors. At the beginning of each interview, we explained our purpose and showed the system of classification (Table IV) to the subject. Subjects were encouraged to take as much time as they wished and to ask questions on any aspect. All questions raised by the subjects were answered during the interview, and in every case the matrix was completed in our presence. Depending on the

subject's time, he or she answered either one, two, or all three questionnaires. The average time for an interview, for one matrix, was 45 minutes.

Each completed matrix (questionnaire) produced a set of three simultaneous equations from which the set of weights (either $W_1 - W_2 - W_3$, $W_3 - W_4 - W_5$, or $W_5 - W_6 - W_7$) could be determined by solving the system, utilizing a Service Bureau Corporation time-sharing computer. Each triplet of values was determined along with their mean values and standard errors and expressed relative to $W_1 - 1.00$.

Table VI. *Typical responses to the three questionnaires**

A.

Case No.	W_1	W_2	W_3
1	(0.1)	(0.01)	0.2
2	(0.01)	0.8	(0.001)
3	0.9	(0.01)	(0.001)

B.

Case No.	W_3	W_4	W_5
1	(1)	(0.1)	0.2
2	(0.1)	0.8	(0.01)
3	0.9	(0.001)	(0.0001)

C.

Case No.	W_5	W_6	W_7
1	(0.1)	(0.1)	0.2
2	(0.01)	0.8	(0.001)
3	0.9	(0.0001)	(0.00001)

*In each questionnaire 3 cases were presented and in each the probability for efficacy of a drug treatment was given. The values in parentheses are the respondent's replies, expressed as the highest probability of adverse reactions whose severities are different from that of the primary illness. (See text for computation from the matrix A.)

Results

Table VI illustrates typical responses to each of the three matrix questionnaires. For example, in Table VI, *A* 3 cases are given, each represented by one of the rows of the matrix. Thus, in Case 1, the primary illness is of severity W_3 and the drug treatment is efficacious with probability 0.2. This value was given to the respondent. The values in parentheses represent the subject's response. We see in this case that the respondent chose probability 0.01 as the highest he would tolerate for the side effect of severity W_2 and the higher probability 0.1 for the least severe side effect W_1. The benefit in Case 1 is, therefore, $E = 0.2 \times W_3$, whereas the risk is $E' = 0.01 \times W_2 + 0.1 \times W_1$. The profit in this case is the difference $E - E' = 0.2W_3 - (0.01W_2 + 0.1W_1)$ (Case 1).

The same respondent, when given Case 2 in which the primary illness was of severity W_2 and the treatment had an efficacious probability 0.8, chose 0.01 as the probability for the mild side effect W_1 and 0.001 as the probability for the more severe side effect W_3. The profit in this case is the difference $E - E' = 0.8W_2 - (0.01W_1 + 0.001W_3)$(Case 2). The entries in row three, representing Case 3, lead to the profit $E - E' = 0.9W_1 - (0.01W_2 + 0.001W_3)$(Case3).

Simultaneous solution of the equations for each of the 3 cases gives relative values for $W_1:W_2:W_3 = 1.00:1.12:4.98$. We obtained 24 completed matrices of this type, "W_1-W_2-W_3," 31 completed matrices of the type "W_3-W_4-W_5," of which Table VI, *B* is typical, and 25 completed matrices of the type "W_5-W_6-W_7," of which Table VI, *C* is typical. These were completed by 53 physicians. The mean ($\pm$SE) values of the severities, $W_1, \ldots, W_7$, along with the standard error of the mean, are given in Table VII. The values range from 1 to 817. The responses of three physicians are not included because their replies contained inconsistencies that led to one or more negative W1 values.

Table VII. *Means ($\pm$ SE) for W_1 to W_7*

	W_1	W_2	W_3	W_4	W_5	W_6	W_7
Mean	1.00	1.59	9.22	24.5	60.1	124	817
SEM		(0.183)	(1.41)	(6.49)	(14.6)	(23.4)	(152)

Application of the severity scale: Calculation of benefit and risk

To illustrate how a severity scale is used to assess benefit and risk, consider the following typical[*] (but hypothetical) example (Table VIII) in which efficacy and ADR statistics are

---

[*] The use of a particular drug is avoided in this example since the purpose here is illustrative. A benefit-risk computation for a particular drug requires accurate frequency data on drug-linked adverse reactions based on a large patient sample.

given after long-term clinical trials. The last column of the table contains the ADR and their frequency grouped according to the severity scale ($W_1, \ldots, W_7$) used here. The drug is an antihypertensive agent which is effective in 90% of the cases. Mild and serious side effects occur with the frequencies shown.

We compute benefit, E, using the severity score for W_5 from Table VII, and multiplying by the frequency 0.9: $E = 0.9W_5 = 0.9 \times 60.1 = 54.1$. Risk is computed using the severity scores for W_1, W_2, and W_6, each multiplied by its frequency: $E' = 0.15W_1 + 0.17W_2 + 0.02W_6 = 0.15(1.00) + 0.17(1.59) + 0.02(124) = 2.9$.

As a second example, consider Table IX which contains data on a drug treatment for rheumatoid arthritis of severity W_4 with a drug that is efficacious in 100% of the cases, but which produces moderate (W_3) and severe (W_6) side effects as shown. Again, the data (though typical) are used only to illustrate. Benefit, $E = 1.0 \times W_4 = 1.00 \times 24.5$. Risk, $E' = 0.25W_3 + 0.025W_3 + 0.10W_6 + 0.05W_6 = 0.275W_3 + 0.15W_6 = (0.275)(9.22) + (0.15)(124) = 2.54 + 18.6 = 21.1$.

Table VIII. *A drug used to treat moderate hypertension (severity, W_5) effective in 90% of cases**

_____

Therapeutic application	Probability of effectiveness	ADRs (frequency)
Antihypertensive agent for moderate hypertension (severity: W_5)	90%	W_1: Headache, drowsiness (15%) W_2: Dizziness, general weakness (17%) W_6: Severe cardiovascular effects (2%)

*ADRs severities W1, W2 and W6 occur with respective frequencies: 15%, 17% and 2%.

Table IX. *Incidence of efficacy and adverse reactions for a drug used in rheumatoid arthritis from which benefit and risk calculations are made*

_____

Therapeutic application	Probability of effectiveness	ADRs (frequency)
Rheumatoid arthritis (severity: W_4)	100%	W_3: Skin eruptions (25%) W_3: GI disturbances (2.5%) W_6: Bone marrow depression (10%) W_6: Proteinuria (5%)

A severity scale can be used to account for partial response. Our questions were based on the premise that a complete cure or alleviation of symptoms of severity W_1 could be achieved, from which we could determine a severity scale. Suppose, however, that a

completely incapacitating illness of severity W_4 (=24.5) responded partially to treatment, so that now the patient's symptoms occurred intermittently (e.g., W_3 = 9.22). The "benefit" in this case is the probability P times the decrease $(W_4 - W_3)$: $P \times (24.5 - 9.22)$ = $P \times (15.3)$ whereas a complete alleviation of the W_4 symptoms with this same probability P would yield the greater benefit $P \times (24.5)$.

Discussion

The concern over ADRs has stimulated studies which deal with their identification, their frequency, and the design of appropriate animal models that can predict their occurrence in man. Equally important in an overall assessment that includes benefit as well as risk is a scale that measure severity of a disease state, whether that state is the purpose of initiating drug treatment or is the result of the treatment. Such a scale, when combined with data on frequency of benefit and ADRs, yields an expectation of values for benefit and risk for the drug.

The responses that we received (the tolerable frequency figures which each respondent entered into the matrix) demonstrate that the respondents have a general notion for a numerical ranking of severity. In other words, they could assign tolerable limits for the frequency of ADRs, viewed against their assessment of the severity of the primary illness being treated. Our efforts have been directed toward extracting the numerical values of severity that follow from their responses. This was accomplished by challenging the subject with difficult cases, each of which involved a different permutation of the same three degrees of severity. In solving for the weights we used the criterion that in each of the three cases the same "profit," the differences between the expectations of benefit and harm, is desired from the treatments. Other criteria might be applied in special cases, but for the general results sought we thought our criterion to be reasonable.

Initially we tried to have the respondents assign severity scores directly. We found that they could not (or would not) make such assignments. The questionnaire matrix approach drew responses, however. The reason is that the case presentations which constitute the matrix mimic the kind of information that physicians and regulators get from clinical trials and from the general literature. This information consists of data on frequency and severity of side effects; usually two or three groupings of severity are used.

Because a scale of severity that is widely accepted will require input from much larger samples, including patients and other responsible citizens who are not physicians, and who will presumably consider more details, the results of this study must be regarded as preliminary. Another limitation of this study is that we have not yet demonstrated intra-observer reproducibility (because of the availability of the physicians). Hence, we do not yet recommend the practical use of the scale derived. Its main value is in the demonstration that the method yields values that agree generally with standards arrived at intuitively. The values obtained are very difference from the rather arbitrary weights 1, 2, 3, and 4 for mile, moderate, serious, and very serious, respectively, that have sometimes been applied.[4] As this scale has been derived from value judgments of

physicians, whose judgments use standards that are considered acceptable, it can, with suitable modifications, be a useful starting model for regulatory agencies, pharmaceutical manufacturers, and physicians in the overall assessment of benefit and risk. Our scale is not intended to be a substitute for sound clinical judgment applied to the individual patient.

We divided the severity into 7 classes, a number large enough to cover most cases, yet small enough that it did not force the respondents to make ultrafine distinctions that would be inconsistent with their realistic prescribing habits. It is interesting that most of the physicians interviewed did not question the use of these seven categories, nor did they disagree with the descriptions associated with each. Our presence during the interviews gave the physicians the opportunity to ask questions, which would not have been possible had we sent them written questionnaires.

There are two points that should not be misunderstood. The first relates to our use of examples along with descriptions. We do not here intend to classify diseases into tight categories. For example, our use of rheumatoid arthritis as an example of a clinical situation of severity W_4 is meant to be a guide. It is an illness that can have this degree of severity. Certainly mild, inactive rheumatoid arthritis would not be put into W_4. There is, thus, potential for overlap. The second point relates to the gap betweens rates of severity W_4 and W_5. Our results are compatible with the average judgment that a life-shortening condition is more severe than a completely incapacitating condition which is usually not life-shortening. Understandably, judgments between these states are highly personal. If one overemphasizes our examples of conditions that can match the stated descriptions for these cases (crippling arthritis, W_4; moderately severe hypertension, W_5), rather than concentrating on the descriptions themselves, the subjective nature of the decision increases dramatically. Individual decision-making often leads to apparent inconsistencies between intuition and theoretical results, as pointed out by Savage.[10] Patients are especially likely to present enormous variation in the assessment of severity, a fact reported by Berg.[2] We did not, in this study, interview patients, yet it is appreciated that the patient viewpoint has to be considered in any comprehensive determination of benefit and risk.

The question most often asked by the subjects was whether or not the ADRs were reversible. For the sake of uniformity, we replied that these untoward reactions should be regarded as "treatable," without giving a precise definition of the treatment which could, of course, include withdrawing the drug. We offered no advice on how decisions are made, but it may be reasonably assumed that each doctor used some set of factors equivalent to those used in representative cases.

Since what we sought were value judgments, we asked our questions without specifying any particular drug or any particular illness. This approach makes the question more challenging and thus taps the respondent's real value judgments to a greater extent than can be achieved if he or she answered about specific clinical situations which, because of their familiarity, might have led to answers based on habit or dogma. Our objective was to determine the physicians' own scale of values. The probability values which we gave

(the diagonal elements of the matrix) were used because the numbers presented a real challenge to the physician, requiring a good deal of thought. In theory, any other set of values could be used and should, in a completely consistent set of replies, produce the same severity scores, but a degree of inconsistency is to be expected whenever a subject is asked about probabilistic clinical situations. Such inconsistencies have been demonstrated in other areas of human behavior as shown by Mosteller and Nogee.[8] We did not check intra-observer reliability in this study because of the practical difficulty (and the imposition on the involved physician) that such a test would entail. Our objective was not to produce a guide for individual patient management, but to extract judgments and test our model, and this omission does not seriously detract from our objective. Intra-observer reliability must be checked before this or any other scale is put to practical use. We attempted to minimize error by using numbers which were thought – provoking. Answering a question about a particular drug that is effective, say in 90% or the cases, and that induces an ADR in less than 0.01% is no challenge to the physician respondent; the challenge comes when the ADR incidence rises to 1%, 5%, or 50%.

A model for assessing severity can ask questions about probabilities for benefit and risk before the initiation of treatment, or it can continually assess probabilities during treatment, gauging the response to the treatment. We might call the former "a priori" and the latter "a posteriori." We have used the former because the objective in an initial study is to develop weights or indices of severity of the symptoms presented. It would seem that as the physician examines the patient's response to a course of treatment, he is presented with new sets of symptoms and changing probabilities, and he would adjust the therapy accordingly. Thus, the importance of time on the calculations should be noted. For example, ADRs that occur only after sustained use, and which add to the risk, can be neglected in an initial assessment of benefit and risk. Ideally, we would like to have a time-dependent frequency table of ADRs. Such dynamic modeling is a subject for further research.

The approach which we have taken, because it extracted opinions of severity based on judgments in general situations (rather than familiar situations whose treatments might be determined by habit or dogma), represents, we think, a sensible approach to the establishment of a scale of severity since our results show a reasonable progression of increasing severities. More complex models can be made, but we emphasize that no broadly constructed scale can be a substitute for individual clinical situations which can be judged only by the doctor and the patient. A severity scale, when used in conjunction with accurate data on incidences of beneficial and adverse effects of drugs, is, we believe, the proper approach for the broad determination of drug benefit and risk and, consequently, for better regulation and usage.

_____

_____

References

1. Alder HL, Roessler EB: Introduction to probability and statistics, ed. 2, San Francisco, 1962, W.H. Freeman & Co., p. 54.

2. Berg RL: Health status index, Chicago; 1973, Hospital Research and Educational Trust.

3. Green DM: Pre-existing conditions, placebo reactions, and "side effects," Ann Intern Med **60**:255-265, 1964.

4. Guy W, Bonato RR: Manual for the ECDEU Assessment Battery N.I.M.H., July, 1970.

5. Jick H: Drugs – remarkably nontoxic, N Engl J Med **291**:824-828, 1974.

6. Karch FE, Lasagna L: Evaluating adverse drug reactions, Adverse Drug Reaction Bulletin **59**:204-207, 1976.

7. Karch FE, Smith CL, Kerzner B, Mazzullo JM, Weintraub M, Lasagna L: Adverse drug reactions – a matter of opinion, Clin Pharmacol Ther **19**:489-492, 1976.

8. Mosteller F, Nogee P: An experimental measurement of utility, J Political Econ **59**:371-404, 1951.

9. Reidenberg MM, Lowenthal DR: Adverse non-drug reactions, N Engl J Med **279**:678-679, 1968.

10. Savage LJ: The foundations of statistics, New York, 1954, John Wiley & Sons, Inc.

11. Wardell WM: Therapeutic implications of the drug lag. Clin Pharmacol Ther **15**:73-96, 1974.