



# Medical progress depends on animal models – doesn't it?

Robert AJ Matthews

Department of Information Engineering, Aston University, Birmingham B4 5DT. E-mail: rajm@physics.org

## DECLARATIONS

### Competing interests

None declared

### Funding

Not applicable

### Ethical approval

Not applicable

### Guarantor

RAJM

### Contributorship

RAJM is the sole contributor

## Acknowledgements

The author is indebted to Harald Schmidt for his constructive comments on the original manuscript.

## Introduction

Animal models are widely recognized as being essential to the progress of medical science. In countering the critics' arguments of the use of animals in medicine, one statement has acquired almost talismanic importance:

*'Virtually every medical achievement of the last century has depended directly or indirectly on research with animals.'*

In this essay, the origins and justification of this oft-repeated statement are examined. Despite its endorsement by leading academic bodies, it is far from clear that the statement has been, or even could be, formally validated.

## Origins of the statement

The use of animal models is a long-standing and deeply controversial aspect of medical research. In the face of the increasing – and increasingly vociferous – protest against the practice, academic institutions and individual scientists have responded with a variety of arguments, notably a statement summarizing the value of animal models as follows: 'Virtually every medical achievement of the last century has depended directly or indirectly on research with animals'. This has been endorsed essentially verbatim by many eminent bodies, including the US Public Health Service, the Royal Society, and the UK Department of Health.<sup>1–3</sup> In 2005, over 500 eminent academics signed a public petition supporting the statement, among them three Nobel laureates and over 250 professors.<sup>4</sup>

It is rare for so unequivocal a statement to command such unqualified support from the scientific community. It would not be unreasonable to assume that this is because it is demonstrably true. There are certainly specific instances consistent with the statement, for example the role of animals in the development of blood transfusions and the identification of insulin.<sup>2</sup> It is also undeniable that every blockbuster drug developed in recent years has involved the use of animal models, such

testing being mandatory in the wake of the Thalidomide disaster.

However, the statement goes further than anecdotes taken from a few areas of medicine. It makes two substantive claims: first, that *virtually every* medical achievement of the last century has involved animal models, and secondly that these achievements have *depended* on the use of these animals.

The claim appears to have originated in a one-page statement by the US Public Health Service, dated February 1994 and published in *The Physiologist* under the title 'The Importance of Animals in Biomedical and Behavioral Research'.<sup>2</sup> It contains no citations to the literature supporting the claim; this is simply asserted. Subsequent reiterations of the claim either cite this original unreferenced source, or merely assert it in turn essentially verbatim (e.g. the Royal Society report states '*...in the past century*', rather than '*...of the last century*').

## Validating the claims

Whether, when or how the claim has ever been validated is thus unclear. Certainly published lists of achievements stemming from animal models (see for example the Research Defence Society's Timeline<sup>5</sup>) fall far short of representing 'virtually all' medical achievements of the last century. Indeed, it is far from clear how such a claim could ever be validated. As the well-known study by Comroe and Dripps shows,<sup>6</sup> identifying even basic features of the most significant advances in a single area of medicine is a process mired in subjectivity, and is prone to reaching conclusions that are 'not repeatable, reliable, or valid'.<sup>7</sup>

Demonstrating the role of animal models in a specific medical advance presents peculiar difficulties characterized by Paton,<sup>8</sup> and exemplified by the fact that in his own comprehensive analysis, he *explicitly* excludes any attempt to estimate the proportion of medical achievements benefiting from animal models.

Demanding validation of the statement that 'virtually all' medical achievements of the last century have involved animal models may seem pedantic, but there is a point of principle here. The eminence of many of those who have repeated this claim, and in particular their scientific eminence, places an obligation upon them to be able to substantiate it. The failure – and, in all likelihood, inability – to do so exposes some of our most respected academic institutions to a charge of abuse of authority.

### Predictive value of animal models

Even if the first claim were capable of validation, this would still not justify the second substantive claim: that virtually all such achievements have *depended* on animal models – that is, such models were not merely included in the research process, but provided *demonstrable evidential weight*, leading to a positive outcome. The use of animals has long been de rigueur in medical research, and is mandatory in drug development. As such, the statement that virtually every medical breakthrough has involved animal experiments says nothing about the inferential value of those experiments – any more than the equally ubiquitous use of animal experiments in *failed* breakthroughs proves their futility. Arguing otherwise constitutes a well-known inferential fallacy known as transposition of the conditional (see for example Paulos<sup>9</sup>), which in this case takes the form of wrongly assuming the (unknown) probability of a medical advance taking place *given* the use of animals is equivalent to the (very high) probability that animals were used *given* a medical advance has taken place.

To move beyond this fallacy, we can follow Paton who urges the use of quantitative measures, as 'There is nothing like quantitative measurement for sharpening the wits, calling of bluffs and setting things in proportion'.<sup>8</sup> Such an approach is also valuable in revealing gaps in extant knowledge, and providing a clear resolution of the problem of transposition of the conditional. This can be made precise and quantitative via the familiar concepts of sensitivity (i.e. the true positive rate) and specificity (i.e. true negative rate). These lead to various ways of quantifying evidential weight, of which the most direct and transparent is the so-called likelihood ratio (LR), whose definition is such that only tests producing  $LR > 1$  can be

deemed to have contributed any weight of evidence.<sup>10</sup> More specifically the positive likelihood ratio,  $LR_+$  in support of the hypothesis that a specific effect will obtain, given a positive test result, is given by the ratio

$$LR_+ = \text{sensitivity} / (100\% - \text{specificity})$$

(an analogous definition exists for *negative* likelihood ratio,  $LR_-$ , in support of the hypothesis that a specific effect will *not* obtain, given a negative test result). Thus for any putative source of evidential weight to be deemed useful, its specificity and sensitivity must be such that  $LR_+ > 1$ . Tossing a coin contributes no evidential weight to a given hypothesis as the sensitivity and specificity are the same – 50% – and thus the  $LR_+$  is equal to 1.

All animal models possess both sensitivity and specificity values, and thus lead to values for the evidential weight provided by each such model. As such, they provide the quantitative underpinning for statements about the value of animal models to medical progress. Or, rather, they would if they existed. As has been pointed out repeatedly by authors for several decades, there is a striking paucity of quantitative comparative data for animal models.<sup>11–13</sup>

Various explanations for this can be offered. First and most obviously, compounds that produce unacceptable effects in animal models will not progress to human trials, making studies capable of giving sensitivity/false positive rates for animal models ethically problematic. Secondly, it is frequently difficult to establish end-points sufficiently clear-cut to allow categorization as true positives or true negatives. Thirdly, much of the comparative animal-human data is obtained under conditions of commercial confidentiality. These are all serious difficulties for those seeking to show that the value of animal models is supported by quantitative evidence rather than anecdote.<sup>14</sup>

### Published evidence of predictive value

None of these difficulties is, however, insuperable. While there may be relatively few quantitative studies of the predictive abilities of animal models, they do exist. The principal source of such studies is in an area where both critics and advocates agree there is a pressing need for the validation of animal models: toxicity testing.<sup>15</sup> Despite the difficulties, several authors have succeeded in acquiring comparative data from

animal and human studies with a view to estimating the evidential weight provided by the animal models in relation to specific organ toxicities. Regrettably, the data provided by these studies is typically incomplete, ambiguous, and subjected to inadequate or incorrect analysis. As a result, the estimates for the evidential weight of animal models that emerge are at best inconclusive, and sometimes wholly misleading.

For example, the largest review of the predictive performance of animal toxicity studies covers 150 drugs specifically associated with adverse events or toxicity in humans in testing by pharmaceutical companies.<sup>16</sup> Using unpublished data, it found a figure for the sensitivity for rodent and non-rodent species collectively of 71%. However, as its authors make clear, the review did not attempt to estimate the corresponding specificity, stating that 'a more complete evaluation of this predictivity aspect will be an important part of a future prospective survey'. Yet without this, it is simply impossible to assess the evidential weight provided by the animal models.

In his review of the predictive power of seven animal models for toxic lesions in humans,<sup>17</sup> Hottendorf provides explicit values for both the false positive and false negative rates. Unfortunately, these are based on incorrect definitions, while other data are stated in a format that precludes calculation of unambiguous values for the LRs (see Appendix). Despite these shortcomings, it is certainly possible to agree with Hottendorf's conclusion that 'The predictive value of toxicological studies performed in animal species and the incidence of species differences in toxicity *could and should* be placed in sharper perspective with an expanded data base' (emphasis added).

The review of comparative anticancer drug toxicity by Schein *et al.*<sup>11</sup> is exceptional in providing sufficient data to allow direct calculation of sensitivity, specificity and LRs provided by the animal models studied. Once again, however, the values quoted by the authors cannot be used directly, as they are based on incorrect definitions (regrettably, this has not prevented the values being cited directly by other authors<sup>13</sup>). When calculated correctly (see Appendix), the LRs of the animal models examined by Schein *et al.*<sup>11</sup> have 95% confidence intervals that fail to exclude unity for all ten of the organ system toxicities considered. In other words, the data provide no statistically credible evidence that these animal models contribute any predictive value, either separately or in combination.

## So do animal models have predictive value?

The debate over the use of animals in medical experiments has a long and often bitter history. Researchers and those associated with them can now find themselves targeted by verbally if not physically abusive 'activists'. It is therefore perhaps not surprising that the research community has responded by becoming more assertive in its claims. In 1990, many eminent researchers, including six Nobel Prize winners, signed a declaration prepared by the Research Defence Society that made the wholly unexceptionable claim that experiments on animals have made 'an important contribution' to advances in medicine and surgery. By 2005, with intimidation against those involved in animal experiments regularly making headlines, the declaration had been extended to include the statement discussed in this essay: that 'virtually every medical achievement in the past century' has relied on animal models in some way.

As we have seen, despite its now routine use by the scientific community, it is far from clear that this statement has been, or even could be, formally validated. This is not to say that animal models do not provide evidential weight, still less that they have no role in research. There are many examples of research on animals providing insights that have transformed medical science. Regrettable as it might be, however, it is not possible to go beyond these anecdotal examples to the altogether more impressive statement now being promoted by various prestigious academic bodies and individuals.

The scientific community can choose to deal with the current situation in one of three ways. The simplest is to replace the current statement with one which can be formally validated. This need not be a vapid platitude: there is a wealth of evidence to support a statement such as 'Animal models can and have provided many crucial insights that have led to major advances in medicine and surgery'.

The second and most valuable course of action would be to embark on a systematic study of the use of animal models with a view to establishing the weight of evidence they provide. This would undoubtedly be a major undertaking, but it would also bring many benefits – not the least of which would be quantitative support for the claims made for animal models.

The third option is simply to turn a blind eye to the continued promulgation of a statement about the importance of animal experiments lacking in logical or evidential support.

## References

- 1 US Department of Public Health. *The Physiologist* 1994;**37**:107
- 2 Royal Society. *Statement of the Royal Society's Position on the Use of Animals in Research*. London: Royal Society, 2002. See <http://www.royalsoc.org/displaypagedoc.asp?id=11471>
- 3 UK Department of Health. *Memorandum for House of Lords Select Committee on Animals In Scientific Procedures*. London: DoH, 2001
- 4 RDS petition. See <http://www.rds-online.org.uk/upload/docs/Declaration%202005.pdf>
- 5 RDS Timeline. See [http://www.rds-online.org.uk/pages/page.asp?i\\_ToolbarID=3&i\\_PageID=37](http://www.rds-online.org.uk/pages/page.asp?i_ToolbarID=3&i_PageID=37)
- 6 Comroe JH, Dripps RD. Scientific basis for the support of biomedical science. *Science* 1976;**192**:105–11
- 7 Grant J, Green L, Mason B. *From Bedside to Bench: Comroe and Dripps Revisited*. London: Brunel University, 2003. (Health Economics Research Group Research Report No 30). See <http://www.brunel.ac.uk/3289/Herg1/RR30.pdf>
- 8 Paton W. *Man & Mouse: Animals in Medical Research*. 2nd edn. Oxford: Oxford University Press, 1993
- 9 Paulos JA. *Innumeracy: Mathematical Illiteracy and its Consequences*. London: Viking, 1989
- 10 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;**329**:168–9. See <http://www.bmj.com/cgi/content/full/329/7458/168>
- 11 Schein PS, Davis RD, Carter S, Newman J, Schein DR, Rall DP. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 1970;**11**:3–40
- 12 Heywood R. Target organ toxicity. *Toxic Lett* 1981;**8**:349–58
- 13 Greaves P, Williams A, Eve M. First dose of potential new medicines to humans: how animals help. *Nature Rev Drug Discov* 2004;**3**:226–36
- 14 Nuffield Council on Bioethics. *The Ethics of Research Involving Animals* 265. London: Nuffield Council on Bioethics, 2005. See [http://www.nuffieldbioethics.org/fileLibrary/pdf/RIA\\_Report\\_FINAL-opt.pdf](http://www.nuffieldbioethics.org/fileLibrary/pdf/RIA_Report_FINAL-opt.pdf)
- 15 House of Lords Select Committee on Animals in Scientific Procedures Report 2002. *Chapter 4, sections 4.28*. London: TSO, 2002. See <http://www.parliament.the-stationery-office.co.uk/pa/ld200102/ldselect/ldanimal/150/15001.htm>
- 16 Olson H, Betton G, Robinson D, *et al.* Concordance of the toxicity of pharmaceuticals in humans and animals. *Regul Toxicol Pharmacol* 2000;**32**:56–67
- 17 Hottendorf GH. Species differences in toxic lesions in human risk assessment. In: Roloff MV, (ed.) *The Role of Animal Selection and Extrapolation*. London: Taylor & Francis, 1987, 87–96

## Appendix: Calculation of evidential weight provided by animal models

To assess the evidential weight of an animal model, we require data from a comparative study capable of giving values for the sensitivity and specificity of the model. Schein *et al.* provides such data in the specific case of organ toxicity for anticancer drugs; however, their values for the sensitivity and specificity are based on incorrect definitions, and so cannot be used to calculate evidential weight as reflected in the Likelihood Ratio (LR), where  $LR = \text{sensitivity} / (100\% - \text{specificity})$ . Similar problems affect the values quoted by Hottendorf.

To illustrate the consequences, the following contingency table contains comparative data taken from Schein *et al.* for cases of injection site toxicity as observed in monkeys and humans.

	Toxicity observed in human	Toxicity not observed in human	Total
Toxicity observed in animal	3	6	9
Toxicity not observed in animal	2	12	14
Totals	5	18	23

Based on this data, we then calculate the various measures of evidential weight according to the definitions adopted by Schein *et al.* and Hottendorf, and compare them to the values obtained using the correct definition.

Quantity	Schein <i>et al.</i>	Hottendorf	Actual
True positive rate (Sensitivity)	13%	Not calculable	60%
True negative rate (Specificity)	52%	Not calculable	67%
False positive rate	26%	26%	33%
False negative rate	9%	9%	40%
LR (95% CI)	0.27	Not calculable	1.8 (0.6, 3.4)

The final row shows that the LR value calculated using Schein *et al.*'s definitions is incorrect, the value from Hottendorf's definition is not calculable, while the correct LR fails to exclude 1.0 at the 95% confidence level, and thus fails to supply evidence that the animal model has any evidential value.