FIRST DOSE OF POTENTIAL NEW MEDICINES TO HUMANS: HOW ANIMALS HELP

Peter Greaves, Andrew Williams‡ and Malcolm Eve§*

The need for careful testing of new drugs in animal models before study in humans has been recognised by physicians since the First World War. Now, first human studies on new drugs are subject to detailed government guidelines, which in the European Union are presently being reinforced through the wide-ranging *Clinical Trials Directive*. However, despite their long history and widespread application, these guidelines are empirical and have been formulated with a paucity of critical scientific evidence. Here, we review the principles and the available, albeit limited, evidence that support the design and conduct of preclinical studies in a way that permits effective and safe first-dose studies of potential new medicines in humans.

SULPHANILAMIDE
An old antimicrobial drug that is still occasionally used therapeutically.

**Medical Research Council Toxicology Unit, Hodgkin Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK.*
‡Marix Drug Development, Rhodfa Marics, Llantrisant, CF72 8UX, UK.
§Pfizer Clinical Research Unit, Kent and Canterbury Hospital, Ethelbert Road, Canterbury, Kent CT1 3NG, UK.
Correspondence to P.G.
e-mail: pg29@le.ac.uk
doi:10.1038/nrd1329

At the end of the First World War it was recognised that close collaboration between pharmacologists, physiologists, physicians and medicinal chemists was required to test new medicines that were then being derived from coal tar[1]. In the years between the First and Second World Wars, the need for careful testing of the mode of action and toxicity of new drugs in animal models prior to testing in humans was widely accepted by physicians[2]. However, it was not until at least 76 people died from poisoning with an elixir of SULPHANILAMIDE containing 72% diethylene glycol that this need was legislated in the United States. In their analysis of these deaths, Geiling and Cannon[3] summarized the key principles for testing new drugs, which are still relevant today (BOX 1). After the Second World War, these principles were widely applied to the study of new drugs. Added impetus for appropriate preclinical testing came from the Nuremberg Code[4] (BOX 2). This important document still serves as a blueprint for today's principles that ensure the rights of subjects in medical research.

At the end of the First World War, chemists recommended that preclinical testing of new drugs be done by well-funded, independent multidisciplinary institutes, primarily owing to the lack of funding and interdisciplinary cooperation in academia[1]. Non-commercial testing has not occurred to any significant extent; now, almost all testing is done or commissioned by the pharmaceutical industry. Within the pharmaceutical industry, resources and skills have been made available for collaboration between pharmacologists, chemists, kineticists, pathologists, physicians and toxicologists to permit understanding of any adverse effects and safe testing in humans. The less desirable corollary to this testing being done or sponsored by industry is the development of 'guidelines' by government agencies which aim to assure compliance with acceptable standards. This has given rise to a whole industry of 'regulatory' toxicology in which much of the growth has been driven by demands for protocols to submit to government regulatory authorities. In 1963, J. M. Barnes, a director at the Medical Research Council's Toxicology Unit in the United Kingdom, foresaw the negative effect of guidelines laid down by government regulatory agencies. He noted that, whilst these agencies often emphasize that they are not laying down precise rules for toxicity testing, potential vendors of new drugs who must satisfy governments are inclined to follow such recommendations closely as a means of attaining official acceptance and a marketing licence[5]. Moreover, he commented that scientific study of the hazards from new drugs would dwindle if the tests recommended by authorities were too detailed and were performed

Box 1 | **Principles of drug testing prior to trials in humans[3]**

- Exact composition of drug should be known; if not, method of preparation
- Acute toxicity studies in animals of different species
- Chronic toxicity experiments at varying doses in different species for cumulative effects
- Careful and frequent observations of animals, to develop a composite picture of clinical effects
- Careful pathological examination of tissues with appropriate stains
- Effects of drugs on excretory or detoxifying organs, especially kidney and liver
- Rate of absorption and elimination, path and manner of excretion, concentration in blood and tissues at varying times
- Possible influence of other drugs and foodstuffs
- Careful examination for any idiosyncrasies or untoward reactions

unthinkingly, simply to supply a mass of data to regulatory authorities. This danger persists, particularly under the obligations of the *Clinical Trials Directive*[6] of the European Union.

The clinical challenge

The Nuremberg Code[4,7] (BOX 2) and the amended Declaration of Helsinki[8] underline the level and nature of the information needed because they define the ethical constraints of experimental studies in humans. A study must be designed to yield results that are valuable to society, thereby justifying any risk. Risk should be clearly defined, minimal and clearly explained to volunteers.

The aim of the first study of a new drug in humans is to explore the dose and exposure range that is well tolerated and, if possible, to identify any dose-limiting adverse events. Detailed physiological monitoring, particularly of the cardiovascular system, permits assessment of any drug-induced alterations in the function of important organ systems. Laboratory analysis of body fluids is also used to assess cellular toxicity. Insight into pharmacokinetics is obtained through measurement of circulating or excreted drug and drug-related products. With some classes of agent it is also possible to assess whether the intended pharmacodynamic effects occur at doses that are well tolerated, prior to more extensive testing in patients. Achievement of these aims represents a major leap from the laboratory bench to humans, and requires a substantial body of information characterizing the drug substance.

Agents with high potential toxicity for volunteers need to be excluded, although a level of toxicity that is unacceptable for healthy volunteers might be acceptable for volunteer cancer patients. Information that permits monitoring of significant organ toxicities should be provided. Quantitative upper safety limits need to be set at a level that is appropriate to the type of experiment, or dose-limiting toxicity must be clearly defined. Selection of the starting dose is important as a dose that is too low will provide no information and will unnecessarily prolong the experiment. As the main aim of human experiments is to define primary pharmacological effects, these effects must first be stringently characterized in appropriate animal models. Exaggerated or secondary pharmacological effects must also be identified, as the most common type of drug toxicity in humans is that of a pharmacological nature[9]. Moreover, a coherent package of preclinical information requires an integrated understanding of pharmacological and toxicological responses; that is, dose and exposure relationships across the various preclinical models used.

All of these activities are now dominated by stringent standards of Good Clinical Practice as well as GOOD LABORATORY PRACTICE (GLP) and GOOD MANUFACTURING PRACTICE (GMP). Proposals for the testing of new agents in humans are also subject to independent ethical review.

The evidence for preclinical testing

Design of studies. Preclinical studies have been conventionally divided into those of primary pharmacology; secondary (or safety) pharmacology; toxicology; and drug kinetics (or toxicokinetics). Primary pharmacological studies are the most variable in nature, being dependent on the particular type of agent under study. They can be carried out *in vivo* or, increasingly, *in vitro*. Safety or secondary pharmacology studies are generally more standardized animal studies using mainly physiological monitoring of vital organs or organ systems. Toxicology studies have been standardized by GLP guidelines that embody daily dosing of animals, general clinical examination and monitoring, and clinical pathology testing of blood and urine, followed by extensive histopathological examination of tissues after detailed NECROPSY. These studies are accompanied by the measurement of drug or metabolites in body fluids termed drug kinetics, or toxicokinetics if more specifically related to toxicology studies.

The available data. Data comparing the power of preclinical studies to predict effects in humans remain limited, and are dominated by toxicology data obtained in compliance with GLP and government regulations. This is reflected in the small number of published reviews on this topic.

One exception is in Japan, where greater emphasis has been traditionally placed on secondary pharmacology studies and the Japanese Ministry of Health and Welfare have guidelines for their conduct. As such, a review of the predictability of adverse reactions based on secondary pharmacology studies has been published[10]. In addition, the published review of 45 drugs in the database of the Committee on Safety of Medicines (CSM) in the United Kingdom[11] includes both toxicity and preclinical pharmacology studies.

The largest and most recently published review of the performance of animal toxicity studies is by Olson and colleagues[12] who analysed data on 150 drugs that caused adverse events or toxicity in humans. Although the Japanese Pharmaceutical Manufacturers Association reviewed the animal and human test data of 139 new drugs in 1994 (REF. 13), this review was limited to publications in the Japanese literature. The Centre for Medicines Research in the United Kingdom has also completed a number of questionnaire-based reviews of pharmaceutical companies about the fate of new drugs

GOOD LABORATORY PRACTICE (GLP). This defines a set of rules and criteria for the organizational processes and the conditions under which preclinical safety studies are planned, performed, monitored, recorded, reported and archived.

GOOD MANUFACTURING PRACTICE (GMP). This defines an assurance process that is similar to GLP. It ensures that products are consistently manufactured and controlled to the quality standards that are appropriate to their intended use.

NECROPSY
A detailed post-mortem examination. Also referred to as autopsy.

Box 2 | **The Nuremberg Code**

- The voluntary consent of the human subject is absolutely essential (for full text of this article see REF. 7).
- The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.
- The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.
- The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
- No experiment should be conducted where there is an *a priori* reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
- The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
- Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability or death.
- The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.
- During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.
- During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him, that a continuation of the experiment is likely to result in injury, disability or death to the experimental subject.

tested in humans and the reasons for termination of drug development, including toxicity[14,15]. Few, if any, of these retrospective reviews contain data from newer biotechnology-derived products.

Other reviews are much less recent and have generally focused on anticancer drugs. However, these peer-reviewed publications are particularly insightful as they critically reviewed and compared preclinical and human data on more severe toxicities. Particularly pertinent in this respect are the reviews of 25 anticancer drugs in dog, monkey and human studies[16], 21 diverse cancer chemotherapeutic drugs in rodents, dogs, monkeys and humans[17] and the highly cited quantitative comparison of the toxicity of 18 anticancer drugs in mice, rats, hamsters, dogs, monkeys and humans by Freireich and colleagues[18].

There is an almost complete lack of detailed published reviews of proposed new chemical entities that were intended for testing in volunteers but failed to reach this stage because of serious toxicity in preclinical studies.

Organ systems

Nervous system. General pharmacological tests for effects on the nervous system are usually observational studies of rodent general activity or multidimensional functional assays of motor activity. For a series of 84 new drugs (excluding anticancer agents) studied in Japan, an evaluation of their capacity to predict adverse reactions in humans showed a general nonspecific correlation. For example, changes in locomotor activity in rodents correlated with dizziness in humans[10]. A degree of overprediction was reported, particularly from studies that used high doses. Similarly, in the study of 45 miscellaneous drugs by Fletcher[11], high-dose effects such as ataxia and convulsions in animals did not occur in humans, and subjective effects such as dizziness, headache, dry mouth and sweating in humans were not predicted by animal studies. The correlation was stronger for other effects on the central nervous system (FIG. 1).

Where effects on the central nervous system have been assessed in conventional toxicity studies using both clinical monitoring and histopathological examination of the brain and nervous tissue, a reasonable degree of concordance has been shown. Evaluation of the effects of up to 25 diverse anticancer drugs in dogs, monkeys and humans showed a reasonable degree of concordance (nearly 40%) in neurological and neuromuscular toxic effects[16]. Dogs and monkeys had similar predictive value and high doses were needed to achieve the best correlation (FIGS 2 and 3). Specific symptoms correlated poorly. The earlier study of 21 anticancer drugs by Owens[17] indicated only a moderate correlation between neurotoxic effects in humans and animals (TABLE 1). The correlation was strong for alkylating agents, but less so for other classes of drug studied. Interestingly, the study of 150 miscellaneous drugs by Olson and colleagues[12] showed that, overall, the non-rodent data were better correlated with adverse neurological effects in humans than the rodent data.

So, whilst the data indicate poor prediction of subjective neurological effects, the information on the significant toxicities of anticancer drugs indicates that the conventional approach using histopathological examination detects potentially serious neurotoxic effects.

Special sense organs. Relatively few instances of visual, auditory or vestibular disturbances are reported in early clinical studies with new drugs. As such, there is a paucity of data comparing effects on these functions between laboratory animals and humans. However, ophthalmoscopic examination is usually performed in toxicity studies and is routinely accompanied by histopathological examination of the eye, prior to dosing of humans with a new drug. Consequently, it is probable that any agent that provokes severe ocular damage in animals after relatively short periods of dosing would not progress to clinical studies. Moreover, agents that have potent cataractogenic properties or that are severely toxic to the retina would be identified in relatively short, repeat-dose studies.

Specific tests of auditory function are seldom done routinely. But careful clinical observation of animal comportment probably eliminates agents that produce acute and severe auditory or vestibular damage.

Cardiovascular. Cardiovascular effects have been assessed in the dog model for many years. Moreover, the primary pharmacological effects of most cardiovascular drugs in use today were initially evaluated in the dog

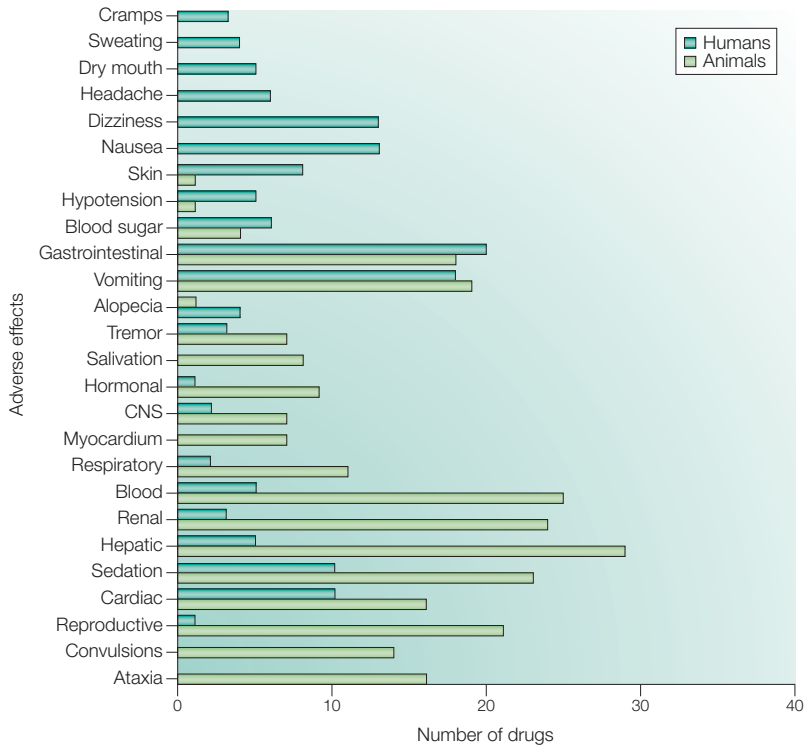

Figure 1 | **Animal and human toxicities of 45 drugs assessed by the Committee on Safety of Medicines in the United Kingdom during the eight or nine months prior to publication in 1978** (REF. 11)**.** Data are for drugs of diverse therapeutic classes, including several cardiovascular and central nervous system drugs but only one anticancer agent. The six uppermost adverse effects were observed in humans but not in animals; the two adverse effects at the bottom of the graph were observed in animals but not in humans. For most adverse effects there is a degree of over- or under-prediction. CNS, central nervous system.

model. In the review of 25 anticancer drugs by Schein and colleagues[16], where evaluation was based on general clinical assessment and pathology rather than on detailed physiological monitoring, studies that used both primates and dogs failed to predict cardiovascular toxicity in one out of ten cases in which it was observed in humans. The authors suggested that physiological measurements could have improved predictive capacity for this series of compounds. More recently, the study by Olson *et al.*[12] demonstrated good concordance between cardiovascular findings in dogs and humans. The correlation seems less robust between humans and rodents, possibly due in part to the technical difficulties associated with monitoring cardiovascular function in rodents.

Following a substantial number of life-threatening cardiac arrhythmias linked to electrocardiographic QT INTERVAL PROLONGATION in patients treated with a range of non-cardiovascular therapeutic drugs, a group of experts convened by the European Agency for the Evaluation of Medicinal Products suggested that *in vitro* electrophysiological studies should be undertaken for all new drugs[19]. However, this suggestion has been challenged on the basis of species differences in ion channel expression, pharmacology and kinetics, which make the extrapolation of findings from *in vitro* electrophysiological studies difficult[20–22]. A review of published data indicated that *in vitro* studies do not confer additional reliability on data obtained in a more traditional *in vivo* model[23]. So, for the physician responsible for first dose to human studies, electrocardiographic assessment in dogs at crucial time points (notably at peak plasma concentrations) provides the most useful data for translation into clinical study design and effective monitoring.

Pulmonary tract. A large body of data has accumulated on experimental methodology for examination of the effects of environmental and occupational chemicals on the respiratory tract. This is because inhalation is a primary mode of human exposure to foreign materials[24]. The effects of drugs are often evaluated preclinically in specific pharmacology studies, usually alongside cardiovascular assessment in anaesthetized dogs. In the comparison of 104 investigational new drugs by Igarashi and colleagues[10] in which this approach was used, respiratory disturbance was not frequently reported in humans. However, those cases that were reported were not predicted by safety pharmacology testing.

In toxicity studies, effects on respiration are usually evaluated by clinical observation and histopathology of lungs and air passages. In the study of 45 drugs by Fletcher[11], both toxicology and pharmacology animal studies over-predicted respiratory effects in humans (FIG. 1). Similarly, Schein and colleagues[16] noted that this form of screening in non-rodents predicted respiratory signs or respiratory pathology in four out of five cases, but with a high percentage of over-prediction (FIGS 2 and 3).

Gastrointestinal tract. The review of safety pharmacology studies performed in Japan on 88 non-cancer drugs showed a good correlation between rodent intestinal transport and general adverse effects such as anorexia and constipation in humans[10].

In the review of conventional toxicology studies that included histopathology of the gastrointestinal tract, Olson and colleagues[12] showed good concordance between gastrointestinal effects in animals and humans, particularly for non-steroidal anti-inflammatory drugs, anti-infective and anticancer agents (FIG. 4). In that review, large animal data were a better predictor than data obtained from rodents. The CSM data also showed good correlation between animal toxicology studies and humans for 45 diverse drugs[11].

The rodent, dog, monkey and human gastrointestinal toxicity data also showed a strong correlation in the study of 21 anticancer drugs by Owens[17] (TABLE 1). Surprisingly, in the study of 25 anticancer drugs by Schein[16], the dog was superior to the monkey as a predictor of adverse gastrointestinal effects in humans. For example, monkeys were remarkably resistant to vomiting, an adverse event that was observed in humans with 21 of the 25 compounds. Gastrointestinal tract toxicity was a significant contributor to the remarkably good quantitative correlation of toxicity across species based on dose/body surface area for the 18 anticancer drugs studies by Freireich and colleagues[18].

QT INTERVAL PROLONGATION
Increase in the total time of ventricular polarization as measured from the onset of the Q wave to the end of the T wave on the electrocardiogram of the heart.

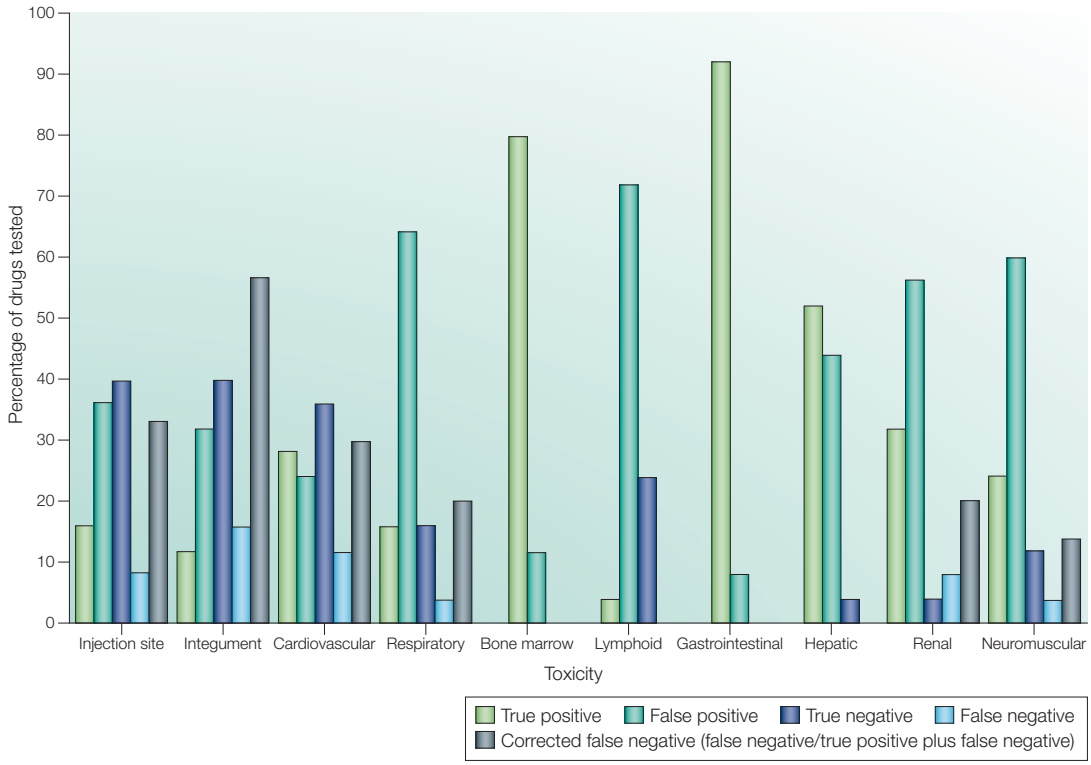

Figure 2 | **The dog as a predictor of organ-specific toxicity (*n* = 25)**[16]**.** True positive, toxicity observed in both dog and human; false positive, toxicity observed in dog but not in human; true negative, toxicity observed in neither dog nor human; false negative, toxicity not observed in dog but recorded in human; corrected false negative (an index of false negatives for only those compounds that produced a specific toxicity in human), false negatives/true positives plus false negatives.

It has been suggested that the gastrointestinal tract of dogs is highly physiologically similar to that of humans, in terms of motility patterns, gastric emptying and pH, particularly in the fasted state[25]. This observation, coupled with the ability to use a formulation similar to that used in humans, makes the canine gastrointestinal tract a most relevant model.

Hepatic. Hepatotoxicity is an important adverse drug effect and a relatively common reason for termination of the development of a new drug[12,15]. At present, drug-induced hepatic injury accounts for more than 50% of cases of acute liver failure in the United States[26]. In conventional preclinical studies of toxicity, the cornerstone of the assessment of hepatotoxic potential is measurement of circulating liver enzymes and hepatic histopathology[27]. A review of 38 chemicals, 24 of which were drugs that produce hepatic toxicity in humans, showed a concordance of 80% with findings in conventional toxicity studies[28]. Hepatic toxicity was not under-predicted in the study of 25 anticancer drugs in dogs and monkeys that used conventional hepatic enzyme measurements and histopathology[16]. The study of anticancer drugs by Owens[17] showed a similar good correlation (TABLE 1).

Conversely, the study of data on 150 drugs exhibiting human toxicities showed that the concordance between hepatotoxicity found in animal studies and that observed in clinical practice was little more than 50%[12] (FIG. 4). This larger study undoubtedly included agents that developed IDIOSYNCRATIC RESPONSES, which are not usually detected in early clinical trials because of their rarity. This is a significant problem — in recent years there have been notable examples of hepatic toxicity of a poorly understood, idiosyncratic nature that have caused the withdrawal of marketed drugs despite extensive and essentially negative preclinical testing and large clinical trials. The thiazolidinedione troglitazone, an antidiabetic drug, was associated with serious hepatic injury in patients despite its lack of hepatic toxicity in preclinical studies[29]. Another example is bromfenac, a non-steroidal anti-inflammatory drug[26]. Hepatic failure also occurred in clinical trials with the nucleoside analogue fialuridine as a result of mitochondrial disturbance and STEATOSIS. Despite long-term treatment of monkeys, dogs and rats with fialuridine, the only hepatic effects observed were increases in apoptosis and nuclear atypia in rats[29].

It is probable that most new drugs that produce severe hepatotoxicity in animals are not tested in humans, so that the true level of concordance is likely to remain obscure. However, overall, the data seem sufficiently robust to conclude that overt liver damage observed in animal toxicity studies indicates potential risk of hepatic toxicity in humans. This underlines the prudence of a critical histopathological examination of liver tissue in preclinical studies and careful patient monitoring in response to any hepatic alerts from animal studies.

IDIOSYNCRATIC RESPONSES
Infrequent adverse responses to drugs that differ from predictable, dose-dependent toxicities. They are characterized by a variable delay or latency period and might cause severe injury and death.

STEATOSIS
A cellular change characterized by an increase in lipid, usually seen as cytoplasmic droplets.

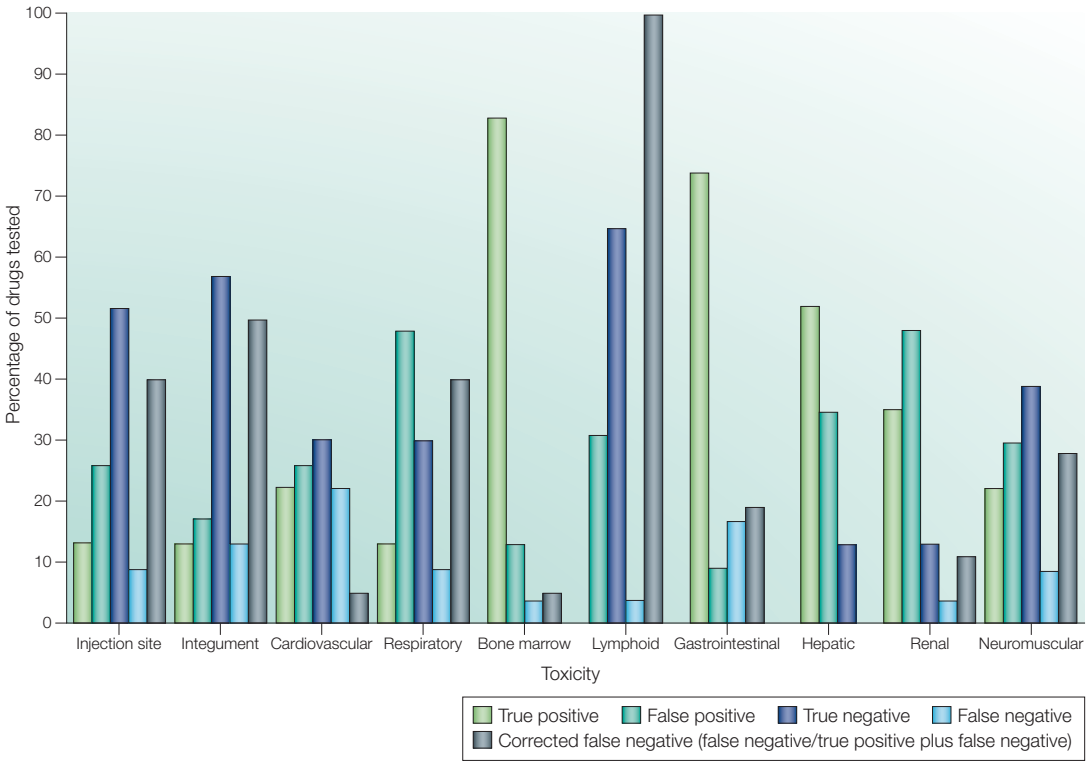

Figure 3 | **The monkey as a predictor of organ-specific toxicity (*n* = 23)[16].** True positive, toxicity observed in both monkey and human; false positive, toxicity observed in monkey but not in human; true negative, toxicity observed in neither monkey nor human; false negative, toxicity not observed in monkey but recorded in human; corrected false negative (an index of false negatives for only those compounds that produced a specific toxicity in human), false negatives/true positives plus false negatives.

Renal and urinary outflow tract. Renal toxicity is assessed by conventional histopathology, measurement of blood urea and electrolytes, and examination of urine volume and the sediment it contains. Concordance in the database of 150 drugs reviewed by Olson and colleagues[12] was fair. A good correlation was noted among the 21 anticancer drugs reviewed by Owens[17], with rodents and dogs performing equally well. However, in the CSM review of 45 drugs, renal toxicity was correctly predicted by animal studies in three instances but over-predicted in 22 others[11]. Similarly, the study of 25 anticancer drugs in dogs or primates correctly predicted renal toxicity in nine cases, under-predicted in one and over-predicted in 14 (REF. 16).

Genital tract. Reproductive changes are seldom reported in early clinical studies, largely due to exclusion of women of childbearing potential from these experiments[11]. Conventional toxicology studies carried out prior to the first human studies include histopathological examination of both the female and male reproductive tracts, incorporating examination of testes and ovaries. It is probable that this approach excludes agents that have potent ovarian or testicular toxicity in humans. Studies of ovarian toxicity in animals and female patients indicate that alkylating agents, such as cyclophosphamide, which can induce ovarian failure in patients also produce histological evidence of follicular damage in animals after relatively short periods of dosing[30]. Similarly, a collaborative study in Japan of 16 drugs, 12 of which are associated with infertility in humans, showed that histopathological study of the testes of rats treated for 2 or 4 weeks was the most sensitive method for preclinical detection of drugs with anti-fertility properties[31].

Endocrine system. Endocrine changes during preclinical studies are routinely assessed only by histological examination of endocrine organs, unless there are particular reasons to suspect endocrine effects. Olson and colleagues[12] noted only moderate concordance (60%) between animals and humans (FIG. 4). As might be expected from the way in which the endocrine system responds to stimuli, these effects were not common in humans and generally occurred after Phase I studies (four out of five cases reported in the database). The review by Fletcher[11] indicated that endocrine findings in preclinical studies significantly over-predict effects in humans.

Endocrine effects, particularly those involving the adrenal gland, are commonly reported in toxicology studies[32]. These findings often represent adaptive alterations to repeated doses of drugs and usually manifest as changes in glandular weight and cellular atrophy or hypertrophy. These changes might not have significant implications for human safety in single-dose studies, but they characterize possible endocrine effects that need to be assessed in clinical trials.

Table 1 | **Cross-species comparison of adverse effects of 21 anticancer drugs[17]**

Type of toxicity	Number showing toxicity in humans	Number showing toxicity in rodent/ number tested	Number showing toxicity in dogs/ number tested	Number showing toxicity in monkey/ number tested
Gastrointestinal	13	9/11	12/13*	6/7
Bone marrow including thrombocytopaenia	13*	9/12*	11/13*	6/6
Hepatic	6	5/6‡	6/6	None tested
Renal	3	3/3	3/3	3/3
Nervous system	7	2/6*	2/7	2/3
Alopecia or dermatitis	6	0/6	0/6	0/1

*One positive finding deemed borderline; ‡two positive findings deemed borderline.

Haemopoietic. Haemopoiesis is routinely assessed by examination of peripheral blood, bone marrow smears, and histopathology of the blood-forming and lymphoid organs. Theus and Zbinden[33] reviewed prior industry practice for the assessment of coagulation in 1984, and found substantial deficiencies. The screening practice that they proposed for animal studies is similar to that used in humans and has now been almost universally adopted for pharmaceuticals testing.

There is substantial data on the concordance of adverse effects on haemopoietic tissue due to anticancer and antimitotic drugs between animals and humans. The evidence indicates good correlation for both rodents and humans and dogs and humans for myelotoxicity, although the particular cell series affected sometimes differs[16]. Thrombocytopaenias were correctly predicted for 13 of 18 anticancer drugs that produced this toxicity in humans. Moreover, in the series of 18 anticancer drugs studied by Freireich and colleagues, haemopoietic toxicity was one of the most significant contributors to the remarkably good quantitative correlation across species based on dose/body surface area[18].

A reasonable correlation between animals and humans was also noted for decreases in white blood cell counts in the study of 139 drugs by the Japanese Pharmaceutical Manufacturers Association[13]. Anticancer drugs and antibiotics did predominate in this series, but the authors also detected a considerable number of false negatives and false positives in their data.

Immunological system. Specific tests of immune function are not routinely performed for conventional new drugs prior to their use in humans. An international collaborative study showed that examination of peripheral blood white cells, histological examination of thymus and spleen and, in particular, careful histological examination of lymphoid tissue in the rat is a good primary method of identifying agents that are significant direct-acting immunotoxins[34]. New screens of immune function in animals are sometimes proposed for drug assessment[35,36] but more sophisticated tests of immune function might be more appropriate and safely conducted in human studies. Coping with the potential impact of biotechnology-derived pharmaceuticals on immune status and immunogenicity is a special challenge for which careful attention to the principles of immunology is needed[37].

Skin. Of all tissues, skin shows the least concordance between effects in animal studies and human patients. A general lack of predictive reliability for skin reactions in humans has been noted in the reviews of anticancer and other drugs[11,16,17]. Adverse skin hypersensitivity effects have caused the development of a relatively large number of potential new drugs to be terminated[12,15,38].

Injection site. Dogs and monkeys correctly predicted human injection site toxicity for four out of six anticancer compounds, although some over-prediction (36%) was reported for the 25 drugs reviewed by Schein and colleagues[16]. However, the animal studies produced no evidence of injection site pain for four out of five compounds that produced this effect in humans. This broad comparability between animals and humans was also evident in a study of several intramuscular preparations that are used clinically. This latter study used serum creatinine phosphokinase activity in rabbits, pigs and humans with concomitant histopathology in rabbits and pigs to compare local damage caused by intramuscular injection[39].

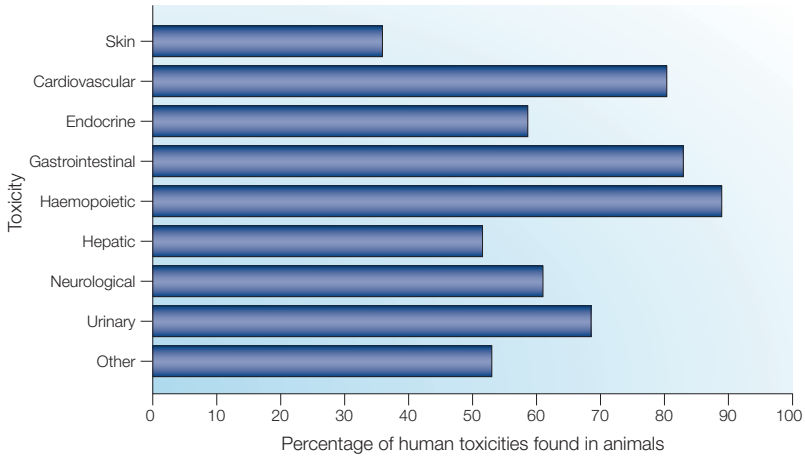

Figure 4 | **Percentage concordance between animal and human toxicities, grouped by organ.** Similarly to data on anticancer drugs, correlation is better for toxicities in the gastrointestinal tract, and haemopoietic and cardiovascular systems. Modified, with permission, from REF. 12 © (2002) Elsevier Science.

Genotoxicity

There has been much contention about the relevance of these assays to the testing of pharmaceutical agents since their introduction more than 30 years ago[40]. However, extensive study has led to a better understanding of the chemical determinants that provoke genotoxic effects through electrophilic attack of biological macromolecules[41]. As a consequence of this understanding, mutagenic activity is often simply avoided in the drug discovery process.

Nevertheless, prior to first human exposure, *in vitro* tests for mutations and chromosomal damage are routinely carried out according to internationally agreed technical guidelines that are based on a large body of historical data for diverse chemicals. However, it can be difficult to assess human risk when unexpected or unexplained activity in these bacterial or mammalian cell tests occurs. Such activity usually precludes dosing to healthy volunteers at least until further work elucidates the mechanism of activity and characterizes any hazard. Subsequently, *in vivo* assays of bone marrow micronucleus, peripheral blood cytogenetics and liver unscheduled DNA synthesis in rodents are usually done. The technical performance of these tests has also been the subject of international collaborative studies.

Standards

Whilst there might be excessive bureaucracy associated with regulation of GLP by government agencies, this now widely accepted international standard provides assurance to the physician that preclinical data have been generated according to an acceptable standard. This is particularly useful when studies are carried out in different laboratories in different countries and continents.

GMP provides a degree of assurance regarding the purity and stability of the material being dosed to humans. However, any significant deviation in quality from that of the test material used in the preclinical studies necessitates additional safety assessments.

Duration of animal studies

Most of the data reviewed here compare the adverse effects of drugs in animals with those in humans based on the empirical approach using repeat-dose toxicity testing of at least 2 weeks duration together with pharmacological testing that is usually based on single doses.

A proposal made by Monro and Mehta[42] that single-dose toxicity studies in animals are sufficient to support single doses of a new drug in humans has been accepted in certain instances; most notably, when applied to a series of investigational compounds with closely related chemical structures[43]. This proposal was based on the concept that correlation of toxicity between animals and humans is poor and that many of the adverse effects following single doses in volunteers are pharmacological in nature. Olson and colleagues[12] showed that a significant number of neurological toxicities were detected after single doses, either in specific safety pharmacology studies or as clinical observations after the first dose in multiple-dose toxicity studies. A general correlation between secondary pharmacology studies of neurological

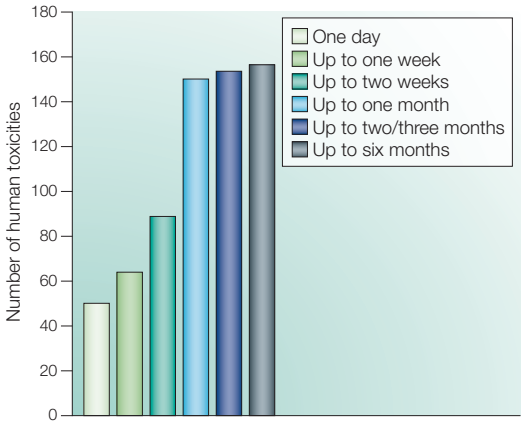

Figure 5 | **Time to first detection of animal toxicity.** The number of toxicities that can be detected in animal systems reaches a plateau at the one-month stage of the study. By this time, 94% of toxicities were detected, but prior to this time some toxicities were not apparent. On the first day, 25% of these observations were from safety pharmacology rather than from toxicology studies. Modified, with permission, from REF. 12 © (2002) Elsevier Science.

toxicity, often of short duration, and human adverse reactions has also been reported[10]. Moreover, a single-dose approach has been used for life-threatening conditions such as cancer.

Nevertheless, whilst the data on 150 drugs[12] indicated that a significant number of adverse effects in humans were predicted by single-dose data, a significant proportion were only detected following studies of up to one month. However, the 90% of human toxicities that can be detected in animals were increasingly detected as the duration of studies increased, up to one month. Few additional toxicities were identified in longer experiments (FIG. 5).

The data discussed above indicate that characterization of certain serious organ toxicities can only be reliably based on histopathological examination of tissues. This requires repeat-dose studies because several days are often required for the expression of pathological change in tissues.

In addition, Rozman and Doull[44] have highlighted the quantitative relationship between dose and time that might be particularly important at the lower end of the dose–response curve.

Dose levels

The use of a maximum tolerated dose in toxicity studies is sometimes contested because it reveals toxicities that are deemed irrelevant to the use of a new drug at pharmacologically active doses in clinical practice. Indeed, on the basis of almost no comparative data or evidence in the peer-reviewed literature, 'microdose' studies at sub-pharmacologically active doses in volunteers have been proposed. This proposal is based on a rat study at non-toxic doses for 7 days and a cardiovascular safety pharmacology study in three dogs[45,46]. Moreover, guidelines utilizing this proposal have emerged

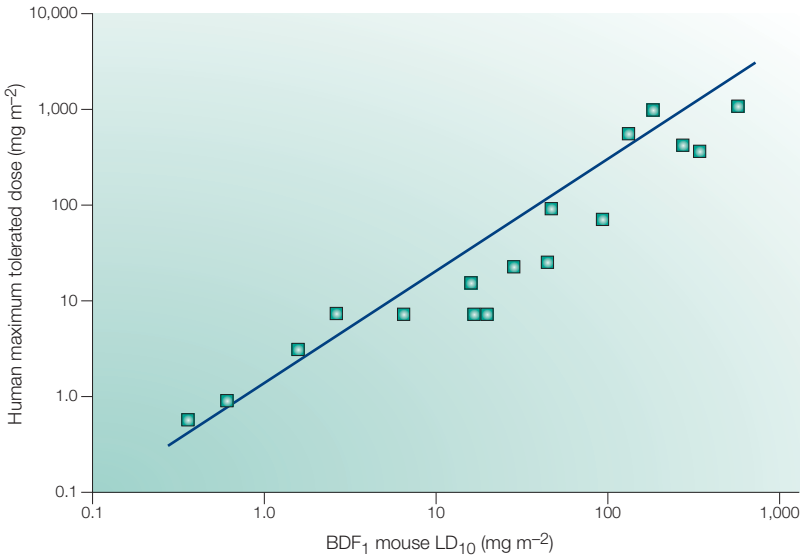

Figure 6 | **Correlation of lethal dose 10% (LD_{10}) in BDF_1 mice with the human maximum tolerated dose for anticancer drugs[18].** The relationship between humans and mice (as well as rats, dogs and monkeys) are close to unity when compared on the basis of mg per m^2 rather than mg per kg.

from the European Committee of Proprietary Medicinal Products[47]. Such an approach might be justified under certain circumstances for series of relatively well-characterized new compounds, but the design of a normal clinical study protocol requires full knowledge of dose–response, therapeutic ratio, and when and at what exposure levels potential adverse effects are likely to occur.

The highly cited quantitative comparison of the toxicity of 18 anticancer drugs in the mouse, rat, hamster, dog, monkey and human by Freireich and colleagues[18] indicated that there is a substantial correlation of maximum tolerated dose between species (humans with rodents, dogs and monkeys) (FIG. 6). Although most toxicity in this study was either haemopoietic or gastrointestinal, it revealed that the ratios of animal to human toxicity on a mg per m^2 basis in these species were remarkably close to unity, supporting the concept that high doses permit characterization of possibly dose-limiting toxicities in humans.

Another consideration is the dose–time relationship. Prolonged dosing tends to show toxicity at lower levels of exposure relative to short-term studies, so there is a trade off between dose and duration of study. It has been suggested that many toxicities do not adhere to a linear dose–response, but show a hormetic or biphasic dose–response[48]. This emphasizes the need for exploration of the entire dose–response curve to adequately characterize a new drug prior to clinical testing in humans.

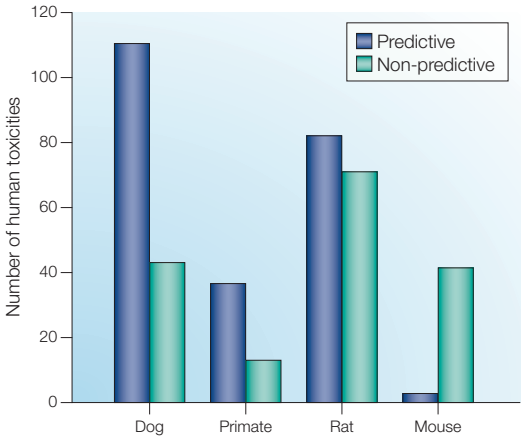

Figure 7 | **Prediction by species.** The dog is a better predictor of human toxicities than rodents, and possibly better than primates, although a number of the primate studies reviewed were small. Modified, with permission, from REF. 12 © (2002) Elsevier Science.

Drug disposition

Closely linked to dose is the study of drug absorption, distribution, metabolism and elimination. It was argued more than 25 years ago that studies of drug metabolism and kinetics should be closely integrated with many other aspects of drug development[49]. This is still not always carried out in a manner that permits an integrated understanding of exposure–pharmacological and toxicological response relationships across the range of preclinical studies[50].

Species

One trend in the studies reviewed above is that the dog is a better predictor of human adverse effects than rodents or, surprisingly, monkeys.

In a detailed study of six miscellaneous non-cancer drugs (antibacterial, central nervous system, glucocorticoid and anti-alcoholic), data showed that effects in humans could be better predicted by observations in dogs than by those in rats[38]. The dog data were better correlated with human adverse effects in the study of 25 anticancer drugs by Schein *et al.*[16]. A review of seven anticancer drugs[51] showed that the dog was a better predictor of toxicity — largely myelosuppression and hepatotoxicity — than the mouse. Data on the 150 compounds studied by Olson and colleagues[12] also indicate that the dog is a better qualitative predictor of toxicity than rodents (FIG. 7).

Conclusions

The widespread perception that preclinical tests, particularly toxicity studies on new drugs, are an exercise in compliance with government regulations disguises the fact that the progression of a potential new therapy from the laboratory to the clinic is still a significant scientific, medical and organizational challenge operating within ethical constraints.

Although largely empirical, the present principles of preclinical safety testing of new drugs are those elaborated by Geiling and Cannon more than 60 years ago[3]. Despite the experience-driven evolution of practical methods through the use of new analytical techniques, better laboratory animals and improved study design and data management, the actual data to support current practice remains limited. However, it is now

unlikely that different laboratory animals that might be more effective predictors of human toxicity, such as the ferret or guinea-pig, will be used.

Although the available performance data on preclinical drug testing are fragmentary, they indicate that the conventional approach of experimental pharmacology together with repeat-dose toxicity studies of up to one month's duration predicts adverse events in first dose to human studies with reasonable success (that is, it identifies more than 90% of the toxicities that can be detected in animal models). There is significant over- and under-prediction of adverse effects from animal studies that varies with the particular organ or system. Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities not predicted by safety pharmacology or toxicity studies[12].

There is evidence that single, high-dose studies will detect many of the important adverse effects in humans, because most of these events are of a pharmacological nature. This supports the concept of dosing humans following single-dose preclinical studies. However, good data on the number of agents eliminated before human dosing because of major organ toxicity are unavailable, so it is impossible to assess how often serious organ toxicities might be missed using the single-dose approach. There are, however, individual reports of significant damage to eye, liver, kidney and testes that have precluded testing in humans, notably in the toxicological pathology literature[52]. So, careful consideration of the nature of the compound, prior experience, clinical study objectives and study design are important when safety studies shorter than one month are proposed in support of human testing. Repeat-dose toxicity testing at high doses for at least a few days to permit the development, expression and identification of organ pathology is clearly advisable where there is no prior knowledge of organ toxicity.

In light of the need for first dose to human studies to provide meaningful data to justify the experiment, to manage adverse events and to avoid serious organ toxicity, integrated primary and secondary pharmacology data and repeat-dose toxicity information across the experimental models used are required.

In general, information obtained from dogs better predicts adverse effects in humans, relative to data from rodents and even monkeys. Indeed, because of the potential overlap between data obtained from clinical observation and that gleaned from physiological measurement in dog toxicity and safety pharmacology experiments, there is an opportunity to rationalize and reduce animal usage without reducing the quality of preclinical evaluation. This review also highlights the reliance of the conventional approach on good clinical observation in toxicity studies for the detection of neurological changes, respiratory perturbations and alterations to the special sense organs. Changes in these systems are often more easily assessed in dogs.

It is important to recognize that there have been many exciting new developments in the preclinical study of drug toxicity, such as those using genomics technologies and structure–activity relationships. These developments might lead to more rapid or earlier selection of new targets and new drugs, as well as investigation of basic mechanisms in toxicity, but there is little evidence that they have had a major and direct impact on the safety assessment which supports the very first studies of a new drug in humans.

Serious and uncommon idiosyncratic reactions involving liver, skin and haemopoiesis are major problems in drug development. Unfortunately, they are difficult to predict from animal studies or to detect in early clinical studies with few subjects, and new methods are clearly required. Using new animal models is one possibility, but most idiosyncratic reactions result from a succession of events that are probably related to genetic variation. This raises the possibility of prevention by pharmacogenomic study in humans[26].

Finally, much pertinent information is held by government regulatory authorities and pharmaceutical companies, and is not available in the peer-reviewed scientific literature. As noted by Barnes in 1963, in this field, as in so many others, we can only learn from experience and only then if we have access to information[5].

1. Editorial. War chemistry in the alleviation of suffering. *J. Ind. Eng. Chem.* **10**, 673–674 (1918).
2. Leake, C. D. The pharmacologic evaluation of new drugs. *JAMA* **93**, 1632–1634 (1929).
Report of an early symposium on the testing of new drugs, which outlines some of the key principles for preclinical and clinical evaluation.
3. Geiling, E. M. K. & Cannon, P. R. Pathologic effects of elixir of sulphanilamide (diethylene glycol) poisoning. *JAMA* **111**, 919–926 (1938).
Detailed report on the effects and cause of death of patients treated with the elixir of sulphanilamide, and the principles that should be adopted for the study of new drugs.
4. Shuster, E. Fifty years later: the significance of the Nuremberg Code. *N. Engl. J. Med.* **337**, 1436–1440 (1997).
5. Barnes, J. M. Are official recommendations for the testing of drugs for toxicity dangerous? *Proc. Eur. Soc. Study Drug Toxicity* **2**, 57–63 (1963).
6. Committee for Proprietary Medicinal Products. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. (The European Agency for the Evaluation of Medicinal Products, London, 2001).
7. Schuster, E. The Nuremberg Code: Hippocratic ethics and human rights. *Lancet* **351**, 974–977 (1998).
8. World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA* **277**, 925–926 (1997).
9. Lawson, D. H. Adverse drug reactions. *Human Toxicol.* **4**, 122–126 (1985).
10. Igarashi, T., Nakane, S. & Kitagawa, T. Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J. Toxicol. Sci.* **20**, 77–92 (1995).
11. Fletcher, A. P. Drug safety tests and subsequent clinical experience. *J. Roy. Soc. Med.* **71**, 693–696 (1978).
12. Olson, H. *et al.* Concordance of the toxicity of pharmaceuticals in humans and animals. *Regul. Toxicol. Pharmacol.* **32**, 56–67 (2000).
A review of multinational pharmaceutical data and the outcome of a workshop which compared adverse effects of a large series of novel drugs in humans with toxicity in animals.
13. Japanese Pharmaceutical Manufacturers Association. Literature survey on the correlation between animal toxicity findings and clinical adverse effects. (Japanese Pharmaceutical Manufacturers Association, Tokyo, 1994).
14. Prentis, R. A. & Walker, S. R. Trends in the development of new medicines by UK-owned pharmaceutical companies (1964–1980). *Br. J. Clin. Pharmacol.* **21**, 437–443 (1986).
15. Lumley, C. In *Animal Toxicity Studies: Their Relevance for Man* (ed. Walker, S. R.) 49–56 (Quay, Lancaster, 1990).
16. Schein, P. S. *et al.* The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin. Pharmacol. Ther.* **11**, 3–40 (1970).
Critical evaluation of the qualitative toxicity of anticancer drugs in dogs and monkeys compared with humans.
17. Owens, A. H. Predicting anticancer drug effects in man from laboratory animal studies. *J. Chron. Dis.* **15**, 223–228 (1962).
18. Freireich, E. J., Gehen, E. A., Rall, D. P., Schmidt, L. H. & Skipper, H. E. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother. Reports* **50**, 219–244 (1966).
Highly cited article that provides a quantitative comparison of the toxicity of anticancer drugs in several animal species and humans.

19. Committee for Proprietary Medicinal Products. Points to consider in the assessment of the potential for QT interval prolongation by non-cardiovascular medicinal products. (The European Agency for the Evaluation of Medicinal Products, London, 1997).
20. Crumb, W. & Cavero, I. QT interval prolongation by non-cardiovascular drugs: issues and solutions for novel drug development. *Pharm. Sci. Technol. Today* **2**, 270–280 (1999).
21. De Ponti, F., Poluzzi, E. & Montanaro, N. QT-interval prolongation by non-cardiac drugs: lessons to be learned from recent experience. *Eur. J. Clin. Pharmacol.* **56**, 1–18 (2000).
22. Fermini, B. & Fossa, A. A. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nature Rev. Drug Discov.* **2**, 439–447 (2003).
23. Davis, A. S. The preclinical assessment of QT interval prolongation: a comparison of *in vitro* and *in vivo* methods. *Human Exp. Toxicol.* **17**, 677–680 (1998).
24. Nemery, B., Dinsdale, D. & Verschoyle, R. D. Detecting and evaluating chemical-induced lung damage in experimental animals. *Bull. Eur. Physiopathol. Respir*. **23**, 501–528 (1987).
25. Dressman, J. B. Comparison of canine and human gastrointestinal physiology. *Pharmacol. Res.* **3**, 123–131 (1986).
26. Lee, W. M. Drug-induced hepatotoxicity. *N. Engl. J. Med.* **349**, 474–485 (2003).
27. Amacher, D. E. Serum transaminase elevations as indicators of hepatic injury following administration of drugs. *Regul. Toxicol. Pharmacol.* **27**, 119–130 (1998).
28. Hayes, A. W. *et al.* Correlation of human hepatotoxicants with hepatic damage in animals. *Fund. Appl. Toxicol.* **2**, 55–66 (1982).
29. Schwartz, S., Raskin, P., Fonseca, V. & Graveline, J. F. Effect of troglitazone in insulin-treated patients with type II diabetes mellitus. *N. Engl. J. Med.* **338**, 861–866 (1998).
30. Mok, C. C., Lau, C. S. & Wong, R. W. S. Risk factors for ovarian failure in patients with systemic lupus erythematosus receiving cyclophosphamide therapy. *Arthritis Rheum.* **41**, 831–837 (1998).
31. Takayama, S., Akaike, M., Kawashima, K., Takahashi, M. & Kurokawa, Y. A collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders induced by drugs in rats. *Regul. Toxicol. Pharmacol.* **14**, 266–292 (1984).
32. Ribelin, W. E. The effects of drugs and chemicals upon the structure of the adrenal gland. *Fund. Appl. Toxicol.* **4**, 105–119 (1984).
33. Theus, R. & Zbinden, G. Toxicological assessment of the hemostatic system, regulatory requirements, and industry practice. *Regul. Toxicol. Pharmacol.* **4**, 74–95 (1984).
34. Dayan, A. D. *et al.* Report of a validation study of assessment of direct immunotoxicity in the rat. *Toxicology* **125**, 183–201 (1998).
35. Dean, J. H. Issues with introducing new immunotoxicology methods into the safety assessment of pharmaceuticals. *Toxicology* **119**, 95–101 (1997).
36. Dean, J. H., Hinks, J. R. & Remander, B. Immunotoxicology assessment in the pharmaceutical industry. *Toxicol. Lett.* **102–103**, 247–255 (1998).
37. Cavagnaro, J. A. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. *Nature Rev. Drug Discov.* **1**, 469–475 (2002).
38. Lichfield, J. T. Forecasting drug effects in man from studies in laboratory animals. *JAMA* **177**, 104–108 (1961).
39. Steiness, E., Rasmussen, F., Svendsen, O. & Nielsen, P. A comparative study of serum creatine phosphokinase (CPK) activity in rabbits, pigs and humans after intramuscular injection of local damaging drugs. *Acta Pharmacol. Toxicol.* **42**, 357–364 (1978).
40. Clive, D. Mutagenicity tests in drug development: interpretation and significance of test results. *Regul. Toxicol. Pharmacol.* **5**, 79–100 (1985).
41. Benigni, R. & Zito, R. Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens. *Curr. Topics Med. Chem.* **3**, 1289–1300 (2003).
42. Monro, A. & Mehta, D. Are single-dose toxicology studies in animals adequate to support single dose of a new drug in humans? *Clin. Pharmacol. Ther.* **59**, 258–264 (1996). **Provides the key arguments for testing new drugs in humans based on single-dose toxicity studies in animals.**
43. Choudary, J. *et al.* Response to Monro and Mehta proposal for use of single-dose toxicity studies to support single-dose studies of new drugs in humans. *Clin. Pharmacol. Ther.* **59**, 265–267 (1996).
44. Rozman, K. K. & Doull, J. The role of time as a quantifiable variable of toxicity and the experimental conditions when Haber's c x t product can be observed: implications for therapeutics. *J. Pharmacol. Exp. Ther.* **296**, 663–668 (2001).
45. Combes, R. D. *et al.* Early microdose drug studies in human volunteers can minimise animal testing. Proceedings of a workshop organised by volunteers in research and testing. *Eur. J. Pharm. Sci.* **19**, 1–11 (2003).
46. Lappin, G. & Garner, R. C. Big physics, small doses: the use of AMS and PET in human microdosing of development drugs. *Nature Rev. Drug Discov.* **2**, 233–240 (2003).
47. Committee for Proprietary Medicinal Products. Position paper on non-clinical safety studies to support clinical trials with a single microdose. (The European Agency for the Evaluation of Medicinal Products, London, 2003).
48. Calabrese, E. J. & Baldwin, L. A. Toxicology rethinks its central belief. *Nature* **421**, 691–692 (2003).
49. Monro, A. M. The role of metabolism studies in drug safety evaluation. *Drug Dev. Comm.* **2**, 377–396 (1976).
50. Jorkasky, D. K. What does the clinician want to know from the toxicologist? *Toxicol. Lett.* **102–103**, 539–543 (1998).
51. Grieshaber, C. K. & Marsoni, S. Relation of preclinical toxicology to findings in early clinical trials. *Cancer Treat. Rep.* **70**, 65–72 (1986).
52. Greaves, P. *Histopathology of Preclinical Toxicity Studies: Interpretation and Relevance in Drug Safety Evaluation* (Elsevier, Amsterdam, 2000).

Competing interests statement

The authors declare competing financial interests: see Web version for details.

Online links

FURTHER INFORMATION

European Union Clinical Trials Directive (2001/20/EC):
http://medicines.mhra.gov.uk/ourwork/licensingmeds/types/clintrialdir.htm

Guidance Documents, Center for Drugs Evaluation and Research, US Food and Drug Administration:
http://www.fda.gov/cder/guidance/guidance.htm

Guidelines from the European Union including those of the International Conference on Harmonisation:
http://www.emea.eu.int/index/indexh1.htm

Japanese Pharmaceutical Manufacturers Association:
http://www.jpma.or.jp/12english/

Access to this interactive links box is free online.