

ORIGINAL ARTICLE

Replication – why we need to publish our findings

Michael J. Curtis¹ & Darrell R. Abernethy²¹Cardiovascular Division, King's College London, Rayne Institute, St Thomas' Hospital, London SE17EH, United Kingdom²US FDA, Office of Clinical Pharmacology and Office of Translational Science Sci, Silver Spring, MD 20903 USA**Keywords**

Experimental design, replication

Correspondence

Michael J. Curtis, Cardiovascular Division, King's College London, Rayne Institute, St Thomas' Hospital, London SE17EH, United Kingdom. Tel: +44 (0) 20 71991085; Fax: +44(0)2071883902; E-mail: michael.curtis@kcl.ac.uk

Received: 10 June 2015;

Accepted: 12 June 2015

Pharma Res Per, 3(4), 2015, e00164,
doi: 10.1002/prp2.164

doi: 10.1002/prp2.164

The article originated from discussions emanating from the Pharmacology 2014 meeting on London (December 2014) at which Pharmacology Research and Perspectives hosted a symposium on replication in research http://www.bps.ac.uk/SpringboardWebApp/userfiles/bps/file/Meetings/meetings%202014/Pharmacology%202014/Pharmacology%202014_Reduced.pdf.

Introduction

Pharmacology journals and societies are aware that the publication landscape is changing. For a history and update of the “transparency agenda” we recommend the following articles: Curtis and Abernethy (2015); Curtis et al. (2013); McGrath and Curtis (2015, 2015); McGrath et al. (2015). Indeed, one of our partner publications has recently introduced new guidance for authors that provides very precise requirements on what constitutes an experiment that is fit for analysis (Curtis

Abstract

The pharmacology research sector is changing to accommodate a need for greater transparency and better standards. The themed articles contained herein explain how Pharmacology Research and Perspectives (PR&P) has responded to this agenda. This issue of PR&P contains three articles that consider the reliability of pharmacological research publications, and approaches to their improvement in this regard. This first article explains the importance of publishing findings that confirm or repudiate published findings (so called “replication” studies). It also emphasizes that PR&P actively encourages submission of such articles, and seeks to oppose the publication bias that favors publication of “positive” findings. The second paper explores some initiatives to publish “negative” clinical findings, including a PR&P initiative. The final paper elaborates a toolkit that can be applied to drug discovery research to facilitate the reliability of findings.

Abbreviations

ALS, amyotrophic lateral sclerosis; IF, impact factor; VF, ventricular fibrillation.

et al. 2015). In the present article, we consider a separate important issue that presents a challenge to pharmacology: replication.

Positive results in psychology can behave like rumors: easy to release but hard to dispel

Ed Young, Nature 485, 298–300, 2012

The quote above applies equally well to pharmacology and toxicology, as the data on positive findings that inform it are almost identical for the disciplines (Fanelli 2010). Why is this important?

Publication Bias

The preponderance of positive findings in the literature is a longstanding and well-recognized phenomenon (Chalmers 1990) and is a particular phenomenon in pharmacology and toxicology (Fanelli 2010). As noted recently (Curtis et al. 2015) there are examples of pharmacological research areas for which there is a long sequence of positive preclinical findings dating back many years, but from which translation (the creation of medicines for human use) has not followed. The sequence was found to exceed 50 in a survey of the amyotrophic lateral sclerosis (ALS) field (Scott et al. 2008). This is a statistically unlikely outcome that reflects either (1) a fundamental flaw in the preclinical research, in that the preclinical data, though correct, simply do not translate to the human disease, or (2) selective reporting of positive study findings (i.e., a publication bias).

Interestingly, “selective outcome and analysis reporting” was proposed as the most plausible basis for the extreme and statistically unlikely array of publication bias identified in another biomedical research field recently (Tsilidis et al. 2013).

If negative findings were being published in proportion to their preponderance, it would mean that nearly all the ideas that we have had are correct (at a rate of around 90% in pharmacology and toxicology, according to the Fanelli survey of 2010). If pharmacologists were so extraordinarily talented, the drug pipeline would be full to bursting. This is certainly not the case (Curtis and Pugsley 2012). This means the preponderance of positive published findings represents either (1) the unchecked publication of false findings (ranging from unfortunate experimental error, or choice of flawed animal models, to fraudulent misrepresentation of data, all of it missed by peer review) or (2) publication bias (there being nothing egregiously wrong with the work, but the equal or larger volume of negative findings obtained by researchers simply has not made it into print).

Given that, statistically, a much greater preponderance of negative publications is expected than is found (Tsilidis et al. 2013), then negative findings must exist. The question for us is not how the literature has become so overburdened by this unfeasible preponderance of positive findings, but: where have all the “negative findings” gone? Clearly, they exist but are simply not being published.

If we are to remedy this publication bias it would help to understand how it has come about. Part of this is “author decision” bias. Consider, the busiest and most productive labs are busy and productive because they have high income. They obtain plentiful and repeated funding because they publish regularly and in high impact factor (IF) journals. A study that will founder in

peer review will not appear in a high IF journal, and busy and productive labs will not spend time and effort seeking to publish lower impact work when they have higher impact priorities. This means that a negative finding will probably not even be written up let alone sent to a journal by the labs most adept at completing and publishing research. This is one mechanism of publication bias, but it is only part of the process. Our experience identifies three different mechanisms of publication bias, with the more obvious “author decision” component in the middle:

- Journals actively encourage referees and editors to reject papers that are “worthy but dull”. A negative finding will often (perhaps usually) receive such a verdict. This generates bias in favor of strongly positive findings, and highly novel findings (i.e., so new they have not had time to be tested by others). This is “journal bias”.
- Authors know this, and often do not waste time writing and submitting a paper they know will not be accepted by a high IF journal, or be well cited. Busy labs prioritize high impact research. Labs that are not busy generate fewer publications by definition, whether they are positive or negative. Together this generates bias in favor of strongly positive and highly novel findings. This is “author decision” bias.
- The readership often fails to recognize the value of a negative finding. People confuse the falsification of a hypothesis with “it didn’t work”. Negative findings are therefore more likely to be left unread and disregarded, and poorly cited owing to a perception bias (that the work is not important). This is the “community bias”.

The Value of Replication Studies

Publication bias is therefore a pervasive phenomenon, and one that is detrimental, giving a falsely positive misrepresentation of the importance of a mechanism or the usefulness of a drug. The preponderance of positive findings in the literature shows that standards of stringency in peer review appear to be related to the message of the paper, leading to easier acceptance of positive novel findings (see Ball 2004). A more recent survey found that approximately 90% of findings in the pharmacology and toxicology literature are positive (Fanelli 2010). Publication bias in favor of the misleadingly positive is bad enough. But, excessive stringency in peer review of negative findings is equally inappropriate given that it makes peer review daunting when findings are negative. With the benefit to the author of publishing a negative finding already low (citation prospects being low – see later) authors are likely to be particularly reluctant to engage

with a daunting and demanding peer review experience. This is hugely saddening, and we deplore it.

This is especially so because there are two types of negative findings that have great potential value, and publication bias leading to their nonpublication is dangerous. These are:

- Studies that falsify a hypothesis. Such studies show that an idea, perhaps a mechanism for treating a disease, is false and would not work.
- Studies that show that an idea shown previously to be likely to be true, is in fact incorrect.

The latter are *replication studies* that identify “failed replication”. These studies are extremely difficult to publish, in part because the “prior publication” has precedence and the new data are seen as a challenge to established “fact”. This is a misunderstanding of scientific precedence; being first does not guarantee being correct, and the prior publication has no special merit just because it has been published first, as is evident from appraisal of the literature (Mullard 2011).

Unfortunately the article of “priority” is usually the one that is most highly cited, even if it makes conclusions that turn out to be false. The field of animal model validation is one that shows heavy citation bias. In 1980, Sheridan et al. published a paper showing that alpha adrenergic blockade prevents ischemia-induced ventricular fibrillation (VF) in cats (Sheridan et al. 1980). This paper has been cited 408 times (data sourced from Web Of Science on May 15 2015). Four years later another paper was published showing that in a different species (dog), alpha blockers in fact do not suppress ischemia-induced VF (Bolli et al. 1984). This paper has been cited a mere 45 times (data sourced from Web Of Science on May 15 2015). In 1985 Trolese-Mongheal et al. showed that coronary artery ligation in dogs produces rhythm outcomes so variable that group sizes of more than 50 would be needed to reliably detect the ability of a drug to suppress VF. This paper has been cited a paltry 13 times (data sourced from Web of Science on May 15 2015). Thirty years later we are in a position to test the impact of these three studies. The facts are that alpha antagonists are not used to suppress human ventricular arrhythmias (Zipes et al. 2006), and coronary ligation in dogs is no longer used as an assay system to detect new drugs to prevent ischemia-induced VF. Of the three studies cited, the one with the real impact is the final paper (it helped reveal a major limitation of a model that renders the model unreliable, resulting in the model no longer being used). Yet it is by far the least well cited of the three. The “negative finding” with the alpha antagonist study fared little better. In contrast, the paper that found the positive effect of a drug class that has not achieved translation for prevention

of VF (Zipes et al. 2006), attracted 30-fold the number of citations of the least well cited of the three papers, and almost 10 times the number of citations that obtained a different outcome in an animal model. The importance of this example is that, for a range of reasons, preclinical study outcomes may vary, and although the positive finding garners the most attention and citations, it is not always obvious at the time which outcome correctly predicts clinical relevance.

It has become commonplace for authors to obtain a positive finding in a disease model then do several immunoblots to identify “the mechanism”. The paper is likely to be accepted for publication. Let us imagine a situation where the immunoblots identify an apparent cascade of “mechanism,” but the phenotypic effect is no benefit against the disease modeled. This data set falsifies a hypothesis. However, during peer review of the submitted manuscript, the authors may be required to identify the mechanism accounting for the lack of benefit. This may seem reasonable, but if the action of the drug (on a cell process) is simply not effective as a mechanism to achieve the desired outcome, how does one provide a mechanistic explanation? At this point many journals will reject the paper on grounds that the “negative” finding is of “low priority”. The work may never be published if the author loses heart. This represents a potentially detrimental loss of information to the community. Inevitably other research groups will pursue the same hypothesis. Perhaps several will identify the same hypothesis falsification, and again may struggle to publish their findings. Eventually, another group may independently find the opposite: the drug “works” and all the immunoblot changes, which are coincidentally similar to those found by the first investigator, now “fit” the phenotypic change, and the paper is submitted and accepted for publication. Fortune favors the bringer of the news everyone wants to hear. Regardless of who is correct (the author who finds a positive outcome, or the authors who find different), it is much more likely that the positive finding will be published first, and indeed that the positive finding is the *only* finding among the set that enters the literature.

A Partial Solution to Publication Bias

The most obvious solution to the problem of publication bias is to encourage authors to not throw away their negative findings, but to publish them. It is certainly arduous to publish data that falsify a hypothesis when the hypothesis has not previously been tested. However, it is even harder when there is a positive paper published that is perceived by journal referees to have precedence. There is a natural tendency for referees to regard the credence of submitted work differently to the credence of published

work, since the latter has presumably already been subjected to rigorous peer review, whereas the former has not. Therefore, when work is submitted that contradicts published findings, referees are inclined to expect the authors of the new work to explain the discrepancy. The authors of the published work, the original discovery which set the precedent, were of course never required to explain why their findings contradict the new work. If it were the case that all published findings are correct, and that all new work builds on the foundations of established fact, this would be a justifiable approach. If not, giving credence preference to published work represents a form of “peer review bias” (a subject beyond the scope of the present article).

A partial solution, therefore, is for journals to actively encourage negative findings, and treat submitted papers with the same level of scrutiny as would be appropriate for a paper showing positive findings, not a disproportionately greater level. Moreover, referees should be encouraged to not assume that published work necessarily sets a precedent, and if authors of a new submitted article that contradicts published findings adhere to good practice (Curtis and Abernethy 2015) then the work should be evaluated on its own merit. Of course the authors should comment on differences between the study outcomes, and perhaps offer possible explanations, but they should not be *required to prove* the basis for the difference.

As we argued recently (<https://youtu.be/IS8NY6Hx-hAQ>), data that falsify a hypothesis may be of greater value than data that supports a hypothesis, and this needs to be recognized and embraced. Certainly, such data when published, may not attract the cites afforded to a paper that has novel positive findings, but numbers of cites is not linearly correlated with value. *Pharmacology Research and Perspectives* is happy to publish falsification of hypotheses (negative findings) and will not engage in IF calculations when rendering a decision on a manuscript. Scientific merit is our only arbiter of acceptability, with that determined by peer review. Every effort needs to be made to reduce the palpable publication bias towards positive findings that is undermining credibility in preclinical research (Mullard 2011; Scott et al. 2008; Tsilidis et al. 2013). We hope that by making it clear that we value negative findings at least as highly as positive, we will contribute to the culture change that is necessary if preclinical drug discovery research, proof of concept research, and target identification and validation are to obtain a higher rate of translation.

Conclusion

Replication studies are undertaken by laboratories who work at a level of highest standard. Repeating first the key

finding that informed your own hypothesis, before embarking on next stage research, is thankfully commonplace. However, the outcome of such work is rarely if ever published. When the outcome is failed replication, the pressure on you to not publish is even more powerful, especially if the project and its funding are predicated to a large extent by the work you have been unable to replicate. You may be forced to change research direction if daunted by the practicalities of confronting the problem. Replication, and publication of the findings of replication studies, should therefore be facilitated.

In addition to the time, money and animal lives saved by showing a hypothesis to be false and publishing the results, replication has other benefits. In particular the consequence of never attempting to replicate findings is that no false discovery will ever be recognized as such. A recent article in a national newspaper in the UK alludes to this.

Replication is the only solution to scientific fraud

Anonymous, *The Guardian*, 2012 www.theguardian.com/commentisfree/2012/sep/14/solution-scientific-fraud-replication

Pharmacology Research and Perspectives encourages replication studies, and we promise to treat submissions on an equal footing with hypothesis-supporting “positive” findings. The sector is changing, and we must all engage.

Acknowledgements

We thank Ann Hayes and Kat Gore for their contribution to the symposium, and to this mini themed section in the journal.

Disclosure

None declared.

References

- Anonymous (2012). Replication is the only solution to scientific fraud. *The Guardian*. Available at www.theguardian.com/commentisfree/2012/sep/14/solution-scientific-fraud-replication. (accessed 8 June 2015).
- Ball P (2004). The memory of water. The life and work of Jacques Benveniste taught us valuable lessons about how to deal with fringe science. *Nature*. doi: 10.1038/news041004-19 Available at <http://www.nature.com/news/2004/041004/full/news041004-19.html>. (accessed 15 June 2014).
- Bolli R, Fisher DJ, Taylor AA, Young JB, Miller RR (1984). Effect of alpha-adrenergic blockade on arrhythmias induced by acute myocardial ischemia and reperfusion in the dog. *J Mol Cell Cardiol* 16: 1101–1117.

- Chalmers I (1990). Underreporting research is scientific misconduct. *JAMA* 263: 14058.
- Curtis MJ, Abernethy DR (2015). Revision of instructions to authors for pharmacology research and perspectives: enhancing the quality and transparency of published work. *Pharmacol Res Perspect* 3: e00106.
- Curtis MJ, Pugsley MK (2012). Attrition in the drug discovery process: lessons to be learned from the Safety Pharmacology paradigm. *Expert Rev Clin Pharmacol* 5: 237–240.
- Curtis MJ, Hancox JC, Farkas A, Wainwright CL, Stables CL, Saint DA, et al. (2013). The Lambeth Conventions (II): guidelines for the study of animal and human ventricular and supraventricular arrhythmias. *Pharmacol Ther* 139: 213–248.
- Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Gienbycz MA, et al. (2015). Experimental design and analysis and their reporting: new guidance for publication in *BJP*. *Br J Pharmacol* 14: 3461–3471.
- Fanelli D (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE* 5: e10068. doi:10.1371/journal.pone.0010068
- McGrath JC, Curtis MJ (2015). *BJP* is changing its requirements for scientific papers to increase transparency. *Br J Pharmacol* 172: 2671–2674.
- McGrath JC, McLachlan EM, Zeller R (2015). Transparency in Research involving Animals: the Basel Declaration and new principles for reporting research in *BJP* manuscripts. *Br J Pharmacol* 172: 2427–2432.
- Mullard A (2011). Reliability of ‘new drug target’ claims called into question. *Nat Rev Drug Discovery* 10: 643–644.
- Scott S, Ktranz JE, Cole J, Lincecum JM, Thompson K, Kelly N, et al. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotrophic Lateral Sclerosis* 9: 4–15.
- Sheridan DJ, Penkoske PA, Sobel BE, Corr PB (1980). Alpha-adrenergic contributions to dysrhythmia during myocardial ischemia and reperfusion in cats. *J Clin Invest* 65: 161–171.
- Trolese-Mongheal Y, Duchene-Marullaz P, Trolese J-F, Leinot M, Lamar J-C, Lacroix P (1985). Sudden death and experimental acute myocardial infarction. *Am J Cardiol* 56: 677–681.
- Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et al. (2013). Evaluation of excess significance bias in animal studies of neurological disease. *PLoS Biol* 11: e1001609.
- Young E (2012). Replication studies: bad copy. *Nature* 485: 298–300.
- Zipes DP, Camm AJ, Borggrefe M, Buxton AE, Chaitman B, Fromer M, et al. (2006). ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. *J Am Coll Cardiol* 48: e247–e346.