A Response to the ABPI’s Letter to the Editor on the Use of Dogs in Predicting Drug Toxicity in Humans

Jarrod Bailey, Animal Studies Repository

Available at: https://works.bepress.com/jarrod-bailey/2/
Dear Sir

The major conclusion of our paper on the use of dogs in predicting human drug safety, published in *ATLA* in November 2013,1 was that the absence of toxicity in dogs for any new drug provides essentially no evidential weight regarding its likely lack of toxicity in humans. This conclusion, based on our study of unprecedented scale and statistical rigour, should be of grave concern for those involved in drug development, as the critical observation for deciding whether a candidate drug can proceed to testing in humans is the absence of toxicity in tests on animals. For the second, non-rodent species to contribute essentially no additional confidence in the outcome of pre-clinical testing, but at considerable extra cost both in monetary terms and in terms of animal welfare, is indefensible, both scientifically and ethically.

The Association of the British Pharmaceutical Industry (ABPI) responded to our publication via a Letter to the Editor, published in this issue of *ATLA*.2 The authors of the original paper are pleased that our work is being treated seriously by the industry, and wish to respond to each point made by the APBI, as a further contribution to this important debate.

Each major point is considered in turn, with italicised headings indicating the comments made by the APBI.

**Data quantity:** data in the paper by Bailey et al. are from relatively few studies where toxicity in dogs has been observed. This, together with the use of ratios in its analysis, has created indexes with low reliability, hence the wide “ranges”.

These ‘relatively few studies’ provided data on 2,366 compounds and 3,275 comparisons of effects between humans and dogs. The scale of our study was therefore unprecedented, being considerably larger than any previously published analysis, at least to our knowledge. It is clear that the data we analysed were from numerous separate, studies where toxicity in dogs has been observed. We concede, as we did in our paper, that our study was limited to published and publicly available data. Relative to all the data generated in dog studies over the 50–70 years that dogs have been used in drug development, it is true that our dataset may be small. However, the data set we used inarguably remains sufficiently large to generate conclusions that are significant. Additionally, that more-comprehensive studies would be possible, if proprietary data were made available by industry (i.e. the industry the ABPI represents), should not mitigate the serious concerns we highlighted, and this should encourage the APBI to facilitate further studies.

We disagree that our analysis is further weakened by ‘indexes with low reliability’, evidenced by ‘wide ranges.’ Firstly, we contend that the wide ranges of likelihood ratios (LRs) actually indicate the low reliability of the dog as a pre-clinical model, not the low reliability of our dataset or our analysis. Yet, in any case, the ranges alone are misleading. While the range of PLRs, for instance, is wide, this is mainly due to outliers. Very shortly, we will publish a follow-up paper that addresses this very issue. An appreciation of quantiles and the inter-quartile range of the PLRs limits the ‘skew’ of the data caused by these outliers, and we also address the effect of rare adverse events on the data. Crucially, there is nothing in this follow-up study that weakens the case we make in our original publication — indeed, it augments it.

**Statistical bias:** Data on compounds that have been shown to be significantly toxic in animal studies are unlikely to have been tested in humans. Thus, the dataset in the paper by Bailey et al. does not include clinical data where toxicity data from dogs have prevented a compound progressing into clinical development. As mentioned by the authors, the lack of this data type unfortunately significantly biases the statistical analysis and underestimates the value of dog data in predicting human safety. Nonetheless, Bailey et al. feel justified to conclude that compounds that are toxic in dog are likely to be toxic in humans.

We analysed our data by using LRs, which, as we argue at some length, are the most relevant and appropriate metric to use. LRs cannot be calculated when there are no human data for a given compound, such as when the development of a new drug has been terminated prior to human exposure due to adverse drug reactions (ADRs) detected preclinically. Indeed, none of the statistical metrics associated with assessing the predictive nature of
animal tests for humans, such as sensitivity, specificity, and positive predictive value, can be calculated for any drug, without human data as well as animal data. This point, as made by the ABPI, is therefore of no relevance to our study. Further, we cannot know which of the drugs terminated in this way would have gone on to cause human toxicity, and which would not; we cannot know which drugs showed false positive results (FPs) in dogs, and which showed true positive results (TPs). Such drugs cannot be part of a dataset used to calculate how much evidential weight an animal model is providing to the probability of human toxicity/non-toxicity. The omission of data from such drugs is therefore inevitable, and cannot contribute to the ‘underestimation of the value of dog data in predicting human safety.’

**Challenges in interpretation:** Contrary to the conclusions of the paper by Bailey et al., the range of calculated inverse negative likelihood ratios (iNLRs) are all > 1, which suggests that dog studies do add some evidential weight. Furthermore, this value would be greater, if those dog studies that had prevented a compound progressing into clinical development had been included.

Our conclusions are not contrary to the iNLR values we calculated. The median iNLR is just 1.11 — it could barely be more marginally greater than 1 — which validates our conclusion that ‘...dogs provide essentially no evidential weight to this aspect of toxicity testing. Specifically, the fact that a compound shows no toxic effects in dogs provides essentially no insight into whether the compound will also show no toxic effects in humans.’ The most trivial and insignificant amount of evidential weight (‘benefit’) added by dogs in this respect cannot be justified by the associated welfare and financial costs. To claim, in addition, that the inclusion in our analysis of ‘dog studies that had prevented a compound from progressing into clinical development’ is erroneous, for the reasons stated above.

**Challenges in handling complex and diverse data sets:** The paper by Bailey et al. groups together adverse effects impacting on different systems — although it may be more challenging, it would be more informative to analyse individual systems. Dog data are likely to be more predictive when this precision is introduced.

Undoubtedly, the dog data would be more predictive — or ‘less poorly predictive’ — for some classifications of ADRs, if individual systems were analysed. However, this was not what we set out to do, and glosses over the salient point, which is ‘that a compound shows no toxic effects in dogs provides essentially no insight into whether the compound will also show no toxic effects in humans.’ Importantly, we stated in our paper that there was ‘no overall pattern regarding the form of toxicity’ with regard to LRs. Overall, our major conclusion stands.

**Dose/exposure considerations:** Consideration of dose/exposure is fundamental to the risk assessment of chemicals/pharmaceuticals. As Paracelsus stated, ‘Poison is in everything, and nothing is without poison. The dosage makes it either a poison or a remedy’. If a drug candidate showed significant toxicity in animals (be it in the rodent or non-rodent species), typically, the maximum dose used in subsequent clinical trials is capped to ensure that the exposure achieved in humans stays below that shown to produce the toxicity in animals. Thus, the fact that a toxicity seen in dog does not translate into humans might be because the toxic exposure was not achieved in humans, rather than indicating the dog is not a predictive model. Toxicity studies in both dogs and rodents serve a major role in preventing toxicities in humans by predicting the type of toxicity, the dose-exposure response of the toxic effect, and by determining how to monitor for the toxicity with biomarkers.

It is, of course, undeniable that any substance may be toxic at some dosage. However, this seems a desperate attempt by those who wish to defend the use of animal studies to balk at our analysis. Asserting that many compounds that were significantly toxic in animals were not toxic in humans because the dose was too low, is extraordinary. Firstly, this was not in the remit of what we were attempting to do. Secondly, it was not possible in any case, given the available data. Thirdly, the ABPI betrays the worth of animal tests in its claim, by accepting that compounds proceed into human trials even when they ‘showed significant toxicity in animals.’ We return to what we set out to do: given the data that were available to us, we wanted to assess the evidential weight provided by dog tests to the probability that a new drug would be toxic or non-toxic to humans.

**Incidence considerations:** For compounds that produced toxicities/adverse events in humans that were not seen in dogs, differences in dose/exposure are unlikely to be the explanation (since typically higher doses/exposures are used/achieved in toxicity studies than in a clinical setting). However, in such cases there appears to have been no consideration of incidence. For example, for a human adverse event that occurs at low incidence, say one case in ten thousand patients, it might be unrealistic to expect that toxicity to be seen in a dog study that would typically only use 6 animals (3 males, 3 females) per dose group.
Such scenarios — very rare human toxicities that are not detected in animal studies, or even in clinical trials, inescapably represent an enormous problem for the pharmaceutical industry, are not a valid objection to our study. Once again, this is not part of what we set out to do, and should not detract from our results and conclusions. Nevertheless, our soon-to-be-published follow-up paper does address this issue, by examining, as much as was possible, the effects of rareness of ADRs on the magnitude and variability of LRs. Notably, first indications suggest that, if rare ADRs in humans and/or dogs are discounted, median PLRs decrease in value significantly, lessening the evidential weight provided by dog tests, while median iNLRs increase slightly, only marginally improving on their general failure to add significant evidential weight in this respect. Once again, the case made in our original paper remains.

What initiatives are on-going to address the data gap?

In this section of their Letter, the ABPI authors cite several previous studies that they believe highlight the value of the non-rodent as a second species in toxicity testing. Firstly, this is not what we set out to do — our analysis, on which they are commenting, expressly focused on the dog, and not ‘non-rodents’ collectively. Although the dog is the most commonly used non-rodent in this respect, this is an important distinction. Notably, we were aware of most of the publications that they cite, and, indeed, cited some of them in our paper. We contend that the ABPI is overestimating the support these papers lend to their position, for the following reasons:

— Broadhead et al.3: This paper was far less supportive of dog studies than the ABPI authors suggest, and indeed, was quite critical of them. While it was noted that 37% of unpublished repeat-dose toxicity studies in rats and dogs (involving 115 new chemical entities [NCEs]) showed ‘new findings’ in the dog, and that, of 159 NCEs, 18 (11%) were terminated due to ADRs seen uniquely in the dog, they cautioned that ‘...the survey did not include information on the relevance of adverse effects in animals to humans; therefore, no firm conclusions can be drawn with regard to the need for dog studies on the basis of these data.’

— Olson et al.4: This study did not estimate specificity. Without it, the evidential weight provided by the animal models cannot be calculated. As the authors acknowledged, ‘a more complete evaluation of this predictivity aspect will be an important part of a future prospective survey.’

— Greaves et al.5: Much of the evidence in this review is from a paper by Schein et al.,6 which is referenced 13 times. However, the data analysis in this study is based on incorrect statistical definitions, which, when calculated correctly, actually show that non-rodents (primates and dogs) fail to contribute any human-predictive value, either separately or in combination.7

— Baldrick8: This paper looked at supporting information for 34 first-in-man drug trials. It noted that, in general, non-rodents (mainly, though not exclusively, dogs) were more sensitive to target organ toxicity than rodents in 17 cases, though the opposite was true in another eight cases. No human relevance was noted, as none of these toxicities prevented the drugs from proceeding to human trials.

— Horner et al.9: This paper looked at only 75 drugs with rodent and non-rodent data, compared to our 2,366 with dog and human data. Further, it considered only small molecule candidate drugs, as opposed to the varied classes of drugs included in our analysis, which also had reached market and therefore had significant associated human toxicity data. Notably, 18 of the compounds referred to in the paper did not progress to human trials. Further, while the dog was the most commonly used non-rodent species, it was not the only one: the authors also included non-human primates in their consideration. They concluded simply that new target organs were identified by the non-rodent tests for 43 of the drugs. The human relevance of this is unknown.

— Tamaki et al.10: This paper looked at only 142 approved drugs, and did not make a comparison between rodents and non-rodents. It noted that animal toxicities correlated with human toxicities, on average, in just 48% of cases. However, these correlations varied substantially from less than 30% to more than 70%, depending on the target organ. Further, the paper looked only at toxicity correlations: our study used the more-appropriate and comprehensive LRs.

— ABPI unpublished data: The human relevance of the 42/221 drug development termination decisions is not provided, probably because it cannot be determined (similarly, see Broadhead et al., summary above).

The evidence provided by the ABPI in this section of its response constitutes an inadequate argument against the analysis and related conclusions in our paper. Compared to our study: the data are more limited in scope; the analyses are more limited; in some cases, the analyses are incorrect; the human relevance of the findings is not known; the significance of the papers is exaggerated; and, in no case
have the authors attempted the type of study and analysis that we conducted.

While any efforts by the industry to further alternatives, for example, by developing various in silico and in vitro human-specific approaches, are extremely welcome, and indeed, crucial, for the future benefit of both humans and animals, there is no evidence to support the ABPI’s assertion that ‘currently available in vitro and in silico tests are inadequately predictive for general toxicity testing of medicines to ensure patient and volunteer safety.’ An important part of furthering alternatives is to be as aware as possible of the performance and human relevance of what those alternatives are designed to replace, i.e. animal tests. Critical assessment of animal toxicity is therefore paramount. Unfortunately, dogged and desperate efforts by those who practise animal toxicology to discredit such assessments if they are critical of animal models, are unhelpful. This is especially so, when parallel attempts are made to fortify such arguments by citing, and overstating the importance of, inferior and inadequate reports.

**Summary**

We are in total agreement with the ABPI that ‘informed debate amongst key stakeholders’ is important. However, this should include animal protection groups, addition to their suggested partnership between ‘the pharmaceutical industry, academia and the regulatory authorities.’

Unfortunately, as we have tried to explain, we believe that the ABPI authors have been too eager to play down the importance of our study. Far from being constructive, their critique is overly critical to the point of being unhelpful. Criticism must be fair and robust. When it has a weak scientific basis, it is unacceptable, especially when it is used as a basis to persist with the status quo, for which there is weak supportive evidence.

We acknowledge (as we did in our paper) that there are inherent and unavoidable caveats to our analysis. Notwithstanding this, our study is unprecedented in scale and type. Notably, it is unacceptable, especially when it is used as a basis to persist with the status quo, for which there is weak supportive evidence.

We are confident about, and as certain as we could be, of our analysis and our conclusions. They should not be overlooked because of their source, and because they are inconvenient to any stakeholders (including the Federation of Laboratory Animal Breeders Associations, which, not surprisingly, supports the ABPI’s arguments). Should the ABPI be genuinely concerned that our analysis, using all the data and correct statistical methods at our disposal, is not sufficient, we refer it to our statement within our paper: ‘…if any pharmaceutical industry stakeholders have issues or concerns with our conclusions, we would encourage them to conduct further analyses by using their own proprietary data, and/or to facilitate such investigations by making available anonymised data, in accordance with the promotion of transparency encouraged by EU Directive 2010/63/EU, as well as to engage fully in constructive discussion and debate with us and our colleagues in animal protection organisations.’ The only acceptable response by industry and regulators should be, at the very least, a desire for positive action to facilitate, or conduct, if possible, another study on a greater scale, involving the use of proprietary data. To ignore the most relevant findings, and/or focus on other inferred shortcomings of other elements of the study, is unacceptable, particularly when the upshot is to defend and perpetuate seriously flawed dog toxicology, with little or no substantiation.

Yours faithfully

Jarrod Bailey

**British Union for the Abolition of Vivisection (BUAV)**

16a Crane Grove

London N7 8NN

UK

E-mail: jarrod.bailey@mac.com
References


