Author’s response to reviews

Title: Recent Efforts to Elucidate the Scientific Validity of Animal-Based Drug Tests by the Pharmaceutical Industry, Pro-Testing Lobby Groups, and Animal Welfare/Three Rs Organisations.

Authors:

Jarrod Bailey (jarrod.bailey@crueltyfreeinternational.org;jarrod.bailey@mac.com)

Michael Balls (Michael.Balls@btopenworld.com)

Version: 1 Date: 30 Nov 2018

Author’s response to reviews:

Cruelty Free International

16a Crane Grove

London

N7 8NN

Email: jarrod.bailey@crueltyfreeinternational.org

Tel: 07939 036460

November 30th, 2018

METH-D-18-00055 - Recent Efforts to Elucidate the Scientific Validity of Animal-Based Drug Tests by the Pharmaceutical Industry, Pro-Testing Lobby Groups, and Animal Welfare/3Rs Organisations.

Jarrod Bailey; Michael Balls

Dear Sir/Madam,

Thank you for the opportunity to resubmit our manuscript, which has been substantially revised in light of the reviewers’ thoughtful and constructive comments.
Their salient points are copied below, in italics and underlined, with our responses beneath each one. All changes are tracked in the new draft of the manuscript.

We hope they are helpful, and look forward to hearing from you in due course regarding acceptance and publication.

Reviewer 1

the authors spend quite a bit of time up front discussing a back and forth about their previous work and its criticism, which is frustrating for the reader. It would be better to set the stage more simply and directly for the current paper and argument.

We have revised the manuscript in an attempt to make it less of a response to previous criticisms of our work in the area, as suggested. Some defence of our work and our approach is still there, as we feel it is important, as well as relevant because of how the two recent studies by Monticello et al., and Clark & Steger-Hartmann were conducted and reported, which are discussed in our manuscript. However, we believe the paper is now less defensive and reactive, which is helpful.

In general, the structure and presentation of the argument would be improved by a more nuanced and accessible approach to the reasons on each side of the debate.

We never wished to make our manuscript a discussion of both sides of the debate! Its importance, we believe, lies in its critical attitude to the status quo (i.e. that animal safety testing is valuable and predictive for humans, and therefore indispensable). The rationale for this is that it is urgently needed: the value of animal tests is often stated and accepted without evidence, i.e. the literature is replete with efforts to reinforce the status quo; yet conversely, at the same time, criticisms of those animal tests are held to different standards, which must be substantiated with voluminous evidence, which will never be sufficient for many. That said, in the interests of balance, we discussed some of the studies often proffered as evidence of the predictive nature of
animal tests for humans, as well as findings from our own and others’ analyses suggesting that there are some specific areas in which animal tests predict human toxicity quite well.

is it not still the case that drug testing in animals is valuable for human safety if it shows toxicity in animals?

Possibly, but there are caveats. First, the reliability and value of a test is seriously impacted if one significant aspect of it fails—and this is the case for animal testing of drugs (as we argue in the paper), when apparent lack of toxicity is the result in animals. Secondly, data suggest animal tests may be predictive of human toxicity only in some specific areas—i.e. not in others. This must be examined and elucidated in more detail in the future. At the very least, animal tests should not be used in areas of toxicity for which they are not predictive for humans, and the issue of apparent lack of toxicity must be acknowledged for its serious nature, and addressed. The areas in which animals may be predictive for human toxicity should, in the first instance, be the only areas in which they could be considered for use in testing…though this leads to the third caveat: could an array of alternative (human, in vitro and in silico) methods be used instead?

Some of the above was already included in the manuscript; some has been expanded on; others have been added.

the authors seem so focused on defending their previous work and undercutting the arguments of the other authors they examine that the paper fails to paint a broader, more approachable, picture of the main issues at stake in animal drug testing.

These numerous and complex broader issues are discussed in cited references, including our own. For reasons of brevity (and limited space to do so in the manuscript), they could not be adequately discussed in this paper.
Overall, while not doubting their scientific objectivity in the least, it seems the authors are invested in a particular perspective on this issue that will appear similarly 'biased', in opposing valence, to that of the drug industry whose work they examine.

We understand this point, but believe we are open and honest about our viewpoints—we have serious reservations about the scientific validity of many, if not all, types of animal research intended to benefit humans. While I (Jarrod Bailey), personally, object to animal experiments ethically, I am objective in terms of how I see the evidence available, and so my scientific perspective is not biased. I am pleased that this reviewer appreciates this!

The main objective of this paper is to argue that some of the major conclusions of two papers by other authors, discussed in it, are not supported by the available evidence, which has serious ethical implications. It is an important case that needs to be made, and we simply hope to highlight the argument, and further debate and discussion.

Reviewer 2

The authors should emphasise the limits of the paper i.e. confined to the safety testing of drugs for use in humans. I acknowledge that while this is obvious to the authors and others in the field, it may not be to the general readership of BMC ME and so it may be helpful to make this distinction more clear. For example, it may not apply to many other areas of animal testing e.g. vaccine safety and development, other medicinal uses in surgery and medicine e.g. anaesthesia, analgesia, anthelmintics, antimicrobials, drugs to be used in veterinary medicine. The authors are concentrating on toxicity and safety not efficacy (see e.g. page 4 lines 1-8, page 5 line 33), which is clearly important as that is when animals are most likely to suffer pain, distress and lasting harm.

We have added some text to clarify this. Particularly in the ‘Background’ section and some sub-headings. We have also mentioned the ethical angle of the high degree of suffering for animals involved, in the second paragraph of the Conclusion.
There is no discussion of the ethical issues that are raised by the use of animals in drug testing, which are numerous and would be worth making those points in the paper, albeit briefly.

We have added some text to highlight the ethical issues with animal drug testing, both with regard to animal suffering and death, as well as consequences for human consumers of drugs if these animal tests are not sufficiently predictive. See, in particular, the second paragraph of the Conclusion.

As the debate hinges around statistics, I wondered about some input from a disinterested professional statistician input: eg the best predictors of human toxicity (and efficacy) being LRs and not PVs?

Please see added information in the first paragraph of the Conclusion. References to this end are also provided in the third paragraph under the ‘Monticello et al., 2017’ sub heading.

Sub-Editing points:

Page 13, Line 59: reads ".. as is the case in drug development — this excuse is cannot be valid." Should it not read "…this excuse is not valid. Or this excuse is not valid”?

Corrected.

Reviewer 3

the “initial guess” of toxicity in a specific drug under development would be different according to their supposed mechanisms. Moreover, one cannot obtain “prior” belief in toxicity for new
compounds in advance. Hence I think the authors’ calculation of LR in their data lacks such Bayesian interpretation of evidence that would be appropriate in diagnostic context.

We have added some brief text to describe what we mean by this in the last paragraph of the ‘Background’ section, “Quantitatively, if, for example, a new drug has (based on prior information, such as similarity to other drugs, data from in vitro or in silico tests, and so on) a 70% chance of not being toxic in humans…” (added text in brackets and bold). This was approved by the statisticians’ input to our own analyses.

likelihood is a relative measure of different models: specific cut-point such as “positive LR > 10 and/or negative LR < 10 is significant” should not be referred to.

This was a cut-off point used by Monticello et al., one of the two main papers we discussed in the manuscript—not in my own work.

All statistical measures should be always used in reference to “cost” in some sense in scientific practice. So it is insufficient for one to judge scientific values referring to just the statistical measures including LR and PPV/NPV only from a “prediction” viewpoint (page 13), even if LRs are measures of statistical evidence. I am not familiar with practical concerns in preclinical studies and cannot make additional comments here; however, the future discussion on this topic would help the stakeholders fill up a gap between them.

This is a good point, which we have addressed in the second paragraph of the Conclusion, with the following text:

“All suffering in animal drug testing is often severe and prolonged: animals used in chronic toxicity and carcinogenicity studies, for instance, receive the test substance daily, seven days a week, for two years with no recovery periods [National Toxicology Program (NTP)], and The Organisation for Economic Co-operation and Development (OECD) [OECD, 2000] and the Nuffield Council on Bioethics [Nuffield council on bioethics, 2005] list the following as common conditions and clinical signs that may occur during such tests, which indicate an animal is experiencing pain
and/or distress, and suffering: Gasping, difficulty breathing, excess salivation and nasal discharge, tremor, changes in blood pressure, seizures, convulsions, coma, abnormal vocalization, aggression, diarrhoea, vomiting, bleeding from any orifice, oedema, abdominal rigidity, rectal or vaginal prolapse, swollen joints, and paralysis.”

It is of course my speculation, but under this conditions, true positive (= toxic both in human and animals) would have been underestimated and false positive (toxic in animals but not in human) would have been overestimated. This leads both positive and negative LRs’ underestimation.

As we discussed in some detail in each of our own three papers on the subject, this is true, but will be true of any dataset, given the precautionary principle of not progressing to human testing with any drug significantly toxic in rodents, for example. To illustrate briefly, as we opined in our second paper of three:

“Naturally, there must be caveats. Our analysis was limited to data that are published and publicly available. It is widely acknowledged that many animal experimental results/preclinical data remain unpublished and/or proprietary, for a variety of reasons (e.g. 15, 27–30). Such publication bias is a major problem (e.g. 31–34), and, compounded by other factors such as size and quality of the animal studies, variability in the requirements for reporting animal studies, ‘optimism bias’, and lack of randomisation and blinding (28, 35), it means that gauging the true contribution of animal data to human toxicology is impossible — at least for third parties without access to pharmaceutical company files. All datasets are imperfect to varying degrees. However, it is only possible to use data which are available, and to ensure that, as far as feasible, those data are of good quality and as free from biases as possible, and that their analysis and derived conclusions are as objective as possible.”

LR is a good measure for statistical evidence but is not the only measure of scientific value in human drug development, and I believe that no single statistical measure can be sufficient for judging the adequacy of scientific activities. The discussion on the topic is welcome; to make a compromise between different positions, broader range of statistical measures (e.g. possibly including more complex Bayesian modeling or cost-benefit analysis) as well as drug-specific circumstance may be necessary.
We have included a response to this in the first paragraph of the Conclusion, and also some of the suggested future analysis/work!

Dr. Jarrod Bailey (and on behalf of co-author, Prof. Michael balls).