AATEX **14**, Special Issue, 125-130
Proc. 6th World Congress on Alternatives & Animal Use in the Life Sciences
August 21-25, 2007, Tokyo, Japan

Systematic reviews of animal experiments demonstrate poor human utility

Andrew Knight

Director, Animal Consultants International
91 Vanbrugh Ct., Wincott St., London SE11 4NR, UK
info@AnimalConsultants.org

Abstract
The assumption that animal models are 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. To investigate the validity of this assumption, the comprehensive 'Scopus' biomedical bibliographic databases were searched for published systematic reviews of the human clinical or toxicological utility of animal experiments. Within 20 reviews examining clinical utility, animal models demonstrated the potential to contribute significantly toward the development of human clinical interventions in only two cases, one of which was contentious. Included were reviews of the clinical utility of experiments expected by ethics committees to lead to medical advances, of highly-cited experiments published in major journals, and of chimpanzee experiments—the species most likely to be predictive of human outcomes. Seven additional reviews failed to clearly demonstrate utility in predicting human toxicological outcomes such as carcinogenicity and teratogenicity. The poor human clinical and toxicological utility of most animal models for which data exists, in conjunction with their generally substantial animal welfare and economic costs, justify a ban on animal experiments lacking scientific data clearly establishing their human predictivity or utility.

Keywords: animal experiment, animal model, animal study, clinical trial, systematic review

Introduction

Trends in laboratory animal use

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 (GM) animals and several large scale chemical testing programs within the US and Europe are increasing laboratory animal use.

Claims supporting laboratory animal use

Biomedical research using laboratory animals is highly controversial. Advocates frequently claim such research is vital for preventing, curing or alleviating human diseases (e.g., Brom 2002, Festing 2004), that the greatest achievements of medicine have been possible only due to the use of animals (e.g., Pawlik 1998), and that the complexity of humans requires nothing less than the complexity of laboratory animals to effectively model during biomedical investigations (e.g., Kjellmer 2002). They even claim that medical progress would be *"severely maimed by prohibition or severe curtailing of animal experiments,"* and that *"catastrophic consequences would ensue"* (Osswald 1992).

The necessity of systematic reviews

The premise that laboratory animal models are generally predictive of human outcomes is the basis for their widespread use in human toxicity testing, and in the safety and efficacy testing of putative chemotherapeutic agents and other clinical interventions. However, numerous cases of discordance between human and laboratory animal outcomes (Knight, 2007a) suggest that this premise may well be incorrect, and that the utility of animal experiments for these purposes may not be assured. On the other hand, 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. 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. Their identification and examination was the purpose of this review.

Methods

The 'Scopus' biomedical bibliographic databases—among the most comprehensive internationally—were searched for systematic reviews of the human clinical or toxicological utility of animal experiments, published in the peer-reviewed biomedical literature. To minimize bias, reviews were included only when systematically conducted using randomization or similarly methodical and impartial means to select animal studies.

Only reviews examining the human toxicological predictivity or utility of animal experiments, their contributions towards the development of prophylactic, diagnostic or therapeutic interventions with clear potential for combating human diseases or injuries, or their consistency with human clinical outcomes, were examined. Reviews focusing only on the contributions of animal experiments toward 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.

Results

As of 1st March 2007, 27 systematic reviews examining the contributions of animal experiments toward the development of human clinical interventions (20), or in deriving human toxicity classifications (seven), were located. Of the 20 clinical reviews, animal models demonstrated the potential to contribute significantly toward the development of human clinical interventions in only two cases, one of which was contentious. Of the seven toxicological reviews, none clearly demonstrated the utility of animal models in predicting human toxicological outcomes such as carcinogenicity and teratogenicity.

Discussion

Space constraints preclude a detailed examination of these 27 reviews, which is provided elsewhere (Knight 2007a). However, three different approaches that sought to determine the maximum human clinical utility that may be achieved by animal experiments were of particular interest.

Experiments expected to lead to medical advances

Lindl and colleagues (2005 & 2006) 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 animal experiments 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 animal experiments are routinely approved. Instead of approving experiments because of the possibility that benefits may accrue, Lindl and colleagues suggested that where significant doubt exists, laboratory animals should receive the benefit of that doubt, and that such experiments should not, in fact, be approved.

Highly-cited animal experiments

Hackam and Redelmeier (2006) also utilized a citation analysis, although without geographical limitation. Based on the assumption 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. And even in these cases human benefit cannot be assumed, because adverse reactions to approved interventions are sufficiently common that they are the 4th-6th leading cause of death (based on a 95% confidence interval) in US hospitals (Lazarou & Pomeranz 1998).

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 1-15). 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 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 (2007) 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 experiments (Knight 2007b).

I searched three major biomedical bibliographic databases, locating 749 papers published between 1995 and 2004 describing 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 (Conlee *et al.* 2004). To obtain 95% confidence intervals with an accuracy of at least plus or minus 10%, when estimating the proportion of chimpanzee studies subsequently cited by other published papers, a subset of at least 86 chimpanzee studies was required (Morris, n.d., Guenther 1973, Green 1982).

Of 95 published chimpanzee experiments randomly selected, 49.5% (47) were not cited by any subsequent papers, demonstrating minimal contribution toward the advancement of biomedical knowledge generally. This is of particular concern, because much research of lesser value is not published; hence it appears that the majority of chimpanzee research generates data of questionable value, which makes little obvious contribution toward the advancement of biomedical knowledge.

35.8% (34/95) of these 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) of these chimpanzee experiments were cited by 27 papers with abstracts that indicated 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 human or other 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 healthcare 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.

Causes for the poor human utility of animal models

Biomedical research

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 biomedical 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 (Bailey 2005). Despite nucleotide difference between chimpanzees and humans of only 1-2%, the results are differences of around 20% in protein expression (Glazko *et al.* 2005), 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 (Bailey

2005, Knight 2007b). 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 (Knight 2007b).

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 the 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 (Knight 2007a). The stresses incurred during handling, restraint, other routine laboratory procedures, and particularly, the stressful routes of dose administration common to toxicity studies, alter immuno-competence and disease predisposition in ways which are difficult to accurately predict, distorting disease progression and responses to putative toxic and chemotherapeutic agents (Balcombe *et al.* 2004, Knight *et al.* 2006).

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 (Gold *et al*. 1998). 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 may represent a more realistic model of environmental exposure for most potential toxins, might not result in toxic changes at all (Knight *et al.* 2006).

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 attempts to accurately extrapolate human hazards from animal toxicity data profoundly difficult (Knight *et al.* 2006).

Methodological quality

At least 11 systematic reviews (Horn *et al.* 2001, Lucas *et al.* 2002, Roberts *et al.* 2002 and Mapstone *et al.* 2003 (who described a single review), Lee *et al.* 2003, Macleod *et al.* 2005a, Macleod *et al.* 2005b, van der Worp *et al.* 2005, Wilmot *et al.* 2005a, Willmot *et al.* 2005b, Hackam & Redelmeier 2006, Perel *et al.* 2007) demonstrated the poor methodological quality of many of the animal experiments examined, and no systematic review demonstrated good methodological quality of a majority of them. Common deficiencies included lack of: sample size calculations, sufficient sample sizes, appropriate animal models (particularly, aged animals or those with comorbidities likely in diseases under investigation), randomized treatment allocation, blinded drug administration, blinded induction of injury, 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.

Raising standards: evidence-based medicine

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 (Evidence-Based Medicine Working Group 1992). The implementation to animal experiments of EBM standards applied to human clinical trials would make results more robust and broadly applicable (Watters *et al.* 1999, Moher *et al.* 2001, Arlt & Heuwieser 2005, Schulz 2005, Perel *et al.* 2007). 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.

Fundamental constraints on the human utility of animal models

Strategies designed to increase full and impartial examination of existing data before conducting animal studies, to decrease variation in experimental environments, protocols and animal models, and to improve the methodological quality of experiments, 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 theoretically be minimized with concerted effort, given their widespread nature, the poor human clinical or toxicological utility of many animal experiments is unlikely to result solely from such factors alone. As stated by Perel *et al.* (2007), the failure of animal models to adequately represent human disease may be another fundamental cause, which, in contrast, could 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 number of factors, including those reflecting the intrinsic complexity of living organisms, such as the variable redundancy of some metabolic pathways between species (Houdebine 2007). Furthermore, the animal welfare burdens incurred during the creation and utilization of GM animals are particularly high (Sauer *et al.* 2006).

Implications for scientific validation of experimental models

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. These 27 systematic reviews of the human clinical or toxicological 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 relevant factors. Likely benefits would include greater selection of models truly predictive of 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.

Conclusions

The historical and contemporary paradigm that animal models are generally reasonably predictive of human outcomes provides the basis for their continuing widespread use in toxicity testing and 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 (O'Connor 1997), and some even believe animal tests are inherently valid, simply because they are conducted in animals (Balls 2004).

However, most existing 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 in deriving human toxicity assessments. Within 20 reviews examining clinical utility, animal models demonstrated the potential to contribute significantly toward the development of human clinical interventions in only two cases, one of which was contentious. Seven additional reviews also failed to clearly demonstrate utility in predicting human toxicological outcomes such as carcinogenicity and teratogenicity. Consequently, animal data may not generally be assumed to be substantially useful for these purposes. On the contrary, the poor human clinical and toxicological utility of most animal models for which data exists, in conjunction with their generally substantial animal welfare and economic costs, justify a ban on the use of animal models lacking scientific data clearly establishing their human predictivity or utility.

Acknowledgement

This article summarizes the complete, final and definitive study previously published elsewhere (Knight 2007a).

References

Arlt S. & Heuwieser W. (2005) [Evidence based veterinary medicine] [German]. *Dtsch Tierarztl Wochenschr,* 112(4), 146-148.

Bailey, J. (2005) Non-human primates in medical research and drug development: a critical review. *Biogenic Amines,* 19(4-6), 235–255.

Balcombe, J., Barnard, N. & Sandusky, C. (2004) Laboratory routines cause animal stress. *Contemporary Topics in Laboratory Animal Science,* 43(6), 42-51.

Balls, M. (2004) Are animal tests inherently valid? *ATLA: Alternatives to Laboratory Animals,* 32(Suppl. 1B), 755–758.

Brom, F.W. (2002) Science and society: different bioethical approaches towards animal experimentation. *ALTEX: Alternatives to Animal Experimentation,* 19(2), 78-82.

Conlee, K.M., Hoffeld, E.H. & Stephens, M.L. (2004) A demographic analysis of primate research in the United States. *ATLA: Alternatives to Laboratory Animals,* 32(Supp 1A), 315-322.

Evidence-Based Medicine Working Group. (1992) Evidence-based medicine. A new approach to teaching the practice of medicine. *J Amer Med Assoc,* 286, 2420-2425.

Festing, M.F.W. (2004) Is the use of animals in biomedical research still necessary in 2002? unfortunately, "Yes". *ATLA: Alternatives to Laboratory Animals,* 32(Suppl. 1B), 733-739.

Glazko, G., Veeramachaneni, V., Nei, M. & Makalowski, W. (2005) Eighty percent of proteins are different between humans and chimpanzees. *Gene,* 346, 215-219.

Gold, L.S., Slone, T.H. & Ames, B.N. (1998) What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metabolism Reviews,* 30, 359–404.

Green, J. (1982) Asymptotic sample size for given confidence interval length. *Applied Statistics*, 31(3), 298-300.

Guenther, W.C. (1973) A sample size formula for the hypergeometric. *Journal of Quality Technology,* 5(4), 167-170.

Hackam, D.G. & Redelmeier, D.A. (2006) Translation of research evidence from animals to humans. *J Amer Med Assoc,* 296(14), 1731-1732.

Hackam, D.G. (2007) Translating animal research into clinical benefit: poor methodological standards in animal studies mean that positive results may not translate to the clinical domain. *Brit Med J,* 334, 163-164.

Horn, J., de Haan, R.J., Vermeulen, M., Luiten, P.G.M. & Limburg, M. (2001) Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. *Stroke,* 32(10), 2433-2438.

Houdebine, L.M. (2007) Transgenic animal models in biomedical research. *Methods Mol Biol,* 360, 163-202.

Kjellmer, I. (2002) [Animal experiments are necessary. Coordinated control functions are difficult to study without the use of nature's most complex systems: mammals and human beings.]. Djurförsök är nödvändiga. Samordnade kontrollfunktioner låter sig svårligen studeras utan tillgång till naturens mest komplexa system: däggdjur och människa. *Lakartidningen,* 99(11), 1172-1173.

Knight, A. (2007a) Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. *ATLA: Alternatives to Laboratory Animals,* 35(6), 641-659.

Knight, A. (2007b) The poor contribution of chimpanzee experiments to biomedical progress. *J Appl Anim Welf Sci,* 10(4), 281-308.

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

Lazarou, J. & Pomeranz, B. (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *J Amer Med Assoc,* 279, 1200-1205.

Lee, D.S., Nguyen, Q.T., Lapointe, N., Austin, P.C., Ohlsson, A., Tu, J.V., Stewart, D.J. & Rouleau, J.L. (2003) Meta-analysis of the effects of endothelin receptor blockade on survival in experimental heart failure. *J Cardiac Fail,* 9, 368-374.

Lindl, T., Völkel, M. & Kolar, R. (2005) [Animal experiments in biomedical research. An evaluation of the clinical relevance of approved animal experimental projects]. [German] *ALTEX: Alternatives to Animal Experimentation,* 22(3), 143-151.

Lindl, T., Völkel, M. & Kolar, R. (2006) 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,* 23(2), 111.

Lucas, C., Criens-Poublon, L.J., Cockrell, C.T. & De Haan, R.J. (2002) 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,* 17 (2), 110-134.

Macleod M.R., O'Collins T, Horky L.L., Howells D.W. & Donnan G.A. (2005a) Systematic review and meta-analysis of the efficacy of melatonin in experimental stroke. *J Pineal Res,* 38, 35–41.

Macleod, M.R., O'Collins, T., Horky, L.L., Howells, D.W. & Donnan, G.A. (2005b) Systematic review and meta-analysis of the efficacy of FK506 in experimental stroke. *Journal of Cerebral Blood Flow & Metabolism,* 25(6), 1–9.

Mapstone, J., Roberts, I. & Evans, P. (2003) Fluid resuscitation strategies: a systematic review of animal trials. *J Trauma Injury Infect Crit Care,* 55, 571-589.

Moher, D., Schulz, K.F. & Altman, D.G. (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet,* 357, 1191–1194.

Morris E. (Undated) Sampling from small populations. Website http://uregina.ca/~morrisev/Sociology/Sampling%20from%20small%20populations.htm. (Accessed 20 Sep. 2006).

O'Connor, A.M. (1997) Barriers to regulatory acceptance, in *Animal Alternatives, Welfare and Ethics,* ed. by L.F.M. van Zutphen & M. Balls, pp. 1173–1176, Amsterdam, The Netherlands: Elsevier Science B.V.

Osswald, W. (1992) [Ethics of animal research and application to humans.]. Etica da investigação no animal e aplicação ao homem. *Acta medica portuguesa,* 5(4), 222-225.

Pawlik, W.W. (1998) [The significance of animals in biomedical research.]. Znaczenie zwierzat w badaniach biomedycznych. *Folia Medica Cracoviensia,* 39(3-4), 175-182.

Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., Macleod, M., Mignini, L.E., Jayaram, P. & Khan, K.S. (2007) Comparison of treatment effects between animal experiments and clinical trials: systematic review. *Brit Med J,* 334, 197-200.

Roberts, I., Kwan, I., Evans, P. & Haig, S. (2002) Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. *Brit Med J,* 324, 474-476.

Sauer, U.G., Kolar, R. & Rusche, B. (2006) [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,* 23(1), 3-16.

Schulz, K.F. (2005) Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Equine Veterinary Journal,* 37(5), 394-395.

Van der Worp, H.B., de Haan, P., Morrema, E. & Kalkman, C.J. (2005) Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. *J Neurol,* 252, 1108–1114.

Watters, M.P.R. & Goodman, N.W. (1999) Comparison of basic methods in clinical studies and *in vitro* tissue and cell culture studies in three anaesthesia journals. *Br J Anaes,* 82, 295-298.

Willmot, M., Gray, L., Gibson, C., Murphy, S. & Bath, P.M. (2005a) A systematic review of nitric oxide donors and L-arginine in experimental stroke; effects on infarct size and cerebral blood flow. *Nitric Oxide,* 12, 141–149.

Willmot, M., Gibson, C., Gray, L., Murphy, S. & Bath, P. (2005b) 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,* 39, 412–425.