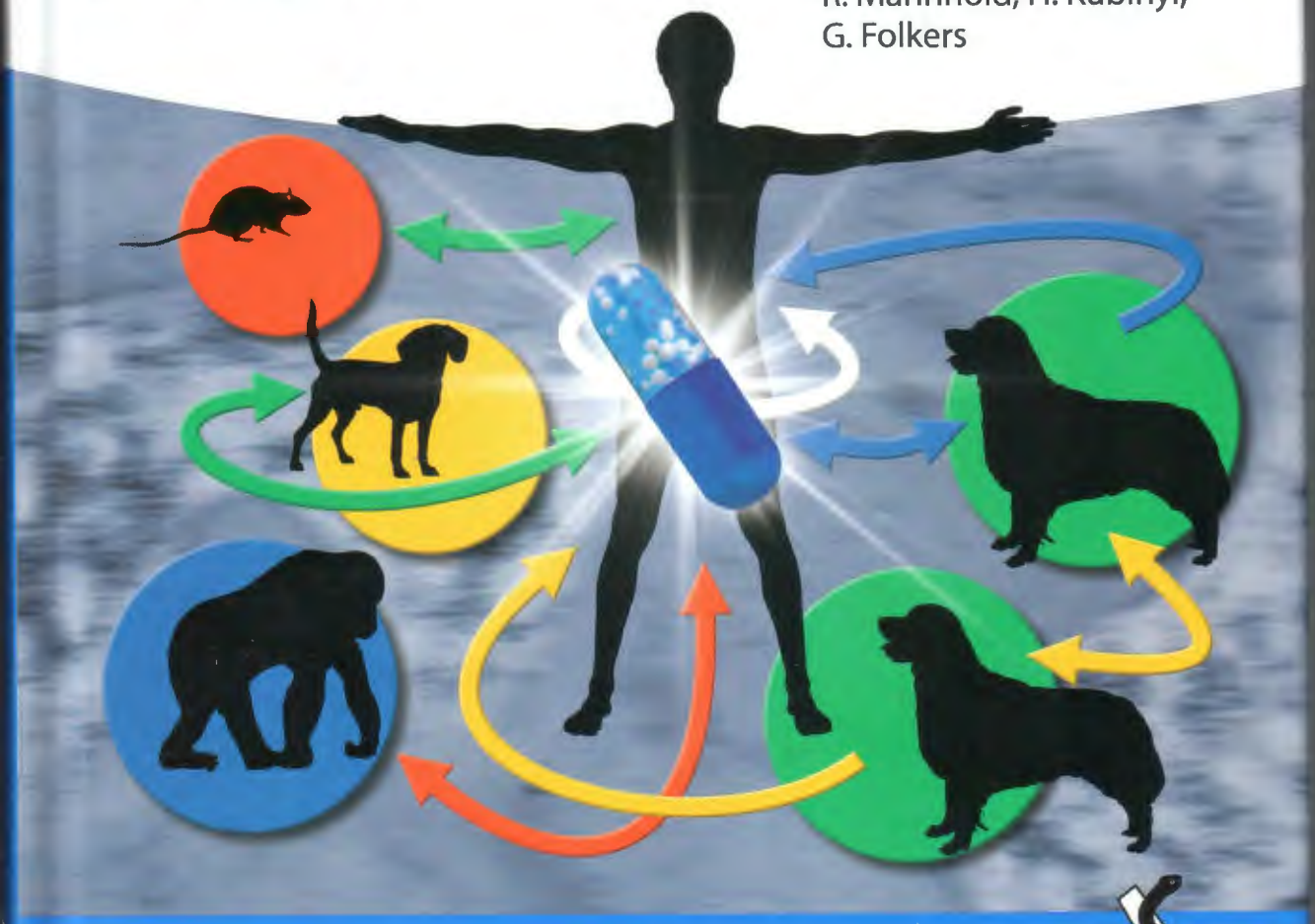
Edited by
Marianne I. Martic-Kehl and P. August Schubiger

Animal Models for Human Cancer

Discovery and Development of Novel Therapeutics

Volume 69

Series Editors:
R. Mannhold, H. Kubinyi, G. Folkers



Methods and Principles in Medicinal Chemistry

5
How to End Selective Reporting in Animal Research

Gerben ter Riet and Lex M. Bouter

> Basic research is like shooting an arrow in the air and, where it lands, painting a target.
>
> *(Homer Adkins,* Nature *1984)*

5.1
Introduction

Would scientific progress not be a lot swifter and cheaper if we published, in some convenient format, all results from our negative studies too? Although convincing evidence is not available, we think the answer would be affirmative. New empirical results appear daily, but it can sometimes take years for *knowledge* to emerge. Isolated studies may be important, but almost all deeper scientific insights evolve at the meta-level; that is, at the level of collections of similar studies around a particular scientific question. Since the 1980s, in clinical medicine and public health, systematic reviews (often including a meta-analysis) of the literature have been increasingly employed to produce ("meta-level") *knowledge* [1]. These systematic reviews ought to be updated when a new piece of evidence comes along. The crucial role of integration of new findings with existing ones is not always appreciated in animal experimental work, although its justification was eloquently expressed over a century ago:

> If, as is sometimes supposed, science consisted in nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight.... Two processes are thus at work side by side, the reception of new material and the digestion and assimilation of the old... The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out. [2]

Timely updating of systematic reviews is needed as evidence keeps accumulating and, at some point, may change the overall picture [3]. The introduction of

Animal Models for Human Cancer: Discovery and Development of Novel Therapeutics, First Edition.
Edited by Marianne L. Martic-Kehl and P. August Schubiger.

systematic reviews has made the clinical scientific community aware that publication bias, the habit of not publishing negative or otherwise unwelcome results, thwarts truth finding and can lead to suboptimal healthcare [4]. It is plausible and there is also some evidence that large portions of the experimental animal literature are also biased because of selective reporting practices [5]. The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES collaboration) is an initiative that brings together data on animal studies and meta-analyzes these where possible. It currently has centers in the UK, Australia, The Netherlands, the USA, and Canada [6]. Its remit is quite similar to that of the Cochrane and Campbell collaborations [7]. Non-publication of complete studies and selective reporting of only a proportion of their results are probably common. Intellectual or financial conflicts of interest along with the widespread misinterpretation and misuse of statistical significance testing appear to be major drivers of selective reporting [8]. Non-publication of "negative" results logically implies that much *wasteful* replication occurs, that is, replication performed inadvertently by investigators unaware of their repeated entry into scientific cul-de-sacs [5c,9]. Conceptually, selective reporting can be viewed as a missing data problem at the meta-level [10]. Therefore, statistical approaches helpful in detecting and repairing bias caused by non-randomly missing data might be relevant to counteract the distortions in the publicly available evidence base [11]. We believe, however, that selective reporting can and must be solved more fundamentally by smart redesign of the research processes [5b,12]. In the field of clinical trials, useful practices such as prospective trial registration, available since 2000 [13] and the promise—as of 2005—of the International Committee of Medical Journal Editors (ICMJE) not to accept any trial-based manuscript for publication unless it has a trial registration number (TRN) [14] were clear signals that major stakeholders wanted to reduce selective reporting. However, it turned out to be difficult for investigators and editors to comply with these initiatives. More specifically, Mathieu and coworkers [15] found that, 5 years down the line, only 45.5% of randomized trials had been pre-registered as intended, and that unregistered trials had nevertheless been published in ICMJE journals. A recent report showed that again, 5 years on, this picture is essentially the same [16]. The US National Institutes of Health has now put in place more carrots and sticks to ensure compliance with the FDA Amendments Act, which requires sharing of summary data within 1 year after completion of data collection [17]. The ICMJE initiative aimed at 100% presence of a TRN for any trial published in an ICMJE-associated journal. However, the more worthwhile goal is the publication of all trials ever performed [18]. Research has shown that in reality, the probability of encountering a TRN in ICMJE journals as well as the probability of a publication being given a TRN both lie around a disappointing 50%, somewhat higher in non-government, non-industry trials, and lower in industry-sponsored trials [15, 19]. Experimental animal research may benefit from the experiences in the field of clinical trials, by copying and by improving procedures with a view to developing a watertight yet efficient system that prevents selective reporting and the ensuing biases in the aggregate literature [20].

In this chapter, after reviewing evidence on magnitude, drivers, consequences of, and solutions to selective reporting, we argue that a future free of selective reporting can be achieved mainly through extending the tasks and jurisdiction of Institutional Animal Care and Use Committees (IACUC) with comprehensive monitoring responsibilities and closer collaboration with sponsors. After all, no animal experiment is allowed to start without ethics approval, making the IACUCs the ideal body to oversee which studies have reached their date of protocol-stipulated completion [21]. A smart and lean system of (electronic) monitoring of the progress of all animal studies started combined with appropriate sanctions could, in principle, put an end to selective reporting. It is a sobering thought that even if we were to end selective reporting practices tomorrow, bias in the publicly available evidence on all hypotheses that are not completely novel will only asymptotically approach zero as the existing, distorted evidence is mixed with new, unbiased, evidence.

5.2
Definition and Different Manifestations of Reporting Bias

Reporting bias occurs if the probability of publication depends on the strength or direction of the results [22]. Thus, the spectrum runs from non-publication of complete studies to non-publication of a selection of the results. Put differently, if we define bias as systematic deviation from the truth, *reporting* bias occurs if the aggregated publicly available evidence (the "pooled estimate") on a particular parameter deviates from the truth because of non-random decisions to publish some research findings but not others. Reporting bias invalidates systematic reviews and meta-analyses and corrupts the cumulative scientific record. Reporting bias in clinical research may lead to errors in clinical practice guidelines and harm patients [23]. In animal studies, reporting bias may cause needless replication attempts and may invite premature first-in-man studies. Reporting bias includes (i) publication bias where whole papers go missing, and (ii) parameter reporting bias where at least one, but not all, measured parameters (risk factor–outcome or intervention–outcome associations) go missing selectively.

5.3
Magnitude of Reporting Biases

There are several ways to learn about the extent of reporting bias. Song *et al.* [22b] distinguished indirect and direct methods. Examples of indirect methods are comparison of the results of large and small studies or assessing the (generally very low) proportion of published studies that do not report any statistically significant finding [8f,24]. Examples of direct methods are, for example, asking scientists [5b,25] or comparing published and unpublished reports [26].

The follow-up of cohorts of study protocols is probably the most robust study design for learning about selective reporting. Possible starting points for follow-up are (i) research protocols in the possession of IACUCs or Medical Research Ethics Committees (MREC) or Institutional review Boards (IRB) as they are called in the USA, (ii) grant applications funded by funding bodies, (iii) entries into web-based trial registries such as clinicaltrials.gov, (iv) abstracts submitted to conferences, and (v) research design papers, such as those published in the BMC series [27]. Reports of study results may be located through dedicated searches of bibliographical databases, such as, for example, Medline and EMBASE, internet searches via Google Scholar, and through contact with researchers. We prefer taking approved submissions to IACUCs or IRBs as a starting point, since these contain the formally approved set of intended measurements that were formulated closest to the date of commencement of studies, whereas plans offered to funding bodies may change after negotiations with sponsors or ethics committees. Follow-up of such cohorts of research protocols has been done for randomized clinical trials [22b,27], but to our knowledge not for animal studies.

Compared with the situation in randomized clinical trials, relatively little is known about the extent of reporting bias in experimental animal research. What we do know are estimates derived from trim-and-fill analyses in the context of meta-analyses [5c] and a survey among animal researchers [5b]. ter Riet *et al.* [5b], in an anonymous web-based survey among 454 Dutch animal researchers, found that respondents believed that overall between 35% and 70% of findings got published and that this was the case for 60–90% of their own work. A subgroup of 21 researchers working for-profit institutes thought that the publication rate was between 5% and 50%, irrespective of whether it concerned their own work or that of others. Size of animals, seniority of researcher, and whether researchers were involved in fundamental research, preclinical research, or both hardly affected these estimates. Survey data on these types of sensitive issues obviously have their limitations. A PubMed search conducted on November 23 2014 located over 25 meta-analyses of animal studies performed by the CAMARADES collaboration. These authors used the statistical *trim and fill* methodology to estimate and repair funnel plot asymmetry [11e] to estimate the relative overestimation of the pooled results in many of their meta-analyses. Across these meta-analyses we calculated a median value of the relative overestimation of intervention effects due to publication bias of 23% (interquartile range from 3 to 45). In a review of 16 reviews comprising 525 animal stroke studies, Sena *et al.* [5c], using trim and fill, estimated that 14% of studies had not been published. Imputing these missing studies lowered the pooled estimate of infarct size reduction across all studies from 31.3 to 23.8%. This was equivalent to a 32% relative bias ((31.3 – 23.8)/23.8). Note that the trim and fill method assumes that forest plot asymmetry is caused by publication bias, which need not be the case; other phenomena may account for (part of) the asymmetry as well. Song *et al.* warned that statistical models to correct for publication bias should be interpreted cautiously: "all statistical methods are by nature indirect and exploratory, and often based on certain strict assumptions that can be difficult to justify in the real world ... the attempt at

identifying or adjusting for publication bias in a systematic review should be mainly used for the purpose of sensitivity analyses" [22a].

5.4 Consequences

To set the scene, we give two examples of the potential harm caused by reporting bias in the area of human randomized trials. Then we will discuss what is known or may be postulated about consequences of reporting bias in experimental animal research.

5.4.1 Consequences of Reporting Bias in Human Randomized Trials

In 1980, a small randomized trial ($N = 95$) showing a 16.6% ($p = 0.015$) excess death rate in men who had a myocardial infarction and were prescribed the anti-arrhythmic drug lorcainide was completed, but remained unpublished. In 1993, the authors, writing about their study, commented that: "It was designed to investigate the effect of lorcainide on arrhythmias, and was never intended to be large enough to allow any conclusions to be reached about an effect of lorcainide on survival. ... The development of lorcainide was abandoned for commercial reasons, and this study was therefore never published; it is now a good example of 'publication bias'. The results described here would have appeared before recruitment to the CAST Study began, and might have provided an early warning of trouble ahead" [28]. Instead of preventing cardiac arrhythmias, lorcainide appeared to trigger them. Only when the CAST trials, testing the drugs encainide, flecainide, and moracizine, in the late 1980s and early 1990s, reproduced these findings were these types of drug withdrawn from the market. In the meantime the number of US patients who had died prematurely due to anti-arrhythmia induced cardiac arrhythmias each year is estimated to be between 20 000 and 70 000.

The Tamiflu (oseltamivir) story may serve as an example of massive economic damage caused by publication bias [29]. In 2008, a Cochrane review on Tamiflu showed the drug's effectiveness against complications of bird flu. Worldwide, developed countries spent billions of dollars (the exact amount is unknown) on stockpiling over 220 million treatments of Tamiflu to protect their populations in case of a bird flu pandemic. After an internet comment by a Japanese physician pointing out that the Cochrane review was mainly based on a manufacturer-sponsored meta-analytic summary of mostly unpublished data, a long struggle over making publicly available all the pertinent trial-based evidence ensued between the Cochrane reviewers and Roche, the manufacturer of Tamiflu [23]. The 2014 version of this Cochrane review, which incorporates much more evidence, shows extremely modest effects of Tamiflu: "For the treatment of adults, oseltamivir reduced the time to first alleviation of symptoms by 16.8 h ($p < 0.0001$). This represents a reduction in the time to first alleviation

of symptoms from 7 to 6.3 days. ... Treatment of adults with oseltamivir had no significant effect on hospitalizations: risk difference (RD) 0.15% (95% CI −0.78 to 0.91). Oseltamivir significantly reduced self-reported, investigator-mediated, unverified pneumonia (RD 1.00%, 95% CI 0.22 to 1.49); number needed to treat to benefit = 100 (95% CI 67 to 451) in the treated population. The effect was not significant in the five trials that used a more detailed diagnostic form for pneumonia. There were no definitions of pneumonia (or other complications) in any trial. No oseltamivir treatment studies reported effects on radiologically confirmed pneumonia" [30]. In this example, the economic damage caused by publication bias was enormous and the pharmaceutical industry's reputation was dealt another blow.

5.4.2
Consequences of Reporting Bias in Experimental Animal Research

The grave consequences of selective reporting in clinical research are clear and a considerable number of horrific stories illustrate the pernicious chain from selectively reporting positive findings, to a biased evidence base, to biased systematic reviews that then impact on clinical (treatment) guidelines finally resulting in flawed decisions in actual healthcare and sometimes massive loss of (quality adjusted) life years [31]. On the other hand, the consequences of selective reporting in animal research are less well understood. To some extent this is caused by the fact that, generally speaking, to many people, the value of animal research for human healthcare is less obvious than that of clinical research [9, 20, 32]. Nevertheless, the general issues are the same: redundancy, misguided follow-up research, and potential harm [33]. The bias that results from over-representation of positive findings (or negative findings when adverse effects are studied) distorts systematic reviews and meta-analyses and leads to overstatement of effectiveness (and understatement of harm) [5c]. Furthermore, the animals used did not contribute to our aggregate knowledge base, and were therefore wasted [33] or played a minor role in some scientist's personal learning curve. Needless repeats of studies are likely, although sometimes at conferences "rumor has it" that certain procedures do not work and at least some investigators know several of the scientific cul-de-sacs and will avoid them. Based on distorted expectations, a decision to perform a first-in-man study may be taken incorrectly or prematurely [34]. And this may lead to useless clinical research that is a waste of resources and a potential risk for the participating patients.

5.5
Causes of Reporting Bias

In an era where many researchers feel pressurized to publish as many papers as possible, publication bias, in the sense that finished work is not even submitted, seems paradoxical. There is some controversy over whether authors are to

be blamed for not submitting or reviewers and editors for blocking publication. There is research on the acceptance decisions of some journals showing that the journals are not to blame [35]. However, from our own experiences, we hypothesize that many scientists anticipate repeated rejections of "negative" results. The survey among animal researchers by ter Riet *et al.* [5b] also seems to support this view. To the question "Who are responsible for non-publication in experimental animal research?," respondents scored a median of 4 on a 5-point scale for the importance of editors, reviewers, and supervisors, whereas the option "lost interest" scored low. In a comprehensive review on the evidence of selective reporting, Song *et al.* take a balanced view and state that "The dissemination profile of a research finding is determined by the interests of research sponsors, investigators, peer-reviewers, and editors. ... publication bias is often due to investigators' failure to write up and submit, although it should be recognized that the investigators' decision to write up an article and then submit it may be affected by pressure from research sponsors, preferences of journal editors, and the requirements of the research award system" [22b].

A useful distinction is that between financial and non-financial conflicts of interest. Conflicts of interest may play a role at the level of sponsors, scientists (including peer reviewers), and editors. Financial conflicts of interest and their role as drivers of reporting bias are easy to understand. Often, the non-financial conflicts of interest will involve pet theories or firmly held methodological beliefs [36].

Here we postulate a few human tendencies that are not always discussed, although they seem relevant in this context. We refer to our common tendency to seek novelty, good stories, and binary classifications as these tendencies may also help to explain the publication pressure–bias paradox. Let us present two of our beliefs. Firstly, people like good stories. Sad tales that only disprove the existence of phenomena do not generally stir our imagination [37], although we may occasionally devour a good story about icons who got it wrong. Until recently, one could still find journals whose instructions for authors stipulated that only findings that were novel or of a certain minimal magnitude would be considered for publication. The ultimate reasons behind this phenomenon are likely to be financial. In the end, even scientific journals are magazines that have to entertain their readers by publishing exciting (new) findings. They have a keen interest in improving their impact factor to keep attracting the "best" papers. After all, the publishers who run these journals are for-profit companies whose shareholders expect revenues produced by subscriptions and, increasingly, by publication fees. We have met a number of animal researchers who explained that they tried to replicate published findings. However, it turned out that publication of replication studies is difficult, since the perception may be that the research is not tackling something novel, is uncreative by only repeating what others did previously and successfully, or that an inability to obtain similar results may be explained by experimental ineptitude. It is hoped that the recent shock caused by a team of industry researchers who were able to replicate only six out 53 published (animal) studies even with help of the original investigators will change attitudes toward replication among cancer scientists [38].

Secondly, medical practitioners are uncomfortable with determinants that follow a continuous distribution. For example, most if not all cardiovascular risks are fairly smooth functions of, for example, blood pressure and serum cholesterol concentrations. In preventive cardiology, we know of no step functions where risks suddenly rise at some threshold value of a risk factor. This does not prohibit most medical practitioners from acting on concepts such as hypotension and hypertension. Thirdly, most people are natural Bayesians. That is, they have a belief; they encounter new evidence, (critically) appraise it, and after assimilating it their updated belief lies somewhere between the old belief and that which the new evidence supports. Thus, depending on the strength of the initial belief and the amount of fresh evidence, gradual shifts in belief seem natural [39]. However, in the planning and the statistical evaluation of scientific studies, most researchers seem to abandon this natural Bayesian inclination. The sample size dogma in essence means that each single experiment by itself should convince everyone irrespective of their initial beliefs [40]. And the evaluation of the evidence, although quantified as a p-value on a continuous scale between 0 and 1, is dichotomized, just like serum cholesterol, into a "Yes, the phenomenon exists" or an "Aw, the study results are negative." Steven Goodman, in an eloquent paper, describes how in the 1930s, Sir Ronald Fisher invented the p-value as quite an informal measure of inference that was to replace its competitors, namely, hypothesis testing [sic!] and Bayesian methods [8a,b,41]. The modern marriage between the p-value and significance testing would have Fisher turning in his grave. Although this issue of the possibility of expressing the evidentiary value of a study into a single number is subtle and complicated, Figure 5.1, based on fictitious data, shows how rigid binary p-value thinking may lead to absurd conclusions about the compatibility of study results. Two studies are pictured that were claimed to be contradictory in the sense that the study by Smith was negative whereas that by Jones was positive. The graph shows that both are in full agreement about the treatment effect (RR = 0.78), but that their precision is different due to different sample sizes of 20 and 2000, respectively. The graph clearly shows the compatibility of these results. However, "concise" binary reporting of the results of these two trials (see last column), omitting a graph or confidence intervals, may easily seduce readers into believing that the results are mutually incompatible.

5.6 Solutions

In this section, we will discuss some methods proposed to counteract selective reporting. This section ends with a proposal for ensuring complete publication.

The idea of submitting to journals manuscripts from which the results section was omitted was first launched in 1970 [42]. Editors and peer reviewers would judge the importance of manuscripts using solely the background, the hypotheses and study objectives, and methods sections. If convinced that the objectives were

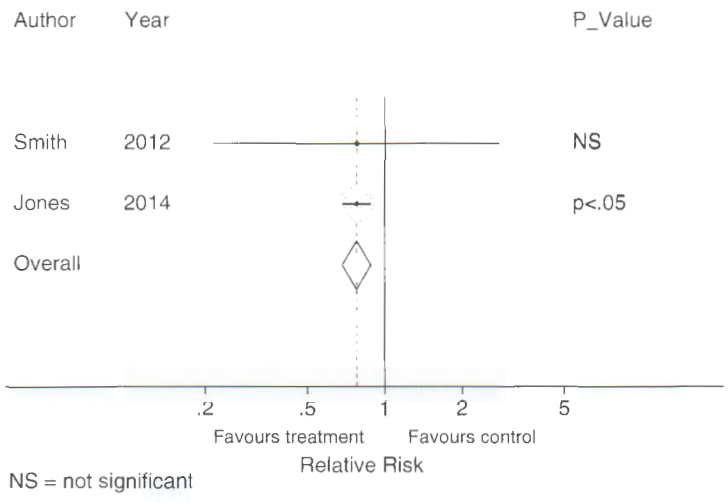


Figure 5.1 Forest plot showing that contradiction of two trial results in terms of statistical significance is fully compatible with exact agreement of their estimated effect size. Here the trials by Smith and Jones both measured a preventive treatment effect of 0.778 relative to control. A huge sample size difference explains the seeming contradiction when a rigid significance testing paradigm is applied.

worth pursuing and the methods appropriate, a document would be signed that was close to a guarantee of publication irrespective of the nature of the results. The next version of the manuscript would then be complete with tables, figures, and other material to describe the results. Such a procedure would ensure that acceptance was not conditional on the nature of the findings. We are not aware of journals that experimented with or adopted this system. Why wouldn't they? The cynical view, of course, is the one we gave earlier: scientific journals are ultimately magazines striving to entertain their readership, with publishers and a commercial market system operating in the background. Let us explore the potential additional administrative burden of the proposal as compared to the current system, where we include the results in the first submission (Table 5.1).

We see that under fairly standard scenarios, the proposed new system is associated with a minor amount of additional administration. Given the deleterious effects on science caused by reporting bias, we strongly recommend the proposed system. Ideally, the decision to adopt a system of publication should be based on cost-effectiveness considerations, explicitly from a societal perspective. We believe that the costs of reporting bias are huge and can easily justify some additional expenditure in the handling of manuscripts.

Other measures that are currently in place to some extent include special journals, journal sections, or repositories for "negative" results, such as the *Journal of Negative Results in Biomedicine* and *The All Results Journals* [43]. In addition,

Table 5.1 Additional administrative steps for editors in a system in which the initial submission contains no results.

Current system (manuscript contains results directly; 10–16 steps)	Proposed system (objective and methods are approved directly; 14–21 steps)
1. Receive manuscript	Receive manuscript without results
2. Decide on straight rejection, if not, then	Decide on straight rejection, if not, then
3. Assign reviewers	Assign reviewers
4. Remind reviewers (mostly automated)	Remind reviewers (mostly automated)
5. Collect reviewers' comments	Collect reviewers' comments
6. Discuss with co-editors, statistician	Discuss with co-editors, statistician
7. Summarize and correspond with authors	Summarize and correspond with authors
8. Receive resubmission	Receive resubmission (now with results)
9. Forward to reviewers	Forward to reviewers
10. Remind reviewers (mostly automated)	Remind reviewers (mostly automated)
11. Collect reviewers' comments	Collect reviewers' comments
12. Discuss with co-editors, statistician	Discuss with co-editors, statistician
13. Summarize and correspond with authors	Summarize and correspond with authors
14. Receive final manuscript	(Repeat steps 6–10 if results section did not match step 1)
15. Posting or printing procedures	Posting or printing procedures
16. Post-publication activities (letters, etc.)	Post-publication activities (letters, etc.)

the *Journal of Cerebral Blood Flow* and *Metabolism and Neurobiology of Aging* feature Negative Results sections with a similar flavor [43b,44]. The *Journal of Cerebral Blood Flow and Metabolism* describes this section as follows: "The Negative Results section of the Journal of Cerebral Blood Flow and Metabolism will provide a platform and raise awareness of a problem with a proven negative impact on scientific progress as well as bench-to-bedside translation. Now researchers must step up to this platform. It is an experiment, but, if successful, it may serve as a role model for other journals and other research fields and thus help to reduce publication bias" [44a].

Since statistical insignificance is probably a main cause of non-publication, its dogmatic use should be discouraged. This may be difficult, however. In our experience, even medical students in their first year already seem indoctrinated with the idea that findings should preferably be statistically significant.

There are a few other plausible candidate solutions to the problem of selective reporting. We mention prospective registration, separate publication of study protocols [45], data sharing, and enhanced carrot and stick approaches by funding bodies [46]. All these are incorporated in the proposal we will sketch below. Figure 5.2 shows our proposal for the organization of scientific (animal) research seeking to eradicate selective reporting. Our proposal is an attempt to integrate different ideas that were launched previously into one coherent system [33, 47].

This system has four key components: (i) early end-user input, (ii) systematic reviews, (iii) the power of funding bodies, and (iv) a central position of IACUCs.

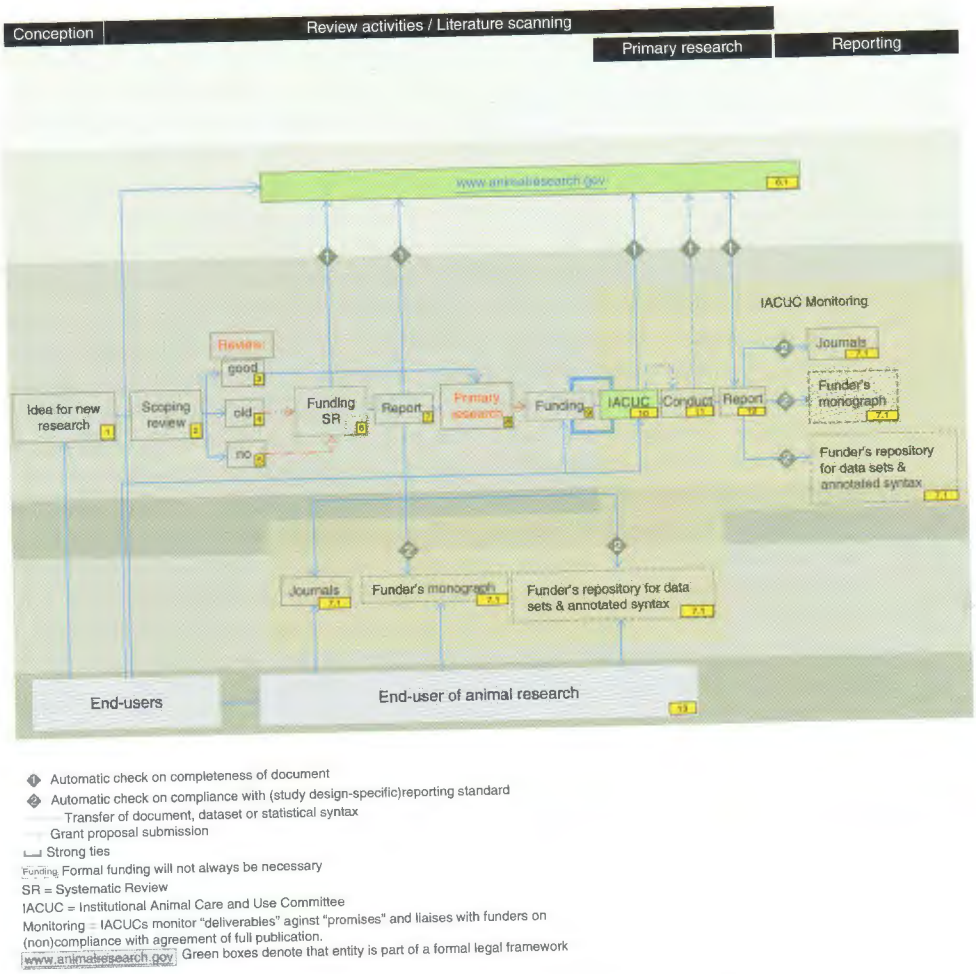


Figure 5.2 Flow chart depicting a system designed to reduce research waste and reporting bias. Note that animalresearch.gov is a non-existing website used to illustrate preregistration of all animal studies. Its name was inspired by clinicaltrials.gov, the US-based website for preregistration of (randomized) clinical trials.

This scheme was inspired by the procedures used by the UK HTA program, which succeeds in having 98% of funded research published [46a,48]. Let us take you through Figure 5.2 and clarify some of its components. Horizontally, the black bar on top shows four activities: the first phase is the conception of ideas and the second systematic review of the relevant literature, a process that also continues through the third phase, the conduct of the primary research project. The final and fourth phase is the reporting of study results. The process starts on the left-hand side, with the box indicated by the yellow label and number 1. A new research idea emerges, potentially developed with help of end-users, which in the case of

animal research may be clinicians or translational specialists or even patients. To comply with strict funders' requirements, a scoping search of the literature is conducted to study the available evidence on the hypothesis at issue, to prevent needless replication, and to learn from predecessors' successes and mistakes. If at least one good and recent systematic review exists, the simplest route goes via box 3 to box 8 and a research proposal is written, submitted for funding (box 9, notice the end-users' influence on the funding decisions (blue arrow)), and if obtained, IACUC approval is sought (box 10). If the IACUC approves the protocol it sends the protocol to animalresearch.gov, where its contents are automatically checked for completeness against an authoritative guideline for protocols using natural language processing [49]. The study is conducted (box 11, which allows for IACUC-approved amendments of the protocol at animalresearch.gov during conduct) and the IACUC monitors progress and reporting. Reporting, which is much more comprehensive than in the current system, has three components: (i) the report is submitted to animalresearch.gov (box 6.1) and checked for completeness against the ARRIVE guidelines (see Section 3.1) for complete reporting using natural language programming software; (ii) the raw data, the cleaned data, and carefully annotated statistical cleaning and analysis syntaxes are submitted to the funder and put in an open access repository with informative meta-data to allow checks and re-use of data for secondary analyses including individual animal meta-analyses (box 7.1); and (iii) irrespective of submissions to animalresearch.gov, researchers will still be allowed to publish (an abbreviated) version of their work in a scientific journal under an open access system of publication. Entries in animalresearch.gov will contain hyperlinks to all open access publications about the study. End-users may access reports via the outlets described above (box 7.1). So the sequence of boxes 10–6.1–11[6.1]–7.1–13 is fixed just like the sequence of boxes 1 and 2. Some variation enters the protocol if a good and recent systematic review is not found and has to be written or updated with its own protocols and reports.

There are some key differences from the current system. The extent of reporting is not the only difference. Note how the funders, using their carrot and stick approach, such as withholding 10–20% of funds unless the requirements for reporting are met, are supposed to cooperate closely with the IACUCs as illustrated by the thick blue lines. The funders are the natural guardians of the data and the syntaxes, since they will eventually receive future submissions for secondary research on existing data. The IACUCs will have an active monitoring role, chasing up researchers who do not deliver within a reasonable time frame. Not delivering without good reasons may lower the chances of subsequent funding because IACUCs and funders cooperate closely and share information. Finally, the academic reward system should also reflect the system depicted in Figure 5.2. Instead of focusing solely on publications and citations, researchers must also be rewarded for the secondary use of their data sets and/or syntaxes and deployment of their work by the end-users. Ioannidis and Khoury [50] have recently proposed the PQRST (P = productive, Q = high-quality, R = reproducible, S = shareable, and T = translatable) reward system to replace the current system, which is

gauged too much toward counting publications and citations. The reward system for scientists is a key element in making our proposal work, since the current high level of competitiveness in science works against pre-registration of research ideas. A key element in the system described here is the fully automated checks of study protocols and study report manuscripts submitted to animalresearch.gov. The sheer numbers of submissions require this. The interplay between the submitted texts and the natural language-processing software based on the guidelines for study protocols and reporting needs to be robust for the system to work appropriately. On the other hand, given the large sums of money wasted in the current system [33,47b,51], the necessary investments in the more sophisticated approach we describe here are likely to make sense. For research that has no explicit funding or projects that fail completely, perhaps due to technical or personnel problems, we hope that a reward system along the lines of the PQRST system will motivate scientists to pre-register projects and submit brief summaries of the reasons why a project may have failed completely. It will by now be abundantly clear that by "failed," we do not refer to the nature of the study results in any way.

In conclusion, selective reporting in clinical research has immense costs in terms of money and health. The economic and scientific impact of selective reporting in animal research is an under-researched topic but is likely to be considerable as well. Increasingly, incomplete reporting of research outcomes is seen as a form of research misconduct, and more attention is being paid to the wasteful aspects of our current system of doing science. The initiatives originating in the clinical field have great potential to improve the state of affairs in animal research as well.

References

1. (a) Hedges, L.V. (1984) Research synthesis: the state of the art. *Int. J. Aging Hum. Dev.*, **19**(2), 85–93. (b) Egger, M., Smith, G.D., and Sterne, J.A., (2001) Uses and abuses of meta-analysis. *Clin. Med. (London)*, **1**(6), 478–484. (c) Chalmers, I., Hetherington, J., Newdick, M., et al. (1986) The Oxford database of perinatal trials: developing a register of published reports of controlled trials. *Control. Clin. Trials*, **7**(4), 306–324. (d) Lau, J., Antman, E.M., Jimenez-Silva, J., et al. (1992) Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N. Engl. J. Med.*, **327**(4), 248–254.
2. Rayleigh, Lord (1885) Address by the Rt. Hon. Lord Rayleigh, http://www.jameslindlibrary.org/illustrating/records/address-by-the-rt-hon-lord-rayleigh-in-report-of-the-fifty-f/title_pages (accessed 24 November 2014).
3. Shojania, K.G., Sampson, M., Ansari, M.T., et al. (2007) How quickly do systematic reviews go out of date? A survival analysis. *Ann. Intern. Med.*, **147**(4), 224–233.
4. (a) Stern, J.M. and Simes, R.J. (1997) Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *Br. Med. J.*, **315**(7109), 640–645. (b) Dickersin, K. (1990) The existence of publication bias and risk factors for its occurrence. *J. Am. Med. Assoc.* , **263**(10), 1385–1389.
5. (a) 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**(4),

225–230. (b) ter Riet, G., Korevaar, D.A., Leenaars, M., et al. (2012) Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. *PLoS One*, **7**(9), e43404. (c) Sena, E. S., van der Worp, H.B., Bath, P.M., et al. (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.*, **8**(3), e1000344.

6. Camarades Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies, http://www.dcn.ed.ac.uk/camarades/default.htm (accessed 25 November 2014).
7. (a) The Campbell Collaboration (2014) http://www.campbellcollaboration.org/ (accessed 25 November 2014). (b) The Cochrane Collaboration—Trusted Evidence. Informed Decisions. Better Health (2014) http://www.cochrane.org/ (accessed 25 November 2014).
8. (a) Goodman, S.N. (1999) Toward evidence-based medical statistics. 2: the Bayes factor. *Ann. Intern. Med.*, **130**(12), 1005–1013. (b) Goodman, S.N. (1999) Toward evidence-based medical statistics. 1: the P value fallacy. *Ann. Intern. Med.*, **130**(12), 995–1004. (c) Deming, W.E. (1976) On probability as a basis for action. *Methods Inf. Med. Suppl.*, **9**, 3–15. (d) Bakker, M. and Wicherts, J. M. (2011) The (mis)reporting of statistical results in psychology journals. *Behav. Res. Methods*, **43**(3), 666–678. (e) van Assen, M.A., van Aert, R.C., Nuijten, M.B., et al. (2014) Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS One*, **9** (1), e84896. (f) Tsilidis, K.K., Panagiotou, O.A., Sena, E.S., et al. (2013) Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol.*, **11**(7), e1001609. (g) Boutron, I., Altman, D.G., Hopewell, S., et al. (2014) Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J. Clin. Oncol.* **32**(36), 4120–4126 (h) Boutron, I., Dutton, S., Ravaud, P., and Altman, D.G., (2010) Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *J. Am. Med. Assoc.*, **303**(20), 2058–2064. (i) Forsyth, S.R., Odierna, D.H., Krauth, D., and Bero, L.A. (2014) Conflicts of interest and critiques of the use of systematic reviews in policymaking: an analysis of opinion articles. *Syst. Rev.*, **3**(1), 122. (j) Norris, S.L., Burda, B.U., Holmer, H.K., et al. (2012) Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. *J. Clin. Epidemiol.*, **65**(7), 725–733.
9. Pound, P., Ebrahim, S., Sandercock, P., et al. Where is the evidence that animal research benefits humans? *Br. Med. J.* 2004, **328**(7438), 514–517.
10. Copas, J. and Shi, J.Q. (2000) Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, **1**(3), 247–262.
11. (a) Copas, J., Dwan, K., Kirkham, J., and Williamson, P. (2014) A model-based correction for outcome reporting bias in meta-analysis. *Biostatistics*, **15**(2), 370–383. (b) Copas, J.B. and Shi, J.Q. (2001) A sensitivity analysis for publication bias in systematic reviews. *Stat. Methods Med. Res.*, **10**(4), 251–265. (c) Langan, D., Higgins, J.P., Gregory, W., and Sutton, A.J. (2012) Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. *J. Clin. Epidemiol.*, **65**(5), 511–519. (d) Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.*, **315**(7109), 629–634. (e) Duval, S. and Tweedie, R. (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **56**(2), 455–463. (f) Terrin, N., Schmid, C.H., Lau, J., and Olkin, I. (2003) Adjusting for publication bias in the presence of heterogeneity. *Stat. Med.*, **22**(13), 2113–2126.
12. Dal-Re, R., Ioannidis, J.P., Bracken, M.B., et al. (2014) Making prospective registration of observational research a reality. *Sci. Transl. Med.*, **6**(224), 224cm1.

13. (a) ClinicalTrials.gov (2014) https://clinicaltrials.gov/ (accessed 25 November 2014). (b) International Clinical Trials Registry Platform Search Portal (2014) http://apps.who.int/trialsearch/ (accessed 25 November 2014).
14. De Angelis, C., Drazen, J.M., Frizelle, F.A., et al. (2004) Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N. Engl. J. Med.*, **351**(12), 1250–1251.
15. Mathieu, S., Boutron, I., Moher, D., et al. (2009) Comparison of registered and published primary outcomes in randomized controlled trials. *J. Am. Med. Assoc.*, **302**(9), 977–984.
16. Viergever, R.F., Karam, G., Reis, A., and Ghersi, D. (2014) The quality of registration of clinical trials: still a problem. *PLoS One*, **9**(1), e84727.
17. Hudson, K.L. and Collins, F.S. (2014) Sharing and reporting the results of clinical trials. *J. Am. Med. Assoc.*, **313**(4):355–356.
18. All Trials Registered | All Results Reported (2014) http://www.alltrials.net/ (accessed 25 November 2014).
19. Ross, J.S., Tse, T., Zarin, D.A., et al. (2012) Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *Br. Med. J.*, **344**, d7292.
20. Hartung, T. (2013) Look back in anger - what clinical studies tell us about preclinical work. *Altex*, **30**(3), 275–291.
21. Edwards, A.S. (2013) Research ethics committees have the power to enforce publication of drug trial results. *Br. Med. J.*, **346**, f1201.
22. (a) Song, F., Eastwood, A.J., Gilbody, S., et al (2000) Publication and related biases. *Health Technol. Assess.*, **4**(10), 1–115. (b) Song, F., Parekh, S., Hooper, L., et al. (2010) Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess.*, **14**(8), iii, ix-xi, 1–193.
23. Goldacre, B. (2012) *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*. London: Fourth Estate.
24. (a) Macleod, M.R., O'Collins, T., Howells, D.W., and Donnan, G.A. (2004) Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke*, **35**(5), 1203–1208. (b) Sterling, T.D. (1959) Publication decisions and their possible effects on inferences drawn from tests of significance-or vice versa. *J. Am. Stat. Assoc.*, **54**, 30–34
25. Song, F., Loke, Y., and Hooper, L. (2014) Why are medical and health-related studies not being published? A systematic review of reasons given by investigators. *PLoS One*, **9**(10), e110418.
26. Turner, E.H., Matthews, A.M., Linardatos, E., et al. (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.*, **358**(3), 252–260.
27. Song, F., Parekh-Bhurke, S., Hooper, L., et al. (2009) Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med. Res. Methodol.*, **9**, 79.
28. Cowley, A.J., Skene, A., Stainer, K., and Hampton, J.R. (1993) The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int. J. Cardiol.*, **40**(2), 161–166.
29. Jefferson, T. and Doshi, P. (2014) Multisystem failure: the story of antinfluenza drugs. *Recenti Prog. Med.*, **105**(5), 187–190.
30. Jefferson, T., Jones, M.A., Doshi, P., *et al.* (2014) Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. *Cochrane Database Syst. Rev.*, **4**, Cd008965.
31. (a) Every-Palmer, S. and Howick, J. (2014) How evidence-based medicine is failing due to biased trials and selective publication. *J. Eval. Clin. Pract.* **20**(6), 908–914. (b) Landewe, R.B. (2014) Editorial: how publication bias may harm treatment guidelines. *Arthritis Rheumatol.*, **66**(10), 2661–2663.
32. (a) Pound, P. (2001) Scientific debate on animal model in research is needed. *Br. Med. J.*, **323**(7323), 1252. (b) Pound, P. and Bracken, M.B. (2014) Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *Br. Med. J.*, **348**, g3387. (c) Pound, P. and Ebrahim, S. (2002) Supportive evidence is lacking for report on animal studies. *Br. Med. J.*, **325**(7371), 1038.

(d) Landis, S.C., Amara, S.G., Asadullah, K., et al. (2012) A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*, **490**(7419), 187–191. (e) Perel, P., Roberts, I., Sena, E., et al. (2007) Comparison of treatment effects between animal experiments and clinical trials: systematic review. *Br. Med. J.*, **334**(7586), 197. (f) Roberts, I., Kwan, I., Evans, P., and Haig, S. (2002) Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. *Br. Med. J.*, **324**(7335), 474–476. (g) Sandercock, P. and Roberts, I. (2002) Systematic reviews of animal experiments. *Lancet*, **360**(9333), 586. (h) Greek, R. and Menache, A. (2013) Systematic reviews of animal models: methodology versus epistemology. *Int. J. Med. Sci.*, **10**(3), 206–221.

33. Chan, A.W., Song, F., Vickers, A., et al. (2014) Increasing value and reducing waste: addressing inaccessible research. *Lancet*, **383**(9913), 257–266.
34. Kenter, M.J. and Cohen, A.F. (2006) Establishing risk of human experimentation with drugs: lessons from TGN1412. *Lancet*, **368**(9544), 1387–1391.
35. van Lent, M., Overbeke, J. and Out, H.J. (2014) Role of editorial and peer review processes in publication bias: analysis of drug trials submitted to eight medical journals. *PLoS One*, **9**(8), e104846.
36. Penrose, L.S. (1948) The problem of anticipation in pedigrees of dystrophia myotonica. *Ann. Eugen.*, **14**(2), 125–132.
37. Lemon, R. and Dunnett, S.B. (2005) Surveying the literature from animal experiments. *Br. Med. J.*, **330**(7498), 977–978.
38. Begley, C.G. and Ellis, L.M. (2012) Drug development: raise standards for preclinical cancer research. *Nature*, **483**(7391), 531–533.
39. (a) Ioannidis, J.P. (2005) Why most published research findings are false. *PLoS Med.*, **2**(8), e124. (b) Ioannidis, J.P. (2007) Why most published research findings are false: author's reply to Goodman and Greenland. *PLoS Med.*, **4**(6), e215. (c) Goodman, S. and Greenland, S. (2007) Why most published research findings are false: problems in the analysis. *PLoS Med.*, **4**(4), e168.
40. (a) Sutton, A.J., Cooper, N.J., and Jones, D.R. (2009) Evidence synthesis as the key to more coherent and efficient research. *BMC Med. Res. Methodol.*, **9**, 29. (b) Bacchetti, P. (2010) Current sample size conventions: flaws, harms, and alternatives. *BMC Med.*, **8**, 17. (c) Bacchetti, P., Deeks, S.G., and McCune, J.M. (2011) Breaking free of sample size dogma to perform innovative translational research. *Sci. Transl. Med.*, **3**(87), 87ps24. (d) Bacchetti, P., McCulloch, C.E., Segal, M.R. (2008) Simple, defensible sample sizes based on cost efficiency. *Biometrics*, **64**(2), 577–585; discussion 586–594.
41. Goodman, S.N. (2005) Introduction to Bayesian methods I: measuring the strength of evidence. *Clin. Trials*, **2**(4), 282–290; discussion 301–304, 364–378.
42. (a) Walster, G.W. and Cleary, A.A. (1970) Proposal for a new editorial policy in social sciences. *Am. Stat.*, **24**, 16–19. (b) Greenland, S. (2007) Commentary: on 'quality in epidemiological research: should we be submitting papers before we have the results and submitting more hypothesis generating research?'. *Int. J. Epidemiol.*, **36**(5), 944–945. (c) Lawlor, D.A. (2007) Quality in epidemiological research: should we be submitting papers before we have the results and submitting more hypothesis-generating research? *Int. J. Epidemiol.*, **36**(5), 940–943.
43. (a) BioMed Central (2014) Journal of Negative Results in BioMedicine, http://www.jnrbm.com/ (accessed 26 November 2014). (b) The All Results Journals (2014) http://www.arjournals.com/ojs/ (accessed 26 November 2014).
44. (a) Nature (2014) Journal of Cerebral Blood Flow and Metabolism, http://www.nature.com/jcbfm/journal/v30/n7/full/jcbfm201051a.html (accessed 26 November 2014). (b) Neurobiology of Aging (2014) http://www.neurobiologyofaging.org/content/authorinfo (accessed 26 November 2014).
45. (a) Booth, A. (2013) PROSPERO's progress and activities 2012/13. *Syst.*

Rev., **2**, 111. (b) Booth, A., Clarke, M., Dooley, G., et al. (2012) The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst. Rev.*, **1**, 2. (c) Booth, A., Clarke, M., Dooley, G., et al. (2013) PROSPERO at one year: an evaluation of its utility. *Syst. Rev.*, **2**, 4. (d) Siebeling, L., ter Riet, G., van der Wal, W.M., et al. (2009) ICE COLD ERIC—International collaborative effort on chronic obstructive lung disease: exacerbation risk index cohorts—study protocol for an international COPD cohort study. *BMC Pulm. Med.*, **9**, 15.

46. (a) Turner, S., Wright, D., Maeso, R., et al. (2013) Publication rate for funded studies from a major UK health research funder: a cohort study. *BMJ Open*, **3**(5), e002521. (b) Chalmers, I., Glasziou, P., and Godlee, F. (2013) All trials must be registered and the results published. *Br. Med. J.*, **346**, f105.
47. (a) Macleod, M.R., Michie, S., Roberts, I., et al. (2014) Biomedical research: increasing value, reducing waste. *Lancet*, **383**(9912), 101–104. (b) Glasziou, P., Altman, D.G., Bossuyt, P., et al. (2014) Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*, **383**(9913), 267–276.
48. Chinnery, F., Young, A., Goodman, J., et al. (2013) Time to publication for NIHR HTA programme-funded research: a cohort study. *BMJ Open*, **3**(11), e004121.
49. (a) Chan, A.W., Tetzlaff, J.M., Gotzsche, P.C., et al. (2013) SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *Br. Med. J.*, **346**, e7586. (b) Mishra, R., Del Fiol, G., Kilicoglu, H., et al. (2013) Automatically extracting clinically useful sentences from UpToDate to support clinicians' information needs. *AMIA Annu. Symp. Proc.*, **2013**, 987–992. (c) Moore, C.R., Farrag, A., and Ashkin, E. (2014) Using natural language processing to extract abnormal results from cancer screening reports. *J. Patient Saf.*
50. Ioannidis, J.P. and Khoury, M.J. (2014) Assessing value in biomedical research: the PQRST of appraisal and reward. *J. Am. Med. Assoc.*, **312**(5), 483–484.
51. Ioannidis, J.P., Greenland, S., Hlatky, M.A., et al. (2014) Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*, **383** (9912), 166–175.