JAMA

THE JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION

EDITED FOR THE ASSOCIATION UNDER THE DIRECTION OF THE BOARD OF TRUSTEES BY

JOHN H. TALBOTT, M.D.

VOLUME 177

JULY 8–SEPTEMBER 30
1961

AMERICAN MEDICAL ASSOCIATION, CHICAGO 10

Forecasting Drug Effects in Man from Studies in Laboratory Animals

John T. Litchfield, Jr., M.D., Pearl River, N.Y.

Experiments on animals are the most important source of data for predicting the effects of administering a new drug to patients. Nevertheless, the predictive value of such experiments is limited. In the present retrospective study of 6 drugs of different types, it is shown that many of the most serious side effects that can result when a drug is given to man were not predictable from observations on dogs or rats. From an initial list of signs of toxicity that could occur in man, 39 physical signs were retained that could also be observed in the dog or the rat, or both. Analysis showed that effects on man could be predicted better from observations on dogs than from those on rats.

HAND IN HAND with the growth of medical research has been the steady increase in the number of new drugs introduced, and in the extent to which each drug is investigated in the laboratory and clinic. The primary objective of much of this investigation is to determine if and how the new drugs may be used safely in man. This is also a legal prerequisite to the sale of the drug, since, under the provisions of the Federal Food, Drug, and Cosmetic Act, a new drug cannot be marketed until its safety in man has been established. The method of study used to determine the safety of a new drug in laboratory animals is influenced by regulatory interpretations of this Act. However, in the face of the pressures created by the tremendous growth of such work, it is easy to lose sight of the real purpose of these studies. Thus, a restatement and reexamination of this purpose is in order.

Studies on the safety of new drugs in laboratory animals are intended to develop knowledge which will help to protect the patients who are to receive the new drug by forewarning the doctor of its possible dangers. If this basic purpose is clear, it is apparent that these studies must permit a prediction of what will occur when the drug is used in man. If this were not so, there would be no reason to conduct these elaborate and costly studies.

One may still ask, "On what evidence is the predictive value of data from laboratory animals based?" Rather astonishingly, there is no good evidence to answer this question; instead, it is generally assumed that such predictive value exists.

It is true that many drugs have been shown to have a particular activity both in animals and in man. This is a very narrow aspect of the problem, however, because no drug has a single action, although it may have a predominant one compared to all of the actions which it can exert. Thus, since every possible action of a drug must be evaluated, the task of showing the predictive value of animal studies is formidable. The literature discloses no example of a critical comparison of the total actions of 2 or more drugs observed in laboratory animals with the total actions later found in man. In fact, the available evidence along these lines led Barnes and Denz,[1] in their comprehensive review of methods for determining chronic toxicity, to conclude that the conventional procedures for carrying out these studies in the laboratory are entirely empirical and have little scientific basis, and that extrapolation of the results of these studies to man is a matter of guesswork. If this indictment is correct, then there is little purpose in most of the toxicity tests on new drugs which are performed on laboratory animals today, because the findings, in essence, are uninterpretable when they are applied to man.

In view of the fact that apparently no one has previously attempted systematically to predict from laboratory data what would happen subsequently in man, a retrospective study of 6 drugs was undertaken.

The drugs were selected because they met the following criteria: 1. Detailed studies in the rat, dog, and man were available (500 case reports, minimum for man). 2. All studies were performed within the past 7 years. 3. All studies were performed under comparable conditions with respect to the standards observed. 4. All drugs were unrelated in chemical structure.

Director, Experimental Therapeutics Research Section, Lederle Laboratories, Division of American Cyanamid Company.

The drugs were chosen from the following classes: antibacterial, tranquilizer or central nervous system depressant, glucocorticoid, and antialcoholic.

The studies in animals comprised acute experiments, as well as chronic experiments of 1 to 3 months', 6 months' (dogs), and 1 year's duration (rats). All studies of 30 days or longer included detailed gross and microscopic pathological examinations. These studies, which were planned and conducted jointly by pharmacologists and pathologists, were directed toward exceeding the tolerance of the animals for the drug, and every effort was made to discover the actions of the drug by careful observation of the animals. The number of dogs used was relatively small compared to rats; consequently, observation of the dogs was more comprehensive than that of the rats. In the case of man, case reports available varied from 800 to 7,500, depending on the drug under investigation (average 3,300).

All the available data on each drug were examined, and every drug effect which had been noted was tabulated according to species studied. At this point, several requirements became apparent:

1. The incidence of drug effects in animals was not comparable to the incidence of drug effects in man because, in the former, dosage was pushed to the point of intolerance while, in the latter, no such procedure was or could be followed. It was necessary, therefore, to disregard incidence; either a drug effect occurred or it did not.

2. A limitation on the vocabulary of drug effects was necessary in order to eliminate synonymous terms. For example, there are many possible kinds of anemia and also a variety of ways in which anemia can be detected. Therefore, for the purpose of this study, it was sufficient to use the general term "anemia" rather than a variety of other descriptive terms.

3. It is possible to observe only physical signs in animals, not symptoms. This required eliminating from the tabulations those symptoms recorded in man. Similarly, any physical sign which could not occur in an animal also was eliminated. For example, the rat cannot vomit, although dog and man may. Physical signs in this study included results of laboratory examinations.

4. It was found to be impossible to judge the relative importance of one physical sign over another. Thus, anemia may be fatal in one case but insignificant in another.

Under these requirements, the data on each drug could be tabulated as shown:

	Rat	Dog	Man
Impaired reflexes, hypotension, ataxia, decreased activity	+*	+	+
Weight loss	+	+	—†
Tremors	..	+	+
Ptosis, urinary incontinence, catatonia	+	..	..
Lacrimation	..	+	..
Diarrhea	..	..	+

*+=occurrence
†—=not observed

From these tabulations a complete list of physical signs was prepared, and an appropriate entry was made under each species to indicate the presence or absence of a sign produced by each drug. Finally, it was required that the list include only those physical signs which were noted at least once in the rat or dog. This eliminated 16 physical signs which were reported in man but which had no counterpart in either the rat or dog. Each one of the remaining 39 different physical signs had been found with one or more drugs in either or both rats and dogs.

Finally, the data on the 39 different physical signs were classified so that an appropriate entry for each physical sign and each drug was made under one of the following classes:

Class	Meaning
Absent	Sign not elicited in rat, dog, or man
Rat	Sign elicited only in rats
Dog	Sign elicited only in dogs
Man	Sign elicited only in man
Rat, dog	Sign elicited in both rats and dogs but not man
Rat, man	Sign elicited in both rats and man but not dog
Dog, man	Sign elicited in both dogs and man but not rat
Rat, dog, man	Sign elicited in rats, dogs, and man

The incidence of physical signs noted in each of these classes was then determined. Tests of significance were made by calculating $(Chi)^2$ as

$$\sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}.$$

Results

When all of the physical signs observed from use of the 6 drugs in rats and dogs were collected, synonymous terms consolidated, and signs which could not occur in all 3 species eliminated, the data were tabulated as shown in Table 1. It will be noted that for each physical sign, the 6 drugs must be accounted for in 2 or more of the classes shown. For example, for the first line (weight loss) 2 drugs caused this sign in both rats and dogs but not in man, while 4 drugs elicited this sign in all 3 species. There are 8 classes in the tabulation because the data were treated as if coming from a 2^3 factorial. The "absent" class, representing those instances in which a drug failed to produce a given sign in any of 3 species, is by far the largest.

The objective of this tabulation is to contrast those findings present in man with those findings not present in man. This is shown in Table 2, which compares the observed incidence of physical signs in the 4 classes excluding man with the corresponding 4 classes including man. The first line of the table shows that physical signs from use of the 6 drugs were not observed 146 times in any of the 3 species, while signs were observed only in man and not in rats or dogs 23 times. The over-all total of 234 represents the product of 6 drugs times 39 physical signs. For each class the contribution to $(Chi)^2$ is given. A large $(Chi)^2$ value permits the inference that a correlation is present between the

different species. The $(Chi)^2$ of 43.5 is significant, and results from the fact that the observed frequencies were quite different in several cases from the expected value, which is calculated on the assumption that all classes are alike in relative frequency. [$(Chi)^2$ computed omitting the "absent" and "man" classes is 9.7, n = 2, and $p<0.01$.]

Having established the fact that a significant relationship among the 3 species is present, the task is to discover the most effective way of utilizing this information to predict the carry-over of physical signs from animals to man. Altogether, 5 different methods are evident. The first is that it could be predicted that either all or none of the physical signs would occur in man. This method represents the "no experimentation" rule: Either no experimental information on the rat or dog is available, or such information is ignored in the belief that there is no correlation between species. As is shown in Table 3, if all signs are predicted to occur in man, 53 out of 234 predictions (or 23%) would have been correct. Conversely, if none of the signs were predicted to occur in man, 181 out of 234 (or 77%) of the predictions would have been correct.

Table 1.—Occurrence of 39 Physical Signs from Six Drugs in Three Species

Physical Signs	Absent	Rat	Dog	Rat and Dog	Man	Rat and Man	Dog and Man	Rat, Dog, and Man
Weight loss	...	...	...	2	...	...	...	4
Weight gain	3	1	...	...	2	...	...	...
Muscle atrophy	4	...	...	...	1	...	...	1
Myositis	5	...	...	...	...	...	1	...
Lymphocytopenia	4	...	1	...	...	...	...	1
Neutropenia	5	...	...	1	...	...	...	...
Leukopenia	3	...	1	...	1	...	...	1
Anemia	1	...	2	...	1	...	...	2
Leukocytosis	3	...	1	...	2	...	...	...
Hyperglycemia	5	...	...	...	...	...	...	1
Liver damage	2	1	2	...	...	...	1	...
Jaundice	5	...	...	...	...	...	1	...
Fatty liver	5	1	...	...	...	...	...	...
Polydipsia	5	...	1	...	...	...	...	...
Polyuria	2	...	...	...	3	...	1	...
Oliguria	4	...	...	...	1	...	1	...
Hematuria	3	...	1	...	...	...	2	...
Crystalluria or renal concretions	5	...	...	...	...	1	...	...
Renal damage	3	1	1	...	...	...	...	1
Gastroduodenal ulcer	4	...	1	...	...	...	1	...
Diarrhea	1	...	1	...	4	...	...	...
Salivation	5	...	1	...	...	...	...	...
Ataxia	2	...	1	1	1	...	...	1
Impaired reflexes	3	...	...	1	1	...	...	1
Decreased activity	3	...	...	2	...	...	...	1
Tremors	3	...	...	...	2	...	1	...
Ptosis	5	1	...	...	...	...	...	...
Catatonia	5	1	...	...	...	...	...	...
Priapism	5	1	...	...	...	...	...	...
Lacrimation	5	...	1	...	...	...	...	...
Urinary incontinence	4	1	...	...	1	...	...	...
Bacterial invasion	5	...	...	...	...	...	1	...
Parasitic invasion	4	...	...	...	1	...	1	...
Decreased thyroid function	4	...	...	1	...	...	...	1
Genital hypoplasia	4	2	...	...	...	...	...	...
Decreased adrenal function (cortical)	5	...	...	...	...	...	...	1
Hypotension	2	1	...	...	2	...	...	1
Lung edema	5	...	...	...	...	...	1	...
Tachypnea	5	...	1	...	...	...	...	...
Totals	**146**	**11**	**16**	**8**	**23**	**1**	**12**	**17**

Table 2.—Summary of Occurrence of Physical Signs by Factorial Classes with Contributions to $(Chi)^2$*

Not Man		(Obs.−Exp.)²/Exp.	Man		(Obs.−Exp.)²/Exp.	Total
Class	Observed		Class	Observed		Observed
Absent	146	1.7	Man	23	5.9	169
Rat	11	0.4	Rat, man	1	1.3	12
Dog	16	1.6	Dog, man	12	6.0	28
Rat, dog	8	6.4	Rat, dog and man	17	20.2	25
Totals	**181**	**10.1**		**53**	**33.4**	**234**

* $(Chi)^2 = 43.5$, n = 3, $p < 0.01$.

Table 3.—Results of Five Different Methods of Predicting Occurrence of Physical Signs in Man

Predict to Carry Over to Man	Class	Class: Not Man	Class: Man	Correct Predictions as % of Total: Incidence in Man	Correct Predictions as % of Total: Predictions Made	$(Chi)^2$
	Total	181	53	...	23	...
All signs						
Signs seen only in rats	Not rat	162	35			
	Rat	19	18	34	49	19
Signs seen only in dogs	Not dog	157	24			
	Dog	24	29	55	55	40
Signs seen only in rat and dog	Not rat and dog	173	36			
	Rat and dog	8	17	32	68	80
Signs seen in rat or dog	Not rat or dog	146	23			
	Rat or dog	35	30	57	46	27

The remaining 4 methods consist of using physical signs seen only in (2) rats, (3) dogs, (4) both rats and dogs, and (5) rats or dogs as the basis for predicting their occurrence in man. Table 3 shows the results when these 5 methods were employed. The first method uses the totals shown at the bottom of Table 2 and mentioned above. The second method utilizes the rat as a predictor. In the "not man" column, the "absent" and "dog" classes of Table 2 are combined to give a total of 162 for the "not rat" class, and the "rat" and "rat, dog" classes are combined to give a total of 19 for the "rat" class. The values for the "man" column are obtained in a similar manner. When the rat is used as the basis for predicting, 18 out of 53 (or 34%) of the physical signs observed in man were predicted correctly, which is a little better than the 23% which can be obtained without experimentation. However, 49% of the positive predictions made (18 out of 37) were correct, and this accounts almost entirely for the significant value of $(Chi)^2$.

The remaining 3 methods given in Table 3 were similarly calculated. It was already noted in Table 2 that 23 physical signs were seen only in man. In this case, no basis existed for predicting these signs, since the drugs involved had failed to elicit the signs in rats or dogs. The objective, therefore, is to predict correctly as many of the remaining 30 signs as possible or, in other words, to maximize the correct predictions as per cent of total incidence in

man. This is not the only objective, however, since it is also desirable to maximize the correct predictions as a per cent of the predictions made. Table 3 shows that both of these percentages most nearly approach simultaneous maxima in the method in which the dog is used as the basis for prediction. This is verified by the value of $(Chi)^2$, which is also maximal for this method. For the data of this study, therefore, the physical signs observed in dogs were of the most value in predicting the drug effects in man.

Very little has been said about predicting the absence of physical signs in man, and this kind of prediction merits comment. The results of predictions may be listed as (1) successful prediction of occurrence of a sign in man; (2) failure to predict occurrence of a sign in man; (3) successful prediction of absence of a sign in man; (4) failure to predict absence of a sign in man. The first outcome represents achieving, while the second one represents failing to achieve, the purpose of toxicity studies in animals or, in other words, warning or failing to warn of a hazard to the patient. The third and fourth outcomes are not comparable to the first 2. There would seem to be about as much value in predicting correctly that a particular physical sign not seen in animals with use of a particular drug will not occur in man as there would be in predicting correctly that sunrise will not occur at midnight tonight. Similarly, failure to make such a prediction is equally inconsequential. However, in some cases, a drug did produce a physical sign in only the rat or dog. In these cases, there is a basis for predicting its occurrence or lack of occurrence in man. In terms of this study, the most efficient use of this information would be to predict that signs seen only in rats (and not in dogs) would not occur in man, while all signs seen in dogs would be predicted to occur in man.

Table 4.—Results of Predictions*

Predictions of Sign	Correct	Incorrect	Total
Occurrence	29	24	53
Absence	11	1	12
Total	**40**	**25**	**65**
Per Cent	62	38	100

*$(Chi)^2 = 6.77$, $n = 1$, $p < 0.05$.

The results obtained on the basis of these predictions are shown in Table 4. Although the per cent of correct predictions is better than would be expected by chance, the reason for this is mainly due to the relatively high score achieved in predicting the absence rather than the presence of physical signs.

Finally, the group of physical signs which were noted only in man should be considered. These consisted of nasal congestion, localized fat deposition, constipation, gastrointestinal irritation and inflammation, aplastic anemia, thrombocytopenic purpura, bradycardia, interstitial myocarditis, anuria, edema, cystitis, vaginitis, trismus, chills, fever, and dermatitis. There is no reason to doubt that most of these are drug-induced effects. Some of these effects can be extremely serious, and animal studies offer little hope in predicting most of these effects in man.

Comment

The purpose of this study was to find a method which could be used to examine the hypothesis that studies in rats and dogs have predictive value in determining the effects of drugs in man. Although this is a generally accepted hypothesis, the literature discloses no report in which the total spectrum of drug effects of 2 or more drugs tested in animals has been examined to establish the extent to which their effects occur in man. An examination of the nature of studies in animals as compared to those in man immediately discloses certain basic differences. In animals, drugs are usually given in intolerable doses in order to elicit physical signs of drug action, and the reality of any physical sign is judged primarily by the criterion of dose dependency. In man, however, no such practice is or can ordinarily be observed. Information obtained from human case reports is almost entirely limited to dosage in the therapeutic range. Aside from the therapeutic effect, all other possible actions of the drug are reported as side effects, and a collection of case reports would give the incidence of each of these. Excluding symptoms which have no counterpart in speechless animals, one might attempt to decide which physical signs in man are drug-related on the basis of their frequency of occurrence. However, this approach fails completely because an infrequent physical sign from a drug may be entirely reproducible in a given patient, while a physical sign that occurs frequently may represent no more than a component of the disease or disorder which is being treated. Therefore in this study it was necessary to adopt the inefficient assumption that all physical signs observed in man were drug-related. Consequently, if the results of this study had failed to show that studies in animals have predictive value for man, it could be argued that this failure was due to the crudeness of the approach. On the other hand, since this crude qualitative approach demonstrated predictive value of animal studies, one may infer that refined quantitative studies would certainly verify this conclusion. Although the desired result was attained, there is no less need for examining the physical signs of drug action in man much more critically in order to establish which signs are drug-related and which are not.

With respect to the animals studied, the dog was found to be considerably more useful in predicting the drug's effect in man than was the rat. This is consistent with today's practice of using a third species, such as the monkey or chicken, in place of, or in addition to, the rat. No data have been pre-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.