

Review Article

Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments

D A Korevaar¹, L Hooft² and G ter Riet¹

¹Department of General Practice, Academic Medical Center (AMC), University of Amsterdam (UvA), PO Box 22700, 1100 DD Amsterdam, The Netherlands; ²The Dutch Cochrane Centre, Academic Medical Center (AMC), Amsterdam, The Netherlands
Corresponding author: D A Korevaar. Email: d.a.korevaar@amc.uva.nl

Abstract

In 2006, Peters *et al.* identified 86 systematic reviews (SRs) of laboratory animal experiments (LAEs). They found 46 LAE meta-analyses (MAs), often of poor quality. Six of these 46 MAs tried to assess publication bias. Publication bias is the phenomenon of an experiment's results determining its likelihood of publication, often over-representing positive findings. As such, publication bias is the Achilles heel of any SR. Since researchers increasingly become aware of the fact that SRs directly support the 'three Rs', we expect the number of SRs of LAEs will sharply increase. Therefore, it is useful to see how publication bias is dealt with. Our objective was to identify all SRs and MAs of LAEs where the purpose was to inform human health published between July 2005 and 2010 with special attention to MAs' quality features and publication bias. We systematically searched Medline, Embase, Toxline and ScienceDirect from July 2005 to 2010, updating Peters' review. LAEs not directly informing human health or concerning fundamental biology were excluded. We found 2780 references of which 163 met the inclusion criteria: 158 SRs, of which 30 performed an MA, and five MAs without an SR. The number of SRs roughly doubled every three years since 1997. The number of MAs roughly doubled every five years since 1999. Compared with before July 2005, more MAs were preceded by SR and reported on (quality) features of included studies and heterogeneity. A statistically significant proportion of MAs considered publication bias (26/35) and tried to formally assess it (21/35).

Keywords: Systematic review, meta-analysis, publication bias, three Rs

Laboratory Animals 2011; 45: 225–230. DOI: 10.1258/la.2011.010121

Many laboratory animal experiments (LAEs) are performed to inform human health. They may play an important role in the identification and development of drugs, medical devices and surgical procedures, in making risk assessments for safe human exposure and increasing biological knowledge.¹ It would seem rational to critically review the relevant LAEs before new LAEs and, in particular, clinical trials in humans are performed. Systematic reviews (SRs) and, where appropriate, meta-analyses (MAs) are suitable tools to summarize the current evidence on a given subject and therefore directly support the 'three Rs' (replacement, reduction and refinement), for example, by preventing unnecessary duplication of animal studies. SRs and MAs play an important role in physics, the social sciences and medicine. However, a few years ago, several studies have shown that, especially compared with other research fields, relatively few SRs and MAs had been performed in LAEs. In 2004, about one in every 1000 Medline records about human research were said to be tagged as an MA, compared with one in 10,000 records about animal research.² In 2006, Mignini and Khan³ tried to collect all

LAE SRs and found 30. By contrast, around the same time, Peters *et al.*,⁴ found 86 SRs. The discrepancy is explained by the fact that Peters *et al.* searched more databases, used a broader definition of LAEs and also included studies reviewing human evidence in addition to LAEs.

Unfortunately, SRs and MAs are sensitive to publication bias. Publication bias is the phenomenon of an experiment's results determining its likelihood of publication, often leading to severe under-representation of negative findings.⁵ Since unpublished study results are hard to find, publication bias is the Achilles heel of any SR and may lead to biased syntheses of the evidence in any given field. Publication bias has often been demonstrated in clinical sciences,⁶ but little is known about its extent in LAE.^{3,7,8} However, many authors speculate that publication bias may be common in LAE^{3,9} and the number of published negative results of LAEs is surprisingly low.¹⁰ Interestingly, authors of SRs and MAs of LAEs often do not (explicitly) consider the fact that their results might be biased by publication bias. Mignini and Khan³ concluded that only five out of 30 SRs assessed the risk of publication

bias. Peters *et al.* found that only 17 out of 46 MAs mentioned and, to some extent, considered publication bias. Only six of these made an effort to assess it by means of visual or statistical methods.⁴ We expect that over the next decade the number of SRs and MAs of LAEs will sharply increase. Therefore, it is useful to see how publication bias may be dealt with.

The objective of this study is to identify all SRs and MAs of LAEs published between July 2005 and 2010, thereby updating the earlier work by Peters *et al.* To facilitate a comparison with the period before July 2005, we also assessed how the MAs scored on several quality features and how they dealt with potential publication bias.

Materials and methods

We systematically searched Medline, Embase, Toxline and ScienceDirect, all from July 2005 to 2010. There were no language restrictions. For details on the search, see Supplementary file 1 [<http://la.rsmjournals.com/cgi/content/full/la.2011.010121/DC1>]. In addition, reference lists of relevant articles were manually checked to identify studies missed by our database searches. We did not search any grey literature (i.e. literature that has not been formally published) since Peters *et al.* found no relevant reports out of 960 hits before July 2005. One author (DK) assessed all search results. When a title and/or abstract could not with certainty be rejected, the full text of the paper was obtained and assessed for inclusion. When there was any doubt of whether an article fulfilled the inclusion criteria, the case was discussed with a co-author (LH/GtR).

To facilitate a formal comparison, we used similar criteria to identify relevant SRs and MAs of LAEs as Peters *et al.* An article was tagged as an SR if it gave details on the source(s) of evidence and some information on at least one of the following: search terms used, any limitation placed on the search, or inclusion and exclusion criteria. To qualify as an MA, a paper had to report on some form of quantitative synthesis of results of more than one experiment.

SRs and MAs were included if they involved *in vivo* LAEs, where the purpose of reviewing animal evidence was to inform human health and could be allocated to one of the following groups:

- (1) A medical intervention was applied, subdivided in measurement of:
 - (a) Efficacy of a medical intervention;
 - (b) Side-effects of a medical intervention, or its toxicity;
 - (c) The mechanisms of action of a medical intervention.
- (2) Risk factor research (epidemiological associations or mechanisms of action of disease).
- (3) Effects of an exposure to a chemical substance were measured.
- (4) An overview of animal models for disease was given.
- (5) The accuracy of a test to diagnose disease was measured.

Articles including other experiments, such as human studies, in addition to the LAEs were also included. We excluded genome-wide association studies and LAEs whose main purpose was to learn more about fundamental

biology, physical functioning or behaviour and was not directly to inform human health.

One author (DK) extracted relevant data of the included articles. Of the SRs, authors, publication year, country, journal, objective, search strategy, assessment of study quality and any comments on publication bias were extracted. From MAs, additionally, data on species, number of experiments included and details on the methods used (effect estimates reported, assessment of heterogeneity and synthesis method used) were extracted. These data were compared with the results before July 2005. Ninety-five percent confidence intervals (CIs) for differences between proportions were calculated using exact binomial methods using STATA version 10.1.

Results

Figure 1 shows the flow of references from electronic searches to final inclusion in our review. The searches identified 2780 references of which 163 fulfilled the inclusion criteria. We excluded 35 studies in whose title, abstract or methods it was claimed that an SR or a systematic search had been performed. However, these papers contained no details on the methods used. Seven studies pertained to SR or MA of LAEs but were excluded because their main purpose was not to inform human health.¹¹⁻¹⁷

So in total we found 158 SRs of LAEs published between July 2005 and 2010, of which 30 also included an MA. We found five MAs of LAEs which did not fulfil the criteria of an SR. Figure 2 shows to what type of animal research,

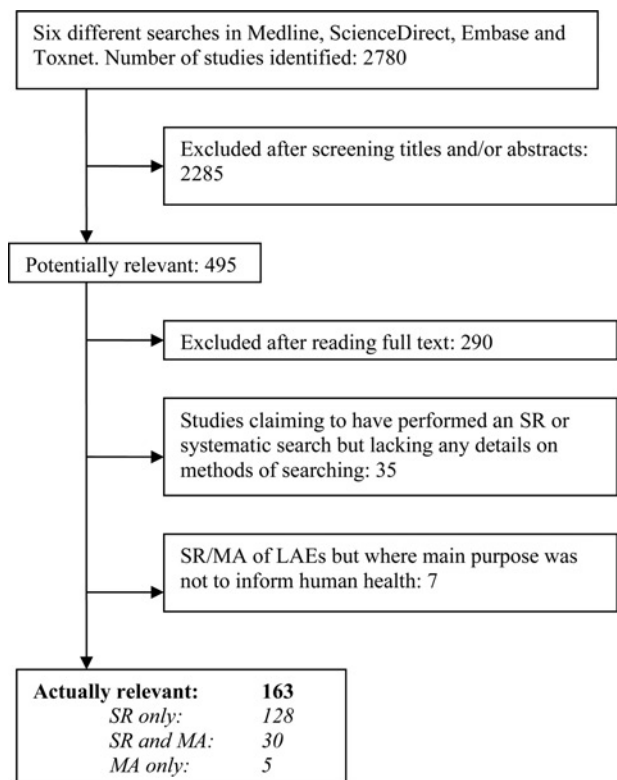


Figure 1 Flowchart showing how the papers entered the review SR: systematic review; MA: meta-analysis; LAE: laboratory animal experiment

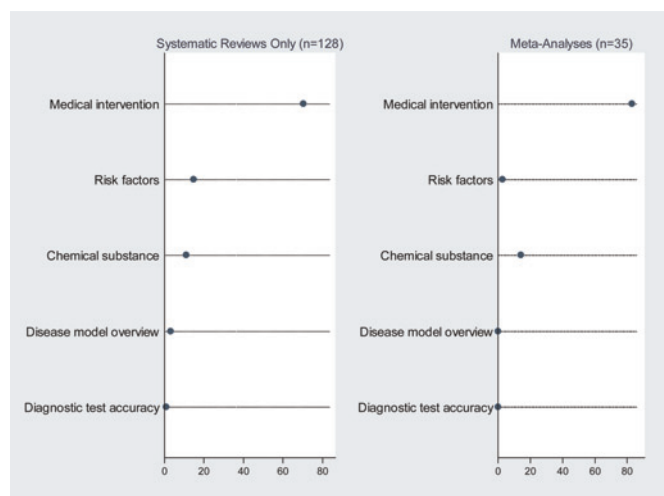


Figure 2 Type of research that was reviewed. The first dot plot shows the type of animal research which the included systematic reviews (without meta-analysis) reviewed in percentages. The second dot plot shows the type of animal research which the included meta-analysis (with or without a systematic review) reviewed in percentages

as defined in the Materials and methods section, the included papers were allocated. Supplementary file 2 [<http://la.rsmjournals.com/cgi/content/full/la.2011.010121/DC2>] shows a reference list of all the included articles.

Systematic reviews

Table 1 compares the SRs performed after July 2005 with those found by Peters *et al.* Individual summaries of all the included SRs are in Supplementary files 3 [<http://la.rsmjournals.com/cgi/content/full/la.2011.010121/DC3>] and 4 [<http://la.rsmjournals.com/cgi/content/full/la.2011.010121/DC4>].

The number of published SRs of LAEs has been growing over the past 15 years. There are now 244 SRs and the number about doubled every three years as of 1997 (Figure 3).

As before July 2005, the majority of the SRs evaluate a medical intervention. In the subgroup of SRs that performed an MA this percentage was even higher. The percentage of SRs incorporating human evidence was comparable before and after July 2005, as is the case with SRs searching more than one source of evidence. Only 15 SRs not performing an MA assessed study quality of included articles and only 10 commented on the possibility that findings might be distorted by publication bias. How SRs performed before July 2005 scored on these items is unknown.

Meta-analyses

Table 2 compares MAs published before July 2005 with those published later. Individual summaries of all the included MAs are in Supplementary file 4.

There are now 81 published MAs of LAEs. As of 1999, the number of MAs of LAEs roughly doubled every five years (Figure 4). MAs performed after July 2005 more often evaluate a medical intervention (group A) than those before. In addition, a larger percentage is preceded by an SR.

Some general features, such as the percentage of studies incorporating human data, the minimum and maximum number of experiments combined and the number of species included, have not changed. All the MAs performed after July 2005 reported the number of studies being combined and the number of MAs providing exact details on animal species has grown. Furthermore, the percentage of MAs assessing study quality and between-study heterogeneity has increased.

Publication bias

The proportion of MAs commenting on the possibility that their results might be affected by publication bias increased from 37% before to 74% after July 2005. Four MAs performed after July 2005 mentioned that their results might be biased by publication bias but made no effort to assess it more formally and one mentioned that the number of publications was too small to test for publication bias statistically. Nine MAs did not mention publication bias. Four of the latter group performed an MA not preceded by an SR.

The proportion of MAs trying to assess publication bias increased from 13% before to 60% after July 2005. The ways in which publication bias was addressed were Egger's test for funnel plot asymmetry, the Fail-Safe Number (indicating the number of negative studies necessary to render the pooled estimate non-significant¹⁸), the 'trim and fill' method (imputes missing studies based on a statistical model¹⁸) and comparison of abstracts with full papers (Table 3).

Discussion

Using SR methods, we updated previous work by Peters *et al.* and found that the number of published SRs of LAEs increased from 86 to 244 between mid-2005 and 2010. The number of LAE SRs roughly doubled every

Table 1 Features of systematic reviews of laboratory animal experiments published before and after July 2005

Systematic review characteristic	Before July 2005 n (%)	July 2005–2010 n (%)	Difference % (95% confidence interval)
Total number	86 (100)	158 (100)	
Evaluating a medical intervention (group A)	71 (83)	115 (73)	−9.8 (−20.4 to 0.8)
Reviewing research on risk factors for disease (group B)	–	22 (14)	
Reviewing research on effects of exposure to chemicals (group C)	–	16 (10)	
Also performing a meta-analysis	29 (34)	30 (19)	−14.7 (−26.4 to −3.0)
of which evaluate a medical intervention (group A)	23 (79)	27 (90)	10.7 (−7.5 to 28.9)
Assessing human evidence in addition to laboratory animal experiments	52 (60)	104 (66)	5.4 (−7.3 to 18.1)
Searching more than one source of evidence	73 (85)	125 (79)	−5.8 (−15.6 to 4.1)

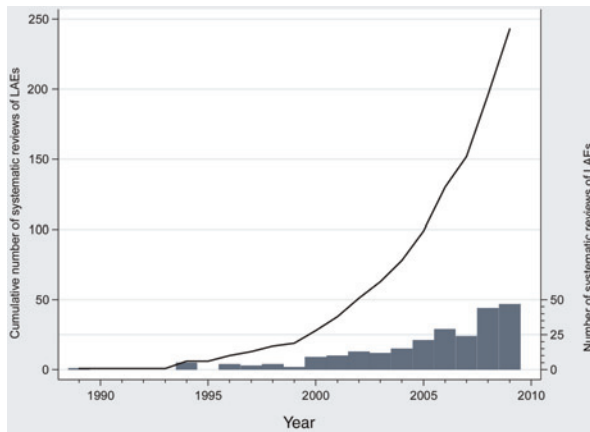


Figure 3 Systematic reviews of laboratory animal experiments. Bar graph shows the number of published systematic reviews (SRs) of laboratory animal experiments (LAEs) for the years 1990–2010. Line graph shows the cumulative number of such reviews. As of 1997, the number of SRs of LAEs about doubles every three years

three years as of 1997. Similarly, the number of published LAE MAs increased from 35 to 66 roughly doubling every five years since 1999. Compared with the number of SRs and MAs of clinical trials and observational studies in humans (over 23,000 through Medline alone), the absolute numbers in LAEs are still modest, but the growth rate is impressive.

The quality of MAs seems to be increasing. A larger percentage is being preceded by an SR (statistically significant improvement by 23% [95% CI 4.5–40.8]), thus ensuring proper searches and data extraction methods, reporting of the number of studies combined (statistically significant improvement by 11% [95% CI 1.9–19.9]) and providing details on animal species included (improvement by 11% not statistically significant [95% CI –14.1–25.7]). Furthermore, study quality of included studies is more often assessed (statistically significant improvement by 27% [95% CI 5.6–47.8]), as in the case for between-study heterogeneity (improvement by 6% not statistically significant [95% CI –15.1–27.8]). Unfortunately, in SRs not performing an MA, study quality and publication bias are seldom considered or assessed.

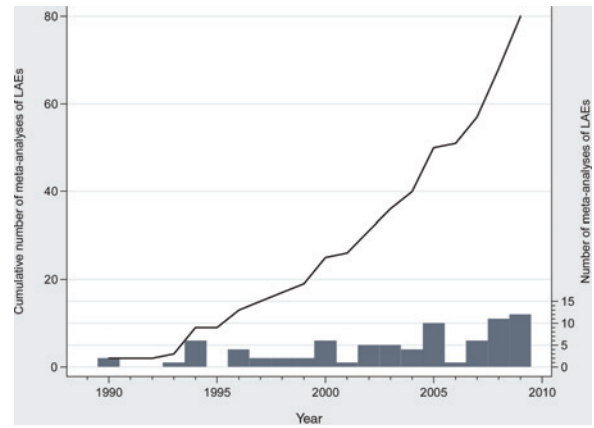


Figure 4 Meta-analyses of laboratory animal experiments. Bar graph shows the number of published meta-analyses (MAs) of laboratory animal experiments (LAEs) for the years 1990–2009. Line graph shows the cumulative number of such meta-analyses. As of 1999, the number of MAs of LAEs about doubles every five years

Among meta-analysts of LAEs the awareness of publication bias appears to have grown over the last five years. Seventy-four percent (26/35) mentioned it, compared with 37% (17/46) before July 2005. This is a statistically significant improvement by 37% (95% CI 17.2–57.4). Furthermore, 60% (21/35) formally addressed publication bias, compared with 13% (6/46) before July 2005. This is a statistically significant improvement by 47% (95% CI 28.0–65.9). In the MAs assessing publication bias, evidence was found in 62% (13/21) (up from 50% [3/6]). However, the value of methods to assess publication bias in MAs should not be overestimated. Funnel plot asymmetry cannot be translated into evidence for publication bias straightforwardly because other phenomena may cause asymmetry, as is the same for the ‘trim and fill’ method.¹⁸ In addition, the power of Egger’s test for funnel plot asymmetry is low unless there is severe bias.¹⁸ Finally, different versions of the Fail-Safe Number produce different results and no statistical criterion exists for its interpretation.¹⁸ Therefore, publication bias may be underestimated as well as overestimated with these methods. However, we think that its existence is plausible.

Table 2 Features of meta-analyses of laboratory animal experiments published before and after July 2005

Meta-analysis characteristic	Before July 2005 n (%)	July 2005–2010 n (%)	Difference % (95% confidence interval)
Total number	46 (100)	35 (100)	
Evaluating a medical intervention (group A)	30 (65)	29 (83)	17.6 (–1.0 to 36.2)
Preceded by a systematic review	29 (63)	30 (86)	22.7 (4.5 to 40.8)
Combining animal with human data	5 (11)	4 (11)	0.6 (–13.3 to 14.4)
Reporting the number of studies being combined	41 (89)	35 (100)	10.9 (1.9 to 19.9)
Minimum number of experiments combined	3	3	
Maximum number of experiments combined	397	484	
Providing exact details on species included	37 (80)	32 (91)	11.0 (–3.8 to 25.7)
Including only one species or strain of animal	14 (30)	13 (37)	6.7 (–14.1 to 27.5)
Assessing study quality of included articles	14 (30)	20 (57)	26.7 (5.6 to 47.8)
Attention to between-study heterogeneity	34 (74)	24 (69)	–5.3 (–25.3 to 14.6)
Assessing between-study heterogeneity	26 (57)	22 (63)	6.3 (–15.1 to 27.8)
Attention to publication bias	17 (37)	26 (74)	37.3 (17.2 to 57.4)
Assessing publication bias	6 (13)	21 (60)	47.0 (28.0 to 65.9)
of which find evidence of publication bias	3 (50)	13 (62)	11.9 (–33.2 to –57.0)

Table 3 Methods for assessment of publication bias used in meta-analyses published before and after July 2005

Method for assessment of publication bias	Before July 2005 n (%)	July 2005–2010 n (%)
Meta-analyses assessing publication bias	6 (100)	21 (100)
Funnel plot only	0 (0)	7 (33)
Egger's test only	2 (33)	0 (0)
Funnel plot & Egger's test	2 (33)	11 (52)
Fail-Safe Number	1 (17)	2 (10)
Trim and fill method	0 (0)	1 (5)
Comparison of abstracts with full papers	1 (17)	0 (0)

Our study has some limitations. Firstly, our definition of SR causes that also several reviews have been included which do not fulfil the definition of SR as defined by the Cochrane Collaboration,¹⁹ an international non-profit organization that promotes the production and accessibility of SRs. Many studies included in this review performed a systematic search but the presentation of their results is more narrative rather than systematic since, for example, study quality of included studies is often not assessed. However, we maintained the definition used by Peters *et al.* for comparability.

Secondly, our definition of LAE was slightly different from Peters *et al.*, with five subgroups instead of three. We added two subgroups since we considered the subgroup of LAEs 'measuring an epidemiological association', as described by Peters *et al.*, rather vague. However, we do not expect this to have affected our results to an important degree since 83% of the included studies fall in subgroup A, in which a medical intervention was applied, or subgroup C, in which effects of an exposure to a chemical substance were measured, which were similar in both studies. A large proportion, if not all, of the remaining 17% may be included in the 'epidemiological association' subgroup used by Peters *et al.*

Thirdly, our results might be biased by publication bias. Although we did a thorough literature search, we did not search any grey literature since Peters *et al.* searched Cancerlit, UK DoH, Graylit, Agricola and British Library and found 0 out of 960 candidate references. However, even if we have missed studies, this would not change our main finding that the number of SRs and MAs in LAEs has increased strongly.

Finally, we were not able to check to which extent the included MAs adhered to the guidelines for good quality reporting of MAs of LAEs as proposed by Peters *et al.* and which were based on the guidelines of the Quality of Reporting of MA (QUOROM) statement.²⁰ Peters *et al.* presented the results of the MAs before July 2005 but did not clearly describe how they applied these guidelines. With the help of the reference list of included MAs provided by Peters *et al.*, we tried to reproduce their methods of scoring. However, when we applied the guidelines to these references exactly, the results were different from those presented by Peters *et al.* Therefore, we concluded that they had used a different method of scoring and exact comparison seemed impossible.

How might SRs and MAs affect the 3Rs? SRs and MAs help to *refine* LAEs since mistakes made by others performing a similar study can be prevented. For example, they can give us important insights into limitations of particular animal models, deficiencies in experimentals' design or execution, and reporting. These insights provide the opportunity to improve the quality of the research process. MAs offer a unique opportunity to explore the reasons behind heterogeneity of findings from related studies. Different animal species and models can be included and compared with each other and with SRs and MAs of human research. This may lead to more knowledge about consistency and generalizability of effects between these different groups. Quality of conduct and reporting can be assessed by scoring included studies using a relevant and standardized checklist. For example, the CAMARADES collaboration published a checklist for study-quality of preclinical stroke research.²¹ There are several examples in which there was an inverse association between studies' methodological quality and the magnitude of their findings.^{22–24}

Results of SRs and MAs provide an overview of all the available evidence on a given subject. Since more data are combined, the precision of estimated effects is usually much increased, resulting in conclusions with more power than those of a single study. This increased precision reduces the number of animals needed in future experiments, directly supporting the principle of *reduction*. On the other hand, it may become clear that larger studies are needed to solve a particular problem. This may imply expansion on the short term, but reduction ultimately. By means of cumulative MA, newer trials can be added to update the MA.^{25,26} The increased precision of estimated effects provides the possibility to select the most promising treatments and therefore ensures a better translation to human research. This way, performing more SRs might also promote a closer cooperation between animal and human research.

However, performing more of these studies also poses some concerns. Firstly, the large number of studies claiming to have performed an SR or systematic search but lack details of their methods of searching is a cause for concern. The credibility of these studies is low since the reader does not know how thoroughly the literature was searched and replication is not possible. Secondly, SRs and, especially, MAs are sensitive to several types of bias. Poor quality of included studies, between-study heterogeneity and publication bias can significantly affect results, and must therefore always be assessed and, where possible, accounted for. Publication bias may lead to unnecessary duplication of research efforts and waste of time, money and animals. If one agrees that SRs and MAs are important to any field of science, it is also important to tackle the issue of publication bias since it may compromise the validity of these studies. Although there are statistical techniques to estimate the extent of the problem, a more fundamental solution is to simply have all the relevant data available avoiding the need for statistical assumptions that cannot always be checked.

The rise of SRs and MAs in the field of applied clinical medicine caused reviewers to become much more aware

of the impact poor study quality, between-study heterogeneity and publication bias may have on their study results. This, in turn, led to a wide debate on these problems and potential solutions, for example trial registries.^{27,28} Since 2005, many top medical journals vowed to no longer publish unregistered trials although this has proven difficult.²⁹ Considering the growing number of MAs that assessed study quality and publication bias, it seems that the performance of SRs and MAs has a similar effect in the LAE field. However, little is currently known about the extent of publication bias in this field. The evidence that does exist indicates that LAEs too are affected by publication bias.¹⁰ An obvious approach would be to, either prospectively or historically, follow up a cohort of ethically approved protocols of LAEs.^{6,30} This way, the proportion of studies that are being finished and published can be established and determinants of (non-)publication studied.

In summary, compared with human research, the absolute number of SRs and MAs of LAEs appears to be low. However, the number has been growing strongly since July 2005. Furthermore, the quality of MAs has increased and publication bias is more often considered and assessed. The majority of the MAs that assess publication bias find evidence of missing 'negative' studies, indicating that this problem should probably not be underestimated. Therefore, we recommend that further research is performed to determine the extent of publication bias in the LAE field. In the meantime, given SRs' potential for reduction and refinement, it seems advisable for animal researchers to consider performing an SR before embarking on new experiments.

REFERENCES

- Musch TI, Carroll RG, Just A, Lane PH, Talman WT. A broader view of animal research. *BMJ* 2007;**334**:274
- Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? *BMJ* 2004;**328**:514–17
- Mignini LE, Khan KS. Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. *BMC Med Res Methodol* 2006;**6**:10
- Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR. A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting. *J Environ Sci Health B* 2006;**41**:1245–58
- Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;**263**:1385–9
- Song F, Parekh-Bhurke S, Hooper L, et al. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol* 2009;**9**:79
- Bracken MB. Why are so many epidemiology associations inflated or wrong? Does poorly conducted animal research suggest implausible hypotheses? *Ann Epidemiol* 2009;**19**:220–4
- van der Worp HB, Howells D, Sena E, et al. Can animal models of disease reliably inform human studies? *PLoS Med* 2010;**7**:e1000245
- Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci* 2007;**30**:433–9
- Sena E, van der Worp HB, Bath P, Howells D, Macleod M. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 2010;**8**:e1000344
- Pfannkuche KA, Bouma A, Groothuis TG. Does testosterone affect lateralization of brain and behaviour?: A meta-analysis in humans and other animal species. *Philos Trans R Soc Biol Sci* 2009;**364**:929–42
- Carrillo M, Ricci LA, Coppersmith GA, Melloni RH Jr. The effect of increased serotonergic neurotransmission on aggression: a critical meta-analytical review of preclinical studies. *Psychopharmacology* 2009;**205**:349–68
- Cornwall GA. New insights into epididymal biology and function. *Hum Reprod Update* 2009;**15**:213–27
- Goldstein DS, Kopin IJ. Adrenomedullary, adrenocortical, and sympathoneural responses to stressors: a meta-analysis. *Endocr Regul* 2008;**42**:111–19
- Baker-Herman TL, Mitchell GS. Determinants of frequency long-term facilitation following acute intermittent hypoxia in vagotomized rats. *Respir Physiol Neurobiol* 2008;**162**:8–17
- Frank ST, Krumm B, Spanagel R. Cocaine-induced dopamine overflow within the nucleus accumbens measured by *in vivo* microdialysis: a meta-analysis. *Synapse* 2008;**62**:243–52
- Isensee J, Ruiz NP. Sexually dimorphic gene expression in mammalian somatic tissue. *Gender Med* 2007;**4**(Suppl. B):S75–95
- Rothstein HR, Sutton AJ, Borenstein M. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester: John Wiley & Sons, 2005
- Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2*. The Cochrane Collaboration, 2009
- Moher D, Cook D, Eastwood S, Olkin I, Rennie D, Stroup D. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;**354**:1896
- Macleod M, O'Collins T, Howells D, Donnan G. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* 2004;**35**:1203–8
- Macleod M, O'Collins T, Horky L, Howells D, Donnan G. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. *J Cereb Blood Flow Metab* 2005;**25**:713–21
- Sena E, Wheble P, Sandercock P, Macleod M. Systematic review and meta-analysis of the efficacy of tirilazad in experimental stroke. *Stroke* 2007;**38**:388–94
- Macleod M, van der Worp B, Sena E, Howells D, Dirnagl U, Donnan G. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* 2008;**39**:2824–9
- Lau J, Antman E, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers T. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;**327**:248–54
- Lau J, Schmid C, Chalmers T. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995;**48**:45–57
- Dickersin K, Rennie D. Registering clinical trials. *JAMA* 2003;**290**:516–23
- Moher D. Clinical-trial registration: a call for its implementation in Canada. *Can Med Assoc J* 1993;**149**:11
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;**302**:977–84
- de Jong J, Ter Riet G, Willems D. Two prognostic indicators of the publication rate of clinical studies were available during ethical review. *J Clin Epidemiol* 2010;**12**:1342–50

(Accepted 26 April 2011)