Environmental Health Perspectives
Vol. 13, pp. 43–50, 1976

Problems in Extrapolating Toxicity Data for Laboratory Animals to Man

by R. L. Dixon*

Some of the problems in extrapolating laboratory animal toxicity data to man are considered. The quantitative predictiveness of preclinical studies of anticancer drugs using dogs and monkeys for man has also been examined. The relationship between the maximum tolerated dose (MTD) in the dog, monkey, and the more sensitive of the two species and clinical observations are discussed. The effectiveness of using doses expressed on the basis of body weight (mg/kg) and body surface area (mg/m²) are compared. A method is introduced to assess the "statistical risk" associated with the extrapolation of the initial clinical (phase I) dose from experimental animal data. The best clinical prediction is obtained when one uses the experimental MTD expressed in mg/kg for the more sensitive of the large animal species (dogs or monkeys). The clinical introduction of a new anticancer agent at a dose 1/10 the MTD in the more sensitive species carries a statistical risk of about 3%; that is, the initial doses of about 3 of every 100 new drugs introduced into the clinic will produce some toxic effects in man. These same data have been extended theoretically to the total population and toxic chemicals in general. Reliable extrapolation from laboratory test models to man requires a much more complete understanding of structure–activity relationships, pharmacokinetic factors, and mechanisms of toxicity.

There appears to be worldwide agreement concerning the fact that extrapolating laboratory animal toxicity data to man remains a major unresolved problem in toxicology. Predictiveness of laboratory models must include both qualitative (clinical signs, chemical, hematologic, and pathologic lesions) and quantitative (dose or exposure level) aspects. Sidorenko and Pinigin (*1*) and other Soviet scientists (*2*) have described the multifaceted questions which remain with regard to predicting biological effects and identifying harmful environmental substances. A complete understanding of the toxicodynamic parameters must be achieved to allow reliable predictiveness. The lack of understanding in this area makes the building of mathematical models for extrapolation unreliable at the present time.

The ever-present need to establish reasonable and safe maximally permissible concentrations for various environmental chemicals adds emphasis to this task. Soviet and American toxicologists agree that greater efforts must be devoted to revealing the relationship between the physicochemical properties of a substance and its biological effect (structure–activity relationship, SAR), to studying the effects of accumulation, potentiation and antagonism, and to evaluating the state of adaptation of test animals.

Safety evaluation is the major goal of toxicology while the reliable estimate of human risk is the greatest challenge. Weil (*3*) has offered the following guidelines for experiments where the results are to be used to predict the degree of safety of a material for man.

1. Wherever practical or possible, one or more species should be used that biologically handle the material qualitatively and/or quantitatively as similarly as possible to man. For this, metabolism, absorption, excretion, storage, and other physiological effects might be considered.

2. Where practical, several dose levels should be used, on the principle that all types of toxicologic and pharmacologic actions in man and animals are dose-related. The only exception to this should be the use of a single, maximum dosage level if the material is relatively nontoxic; this level should be

*National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709.

a sufficiently large multiple of that which is attainable by the maximum applicable hazard exposure route, and should not be physiologically impractical.

3. Effects produced at higher dose levels are useful for delineating mechanism of action, but for any material and adverse effects, some dose level exists for man or animal below which this adverse effect will not appear. This biologically insignificant level can and should be set by use of a proper safety factor and competent scientific judgment.

4. Statistical tests for significance are valid only on the experimental units (e.g., either litters or individuals) that have been mathematically randomized among the dosed and concurrent control groups. It is to be understood that statistical significance may be of little or no biological importance, and, conversely, that important biological trends should be further examined even in the absence of statistical significance.

5. Effects obtained by one route of administration to test animals are not *a priori* applicable to effects by another route of administration to man. The routes chosen for administration to test animals should, therefore, be the same as those to which man will be exposed. Thus for example, food additives for man should be tested by admixture of the material in the diet of animals.

Weil properly emphasizes the importance of pharmacokinetics, dose–response relationships, toxic thresholds, statistical reasonableness, and selection of the appropriate route of administration. Although these aspects should need no further emphasis, it is apparent from the problems which continually arise regarding safety assessments that many toxicologists fail to fully appreciate these important points.

Our increased understanding of pharmacokinetics, especially the role of enzymes which either degrade or activate chemical substances, as well as the recent significant advances in *in vitro* and lower animal biological tests emphasized the complexity of this overall problem.

There are three questions that must be considered regarding the interspecies comparison of a toxic agent: What chemical is the toxic agent? How much of the agent is present at sites of action? How long is that agent present?

The question of identity of the agent must include studies of metabolic alterations of the compound by each of the species considered. The question of amount or concentration of agent is a function of not only the metabolism, but also of distribution and elimination rates as well as various environmental and physiological factors. How long the susceptible cells or receptors are exposed to the toxic agent is a function of environmental exposure, distribution and elimination by metabolism, tissue uptake, and/or excretion. These are all problems which involve the pharmacologic disposition or pharmacokinetics of the toxic agent.

If one is trying to extrapolate from a laboratory experiment to man, it is also important to ask how well the laboratory test situation reflects man and his environment. There are obviously great differences in the genetic make-up of the human population. Every aspect of the handling and elimination of a chemical by the body is potentially involved in this human heterogeneity. It is also necessary to realize the very selective nature of most experimental test populations. Experimentalists tend to select vigorous, well fed, healthy animals to extrapolate to a population which contains subpopulations that have all varieties of illness, weakness, and disease. This problem has been described by Rall as the "median mouse" to "median man" consideration. That is, in a very homogeneous population under strict environmental control, what are the differences in response between a very small mammal with its own peculiar set of metabolic processes and a relatively large mammal, man, with his own peculiar set of physiological, biochemical, and pharmacological processes? It must be kept in mind that the final organism one is attempting to protect is not "median man" but every single individual in a very large and diverse population.

It is readily apparent that there are differences among experimental animals, between experimental animals and man, and among different individuals of the same species, but there are also similarities. Knowledge of these similarities and a proper accounting for differences will eventually lead to rational pharmacokinetic models. Such models will allow an investigator to rapidly synthesize many observations and quantitatively describe the time-course of drug concentrations at various tissue receptors in various species.

Zharako and Dedrick (*4*) have attempted to use existing anatomical, biochemical, and physiological information to construct more meaningful model systems for interspecies comparisons. With respect to the pharmacokinetic aspects of these types of models, the compartments and rate constants have a physiological basis in addition to a drug data basis. These model systems are being developed in an effort to "scale up" data from one species to another. These investigators studying the

pharmacological disposition of methotrexate in mice have been able to predict plasma and several tissue drug concentrations in rats, dogs, monkeys, and humans when the mouse model parameters were adjusted appropriately on the basis of known or measured physiological and pharmacological differences among the species.

Each compartment of their model is identified with an anatomic rather than functional space which permits incorporation of concepts from physiology, membrane transport, and enzyme kinetics at the local site. Furthermore, attempts can be made to correlate metabolism, transport, and cell response *in vivo* with measurements made *in vitro*.

However, one must be continually aware that the mathematically correct solution may not always represent the biological solution. Kinetic models like whole animal tests usually apply only to the average or median animal in the population. Although progress has been made, the complicated nature of interspecies comparison and biological test procedures have largely ruled out any attempt to offer some universal mathematical extrapolation factor which would allow interspecies comparisons (including man) for all environmental substances; neither is there such a factor to equate *in vitro* test results to whole animal studies.

Most of the evaluations regarding the predictiveness of laboratory animal toxicology for man have focused on therapeutic agents in the United states. The publications of Schein (*5*) and Freireich (*6*) and their co-workers regarding antineoplastic agents are especially important in this area. These studies were supported by complete preclinical toxicology and carefully conducted clinical studies. The problems with regard to environmental agents are even more difficult. Environmental toxicology is concerned primarily with the biological effects of chemicals that are encountered by man either incidentally because they are in the atmosphere, by contact during occupational or recreational activities, or by ingestion with water or food. In contrast to therapeutic agents, no one is entirely free of exposure to a variety of chemicals capable of producing undesirable effects on biologic tissues. The real and potential hazards of environmental chemicals are difficult to define, exposure levels are hard to quantitate, and acute toxicity is much less of a concern than are long-term risks such as carcinogenesis and mutagenesis.

The usefulness of animal studies in predicting irreversible toxicity such as mutagenesis and carcinogenesis and the extrapolation of exposure levels to estimate human risk are especially difficult problems and are receiving deserved attention. These areas will be discussed by others at this symposium.

The problems involved in extrapolating to humans results of newly developed test systems applied *in vitro* and lower animals, although not discussed here, is an area needing additional study and perhaps should be considered for future collaborative efforts.

This paper will be concerned with the prediction of toxicity in man by using laboratory models, especially dogs and monkeys. Initially, I would like to describe some of our work with regard to the quantitative relationship of drug-related toxicity, and more specifically with the extrapolation from preclinical studies of the starting dose for the initial (phase I) clinical trial in man (*7*). The data to be described here are the result of applied toxicologic studies performed in support of the clinical introduction of new chemicals for the treatment of cancer. These considerations are somewhat superficial and restricted due to the fact that many questions regarding the chemical mechanisms of action, and the adsorption, distribution, excretion, and metabolism of these drugs are unanswered. We examined the quantitative predictiveness of dog and monkey toxicologic studies for the human patient. Our analyses attempt to answer two questions: (1) What is the quantitative relationship of drug doses based on kilogram of body weight (mg/kg) and square meter of body surface (mg/m^2) for prediction of toxic doses in man? (2) What is the "statistical risk" of toxicity associated with various extrapolations of the initial clinical dose from dog and/or monkey toxicity studies? The application of these data to environmental chemicals and the selection of reasonable safety factors will also be considered.

The primary source of data used are the publications by Schein et al. (*5*) and Freireich et al. (*6*). These papers evaluate the quantitative and qualitative predictiveness of experimental animals for man, and include tabulated toxicity data for many antineoplastic agents. Data are summarized as the maximum tolerated dose (MTD) for each anticancer agent for a variety of treatment regimens based on doses expressed in terms of mg/kg and mg/m^2. All of the large experimental animal data were obtained from studies performed for the Laboratory of Toxicology on contracts with research laboratories. Clinical data were collected by the National Cancer Institute and its cooperating clinical groups. Three important simplifying restrictions were applied to the handling of the

data: (1) only drugs which were administered as multiple daily doses (five or more) in both experimental animals and man were included; (2) evaluation of the data was based on total dose administered; and (3) the individual authors' definitions of the maximum tolerated dose were accepted.

For comparison of dose bases, MTDs were plotted as logarithms, least-squares regression analysis performed, and comparisons made either mathematically or visually.

To estimate the statistical risk associated with the extrapolation of initial clinical doses from dog and/or monkey toxicity data, ratios of the clinical and experimental (clinical/experimental) MTDs were determined and ranked from the lowest to the highest ratio. A cumulative percent figure was calculated for each ranked ratio by dividing and ranking by one more than the total number of drugs being studied ($n + 1$). When the clinical/experimental MTD ratios are expressed as logarithms (or numerically on log-probit paper), the population of ratios (drugs) is normally distributed. If one draws a random sample from a population which is normally distributed, the ordered observations would be expected to approximate a linear function of representative values. This procedure is based on a statistical method known as the empirical cumulative distribution function. Stated simply, the approach is essentially equivalent to the log-probit transformations used routinely in pharmacology to determine drug levels associated with median effects. For the present purposes, it is sufficient to state that this method provides a convenient way of describing and comparing the distribution of values and requires only the assumption of normal distribution of the logarithms of the MTD ratios which can be demonstrated.

This method allows for the determination or estimation of the statistical risk associated with the clinical introduction of any new drug when the clinical/experimental MTD ratios are associated with a cumulative fraction of the total sample of ratios for similar drugs. For example, if the mean ratio of clinical to animal tolerated dose was 1 (perfect predictability) then, depending on variation, one would expect half of the drugs to be less toxic to humans and half to be more toxic if the experimental MTD was administered initially to man. This, of course, is never done. A point of particular interest is the percent probability of exceeding a certain clinical/experimental MTD ratio, e.g., 0.1, for the more sensitive experimental species and man. This would indicate statistically the percent of drugs introduced into the clinic which are more than 10 times as toxic to man than to the experimental species. New anticancer agents are most often introduced into the clinic as 1/10 the mg/kg MTD in the most sensitive experimental species (dog or monkey).

Pinkel (*8*) studied the toxicity of antitumor agents and found interspecies correlation to be very good, provided doses were expressed in terms of milligrams of drug per square meter of body surface area (mg/m^2) rather than milligrams per kilogram body weight (mg/kg). Freireich et al. (*6*), studying a variety of antitumor agents and using several animal species, have extended the work of Pinkel. Our study further analyzes the data of both Freireich et al. (*6*) and Schein et al. (*5*). A tabulation was made of the maximum tolerated doses (MTDs) for monkey, dog, and man for more than 40 antitumor drugs. The predictiveness of dog and monkey MTDs for man has been examined for doses based on both mg/kg and mg/m^2. Since it is the practice to use the MTD of the more sensitive of the two species as the basis for extrapolation of phase I clinical doses, a similar analysis was carried out using the more sensitive species as the basis of the independent variable.

To further investigate the relationship between doses expressed as mg/kg and mg/m^2, data from Schein et al. (*5*) were added to those for the drugs previously studied by Freireich. Figure 1 presents

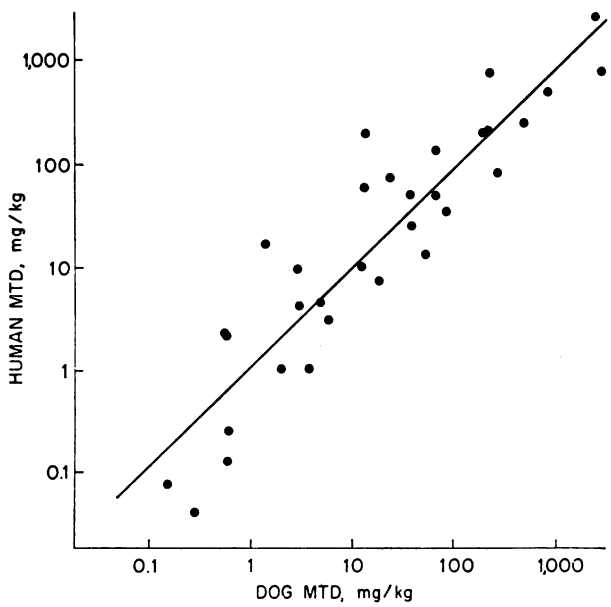

FIGURE 1. Relationship between human and dog MTD based on mg/kg doses.

these data for the dog and compares the MTDs for dog and man based on mg/kg. This representative plot indicates the scatter of the actual data points. Data are expressed as logarithms because of the range of doses encountered. If the relationships between the MTDs in the two species were perfect, then the data would obviously fall along the diagonal line.

Figure 2 presents the calculated best-fitting straight lines for MTDs for the dog, monkey and most sensitive of the two species tested compared to man. Data based on mg/kg and mg/m^2 are represented. It is obvious that there is little difference between the curves.

Table 1 summarizes the important points from this type of analysis. The mean MTD ratios (clinical/experimental) were calculated statistically from the log distributions. The expected median of the ratios of clinical/experimental MTD was derived by regressional analyses. For mg/kg doses, the average calculated clinical MTD is 0.9 times the dose observed in dogs. That is, the clinical MTD is 10% less than the predicted dose. Clinical mg/kg MTDs are, on the average, half the dose predicted by monkey studies and have a clinical/experimental MTD ratio of 0.5. If the most sensitive animal species is used and the dose expressed on mg/kg basis, the prediction is nearly perfect. The clinical/experimental MTD ratio is 1.0. The more sensitive species provides a safety margin over using either dog or monkey data alone.

Using mg/m^2 as the basis for the MTD does little to improve the quantitative predictiveness (Table 1). The calculated clinical experimental MTD ratio for dogs and monkeys is 1.7 and 1.6, respectively. Therefore, the clinical dose associated with minimal toxicity is more than 50% greater than the

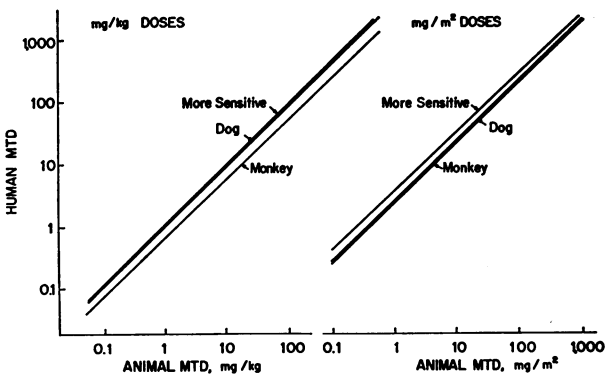

FIGURE 2. Relationship between human and experimental MTD.

MTD in the experimental species. When the most sensitive species is considered on a mg/m^2 basis, the ratio of the clinical MTD to animal MTD is 2.2 indicating that the clinical dose is more than twice that predicted by large animal toxicology. Therefore, although the mg/m^2 extrapolation appears useful with smaller laboratory animals, the conversion adds little to the extrapolation of dog and monkey data to man.

The major effort of our recent data evaluations has been directed to a consideration of the correlation between experimental animal and clinical observations with respect to MTDs and the subsequent extrapolation to the clinical situation. Clinical/experimental MTD ratios were ranked and graphed as previously described. The individual ratios of clinical/animal MTDs were considered to be members of a normalized cumulative distribution and plotted using log-probit transformations. With regression analysis, one can then associate any dose ratio with a cumulative fraction of the sample of ratios. Perfect prediction would be a clinical/experimental MTD ratio of 1.0 corresponding with the 0.5 (50%) level of the cumulative distribution. The cumulative fractions may be considered numerically equivalent to the probability that a given dose extrapolation will exceed the clinical MTD, and thus provide an estimate of the clinical risk for an untried clinical drug candidate based on animal data and patterns of relationships of similar drugs. As in the previous analyses, comparisons were made among dog, monkey, and the more sensitive of these experimental species as sources of the denominator of the MTD ratio and between mg/kg and mg/m^2 as means of dosage expression.

Figure 3 presents a representative plot. The cumulative percent figure is indicated as the probability of exceeding human MTD on the ordinate (probit units); the abscissa is the logarithm of the clinical/experimental MTD ratio. This figure

Table 1. Comparison of observed experimental and clinical maximum tolerated doses.

Basis of comparison	Experimental animal	Mean MTD ratio, clinical/ experimental
mg/kg	Dog	0.9
	Monkey	0.5
	More sensitive	1.0
mg/m^2	Dog	1.7
	Monkey	1.6
	More sensitive	2.2

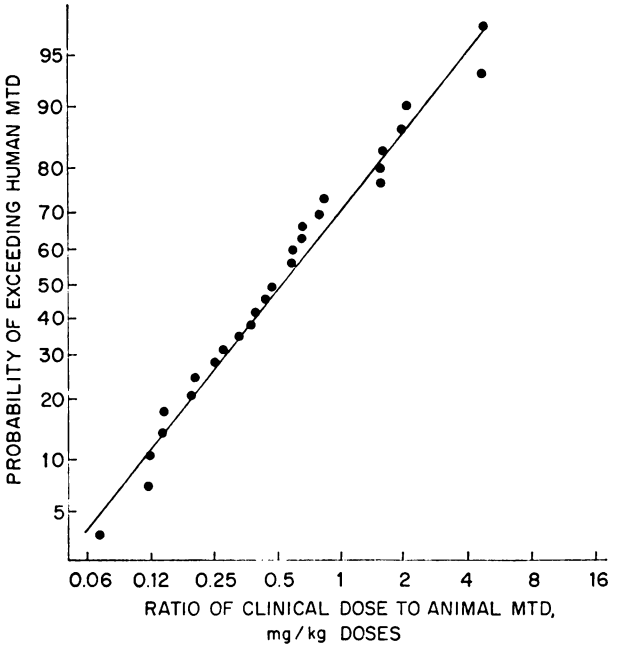

FIGURE 3. Estimation of clinical risk from clinical/monkey MTD ratios.

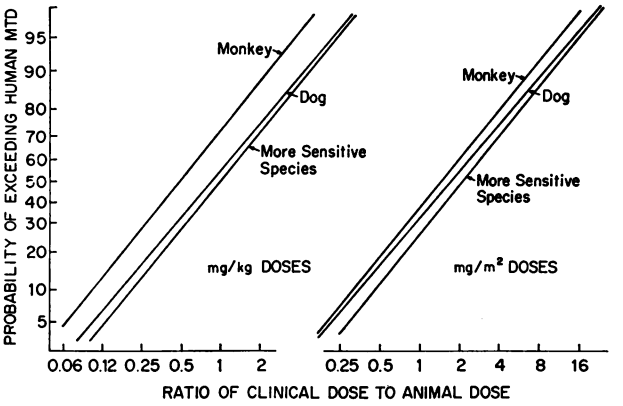

FIGURE 4. Comparison of estimates of clinical risk.

Table 2. Probability of exceeding human MTD with doses based on MTD ratios.

Basis of comparison	Experimental animal	Dose ratio	Probability of exceeding human MTD, %
mg/kg	Dog	0.1	4.1
	Monkey	0.1	8.7
	More sensitive	0.1	2.9
mg/m²	Dog	0.1	1.3
		0.33	10.0
	Monkey	0.1	1.1
		0.33	9.7
	More sensitive	0.1	0.5
		0.33	5.9

presents data for monkeys, and the doses are expressed as mg/kg. The normal distribution of the data is apparent.

Figure 4 is the calculated best-fitting straight lines for dog, monkey, and more sensitive species when compared to man. Data for doses based on mg/kg and mg/m² are included.

Table 2 presents tabulated figures on special interest. The probability of exceeding a clinical/experimental dose ratio of 0.1 (1/10) is indicated. This would indicate that man is more than 10 times as sensitive as the experimental animal and estimates the safety factor achieved when the initial clinical dose is calculated as one tenth the MTD in mg/kg for the most sensitive species. A clinical dose one-tenth the dog mg/kg MTD carries a 4 percent risk of exceeding the clinical MTD. Statistically, 4 of 100 drugs will be introduced at a dose greater than the human MTD. The corresponding risk based on monkey data is 9%, and for the more sensitive species, 3%.

More recently, it has been suggested that the initial clinical dose be calculated as one third the experimental MTD in mg/m² for the most sensitive species. Analysis indicates equivalent risks for dogs and monkeys are about 10% in each case; when the most sensitive species is used, the risk is reduced to approximately 6%. This risk is about twice that of using one tenth the MTD dose expressed as mg/kg.

There is no question that some level of risk must be accepted in the selection of the initial clinical doses to insure that large numbers of seriously ill patients do not receive ineffective drug levels. On the other hand, the dose must not be so high that the patient is subjected to unreasonable toxic hazards.

The percent risk determined statistically in these studies estimates the probable percentage of new drugs that will exceed the MTD with the initial dose during phase I clinical trials. Therefore, two points are clear: clinicians conducting the phase I trial must define an acceptable level of risk for clinical toxicity, and it must be recognized that statistically, with any level of risk selected, eventually a certain percentage of drugs will be introduced into the clinic at doses which will produce toxicity. The difficult question, of course, remains: What level of risk is acceptable?

The acceptable risk for the general population with regard to environmental chemicals is, of course, very different from that just described for

Table 3. Probability of environmental chemicals exceeding human toxic threshold based on extrapolation from experimental animal toxicology.

Experimental toxicological values	Percentage of experiment dose for varying degrees of safety					
	10% [a]	5% [a]	1% [a]	0.1% [a]	0.01% [a]	0.001% [a]
MTD						
Dog (mg/kg)	17.13	9.46	3.05	0.84	0.29	0.114
Dog (mg/m^2)	34.59	19.14	6.19	1.71	0.59	0.233
Monkey (mg/kg)	14.00	8.36	3.13	1.02	0.40	0.179
Monkey (mg/m^2)	32.06	17.97	5.94	1.68	0.59	0.237
LD						
Dog (mg/kg)	1.51	0.65	0.13	0.020	0.004	0.001
Dog (mg/m^2)	2.76	1.16	0.22	0.034	0.007	0.002
Monkey (mg/kg)	0.78	0.35	0.07	0.013	0.003	0.001
Monkey (mg/m^2)	5.48	2.76	0.75	0.168	0.049	0.017

[a]Percentage of total chemicals estimated to produce toxicity upon human exposure.

cancer patients. However, in this area also, some degree of risk must be accepted; few chemicals are without potential harmful effects.

The Environmental Biometry Branch at the NIEHS has been helpful in extending these observations regarding clinical/experimental dose ratios for antineoplastic drugs and the accompanying percent risk of initial toxicity to the at large population and toxic chemicals in general. Both doses maximally toxic and lethal to experimental animals are considered. There is one very important caveat—an assumption is made that the 20–24 drugs studied represent a normally distributed population. This assumption appears valid, but requires further study.

Table 3 summarizes these data. The percentages of the experimental dose for varying degrees of safety with regard to numbers of chemicals are indicated. In this case, the percentage of the experimental MTD associated with a predetermined level of risk was determined. A good example is the MTD dog (mg/kg) data because it represents the NCI's clinical experience. Taking approximately 1/10 (9.46%) of the MTD dose in dogs indicates that statistically 5% of the drugs (or chemicals) will be introduced at a dose in excess of the human MTD; 1/100 (0.84%) of the experimental dose in dogs reduces the risk to 1 drug (or chemical) in 1000. One-thousandth (0.114%) of the experimental dose further reduces the risk of about 1 in 100,000 chemicals which will produce some toxicity in the human population. This appears to be a low estimate of the percentage of chemicals found to have harmful human effects in today's world. A single dose safety factor of 5,000 from the lowest observed effect level has recently been proposed by Weil. This factor probably provides adequate protection of the general population from reversible chronic toxicity. In fact, in many cases it is probably excessive. However, with regard to irreversible chronic toxicity, the validity of this extrapolation factor has yet to be determined; in these cases it might be totally inadequate.

In summary, greater emphasis on achieving an understanding of all aspects of interspecies variations is needed. Reliable extrapolation from laboratory test models to man requires a more complete understanding of structure-activity relationships, pharmacokinetic factors, and the mechanisms of toxicity. Until that future time of greater understanding, safety factors must be determined for each substance considered based not only on extrapolation from animal studies, but supported by the informed judgment regarding as many parameters as possible, and by selecting those factors least likely to underestimate the risk for man.

REFERENCES

1. Sidorenko, G. I., and Pinigin, N. A. Gigiena i Sanit. 39: No. 3; 93 (1972).
2. Cooperation Between the USSR and the United States in the Field of Environmental Health, Problem I: Methodological bases for evaluating the biological effect of chemical pollutants during their entrance by inhalation (Topic I: Studying the quantitative relationship dose–time–effect with the objective of predicting the biological effect). National Survey of Literature, Public Health Ministry of the U.S.S.R., Moscow, 1973.
3. Weil, C. S. Guidelines for experiments to predict the degree of safety of a material for man. Toxicol. Appl. Pharmacol. 21: 194 (1972).
4. Zaharko, D. S., and Dedrick, R. L. Pharmacokinetic models for antineoplastic agents. In: Handbook of Experimental Pharmacology, O. Eichler, A. Farah, H. Herken, A. D.

Welch, Eds. Vol. 38, No. 1 (A. C. Sartorelli and D. G. Johns, Eds), 1974.
5. Schein, P. S., et al. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. Clin. Pharmacol. Therap. 11: 3 (1970).
6. Freireich, E. J., et al. Quantitative comparison of toxicity of anticancer agents in mouse, rats, hamster, dog, monkey, and man. Cancer Chemother. Repts. 50: 219 (1966).
7. Homan, E. R., and Dixon, R. L. Quantitative toxicologic relationship between preclinical and clinical studies of anticancer agents. In: Pharmacology and the Future of Man (Proc. 5th Int. Congr. Pharmacology, San Francisco, 1972) 3: 289 (1973).
8. Pinkel, D. The use of body surface area as a criterion on drug dosage in cancer chemotherapy. Cancer Res. 18: 853 (1958).