5-2008

Systematic Reviews of Animal Experiments Demonstrate Poor Contributions to Human Healthcare

Andrew Knight
Animal Consultants International, drandrewknight40@gmail.com

Recommended Citation
Knight, A. (2008). Systematic reviews of animal experiments demonstrate poor contributions toward human healthcare. Reviews on Recent Clinical Trials, 3(2), 89-96.

Systematic Reviews of Animal Experiments Demonstrate Poor Contributions to Human Healthcare

Andrew Knight, Animal Consultants International

KEYWORDS

animal experiment, animal model, animal study, clinical trial, human healthcare, systematic review

ABSTRACT

Widespread reliance on animal models during preclinical research and toxicity testing assumes their reasonable predictivity for human outcomes. However, of 20 published systematic reviews examining human clinical utility located during a comprehensive literature search, animal models demonstrated significant potential to contribute toward clinical interventions in only two cases, one of which was contentious. Included were experiments expected by ethics committees to lead to medical advances, highly-cited experiments published in major journals, and chimpanzee experiments—the species most generally predictive of human outcomes. Seven additional reviews failed to demonstrate utility in reliably predicting human toxicological outcomes such as carcinogenicity and teratogenicity. Results in animal models were frequently equivocal, or inconsistent with human outcomes. Consequently, animal data may not be considered generally useful for these purposes. Regulatory acceptance of non-animal models is normally conditional on formal scientific validation. In contrast, animal models are simply assumed to be predictive of human outcomes. These results demonstrate the invalidity of such assumptions. The poor human clinical and toxicological utility of animal models, combined with their generally substantial animal welfare and economic costs, demand greater rigor within animal studies, and justify a ban on animal models lacking scientific data clearly establishing their human predictivity or utility.

LABORATORY ANIMAL USE: A MEDICAL CONTROVERSY

Annually, many millions of animals are used worldwide in toxicity testing and biomedical research aimed at developing cures for human diseases. Steady increases in the use of genetically modified animals and several large-scale chemical testing programs within the US and Europe are increasing laboratory animal use [1].

Yet biomedical research using laboratory animals remains highly controversial. Advocates frequently claim such research is vital for preventing, curing or alleviating human diseases [e.g., 2-3], that the greatest achievements of medicine have been possible only due to the use of animals [e.g., 4], and that the complexity of humans requires nothing less than the complexity of laboratory animals, to effectively model, during biomedical investigations [5]. They even claim that medical progress would be “severely maimed by prohibition or severe curtailing of animal experiments,” and that “catastrophic consequences would ensue” [6].

However, such claims are hotly contested [e.g., 7], and a growing body of empirical evidence casts doubt upon their scientific utility as experimental models of humans [e.g., 8-10]. The right of humans to experiment on animals has also been strongly contested philosophically [e.g., 11-12].

SYSTEMATIC REVIEWS OF CLINICAL UTILITY

Numerous cases of discordance between human and laboratory animal outcomes, [e.g., 13], cast doubt on the human predictivity and reliability of animal models, despite their widespread use in human toxicity testing, and in the safety and efficacy testing of putative chemotherapeutic agents and other clinical interventions.

However, only small numbers of experiments are normally reviewed in such case studies, and their selection may be subject to bias. To provide more definitive conclusions, systematic reviews of the human clinical or toxicological utility of large numbers of animal experiments are necessary. Experiments included in such reviews are selected without bias, via randomization or similarly methodical and impartial means.

In recent years a growing number of such reviews and meta-analyses have been published, which collectively provide important insights into the human clinical and toxicological utility of animal models. To locate and analyze them, in 2007 I searched the Scopus Life and Health Sciences biomedical bibliographic databases for systematic reviews of the human clinical or toxicological utility of animal experiments that had been successfully published in the peer-reviewed biomedical literature. Among the world's most comprehensive, the Life Sciences database includes over 3,400 titles, and the Health Sciences database includes over 5,300 titles [14], including all of Medline, the leading medical and allied health profession database, which itself contains over 15 million citations, sourced from more than 5,000 biomedical journals from over 80 countries [15].

All abstracts, titles and key words were searched for (animal experiment OR animal model OR animal study OR animal trial) AND (clinical trial OR human outcome OR human relevance OR human result). Results were limited to articles or reviews, and no other limitations were applied. Additional relevant studies were obtained by examination of reference lists of retrieved papers, and by consultation with colleagues working in this field.

To minimize bias, reviews were selected only when systematically conducted using randomization or similarly methodical and impartial means to select animal studies. For example, in some cases all animal studies within relevant subsets of toxic chemical databases were examined, without exclusion.

Only reviews examining the contributions of animal experiments toward the development of prophylactic, diagnostic or therapeutic interventions with clear potential for combating human diseases or injuries, or their consistency with human clinical outcomes, or their human toxicological predictivity or utility, were examined. Reviews focusing only on the contributions of animal experiments towards increased understanding of a wide range of other biomedical concepts, including the etiological, pathogenesic or other aspects of human diseases, or on the clinical utility of animal experiments in non-human species, for example, were not included.

As of 1st March 2007, 20 systematic reviews examining the contributions of animal experiments toward the development of human clinical interventions, and seven reviews examining their utility during the assessment of human toxicity, were located [1]. Of the 20 clinical reviews, authors concluded that animal models had the potential to be significantly useful in contributing to the development of clinical interventions in only two cases, one of which was contentious. Results in animal models were frequently equivocal, or inconsistent with human clinical outcomes. Of the seven toxicological reviews, none clearly demonstrated the utility of animal models in reliably predicting human toxicological outcomes, such as carcinogenicity and teratogenicity. Three different approaches that sought to determine the maximum human clinical utility that may be achieved by animal experiments were of particular interest.

Animal Experiments Expected to Lead to Medical Advances

Lindl and colleagues [16-17] examined animal experiments conducted at three German universities between 1991 and 1993 that had been approved by animal ethics committees at least partly on the basis of claims by researchers that the experiments might lead to concrete advances toward the cure of human diseases. Experiments were included only where previous studies had shown that the applications of related animal research had confirmed the hypotheses of the researchers, and where the experiments had achieved publication in biomedical journals.

For 17 experiments meeting these inclusion criteria, citations were analyzed for at least 12 years. Citation frequencies and types of citing papers were recorded: whether reviews, or animal-based, in vitro or clinical studies. 1,183 citations were evident; however, only 8.2% of all citations (97) were in clinical publications, and of these, only 0.3% of all citations (four publications) demonstrated a direct correlation between the results of animal experiments and human outcomes. However, even in these four cases the hypotheses that had been verified successfully in the animal experiment failed in every respect when applied to humans. None of these 17 experiments led to any new therapies, or had any beneficial clinical impact during the period examined.

Accordingly, Lindl and colleagues called for serious, rather than cursory, evaluations of the likely benefits of animal experiments by animal ethics committees and related authorities, and for a reversal of the current paradigm in which experiments are routinely approved. Instead of approving experiments because of the possibility that benefits may accrue, Lindl and colleagues asserted that where significant doubt exists, laboratory animals should receive the benefit of that doubt, and such experiments should not, in fact, be approved.

Highly-Cited Animal Experiments

Hackam and Redelmeier [18] also utilized a citation analysis. Based on the premise that findings from highly-cited animal experiments would be most likely to be subsequently tested in clinical trials, they searched for experiments with more than 500 citations published in the seven leading scientific journals when ranked by citation impact factor.

Of 76 animal studies located with a median citation count of 889 (range: 639-2,233), only 36.8% (28/76) were replicated in randomized human trials. 18.4% (14/76) were contradicted by randomized trials, and 44.7% (34/76) had not translated to clinical trials. Ultimately, only 10.5% (8/76) of these medical interventions were subsequently approved for use in patients.

Even in these cases human benefit cannot be assumed, because adverse reactions to approved interventions are sufficiently common that they were recently recorded as the 4-6th leading cause of death in US hospitals [based on a 95% confidence interval; 19], a rate considered by these investigators to be “extremely high.” Such limited predictivity for wider human outcomes of human clinical trials may result from their focus on small groups of healthy young men, and from insufficient study durations. Particularly in Phase III, small cohorts of young men (20-300) are typically used to minimize experimental variability and to eliminate possible endocrinological disruption or other risks to women of reproductive age. Although 1,000-3,000 volunteers may be used in Phase III trials— the final phase before marketing [20], it is nevertheless clear that cohort numbers, study durations or other aspects of protocol design, conduct or interpretation are inadequate to detect the adverse side effects of the considerable number of pharmaceuticals that harm patients after marketing. Longer studies of more broadly representative human populations would be more predictive, but would increase the time and cost of pharmaceutical development, and are resisted by pharmaceutical companies.

The low rate of translation to clinical trials of even these highly-cited animal experiments occurred despite 1992 being the median publication year, allowing a median of 14 years for potential translation. For studies that did translate to clinical trials, the median time for translation was seven years (range one to 15). Frequency of translation was unaffected by laboratory animal species, type of disease or therapy under examination, journal, year of publication, methodological quality, and even, surprisingly, citation rate. However, animal studies incorporating dose-response gradients were more likely to be translated to clinical trials (odds ratio [OR] = 3.3; 95% confidence interval [CI] = 1.1 - 10.1).

Although the rate of translation of these animal studies to clinical trials was low, as Hackam and Redelmeier stated, it is nevertheless higher than that of most published animal experiments, which are considerably less likely to be translated than these highly-cited animal studies published in leading journals. Furthermore, the selective focusing on positive animal data while ignoring negative results (optimism bias), is one of several cited factors that may increase the likelihood of translation beyond that scientifically merited. As Hackam [21] stated, rigorous meta-analysis of all relevant animal experimental data would probably significantly decrease the translation rate to clinical trials.

Additionally, only 48.7% (37/76) of these highly-cited animal studies were of good methodological quality. Despite their publication in leading scientific journals, few included random allocation of animals to treatment groups, adjustment for multiple hypothesis testing, or blinded assessment of outcomes. Accordingly, Hackam and Redelmeier cautioned patients and physicians about extrapolating the findings of even highly-cited animal research to the care of human disease.

Chimpanzee Experiments

Chimpanzees are the species most closely related to humans, and consequently, most likely to be predictive of human outcomes when used in biomedical research. Accordingly, in 2005 I conducted a citation analysis examining the human clinical utility of chimpanzee experimentation [22].

I searched three major biomedical bibliographic databases, locating 749 papers published between 1995 and 2004 that described experiments on captive chimpanzees or their tissues. Although published in the international scientific literature, the vast majority of these experiments were conducted within the US [23].

Within a statistically-significant subset of 95 randomly-selected published chimpanzee experiments, 49.5% (47/95) were not cited by any subsequent papers, demonstrating minimal contribution toward the advancement of biomedical knowledge generally. This is of particular concern, because research of lesser value is not published; hence these results indicate that the majority of chimpanzee research generates data of questionable value, which makes little obvious contribution towards the advancement of biomedical knowledge.

35.8% (34/95) of 95 published chimpanzee experiments were cited by 116 papers that clearly did not describe well-developed methods for combating human diseases. Only 14.7% (14/95) were cited by 27 papers that abstracts indicated described well-developed methods for combating human diseases. However, detailed examination of these medical papers revealed that in vitro studies, human clinical and epidemiological studies, molecular assays and methods, and genomic studies, contributed most to their development. 63.0% (17/27) were wide-ranging reviews of 26-300 (median 104) references, to which the cited chimpanzee study made a very small contribution. Duplication of human outcomes, inconsistency with other human or primate data, and several other causes, resulted in the absence of any chimpanzee study able to demonstrate an essential contribution, or, in most cases, a significant contribution of any kind, toward the development of the medical method described.

Despite the low utility of chimpanzee experiments in advancing human health indicated by these results, it remains true that chimpanzees are the species most closely related to human beings. Hence, it is highly likely that other laboratory species are even less efficacious when used as experimental models of humans in biomedical research and toxicity testing.

Animal Models of Stroke and Head Injuries

Despite the existence of experimental literature describing the efficacy of more than 700 drugs in treating experimental models of stroke (artificially-induced focal cerebral ischemias) [24], only recombinant tissue plasminogen activator (rt-PA) and aspirin have convincingly demonstrated efficacy in human clinical trials of acute ischaemic stroke [25-27]. Macleod and colleagues [24] stated that, *"This failure of putative neuroprotective drugs in clinical trials represents a major challenge to the doctrine that animals provide a scientifically valid model for human stroke."* At least 10 published systematic reviews have described the poor human clinical utility of animal experimental models of stroke [24, 28-36].

In some cases clinical trials proceeded despite equivocal evidence of efficacy in animal studies. Horn and colleagues [29], for example, systematically reviewed 20 animal studies of the efficacy of nimodipine, of which only 50% showed beneficial effects following treatment. They concluded that*, "... the results of this review did not show convincing evidence to substantiate the decision to perform trials with nimodipine in large numbers of patients."* These human clinical trials also demonstrated equivocal evidence of efficacy, and furthermore, proceeded concurrently with the animal studies, despite the fact that the latter should be conducted prior to clinical trials, to allow detection of potential toxicity.

O'Collins and colleagues [30] conducted a very large review of 1,026 experimental treatments for acute stroke, finding that the effectiveness in animals of 114 drugs chosen for human clinical use was no greater than that of the remaining 912 drugs not chosen for clinical use, thereby demonstrating that effectiveness in animal models had no measurable effect on whether or not these drugs were selected for human clinical use. Accordingly, O'Collins and colleagues questioned whether the most efficacious drugs are, in fact, being selected for clinical trials, and called for greater rigor in the conduct, reporting, and analysis of animal experiments.

In many cases animal models did indicate efficacy, but these successes did not translate to humans. In a few cases authors speculated about the possible causes. Jonas and colleagues [31], for example, hypothesized that poor clinical efficacy of neuroprotectants successful in animal models was due to differences in the timing of treatment initiation. Curry [32] hypothesized that the clinical failure of fourteen neuroprotective agents successful in animal models was due to antagonism of glutamate—which may be associated with neuroprotection—by drug treatment in clinically normal patients, and that that clinical trials should be restricted to real patients, who experience elevated cerebrospinal fluid and plasma glutamate levels during progressing stroke. However, despite such speculation, the human clinical record of neuroprotectants successful in animal models remains poor.

The utility of the majority of these animal studies also appears to have been impeded by their poor methodological quality (20 animal studies of the efficacy of nimodipine [29]; animal studies of the efficacy of melatonin [24]; 29 animal studies of the efficacy of FK506 [28]; 45 animal studies of five compounds from different classes of alleged neuroprotective agents—clomethiazole, gavestinel, lubeluzole, selfotel, and tirilazad mesylate [33]; 25 animal studies of the efficacy of nitric oxide (NO) donors and L-arginine [34]; 73 animal studies of the efficacy of NO synthase inhibitors [35]).

Methodological quality of animal studies was typically scored based on the presence of characteristics such as appropriate animal models (aged, diabetic or hypertensive animals are considered to more

closely model human stroke patients), power calculations of sample sizes, random allocation to treatment and control groups, use of a clinically relevant time window for commencement of treatment, blinded drug administration, use of anesthetics without significant intrinsic neuroprotective activity (ketamine, for example, may alter neuroprotective activity), blinded induction of ischemia (given that the severity of induced infarcts may be subtly affected by knowledge of treatment allocation), blinded outcome assessment, assessment of both infarct volume and functional outcome, adequate monitoring of physiological parameters, assessment during both the acute (e.g., one to six days) and chronic (e.g., seven to 30 days) phases, statement of temperature control, compliance with animal welfare regulations, peer reviewed publication, and conflict of interest statements. Typically, one point was given for the presence of each characteristic. The Stroke Therapy Academic Industry Roundtable recommendations for standards regarding preclinical and restorative drug development use an eight point scale, for example [29, 37].

Median quality scores were four out of 10 (range zero to six, 13 studies [24]), four out of 10 (range zero to seven, 29 studies [28]), three out of 10 (45 studies [33]), three out of eight (range one to six, 73 studies [35]).

Common deficiencies included lack of: sample size calculations, aged animals or those with appropriate co-morbidities, randomized treatment allocation, blinded drug administration, blinded induction of ischemia, blinded outcome assessment and conflict of interest statements. Some studies also used ketamine anesthesia, and there was also substantial variation in the parameters assessed.

van der Worp and colleagues [33], for example, concluded that the collective evidence for neuroprotective efficacy that formed the basis for 21 clinical trials was obtained in animal studies of a methodological quality that would not, in retrospect, justify such a decision.

Wilmot and colleagues [34] also found considerable variations in animal experimental protocols, namely: animal species, physiological parameters such as blood pressure, drug administration (timing, dosage, and route), surgical methodology, and duration of ischemia. Statistical analysis (Egger's test) also revealed the likely existence of publication bias (an increased tendency to publish studies in which a treatment effect is apparent, or a decreased tendency to so publish, e.g., resulting from commercial pressures, particularly in the case of patented drugs under development).

Macleod and colleagues [24] commented that, *"these deficiencies apply to most if not all of the animal literature."* This is particularly concerning, because Macleod and colleagues [28] reported that efficacy was apparently lower in higher quality studies, raising concerns that apparent efficacy may have been artificially elevated by factors such as poor methodological quality and publication bias.

A related review not limited solely to stroke exemplified some of these points. Perel and colleagues [36] examined therapeutic interventions with unambiguous evidence of a treatment effect (benefit or harm) in clinical trials of corticosteroidal treatment for head injury, antifibrinolytics in hemorrhage, thrombolysis in acute ischemic stroke, tirilazad in acute ischemic stroke, antenatal corticosteroids to prevent neonatal respiratory distress syndrome, and bisphosphonates to treat osteoporosis. They found that three interventions had similar outcomes in animal models, while three did not, suggesting that animals do not reliably predict human outcomes. Perel and colleagues reported that the animal studies varied in methodological quality and sample sizes, that randomization and blinding were rarely reported, and that publication bias was evident.

Animal Models in Other Clinical Fields

Of seven systematic reviews examining the utility of animal models in other clinical fields identified by this review [38, 39 and 40 (which jointly described a single review), 41-45], authors concluded that animal models had the potential to contribute significantly toward human clinical interventions in only two cases, one of which was contentious.

As in the case of stroke, some clinical trials proceeded despite equivocal evidence of efficacy in animal studies. Upon systematically reviewing the effects of Low Level Laser Therapy (LLLT) on wound healing in 36 cell or animal studies, Lucas and colleagues [38] found that in-depth analysis of studies with the highest methodological quality showed no significant pooled treatment effect. Despite this, clinical trials proceeded. Furthermore, almost from the beginning of LLLT investigations, animal experiments and clinical studies occurred simultaneously, rather than in sequence. The human trials also failed to demonstrate significant benefits.

Roberts and colleagues [39], and Mapstone and colleagues [40], all systematically reviewed 44 randomized controlled animal studies examining the efficacy of fluid resuscitation in bleeding animals. Their systematic review of clinical trials of fluid resuscitation had previously found no evidence that the practice improved outcomes, and the possibility that it might be harmful [46]. In this later review [39-40], they found that fluid resuscitation reduced mortality in animal models of severe hemorrhage, but increased mortality in those with less severe hemorrhage.

After clinical trials in humans failed to provide evidence of benefit, Lee and colleagues [41] conducted a systematic review and meta-analysis of controlled trials of endothelin receptor blockade in animal models of heart failure. Meta-analysis failed to provide evidence of overall benefit, and indicated increased mortality with early administration.

In their investigation of the contribution of human clinical trial results and analogous experimental studies to asthma research—one of the most common and heavily investigated of modern diseases, Corry and Kheradmand [42] demonstrated that failure to conduct and analyze the results of animal studies before proceeding to clinical trials is not uncommon: *"Research along two fronts, involving experimental models of asthma and human clinical trials, proceeds in parallel, often with investigators unaware of their counterpart's findings."*

The clinical utility of animal models is clearly questionable in such cases in which clinical trials proceed concurrently with, or prior to, animal studies, or continue, despite equivocal evidence of efficacy in animals.

As in the case of stroke, the clinical utility of the majority of these animal studies also appears to have been limited by their poor methodological quality (36 cell or animal studies of the effect of LLLT on wound healing [38]; 44 studies examining the efficacy of fluid resuscitation in bleeding animals [39-40]; studies of the efficacy of endothelin receptor blockade in animal models of heart failure [41]). Common flaws included inadequate sample sizes, leaving studies underpowered, and lack of randomization and blinding.

In some cases obvious deficiencies with the animal models were identified. In commenting on the clinical relevance of animal models for testing the effects of LLLT on wound healing, Lucas et al. [38] noted that the animal models excluded common problems associated with wound healing in humans, such as ischemia, infection, and necrotic debris.

Difficulty was also apparent in translating animal outcomes to human clinical protocols in at least one case. Lazzarini and colleagues [43] reviewed experimental studies on osteomyelitis to ascertain their

impacts on the systemic antibiotic treatment of human osteomyelitis. Although they found that most of the animal models reviewed were reproducible and dependable, they also found that the human predictivity of these studies was unclear, and was possibly undermined by difficulties in establishing the right dose regimen in animals. Although they considered that the use of antibiotic combinations was associated with better outcomes in the majority of animal studies, and that these studies did provide indications of appropriate minimum treatment durations, they concluded that these studies had limited impact on clinical practice.

In two cases reviewers reported consistency between animal models and human outcomes that appeared to offer potential in contributing to the development of clinical interventions, although in one case this conclusion was contentious. While reviewing therapeutic approaches to streptococcal endocarditis, Scheld [44] reported good correlation overall among results obtained by in vitro susceptibility testing (especially killing kinetics in broth), experimental animal results, and clinical trials of different antimicrobial regimens in humans with streptococcal endocarditis.

To investigate the efficacy of rodent models of carcinogenesis in predicting treatment outcomes in humans, Corpet and Pierre [45] conducted a systematic review and meta-analysis of colon cancer chemoprevention studies using aspirin, -carotene, calcium, and wheat bran in rats, mice and humans. Controlled intervention studies of adenoma recurrence in human volunteers were compared with chemoprevention studies of carcinogen-induced tumors in rats, and of polyps in Min (Apc(+/-)) mice. 6,714 humans, 3,911 rats and 458 mice were included in the meta-analyses. Corpet and Pierre found that rats and humans experienced comparable results using aspirin, calcium, carotene and wheat bran. Min mice and humans results experienced comparable results using aspirin, but discordant results using calcium and wheat bran (comparative carotene results were not available). Corpet and Pierre concluded that these results suggested that rodent models roughly predict treatment effects in humans, but that the prediction is not accurate for all agents, and that the carcinogen-induced rat model is more predictive than the Min mouse model. However, few agents were tested, and two of three agents tested in mice produced different outcomes in humans, so the conclusion that rodents are predictive of human treatment effects, although only roughly, was contentious.

Toxicological Utility of Animal Experiments

Due to limited human exposure data, the identification and regulation of exposure to potential human toxins has traditionally relied heavily on animal studies. However, systematic reviews have indicated that the utility of animal studies for these purposes is lacking in the fields of carcinogenicity testing (at least five reviews: [47-51]), and teratology testing (one review: [52]). No systematic review demonstrated a contrary result. The sensitivity of animal models to a range of human toxicities (ability to identify them) highlighted by one review [53] generally appears to be accompanied by poor human specificity (ability to correctly identify human non-toxins), resulting in a high incidence of false-positive results.

ANIMAL EXPERIMENTAL CHARACTERISTICS LIMITING HUMAN UTILITY

When evaluated overall, these 27 systematic reviews clearly do not support the widely held assumptions of animal ethics committees and the opinions of advocates that animal experiments are generally beneficial in the development of human therapeutic interventions, and the assessment of human toxicity. On the contrary, they frequently demonstrate that animal experiments are of low utility for these purposes. This results both from limitations of the animal models themselves, and also from the poor methodological quality and statistical design of many animal experiments.

Genetic Differences Between Human and Animals

Chimpanzees are our closest living relatives, and consequently might be expected to have the greatest likelihood among laboratory species of accurately predicting human outcomes during preclinical research. However, despite great similarity between the structural regions of chimpanzee and human DNA, important differences between the regulatory regions exert an “avalanche” effect upon large numbers of structural genes [54]. Despite nucleotide difference between chimpanzees and humans of only 1-2%, the results are differences of around 20% in protein expression [55], resulting in marked phenotypic differences between the species. These differences manifest in altered susceptibility to, etiology and progression of diseases; differing absorption, tissue distribution, metabolism, and excretion of chemotherapeutic agents; and differences in the toxicity and efficacy of pharmaceuticals [22, 54]. Such effects appear to be responsible for the demonstrated inability of most chimpanzee research to contribute substantially to the development of methods efficacious in combating human diseases [22].

Other laboratory animal species are even less similar to humans, both genetically and phenotypically, and are therefore less likely to accurately model the progression of human diseases, or human responses to putative chemotherapeutic agents or toxins.

Toxicity Testing

Rodents are by far the most common laboratory animal species used in toxicity studies. Several factors contribute to the demonstrated inability of rodent bioassays to reliably predict human toxicity. The stresses incurred during handling, restraint, other routine laboratory procedures, and particularly, the stressful routes of dose administration common to toxicity tests, alter immune status and disease predisposition in ways which are very difficult to accurately predict, distorting disease progression and responses to putative toxic and chemotherapeutic agents [56-57].

Additionally, animals have a broad range of physiological defenses against general toxic insults, such as epithelial shedding and inducible enzymes, which commonly prove effective at environmentally relevant doses, but which may be overwhelmed at the high doses common to toxicity assays [58]. Carcinogenicity assays also utilize chronic dosing. These may result in insufficient rest intervals between doses for the effective operation of DNA and tissue repair mechanisms, which, as with the unnatural elevation of cell division rates during *ad libitum* feeding studies, may predispose to mutagenesis and carcinogenesis. Lower doses, greater intervals between exposures, intermittent feeding, or shorter total periods of exposure, which represent a more realistic model of human environmental exposures to most potential toxins, might not result in toxic changes at all [57].

Finally, differences in rates of absorption and transport mechanisms between test routes of administration and other important human routes of exposure, and the considerable variability of organ systems in response to toxic insults, between and within species, strains and genders, render profoundly difficult any attempt to accurately extrapolate human hazards from animal toxicity data [57].

Methodological Quality

At least 11 systematic reviews [18, 24, 28-29, 33-36, 38, 39 and 40 (which described a single review), 41] demonstrated the poor methodological quality of many of the animal experiments examined, and no systematic review demonstrated good methodological quality of a majority of experiments. While omission of study details due to publication space constraints may artificially lower apparent quality [24-28], the prevalence of such deficiencies exceeds that which might reasonably be expected, and is therefore grounds for considerable concern.

Common deficiencies included, lack of: sample size calculations, sufficient sample sizes, appropriate animal models (particularly, aged animals or those with appropriate co-morbidities, as stroke models), randomized treatment allocation, blinded drug administration, blinded induction of ischemia in the case of stroke models, blinded outcome assessment, and conflict of interest statements. Some studies also used anesthetics that may have altered experimental outcomes, and substantial variation was evident in the parameters assessed.

These deficiencies limited the clinical utility of these studies in various significant ways. For example, it is well established that studies that lack randomization or blinding often over-estimate the magnitude of treatment effects [59-61]. Bebarta and colleagues [62] described the impacts of lack of randomization or blinding on estimations of the significance of treatment effects in 389 animal studies and 2,203 cell line studies. They found that studies lacking randomization or blinding but not both were more likely to report a treatment response than studies that utilized these measures (OR = 3.4, 95% CI = 1.7 - 6.9; and OR = 3.2, 95% CI = 1.3 - 7.7, respectively), and that studies lacking both randomization and blinding were even more likely to report a treatment response (OR = 5.2, 95% CI = 2.0 - 13.5).

Statistical Design

Insufficient sample sizes left many studies underpowered, limiting the statistical validity of study conclusions. Animal lives and other resources may also be wasted, if experiments subsequently require repetition. As stated by the UK Medical Research Council [63], *"The number of animals used … must be the minimum sufficient to create adequate statistical power to answer the question posed."*

According to Balls and colleagues [64], however, "…surveys of published papers as well as more anecdotal information suggest that more than half of the published papers in biomedical research have statistical mistakes, many seem to use excessive numbers of animals, and a proportion are poorly designed".

Festing [65] similarly stated that*, "Surveys of published papers show that there are many errors both in the design of the experiments and in the statistical analysis of the resulting data. This must result in a waste of animals and scientific resources, and it is surely unethical."* de Boo and Hendriksen [66] noted the tendency to alter animal numbers based on scientifically irrelevant issues such as availability or cost.

Factors that should be considered when calculating appropriate sample sizes include detectability threshold (the size of the difference between treatment groups considered significant); known or expected data variation; the required significance of the test ('p' or 'α'; the probability of a Type I error—assuming a difference where none exists); the acceptable probability of assuming no difference where one does exist ('β'; a Type II error. The 'power' of an experiment = 1-β; 0.8 is the usual choice); and the type of statistical analysis to which the data will be subjected. Smaller thresholds, greater data variation, smaller acceptable error probabilities (greater power) and certain statistical tests for differences require larger samples.

No universal rule for calculating correct sample sizes exists [66]. Festing [67], for example, describes two methods, the preferred 'power calculation,' and the 'resource equation.' Power calculations utilizing formulae are available within interactive computer programs (e.g., [68-69]), and calculate minimum sample sizes required to detect treatment effects with specified degrees of certainty. Mead's 'resource equation' [70] calculates sample sizes using degrees of freedom, and incorporates statistical parameters such as treatment effects, block effects and error degrees of freedom.

Strategies should also be considered to minimize animal numbers without unacceptably compromising statistical power. Several of these aim to decrease data variability by minimizing heterogeneity in

experimental environments and protocols; by the appropriate use of environmental enrichment to decrease physiological variation, by choosing, where possible, to measure variables with relatively low inherent variability, by the use of genetically homogeneous (isogenic or inbred) or specified pathogen free animal strains, and by screening raw data for obvious errors or outliers [66, 71-73].

Meta-analysis involves the aggregation and statistical analysis of suitable data from multiple experiments. Treatment and control groups may be combined, allowing group numbers to be minimized. Although new information can be derived through meta-analysis, more frequently the results allow refinement of existing knowledge. By designing experiments and reporting protocols to maximize their utility for later meta-analyses, the benefit of individual randomized controlled experiments may be maximized [74].

Strategies such as these, aimed at maximizing the statistical power of small samples, are particularly appropriate when marked ethical, cost or practical constraints limit the number of animals that may be used, e.g., during primate experiments.

Finally, appropriate statistical analysis of resultant data should be closely linked to experimental design and to the type of data produced [75].

The relatively poor statistical knowledge of many animal researchers may be the cause of the high prevalence of poor sample size choices in animal studies. Solutions could include statistical training of researchers, and direct input of statisticians in experimental design and data analysis [66, 76].

IMPROVING EXPERIMENTAL STANDARDS

Evidence-based medicine (EBM) bases clinical decisions on methodologically-sound, prospective, randomized, blinded, controlled clinical trials, and the gold standard for EBM is large prospective epidemiological studies, or meta-analyses of randomized, blinded, controlled clinical trials [77]. The implementation to animal experiments of EBM standards applied to human clinical trials would make results more robust and broadly applicable [36, 78-81].

Mechanisms would be needed to ensure compliance with such standards, however. Compliance could, for example, be made prerequisite for research funding, ethics committee approval, and publication of results. These measures would require the education and cooperation of funding agencies, ethics committees and journals.

The UK Medical Research Council requires researchers who are planning clinical trials to reference systematic reviews of previous related work before proceeding [82]. To facilitate the detection of toxicity and potentially efficacious drugs, such reviews should also include all relevant animal research [36]. A similar requirement to reference, or where necessary, conduct, systematic reviews of relevant animal studies prior to commencing further animal studies, would encourage a more complete and impartial assessment of existing evidence [82].

Mechanisms are also needed to encourage the reporting of negative results. Negative preclinical studies are much more likely to remain unpublished than negative clinical trials [83]. In a systematic review of studies describing the efficacy of nicotinamide in combating experimentally-induced stroke, comparisons published only in abstract form gave a significantly lower estimate of effect size than those published in full, demonstrating publication bias [84]. Commenting on the pressure to obtain and publish positive results, van der Worp and colleagues [33] stated: *"It is therefore conceivable that the career of a preclinical investigator is more dependent on obtaining positive results than that of a clinical trialist."*

Theoretical and Technical Constraints of Animal Models

Strategies designed to increase full and impartial examination of existing data before conducting animal studies, to decrease variation in experimental environments and protocols, and to improve their methodological quality, would minimize consumption of animal, financial and other resources within experiments of questionable merit and quality, and would increase the potential human utility of animal data. However, while these problems might be minimized with concerted effort, given their widespread nature, the poor human clinical or toxicological utility of most animal experiments is unlikely to result solely from such factors alone. As stated by Perel and colleagues [36], the failure of animal models to adequately represent human disease may be another fundamental cause, which, in contrast, is likely to be technically and theoretically impossible to overcome.

Genetic modification of animal models through the addition of foreign genes (transgenic animals), or inactivation or deletion of genes (knockout animals), is being attempted, to make them more closely model humans. However, as well as being technically difficult very to achieve, such modification may not allow clear conclusions, due to a large variety of factors, including those reflecting the intrinsic complexity of living organisms, such as the variable redundancy of some metabolic pathways between species [85]. Furthermore, the animal welfare burdens incurred during the creation and utilization of genetically modified animals are particularly high [86].

CONCLUSIONS

The historical and contemporary paradigm that animal models are generally reasonably predictive of human outcomes provides the basis for their widespread use in toxicity testing, and in biomedical research aimed at developing cures for human diseases. However, their use persists for historical and cultural reasons, rather than because they have been demonstrated to be scientifically valid. For example, many regulatory officials *"feel more comfortable"* with animal data [87], and some even believe animal tests are inherently valid, simply because they are conducted in animals [88].

However, most published systematic reviews have demonstrated that animal experiments are insufficiently predictive of human outcomes to provide substantial benefits during the development of human clinical interventions, or during human toxicity assessments. Of 20 reviews examining clinical utility, animal models offered potential to be of significant use during the development of clinical interventions in only two cases, one of which was contentious. Seven additional reviews failed to demonstrate utility in reliably predicting human toxicological outcomes such as carcinogenicity and teratogenicity. Results in animal models were frequently equivocal, or inconsistent with human outcomes. Consequently, animal data may not normally be assumed to be useful for these purposes.

Non-animal models are generally required to pass formal scientific validation prior to regulatory acceptance. In contrast, animal models are simply assumed to be predictive of human outcomes. However, these 27 systematic reviews of the human utility of animal experiments demonstrate the invalidity of such assumptions, even for animal models in use for long periods.

The consistent application of formal validation studies to all test models is clearly warranted, regardless of their animal, non-animal, historical, contemporary or possible future status, with appropriate consideration also given to animal welfare, ethical, legal, economic and any other appropriate factors. Likely benefits would include greater selection of models truly predictive for human outcomes, increased safety of people exposed to chemicals that have passed toxicity tests, increased efficiency during the development of human pharmaceuticals and other therapeutic interventions, and decreased wastage of animal, personnel and financial resources.

In addition, the poor human clinical and toxicological utility of most animal models for which data exists, in combination with their generally substantial animal welfare and economic costs, justify considerably greater rigor within the design of most animal studies, and a ban on the use of animal models lacking scientific data clearly establishing their human predictivity or utility.

REFERENCES

[1] Knight A. Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. ATLA: Alternatives to Laboratory Animals 2007; 35(6): 641-59.

[2] Brom FW. Science and society: different bioethical approaches towards animal experimentation. ALTEX: Alternatives to Animal Experimentation 2002; 19(2): 78-82.

[3] Festing MFW. Is the use of animals in biomedical research still necessary in 2002? Unfortunately, "Yes". ATLA: Alternatives to Laboratory Animals 2004; 32(Suppl. 1B): 733-39.

[4] Pawlik WW. The significance of animals in biomedical research. [Znaczenie zwierzat w badaniach biomedycznych.]. Folia Medica Cracoviensia 1998; 39(3-4): 175-82.

[5] Kjellmer I. Animal experiments are necessary. Coordinated control functions are difficult to study without the use of nature's most complex systems: mammals and human beings. [Djurfo rso k a r no dva ndiga. Samordnade kontrollfunktioner la ter sig sva rligen studeras utan tillga ng till naturens mest komplexa system: da ggdjur och ma nniska.]. Lakartidningen 2002; 99(11): 1172-3.

[6] Osswald W. Ethics of animal research and application to humans. [Etica da investigac ao no animal e aplicac ao ao homem.]. Acta Medica Portuguesa 1992; 5(4): 222-5.

[7] Greek CR, Greek JS. 4th World Congress Point/Counterpoint: Is Animal Research Necessary in 2002? Los Angeles, US: Americans for Medical Advancement. 2002.

[8] Graham DJ, Campen D, Hui R., et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet 2005; 365: 475-81.

[9] Bhogal N, Combes R. TGN1412: time to change the paradigm for the testing of new pharmaceuticals. ATLA: Alternatives to Laboratory Animals 2006; 34: 225–9.

[10] Coghlan A. Mystery over drug trial debacle deepens. NewScientist.com news service. 2006, 14 Aug. Website http://www.newscientist.com/article.ns?id=dn9734. (Accessed 22 Dec. 2007).

[11] Singer P. Animal Liberation: A New Ethics for our Treatment of Animals. 2nd edn. New York, NY, US: New York Review/Random House. 1990.

[12] La Follette H, Shanks N. Animal experimentation: the legacy of Claude Bernard. International Studies in the Philosophy of Science 1994; 8(3): 195- 210.

[13] Greek CR, Greek JS. Sacred Cows and Golden Geese. New York, NY, US: Continuum. 2002.

[14] Elsevier BV. Scopus in detail: what does it cover? 2006. Website http://www.info.scopus.com/detail/what/. (Accessed 1 Mar. 2007).

[15] National Center for Biotechnology Information. Pubmed overview. 2006. Website http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html. (Accessed 14 Apr. 2007).

[16] Lindl T, Völkel M, Kolar R. [Animal experiments in biomedical research. An evaluation of the clinical relevance of approved animal experimental projects]. [German]. ALTEX: Alternatives to Animal Experimentation 2005; 22(3): 143-51.

[17] Lindl T, Völkel M, Kolar R. Animal experiments in biomedical research. An evaluation of the clinical relevance of approved animal experimental projects: No evident implementation in human medicine within more than 10 years. [Lecture abstract]. ALTEX: Alternatives to Animal Experimentation 2006; 23(2), 111.

[18] Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. JAMA 2006; 296(14): 1731-2.

[19] Lazarou J, Pomeranz B. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998; 279: 1200-5.

[20] National Institutes of Health. Information on clinical trials and human research studies. 2006. Website http://clinicaltrials.gov/ct/info/whatis;jsessionid=B9D601AD55432DBDD59314931CA8385C#phases. (Accessed 17 Apr. 2007).

[21] Hackam DG. Translating animal research into clinical benefit: poor methodological standards in animal studies mean that positive results may not translate to the clinical domain. BMJ 2007; 334: 163-4.

[22] Knight A. The poor contribution of chimpanzee experiments to biomedical progress. J Appl Anim Welf Sci 2007; 10(4): 281-308.

[23] Conlee KM, Hoffeld EH, Stephens ML. A demographic analysis of primate research in the United States. ATLA: Alternatives to Laboratory Animals 2004; 32(Supp 1A): 315-22.

[24] Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and meta-analysis of the efficacy of melatonin in experimental stroke. J Pineal Res. 2005; 38: 35–41.

[25] The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl J Med 1995; 333: 1581–8.

[26] Chinese Acute Stroke Trial Collaborative Group. Randomised placebocontrolled trial of early aspirin use in 20000 patients with acute ischaemic stroke. Lancet 1997; 349: 1641–9.

[27] International Stroke Trial Collaborative Group. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, or both, or neither among 19435 patients with acute ischaemic stroke. Lancet 1997; 349: 1569–81.

[28] Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and meta-analysis of the efficacy of FK506 in experimental stroke. Journal of Cerebral Blood Flow & Metabolism 2005; 25(6): 1–9.

[29] Horn J, de Haan RJ, Vermeulen M, Luiten PGM, Limburg M. Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. Stroke 2001; 32(10): 2433-8.

[30] O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. Ann Neurol 2006; 59: 467–77.

[31] Jonas S, Aiyagari V, Vieira D, Figueroa M. The failure of neuronal protective agents versus the success of thrombolysis in the treatment of ischemic stroke: the predictive value of animal models. Annals of the New York Academy of Sciences 2001; 939: 257-67.

[32] Curry SH. Why have so many drugs with stellar results in laboratory stroke models failed in clinical trials? A theory based on allometric relationships. Annals of the New York Academy of Sciences 2003; 993: 69-74.

[33] van der Worp HB, de Haan P, Morrema E, Kalkman, C.J. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. J Neurol 2005; 252: 1108–14.

[34] Willmot M, Gray L, Gibson C, Murphy S, Bath PM. A systematic review of nitric oxide donors and L-arginine in experimental stroke; effects on infarct size and cerebral blood flow. Nitric Oxide 2005; 12: 141–9.

[35] Willmot M, Gibson C, Gray L, Murphy S, Bath P. Nitric oxide synthase inhibitors in experimental ischemic stroke and their effects on infarct size and cerebral blood flow: a systematic review. Free Radical Biology & Medicine 2005; 39: 412–25.

[36] Perel P, Roberts I, Sena E, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 2007; 334: 197-200.

[37] Stroke Therapy Academic Industry Roundtable. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. Stroke 1999; 30: 2752–8.

[38] Lucas C, Criens-Poublon LJ, Cockrell CT, De Haan RJ. Wound healing in cell studies and animal model experiments by Low Level Laser Therapy; were clinical studies justified? A systematic review. Lasers in Medical Science 2002; 17(2): 110-34.

[39] Roberts I, Kwan I, Evans P, Haig S. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. BMJ 2002; 324: 474-6.

[40] Mapstone J, Roberts I, Evans P. Fluid resuscitation strategies: a systematic review of animal trials. J Trauma Injury Infect Crit Care 2003; 55: 571-89.

[41] Lee DS, Nguyen QT, Lapointe N, et al. Meta-analysis of the effects of endothelin receptor blockade on survival in experimental heart failure. J Cardiac Fail 2003; 9: 368-74.

[42] Corry DB, Kheradmand F. The future of asthma therapy: integrating clinical and experimental studies. Immunologic Research 2005; 33(1): 35-51.

[43] Lazzarini L, Overgaard KA, Conti E, Shirtliff ME. Experimental osteomyelitis: What have we learned from animal studies about the systemic treatment of osteomyelitis? Journal of Chemotherapy 2006; 18(5): 451-60.

[44] Scheld WM. Therapy of streptococcal endocarditis: correlation of animal model and clinical studies. Journal of Antimicrobial Chemotherapy 1987; 20(Suppl. A): 71-85.

[45] Corpet DE, Pierre F. How good are rodent models of carcinogenesis in predicting efficacy in humans? A systematic review and meta-analysis of colon chemoprevention in rats, mice and men. Eur. J. Cancer 2005; 41(13): 1911-22.

[46] Roberts I, Evans A, Bunn F, Kwan I, Crowhurst E. Normalising the blood pressure in bleeding trauma patients may be harmful. Lancet 2001; 357: 385- 7.

[47] Tomatis L, Wilbourn J. Evaluation of carcinogenic risk to humans: the experience of IARC. In Iversen O, Ed. New Frontiers in Cancer Causation. Washington, DC, US: Taylor and Francis. 1993. 371–87.

[48] Haseman K. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. Drug Metabolism Reviews 2000; 32: 169–86.

[49] Huff J. Chemicals studied and evaluated in long-term carcinogenesis bioassays by both the Ramazzini Foundation and the National Toxicology Program. Ann. N.Y. Acad. Sci. 2002; 982: 208-30.

[50] Ennever FK, Lave LB. Implications of the lack of accuracy of the lifetime rodent bioassay for predicting human carcinogenicity. Regulatory Toxicology and Pharmacology 2003; 38: 52–7.

[51] Knight A, Bailey J, Balcombe J. Animal carcinogenicity studies: 1. poor human predictivity. ATLA: Alternatives to Laboratory Animals 2006; 34(1): 19-27.

[52] Bailey J, Knight A, Balcombe J. The future of teratology research is in vitro. Biogenic Amines 2005; 19(2): 97–145.

[53] Olson H, Betton G, Stritar J, Robinson D. The predictivity of the toxicity of pharmaceuticals in humans from animal data - an interim assessment. Toxicology Letters 1998; 102-3: 535-8.

[54] Bailey J. Non-human primates in medical research and drug development: a critical review. Biogenic Amines 2005; 19(4-6): 235–55.

[55] Glazko G, Veeramachaneni V, Nei M, Makalowski W. Eighty percent of proteins are different between humans and chimpanzees. Gene 2005; 346: 215-9.

[56] Balcombe J, Barnard N, Sandusky C. Laboratory routines cause animal stress. Contemporary Topics in Laboratory Animal Science 2004; 43(6): 42- 51.

[57] Knight A, Bailey J, Balcombe J. Animal carcinogenicity studies: 2. obstacles to extrapolation of data to humans. ATLA: Alternatives to Laboratory Animals 2006; 34(1): 29-38.

[58] Gold LS, Slone TH, Ames BN. What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. Drug Metabolism Reviews 1998; 30: 359–404.

[59] Poignet H, Nowicki JP, Scatton B. Lack of neuroprotective effect of some sigma ligands in a model of focal cerebral ischemia in the mouse. Brain Res 1992; 596: 320–4.

[60] Aronowski J, Strong R, Grotta JC. Treatment of experimental focal ischemia in rats with lubeluzole. Neuropharmacology 1996; 35: 689–93.

[61] Marshall JW, Cross AJ, Jackson DM, Green AR, Baker HF, Ridley RM. Clomethiazole protects against hemineglect in a primate model of stroke. Brain Res Bull 2000; 52: 21–9.

[62] Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomisation and blinding affect the results? Academic Emergency Medicine 2003; 10(6): 684-7.

[63] Medical Research Council (MRC). Responsibility in the Use of Animals in Medical Research. London, UK: MRC. 1993.

[64] Balls M, Festing MFW, Vaughan S, Eds. Reducing the use of experimental animals where no replacement is yet available. ATLA: Alternatives to Laboratory Animals 2004; 32(Suppl. 2): 1–104.

[65] Festing MFW. Good experimental design and statistics can save animals, but how can it be promoted? ATLA: Alternatives to Laboratory Animals 2004; 32(suppl. 1A): 133-5.

[66] de Boo J, Hendriksen C. Reduction strategies in animal research: a review of scientific approaches at the intra-experimental, supra-experimental and extraexperimental levels. ATLA: Alternatives to Laboratory Animals 2005; 33: 369–77.

[67] Festing MFW. Experimental design and husbandry. Experimental Gerontology 1997; 32: 39–47.

[68] van Wilgenburg H, van Schaick Zillesen PG, Krulichova I. Sample power and ExpDesign: tools for improving design of animal experiments. Laboratory Animals 2003; 32(3): 39-43.

[69] van Wilgenburg H, Van Schaick Zillesen PG, Krulichova I. Experimental design: Computer simulation for improving the precision of an experiment. ATLA: Alternatives to Laboratory Animals 2004; 32(suppl. 1B): 607-11.

[70] Mead R. The Design of Experiments. New York, NY, US: Cambridge University Press. 1988.

[71] Eskola S, Lauhikari M, Voipio H, Laitinen M, Nevalainen T. Environmental enrichment may alter the number of rats needed to achieve statistical significance. Scandinavian Journal of Laboratory Animal Science 1999; 26(3): 134- 44.

[72] Schauber EM, Edge WD. Statistical power to detect main and interactive effects on the attributes of small-mammal populations. Canadian Journal of Zoology 1999; 77(1): 68-73.

[73] Festing MFW, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR Journal 2002; 43(4): 244-57.

[74] Phillips CJC. Meta-analysis - A systematic and quantitative review of animal experiments to maximise the information derived. Animal Welfare 2005; 14(4): 333-8.

[75] Festing MFW, Baumans V, Combes RD, et al. Reducing the use of laboratory animals in biomedical research: problems and possible solutions. ATLA: Alternatives to Laboratory Animals 1998; 26: 283–301.

[76] Balls M, Goldberg AM, Fentem JH, et al. The three Rs: the way forward: the report and recommendations of ECVAM Workshop 11. ATLA: Alternatives to Laboratory Animals 1995; 23: 838–66.

[77] Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 286: 2420-5.

[78] Watters MPR, Goodman NW. Comparison of basic methods in clinical studies and in vitro tissue and cell culture studies in three anaesthesia journals. Br J Anaes 1999; 82: 295-8.

[79] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001; 357: 1191–4.

[80] Arlt S, Heuwieser W. [Evidence based veterinary medicine] [German]. Dtsch Tierarztl Wochenschr 2005; 112(4): 146-8.

[81] Schulz KF. Assessing allocation concealment and blinding in randomised controlled trials: why bother? Equine Veterinary Journal 2005; 37(5): 394-5.

[82] Pound P, Ebrahim S, Sandercock P, Bracken M, Roberts I. Where is the evidence that animal research benefits humans? British Medical Journal 2004; 328: 514-7.

[83] Brown CM, Calder C, Linton C, et al. Neuroprotective properties of lifarizine compared with those of other agents in a mouse model of focal cerebral ischaemia. Br J Pharmacol 1995; 115: 1425–32.

[84] Oktem IS, Menku A, Akdemir H, Kontas O, Kurtsoy A, Koc RK. Therapeutic effect of tirilazad mesylate (U-74006F), mannitol, and their combination on experimental ischemia. Res Exp Med (Berl) 2000; 199: 231–42.

[85] Houdebine LM. Transgenic animal models in biomedical research. Methods Mol Biol. 2007; 360: 163-202.

[86] Sauer UG, Kolar R, Rusche B. [The use of transgenic animals in biomedical research in Germany. Part 2: Ethical evaluation of the use of transgenic animals in biomedical research and perspectives for the changeover in research to research animal-free methods] [German]. ALTEX: Alternatives to Animal Experimentation 2006; 23(1): 3-16.

[87] O'Connor AM. Barriers to regulatory acceptance. In van Zutphen LFM, Balls M., Eds. Animal Alternatives, Welfare and Ethics. Amsterdam, The Netherlands: Elsevier Science B.V. 1997. 1173–6.

[88] Balls M. Are animal tests inherently valid? ATLA: Alternatives to Laboratory Animals 2004; 32(Suppl. 1B): 755–8.