Session 2: Progress in Preclinical Testing for Translational Science

Toxicologic Pathology, 43: 70-77, 2015

ISSN: 0192-6233 print / 1533-1601 online
DOI: 10.1177/0192623314555162

The Future of Preclinical Animal Models in Pharmaceutical Discovery and Development: A Need to Bring *In Cerebro* to the *In Vivo* Discussions

Jeffrey I. Everitt[1]

[1]*GlaxoSmithKline, Research Triangle Park, North Carolina, USA*

Abstract

Animal models have provided an important tool to help make the decision to take potential therapies from preclinical studies to humans. In the past several years, the strong reliance of the pharmaceutical discovery and development process on the use of animal models has come under increasing scrutiny for ethical and scientific reasons. Several prominent and widely publicized articles have reported limited concordance of animal experiments with subsequent human clinical trials. Recent assessments of the quality of animal studies have suggested that this translational failure may be due in part to shortcomings in the planning, conduct, and reporting of *in vivo* studies. This article will emphasize methods to assure best practice rigor in animal study methods and reporting. It will introduce the so-called scientific 3Rs of relevance, robustness, and reproducibility to the *in vivo* study approach and will review important new trends in the animal research and pharmaceutical discovery and development communities.

Keywords: animal model; study bias; reproducibility; ARRIVE guideline; 3Rs.

Introduction

The use of animals in biomedical research, including those utilized in pharmaceutical discovery and development efforts, has come under increasing public scrutiny. While certain elements in the animal rights community have long questioned the value of preclinical studies for informing human health, more recently serious questioning of the value of animal studies has arisen within the scientific community itself (Perel et al. 2007; Pound et al. 2004; van der Worp 2010). The lack of concordance of some human clinical trial outcomes with preclinical animal efficacy findings has led to a questioning of the methods of how animal studies are conducted (Macleod 2011; van der Worp and Macleod 2011). In addition, the lack of concordance of animal efficacy data with human trial outcomes has led to a questioning of the fundamental role that animal-based studies have played in the preclinical arena for certain therapeutic areas (Pound and Bracken 2014; Sena et al. 2007). There have been several prominent publications questioning the translational value of preclinical animal models (Couzin-Frankel 2013; Seok et al. 2013), as well as a growing number of articles that question the quality of *in vivo* study data (Hackam 2007; Henderson et al. 2013; Howells and Macleod 2013; Kilkenny et al. 2009; Macleod et al. 2009). This combination of factors has led to a storm of controversy concerning the value of animal studies in general and whether they truly are needed to develop therapies to meet unmet human and animal medical needs (Mestas and Hughes 2004; Osuchkowski et al. 2014; Perrin 2014; Sundberg et al. 2013). There is a great opportunity for the bioscience community as a whole, and the drug discovery and development community in particular, to own concerns about the quality of animal experimentation and to improve upon the way in which animal research is planned, conducted, reported, analyzed, and interpreted. It is the contention of the author that veterinary and toxicologic pathologists that work with program and project teams in drug discovery, by virtue of their training, background, and experience, are ideally suited to advance the important cause of bringing a more rigorous approach to preclinical animal studies.

This article is based on an invited talk presented by the author at the 33rd Annual Symposium of the Society of

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Address correspondence to: Jeffrey I. Everitt, Glaxo SmithKline, 5 Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709, USA; e-mail: jeffrey.x.everitt@gsk.com.

Abbreviations: 3Rs, replacement, reduction, and refinement; ADME, absorption, distribution, metabolism, excretion; ARRIVE, Animal Research Reporting *In Vivo* Experiments; CAMARADES, Collaborative Approach to Meta-analysis and Review of Animal Data from Experimental Studies; ILAR, Institute of Laboratory Animal Research; MIBBI, Minimum Information for Biological and Biomedical Investigations; NC3Rs, National Centre for the Replacement, Refinement and Reduction of Animals in Research; PD, pharmacodynamic; PK, pharmacokinetic; SR, systematic review; SYRCLE, Systematic Review Centre for Laboratory Animal Experimentation.

Toxicologic Pathologists in a plenary session dealing with progress in preclinical testing in translational science. A disclaimer is warranted as many issues of animal-based research are both complex and controversial. The thoughts expressed in this article belong to the author alone and may or may not be shared by present and/or previous employers. They are based on the aggregate career experience of a toxicologic pathologist who has spent many years leading animal care and use programs. This article will review some current trends in the pharmaceutical industry that influence the use of preclinical animal efficacy models. It will also review important areas for improvement in methodology to enhance the planning, conduct, and reporting of preclinical efficacy studies. Perhaps most importantly, it will discuss the role for the toxicologic pathologist in advancing these studies in order to bring about needed change in study conduct.

Background

The primary objective of the pharmaceutical industry is to create new therapies to solve unmet medical needs to advance the human condition. To date, this pursuit has involved the use of numerous animals of various species in a wide variety of models and testing paradigms. Detailed discussion of testing paradigms and models exceed the scope of this article, but it should be noted that the majority of animal use (rodents) takes place early in drug discovery in preclinical studies from target validation to the selection of candidates, early in the drug discovery/development process. The last decade has seen numerous advances in refinements and changes in methods in how animals are used in drug discovery and development. These include (but of course are not limited to) advances in noninvasive imaging modalities, telemetric monitoring of animals, microsampling methods, and genetic manipulation of rodents and other species. In general, there has been a trend to multiplex end points to maximize information from fewer and fewer animals with an ever-increasing complexity to the models and experimental designs. Advances in animal model technology have been used in conjunction with a wide array of technological advancements in end point assays, as well as numerous experimental advances using computer simulation, *in silico*, *in vitro*, and other nonanimal approaches alone and in combination.

There are unique challenges for pharmaceutical company animal care and use programs that operate in a global multisite environment across multiple regulatory and oversight systems. Little international agreement exists for harmonization concerning ''best practice'' in the animal care and use world. In addition, companies often combine significant internal operations and capability with extensive interactions with academic, industry, and governmental collaborations as well as work at a myriad of contract research organizations. Animals are cared for and used in a highly regulated environment, but what makes this arena somewhat unique within the larger highly regulated pharmaceutical research enterprise is the fact that public thinking varies around the world concerning how animals should be cared for and used. These differing public attitudes are based on a variety of legal and regulatory frameworks, cultural, religious, and political differences.

There are a number of recent trends in the pharmaceutical industry environment that impact preclinical model usage and practices. One of the most impactful has been the significant increase in the globalization and externalization of discovery activities. There have been increasing academic and intercompany collaborations, as well as connectivity of the industry through public–private partnerships. Working with well-conceived and described standard methods and protocols as well as with excellent reporting methods is increasingly recognized as mission critical. The pharmaceutical industry has recognized the value in working together in the precompetitive space to advance the sharing of knowledge with animal models, as well as the opportunity to work across boundaries to advance translational science. Examples of this are the cooperation and support of the industry with the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) in the United Kingdom and with the Institute of Laboratory Animal Research (ILAR) Roundtable in the United States.

The 3Rs

The cornerstone of international policy on animal-based science is embodied in the replacement, reduction, and refinement (3Rs) set out in the now famous treatise of R. Russell and W. M. S. Burch entitled ''The Principles of Humane Experimental Technique'' (London, 1959). Replacement refers to the preferred use of nonanimal methods over animal methods whenever it is possible to achieve the same scientific aim. Reduction refers to methods that enable researchers to obtain comparable levels of information from fewer animals or to obtain more information from the same number of animals. Refinement is much broader and refers to methods that alleviate or minimize potential pain, suffering, or distress, and enhance animal welfare for the animals still used. Most investigators use the term refinement for the replacement of one animal model with another model of a species lower on the phylogenetic spectrum, whereas others think of this in the category of replacement with nonanimal models. Collectively, these 3Rs have been deemed the so-called alternative approaches, and they are referred to in many of the legal and regulatory frameworks used in animal care and use oversight. There are many instances where 3Rs advancements have been incorporated to enhance preclinical animal model science (Chapman et al. 2013; Emerson 2010; Madden et al. 2012). Despite the connection of the 3Rs to scientific advancement, many North American scientists more readily associate the 3Rs with animal welfare issues rather than with science. A recent survey in Canada showed that many *in vivo* scientists did not believe that replacement was possible at this time. A number of them strongly felt that the focus should be on quality animal data rather than a focus on reduction in animal numbers (Fenwick, Danielson, and Griffen 2011).

Some scientists would suggest that framing the issue of animal use around reduction is simplistic and goes against scientific principles, potentially leading to compromise of study quality. These scientists suggest that the conversation should center on optimization of animal numbers rather than reduction in animal use per se. In this scenario, the scientific conversation would revolve around refinement of experimental design and utilization of robust statistical expertise. In many but not all cases, statistical models and experimental design approaches can be incorporated to reduce animal numbers (Festing and Altman 2002). In some instances, optimizing design means that more animals would be needed to obtain the desired scientific aims. The concern with focus on reduction per se was discussed in a recent *Nature Genetics* editorial (Editorial Board 2012). This editorial called attention to the studies of Scott et al. (2008) that showed that the failure of murine amyotrophic lateral sclerosis treatments to translate to the clinic was due to small group size numbers and underpowered experiments. Another potential point of conflict between the application of 3Rs and scientific principles would be if the pressure to reduce animal usage impeded the ability to repeat studies or inhibited the use of potentially valuable satellite groups.

In recent years, there has been significant diminution in the numbers of animals used within the pharmaceutical industry relative to R&D expenditure, and there has been significant focus on application of 3Rs principles as exemplified by the collaboration of the industry with the NC3Rs (Chapman et al. 2013). In addition, the creation in 2012 of the 3Rs Leadership Group within the IQ Consortium is another indication that the pharmaceutical industry intends to provide cross-industry cooperation in this area. The IQ Consortium is a pharmaceutical and biotechnology association that aims to advance innovation and quality in the biopharmaceutical industry.

In the view of the author, good animal welfare is inextricably linked to good science with preclinical animal models, but they are not one and the same. Invariably, many animal studies are a balance between competing needs. Since some scientists are most interested in their scientific objectives as opposed to animal welfare issues, it is proposed that they would readily identify with a second set of "3Rs" (those being relevance, robustness, and reproducibility) that are more aligned to scientific data associated with animal-based experiments. Relevance refers to how the model translates to address the scientific question at hand. Often in the pharmaceutical industry, it pertains to the ability of the preclinical model to predict the intended clinical outcome. Robustness refers to the strength of the model to remain unaffected by minor variations in test conditions that may occur over the course of experimentation in a single laboratory. Reproducibility pertains to the ability to utilize the model repetitively under the same conditions and obtain the same outcome. It is likely that by focusing on the scientific data of *in vivo* studies, one will get traction and engagement with scientists. In the view of the author, most scientists are quick to engage deeply when they are convinced that the aim is to enhance their scientific endeavor.

The issue of reproducibility in animal studies has been of great concern as evident by an ILAR-sponsored workshop in June 2014 at the U.S. National Academy of Sciences. This workshop was conducted in part to explore the concerns that have arisen because a significant number of peer-reviewed studies have not been able to be reproduced. There is evidence that a variety of factors may contribute to the lack of reproducibility in animal studies, including incomplete reporting of details of study design and conduct, inadequate means to control study bias, and a lack of methodological rigor in study conduct (Kilkenny et al. 2009; Macleod et al. 2009). The finding of lack of animal study reproducibility and in the quality of reporting of animal studies is particularly vexing to the pharmaceutical industry due to the recent trend to greater externalization of discovery research activities, and the trend to increased industry academic collaborations and partnerships. The pharmaceutical industry has experienced a significant disparity between published findings and the ability to validate results internally in both animal and nonanimal preclinical studies (Begley and Ellis 2012; Prinz, Schlange, and Asadullah 2011). One experience that illustrates this concern is a report from the hematology/oncology department at Amgen, where industry scientists chose 53 "landmark" publications, many in top-tier journals, and were only able to confirm the findings completely in 6 of the publications (Begley and Ellis 2012). One of the lessons that one can take away from this lack of reproducibility is the need to critically evaluate studies and not just depend on the journal peer review processes. When evaluating a scientific article, the reader should critically assess the materials and methods section to assure that reporting is robust. Studies should be examined with respect to methods to control bias (randomization, blinding, etc.), whether controls are proper, if all results and repetitions are shown, as well as to verify the use of appropriate reagents and statistical methods (Begley 2013).

INCREASING THE METHODOLOGICAL QUALITY OF *IN VIVO* STUDIES

Planning the Animal Study

The methodological quality of an animal study starts with the study planning process and is best performed with the "end in sight." Many of the methodological details believed to be important for reporting so that a study can be evaluated and reproduced (such as the checklist of factors in the Animal Research Reporting *In Vivo* Experiments [ARRIVE] guidelines), are considered important in the genesis of the animal experimental protocol. In the experience of the author, animal protocol development is often lacking in rigor and thus it should come as little surprise that the reporting end of animal studies often lacks proper detail. Animal studies are complex and difficult to properly plan and conduct due to the need to interface biological studies that have inherent variability, with complex animal care and use factors that can affect study conduct and quality. The quality of the animal care and use program and in-life portion of the study contributes in large measure to the overall quality of an animal experiment. The protocol should have enough detail to outline the major expected causes of variation in an experiment and methods used to minimize.

TABLE 1.—Aspects of preclinical study to be detailed in protocol to help assure study quality.

- What are the statistical methods to be used in the study?
- Sample size calculation—how was the sample size calculated and what assumptions were made?
- What are exclusion criteria and how are animals to be removed while on the study? If humane end points are specified, what is the chain of decision making?
- How are animals to be assigned to study groups to minimize bias?—methods of randomization to be employed
- How are animals to be acclimated and to what are they being acclimated to?
- What, if any, blinding and allocation concealment methods are to be used in the study?
- What are the pharmacodynamic (PD)/pharmacokinetic (PK) considerations for the study—what are dose administration methods and what timing considerations are to be employed?
- What is to be done with animals removed from study unscheduled?

The generation of a high-quality protocol for an *in vivo* study is best effected through discussions of the primary investigative team with experts in ancillary disciplines (statistics, laboratory animal science, pathology, etc.), as well as those directly engaged in the data generation and collection process. The best protocols in the author's experience are those that get robust input early and often from interdisciplinary interactions during the early stages of experimental design. This early engagement cannot be overemphasized, as there is abundant evidence in the literature that experimental design and statistical appropriateness are often subpar (Hess 2011; Kilkenny et al. 2009). Early in the genesis of the protocol, the experimental hypothesis to be tested as well as the experimental aims are discussed in a peer interaction process. The model to be studied as well as the basic experimental design and end points to be employed is discussed and debated. In the next stage, with the help of statisticians and other subject matter experts, a detailed protocol is assembled. At this time, the factors responsible for major anticipated sources of variation are discussed as well as methods to control variation and ameliorate bias. In the final stages of protocol genesis, the logistical concerns of the study are discussed with those that are actually responsible for the study conduct (bench scientists, laboratory managers, animal care technical personnel, etc.). The phrases to remember when designing animal studies are ''the devil is in the details,'' and the study is ''only as strong as its weakest link.'' It is useful at this stage to risk assess the life cycle of the study. This consists of determining the most critical phases, so that proper oversight resource and methods can be assigned. In the view of the author, it is often possible to determine those critical study activities where error and sources of variation can occur. Examples of this are when animals are transported, dosed with substance, bled or biopsied, exposed to novel housing or equipment. Whenever there is movement of animals or equipment, performance of novel handling or procedures, weekend or holiday periods, or hand-off from one investigator or staff member to another, there is potential and likelihood for greater variation in study effects and oversight vigilance is warranted.

Experimental Design Considerations in Preclinical Efficacy Studies

Many aspects of study quality are important to consider in the experimental design for preclinical animal studies (Table 1). Preclinical efficacy studies in pharmaceutical discovery generally do not entail specific considerations that are not important in other animal studies, but there is a special emphasis on understanding pharmacodynamic (PD)/pharmacokinetic (PK) relationships and absorption, distribution, metabolism, excretion (ADME) associated with compound administration and target engagement. Inherent to this is a determination of the best methods, form, and timing of dose administration. It is also critically important to determine what methods are to be used to minimize bias on study. Recently, there have been calls for preclinical animal studies to be conducted much more akin to the methodology of randomized clinical trials using similar methods for bias control (randomization, blinding, sample size justification, data transparency, etc.; Muhlhausler, Bloomfield, and Gillman 2013). There are many potential sources of bias to be considered across all study phases ranging from design to interpretation (Lindner 2007). In a number of therapeutic areas, there are best practices that are being formulated to advance the translational value of preclinical studies, and undoubtedly this effort will continue and accelerate (E. Sena et al. 2007; Scott et al. 2008; Shineman et al. 2011).

It cannot be over emphasized, that while there are excellent default positions for best practice such as the use of appropriate randomization methods, and blinding and allocation concealment techniques, there is no single best practice for all circumstances. Customization of each protocol is required as optimal study design is not a ''one-size-fits-all'' proposition. The author has on many occasions found utility in walking statisticians through animal use procedures to have meaningful dialog on how to minimize and control for bias and variation on study. What is good in theory and what is good in practice sometimes differs markedly. For example, in chronic dosed feed experiments, some study design personnel might ask to have cages of rodents randomly distributed across racks in an animal room to minimize variation due to environmental influences. Animal husbandry reality would dictate that the likelihood of cross contamination across groups might pose far greater risk and impart greater degree of variation on the study.

Sources of Variation in Preclinical Animal Studies

There are numerous sources of variation to consider and control in the design and conduct as well as in the reporting of *in vivo* studies. These include inherent factors of the animal

(e.g., stock/strain/substrain, source, sex, age, weight, source, pathogen status, etc.) as well as in the animal facility environment (diet, bedding, housing, water delivery, lighting, noise, vibration, temperature, humidity, etc.). In addition to the numerous animal care and use factors that can impact the results of preclinical animal studies, study methods themselves can influence outcomes and it has to be determined what level of detail needs to be considered important for reporting and replication. For example, with diet as one factor alone, the investigator might need to consider nutritional content, contaminant levels, amount, dietary form, feeding method, or other factors. In the specification of lighting, the investigator might need to consider light level, length of light period, spectral quality, time of lighting and whether gradual in onset and withdrawal. The scientific objectives of the study ultimately dictate what sources of variation are most important to consider and control.

In drug discovery and development studies, the method, dose form and timing of dose administration, types and preparation of excipients and vehicles, blood and tissue sampling sites and methods, handling of subjects, and a myriad of other experimental conditions can all markedly effect study outcomes. Things that may appear to be minor procedural details can prove important. An example is the finding that significant difference in the serum hepatic enzyme, alanine transaminase, can occur if mice are handled by the body instead of the tail (Swaim, Taylor, and Jersey 1985). In addition to effects of animal handling, there have been significant reported differences in research end points, such as cytokine concentrations, depending on method/site of blood removal (Mella et al. 2014). There have been many reviews of factors that are important to consider when conducting and reporting animal studies (Alfaro 2005; Hooijmans, Leenaars, and Ritskes-Hoitinga 2010; Institute for Laboratory Animal Research 2011; Kilkenny et al. 2010). A recent report that caught the eye of the popular press found that even the sex of the researcher could affect the outcome of certain rodent studies through olfactory cues (Sorge et al. 2014).

Reporting Animal Studies

A significant challenge to being able to reproduce preclinical animal studies is the general lack of transparent and accurate reporting in the literature. Analyses of published animal studies have demonstrated numerous deficiencies in the reporting of details in research methods (Kilkenny et al. 2009; Vesterinen et al. 2011). Despite multiple publications over the past 25 years calling attention to the critical factors and information necessary to enhance such reporting (Alfaro 2005; Ellery 1985), there have been continuing deficiencies in the literature. Lack of sufficient experimental procedural detail concerning animal studies has both scientific and ethical implications. It limits the ability to confirm and build on research findings and may lead to the unnecessary use of animals in studies that fail to reproduce the reported results. In addition, it may mask problems in the quality of the design and conduct of animal studies, as well as limit the ability to use systematic reviews (SRs; Hooijmans, Leenaars, and Ritskes-Hoitinga 2010).

Efforts to improve the reporting of biological and biomedical investigations resulted in the Minimum Information for Biological and Biomedical Investigations (MIBBI) project to enhance reporting in biological experiments (Taylor et al. 2008). Authors, reviewers, and journal editors as well as professional societies all have a role in the effort to increase the quality of publication of animal-based research. Recently, there have been a number of significant reports giving guidance for enhancing the description of animal studies including preclinical efficacy studies (Hooijmans et al. 2011; Hooijmans, Leenaars, and Ritskes-Hoitinga 2010; ILAR 2011; Kilkenny et al. 2010). This guidance helps define requirements for accurate descriptions of the research animal as an experimental test system, the critical elements of the research animal environment, the aspects of experimental design, and animal care and use practices that affect research results. The ARRIVE Guidelines from the U.K. National Center for the 3Rs have been endorsed by a variety of journals and funding sources. Unfortunately, despite the fact that the ARRIVE guidelines have been widely endorsed, the implementation has been much less successful to date in the first 2 years since its publication (Baker et al. 2014). Although the ARRIVE guidelines and other guidance listed earlier are excellent tools, the author recommends that each journal customize guidance based on the types of studies reported and the particular information and details needed to ensure reproducibility. There have been some excellent specialized guidance developed for some highly specialized areas of animal studies. Published guidance for preclinical imaging studies and chronobiology experiments are exemplars (Portaluppi, Touitou, and Smolensky 2008; Stout et al. 2013). The importance of reporting quality for preclinical animal studies was noted in a recent meeting of stakeholders by the U.S. National Institute of Neurological Disorders and Stroke in June 2012 that discussed how to improve the methodological reporting of animal studies in grant applications and publications (Landis et al. 2012).

Difficulty in reproducing published studies reflects some of the biases in scientific publications (Prinz, Schlange, and Asadullah 2011). It is generally felt that there is a significant publication bias in the literature that should be realized by scientists. There is relatively little incentive for journals to publish negative studies or to publish nonnovel findings and repeated studies. The extreme pressure to publish in top-tier journals for academic survival is having effects on the quality of the animal literature. It has been noted that for some therapeutic areas there are far too many animal studies with statistically significant positive results, and there is an increasing number of studies demonstrating publication bias (Korevaar, Hooft, and Ter Riet 2011; Sena et al. 2010; Ter Riet et al. 2012; Tsilidis et al. 2013).

Recently, there have been suggestions that preclinical animal studies should be conducted, reported, and analyzed more like clinical trials (Muhlhausler, Bloomfield, and Gillman 2013). The author agrees that methodological rigor, control of bias, and transparent reporting are warranted, but it would be potentially short-sighted to just assume that preclinical studies and clinical trials should be conducted in the same manner.

The intended purpose, types of bias, and the collection of end points can be quite different.

SRs

An important recent trend in the analysis of preclinical animal efficacy studies is the increasing use of SRs as well as methods of meta-analysis (van Lujik et al. 2014). In contrast to a narrative review that has no standardized methodology, an SR is a type of review that is structured, thorough, and transparent. The SR aims to identify all relevant studies to answer a particular research question and is comprehensive and evidence based. Data from studies used in SRs are often used in meta-analysis (Vesterinen et al. 2014). The use of SRs is growing substantially in the preclinical efficacy literature, and there have been centers set up such as Collaborative Approach to Meta-analysis and Review of Animal Data from Experimental Studies (CAMARADES) and Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) to help foster design of preclinical studies and methods of analysis (Hooijmans et al. 2014; Ritskes-Hoitinga et al. 2014; Sena et al. 2014).

The Value of the Toxicologic Pathologist in Preclinical Efficacy Study Conduct

Toxicologic pathologists, especially those with a veterinary medical background, are uniquely qualified by training and experience to help improve the conduct of preclinical animal efficacy studies in the pharmaceutical industry. These individuals usually have an excellent background in laboratory animal science and animal model characterization. A very useful trait of the toxicologic pathology working environment is the use of standard operating procedures, standard lexicons and protocols, and well-established reporting practices, as well as the expectation of peer review input. Pathologists work in a subject area that is both subjective and quantitative at the same time, and one in which there are many sources of variation and bias that need to be controlled. Most pathologists engage with statistics personnel and interact on issues such as whether or not pathology data should be gathered blinded with dose group concealment. The debate in the professional societies concerning blinded reading of slides is but one example (Dodd 1988; Temple 1988).

Conclusion

The concept of the *3Rs of relevance, robustness, and reproducibility* is put forth as a transformational cultural change for bioscience research personnel to embrace to enhance the translational value of preclinical animal efficacy studies. The present inability to show good concordance between some animal efficacy studies and human clinical outcomes is believed to be due in part to shortcomings in experimental design and conduct, as well as in the reporting of results. It is believed that the reproducibility and value of preclinical animal efficacy studies would be enhanced by increased methodological rigor in both study conduct and reporting. An increased emphasis on protocol development with a focus on sources of variation and methods to control bias would enhance preclinical animal study quality. The toxicologic pathologist is uniquely trained and experienced to help drive the changes needed to enhance *in vivo* study quality in drug discovery programs. It is hoped that this article will serve as a call to arms for pathologists to engage in order to enhance preclinical efficacy studies.

Author Contribution

Jeffrey I. Everitt drafted the manuscript, gave final approval, and agrees to be accountable for all aspects of work ensuring integrity and accuracy.

References

Alfaro, V. (2005). Specification of laboratory animal use in scientific articles: Current low detail in the journals' instructions for authors and some proposals. *Meth Find Exp Clin Pharmacol* **27**, 495–502.

Baker, D., Lidster, K., Sottomayor, A., and Amor, S. (2014). Two years later: Journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PloS Biol* **12**, e1001756.

Begley, C. G. (2013). Reproducibility: Six red flags for suspect work. *Nature* **497**, 433–34.

Begley, C. G., and Ellis, M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531–33.

Chapman, K. L., Holsgrefe, H., Black, L. E., Brown, M., Chellman, C., Copeman, C., Couch, J., Creton, S., Gehen, S., Hoberman, A., Kinter, L. B., Madden, S., Mattis, C., Stemple, H. A., and Wilson, S. (2013). Pharmaceutical toxicology: Designing studies to reduce animal use while maximizing human translation. *Reg Toxicol Pharmacol* **66**, 88–103.

Couzin-Frankel, J. (2013). When mice mislead. *Science* **342**, 922–25.

Dodd, D. C. (1988). Blind slide reading or the uninformed versus the informed pathologist. *Comment on Toxicol* **2**, 81–91.

Editorial Board. (2012). The three "Is" of animal experimentation. *Nat Genet* **44**, 611.

Ellery, A. W. (1985). Guidelines for specification of animals and husbandry methods when reporting the results of animal experiments. Working Committee for the Biological Characterization of Laboratory Animals/GV-SOLAS. *Lab Anim* **19**, 106–8.

Emerson, M. (2010). Refinement, reduction and replacement approaches to *in vivo* cardiovascular research. *British J Pharmacol* **161**, 749–54.

Fenwick, N., Danielson, P., and Griffen, G. (2011). Survey of Canadian animal-based researchers' views on the three Rs: Replacement, reduction and refinement. *PLoS One* **8**, e22478.

Festing, M. F., and Altman, D. G. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J* **43**, 244–58.

Hackam, D. (2007). Translating animal research into clinical benefit. *BMJ* **334**, 163.

Henderson, V. C., Kimmelman, J., Fergusson, D., Grimshaw, J. M., and Hackam, D. G. (2013). Threats to validity in the design and conduct of preclinical efficacy studies: A systematic review of guidelines for *in vivo* animal experiments. *PLoS Med* **10**, e1001489.

Hess, K. R. (2011). Statistical design considerations in animal studies published recently in *Cancer Research*. *Cancer Res* **71**, 625.

Hooijmans, C. R., de Vries, R., Leenaars, M., Curfs, J., and Ritskes-Hoitinga, M. (2011). Improving planning, design, reporting and scientific quality of animal experiments by using the Gold Standard Publication Checklist, in addition to the ARRIVE guidelines. *Br J Pharmacol* **162**, 1259–60.

Hooijmans, C. R., Leenaars, M., and Ritskes-Hoitinga, M. (2010). A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three Rs, and to make systematic reviews more feasible. *Altern Lab Anim* **38**, 167–82.

Hooijmans, C. R., Rovers, M. M., de Vries, R. B. M., Leenaars, M., Ritskes-Hoitinga, M., and Langendam, M. W. (2014). SYRCLE's risk of bias tool for animal studies. *BMC Med Res Method* **14**, 43. doi:10.1186/1471-2288-14-43.

Howells, D. W., and Macleod, M. R. (2013). Evidence-based translational medicine. *Stroke* **44**, 1466–71.

Institute for Laboratory Animal Research, National Research Council. 2011. *Guidance for the Description of Animal Research in Scientific Publications*. The National Academies Press, Washington, DC.

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PloS Biol* **8**, e1000412.

Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F., Cuthill, I. C., Fry, D., Hutton, J., and Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One* **4**, e7824.

Korevaar, D. A., Hooft, L., and Ter Riet, G. (2011). Systematic reviews and meta-analyses of preclinical studies: Publication bias in laboratory animal experiments. *Lab Anim* **45**, 225–30.

Landis, S. C., Amara, S. G., Asadullah, K., Austin, C. P., Blumenstein, R., Bradley, E. W., Crystal, R. G., Darnell, R. B., Ferrante, R. J., Fillit, H., Finkelstein, R., Fisher, M., Gendelman, H. E., Golub, R. M., Goudreau, J. L., Gross, R. A., Gubitz, A. K., Hesterlee, S. E., Howells, D. W., Huguenard, J., Kelner, K., Koroshetz, W., Krainc, D., Lazic, S. E., Levine, M. S., Macleod, M. R., McCall, J. M., Moxley, R. T., Narasimhan, K., Noble, L. J., Perrin, S., Porter, J. D., Steward, O., Unger, E., Utz, U., and Silberber, S. D. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **490**, 187–91.

Lindner, M. D. (2007). Clinical attrition due to biased preclinical assessments of potential efficacy. *Pharmacol Ther* **115**, 148–75.

Macleod, M. (2011). Why animal research needs to improve. *Nature* **477**, 511.

Macleod, M. R., Fisher, M., O'Collins, V., Sena, E. S., Dirnagl, U., Bath, P. M., Buchan, A., van der Worp, H. B., Traystman, R. J., Minematsu, K., Donnan, G. A., and Howells, D. W. (2009). Good laboratory practice: Preventing introduction of bias at the bench. *Int J Stroke* **4**,3–5.

Madden, J. C., Hewitt, M., Pryzbylak, K., Vandebriel, R. J., Piersma, A. H., and Cronin, M. T. D. (2012). Strategies for the optimisation of *in vivo* experiments in accordance with the 3Rs philosophy. *Reg Toxicol Pharmacol* **63**, 140–54.

Mella, J. R., Chiswick, E. L., King, E., and Remick, D. G. (2014). Location, location, location: Cytokine concentrations are dependent on blood sampling site. *Shock* **42**, 337–42.

Mestas, J., and Hughes, C. C. (2004). Of mice and not men: Differences between mouse and human immunology. *J Immunol* **172**, 2731.

Muhlhausler, B. S., Bloomfield, F. H., and Gillman, M. W. (2013). Whole animal experiments should be more like human randomized controlled trials. *PLoS Biol* **11**, e1001481.

Osuchkowski, M. F., Remick, D. G., Lederer, J. A., Lang, C. H., Aasen, A. O., Aibiki, M., Azevedo, L. C., Bahrami, S., Boros, M., Cooney, R., Cuzzocrea, S., Jiang, Y., Junger, W. G., Hirasawa, H., Hotchkiss, R. S., Li, X. A., Radermacher, P., Redl, H., Salomao, R., Soebandrio, A., Thiemermann, C., Vincent, J. L., Ward, P., Yao, Y. M., Yu, H. P., Zingarelli, B., and Chaudry, I. H. (2014). Abandon the mouse research ship? Not just yet! *Shock* **41**, 463–75.

Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., Macleod, M., Mignini, L. E., Jayaram, P., and Khan, K. S. (2007). Comparison of treatment effects between animal experiments and clinical trials. Systematic review. *BMJ* **334**, 197.

Perrin, S. (2014). Preclinical research: Make mouse studies work. *Nature* **507**, 423–25.

Portaluppi, F., Touitou, Y., and Smolensky, M. H. (2008). Ethical and methodological standards for laboratory and medical biological rhythm research. *Chronobiol Int* **25**, 999–1016.

Pound, P., and Bracken, M. B. (2014). Is animal research sufficiently evidence based to be a cornerstone of biomedical research. *BMJ* **348**, g3387.

Pound, P., Ebrahim, S., Sandercock, P., Bracken, M., and Roberts, I. (2004). Where is the evidence that animal research benefits. *BMJ* **328**, 514–17.

Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* **10**, 712.

Ritskes-Hoitinga, M., Leenaars, M., Avey, M., Rovers, M., and Scholten, R. (2014). Systematic reviews of preclinical animal studies can make significant contributions to health care and more transparent translational medicine. *Cochrane Database Syst Rev* **28**, ED000078.

Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., Bostrom, A., Theodoss, J., Al-Nakhala, B. M., Vieira, F. G., Ramasubbu, J., and Heywood, J. A. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler* **9**, 4–15.

Sena, E., van derWorp, H. B., Howells, D., and Macleod, M. (2007). How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci* **30**, 433–9.

Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R., and Howells, D. W. (2014). Systematic reviews and meta-analysis of preclinical studies: Why perform them and how to appraise them critically. *J Cereb Blood Flow Metab* **34**, 737–42.

Sena, E. S., van der Worp, H. B., Bath, P. M., Howells, D. W., and Macleod, M. R. (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* **8**, e1000344.

Seok, J., Warren, H. W., Cuenca, A. G., Mindrinos, M. N., Baker, H. V., Xu, W., Richard, D. R., McDonald-Smith, G. P., Gao, H., Hennessy, L., Finnerty, C. C., Lopez, C. M., Honari, S., Moore, E. E., Minei, J. P., Cuschieri, J., Bankey, P. E., Johnson, J. L., Sperry, J., Nathens, A. B., Billiear, T. R., West, M. A., Jeschke, M. G., Klein, M. B., Gamelli, R. L., Gibran, N. S., Brownstein, B. H., Miller-Graziano, C., Calvano, S. E., Mason, P. H., Cobb, J. P., Rame, L. G., Lowry, S. F., Maier, R. V., Moldawer, L. L., Herndon, D. N., Davis, R. W., Xiao, W., and Tompkins, R. G. (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. *PNAS* **110**, 3507–12.

Shineman, D. W., Basi, G. S., Bizon, J. L., Colton, C. A., Greenberg, B. D., Hollister, B. A., Lincecum, J., Leblanc, G. G., Lee, L. H., Luo, F., Morgan, D., Morse, I., Refolo, L. F., Riddell, D. R., Scearce-Levie, K., Sweeney, P., Yrjanheikki, J., and Fillit, H. M. (2011). Accelerating drug discovery for Alzheimer's disease: Best practices for preclinical animal studies. *Alzheimers Res Ther* **3**, 28–41.

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., Wieskopf, J. S., Acland, E. L., Dokova, A., Kadoura, B., Leger, P., Mappebeck, J. C., McPhail, M., Delaney, A., Wigerblad, G., Schumann, A. P., Quinn, T., Frasnelli, J., Svensson, C. L., Sternberg, W. F., and Mogil, J. S. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods* **11**, 629–32.

Stout, D., Ber, S. S., LeBlanc, A., Kalen, J. D., Osborne, D., Price, J., Schiffer, W., Kuntner, C., and Wall, J. (2013). Guidance for methods descriptions used in preclinical imaging papers. *Mol Imaging* **12**, 1–15.

Sundberg, J. P., Roopenian, D. C., Liu, E. T., and Schofield, P. N. (2013). The Cinderella effect: Searching for the best fit between mouse models and human diseases. *J Invest Dermatol* **133**, 2509–13.

Swaim, L. D., Taylor, H. W., and Jersey, G. C. (1985). The effects of handling techniques on serum ALT in mice. *J Appl Toxicol* **5**, 160–62.

Taylor, C. F., Field, D., Sansone, S. A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P. A., Bogue, M., Booth, T., Brazma, A., Brinkman, R. R., Clark, A. M., Deutsch, E. W., Fiehn, O., Fostel, J., Ghazal, P., Gibson, F., Gray, T., Grimes, G., Hancock, J. M., Hardy, N. W., Hermjakob, H., Julian, R. K., Jr., Kane, M., Kettner, C., Kinsinger, C., Kolker, E., Kuiper, M., Le Novère, N., Leebens-Mack, J., Lewis, S. E., Lord, P., Mallon, A. M., Marthandan, N., Masuya, H., McNally, R., Mehrle, A., Morrison, N., Orchard, S., Quackenbush, J., Reecy, J. M., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Rosenfelder, H., Santoyo-Lopez, J., Scheuermann, R. H., Schober, D., Smith, B., Snape, J., Stoeckert, C. J., Jr., Tipton, K., Sterk, P., Untergasser, A., Vandesompele, J., and Wiemann, S. (2008). Promoting coherent minimum reporting guidelines for

biological and biomedical investigations: The MIBBI project. *Nat Biotechnol* **26**, 889–96.

Temple, R., Fairweather, W. R., Glocklin, V. C., and O'Neill, R. T. (1988). The case for blinded slide reading. *Comment on Toxicol* **2**, 99–109.

Ter Riet, G., Korevaar, D. A., Leenaars, M., Sterk, P. J., VanNoorden, C. J. F., Bouter, L. M., Lutter, R., Elferink, R. P. O., and Hooft, L. (2012). Publication bias in laboratory animal research: A survey on magnitude, drivers, consequences and potential solutions. *PLoS One* **7**, e43404.

Tsilidis, K. K., Panagiotou, O. A., Sena, E. S., Aretouli, E., Evangelou, E., Howells, D. W., Salman, R. A., Macleod, M. R., and Ioannidis, J. P. A. (2013). Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS One* **11**, e1001609.

van der Worp, H. B. (2010). Can animal models of disease reliably inform human studies? *PLoS Med* **7**, e100024.

van der Worp, H. B., and Macleod, M. R. (2011). Preclinical studies of human disease: Time to take methodological quality seriously. *J Mol Cell Cardiol* **51**, 449–50.

Van Luijk, J., Bakker, B., Rovers, M. M., Ritskes-Hoitinga, M., de Vries, R. B., and Leenaars, M. (2014). Systematic reviews of animal studies; missing link in translational research? *PLoS One* **9**, e89981.

Vesterinen, H. V., Egan, K., Deister, A., Schlattmann, P., Macleod, M. R., and Dirnagl, U. (2011). Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. *J Cereb Blood Flow Metab* **31**, 1064–72.

Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., Antonic, A., Howells, D. W., and Macleod, M. R. (2014). Meta-analysis of data from animal studies: A practical guide. *J Neurosci Meth* **221**, 92–102.