The Need for Randomization in Animal Trials: An Overview of Systematic Reviews

Jennifer A. Hirst[1]*[◑], Jeremy Howick[1]*[◑], Jeffrey K. Aronson[1], Nia Roberts[2], Rafael Perera[1], Constantinos Koshiaris, Carl Heneghan[1]

1 Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom, 2 Bodleian Libraries, University of Oxford, Oxford, United Kingdom

Abstract

Background and Objectives: Randomization, allocation concealment, and blind outcome assessment have been shown to reduce bias in human studies. Authors from the Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies (CAMARADES) collaboration recently found that these features protect against bias in animal stroke studies. We extended the scope the work from CAMARADES to include investigations of treatments for any condition.

Methods: We conducted an overview of systematic reviews. We searched Medline and Embase for systematic reviews of animal studies testing any intervention (against any control) and we included any disease area and outcome. We included reviews comparing randomized versus not randomized (but otherwise controlled), concealed versus unconcealed treatment allocation, or blinded versus unblinded outcome assessment.

Results: Thirty-one systematic reviews met our inclusion criteria: 20 investigated treatments for experimental stroke, 4 reviews investigated treatments for spinal cord diseases, while 1 review each investigated treatments for bone cancer, intracerebral hemorrhage, glioma, multiple sclerosis, Parkinson's disease, and treatments used in emergency medicine. In our sample 29% of studies reported randomization, 15% of studies reported allocation concealment, and 35% of studies reported blinded outcome assessment. We pooled the results in a meta-analysis, and in our primary analysis found that failure to randomize significantly increased effect sizes, whereas allocation concealment and blinding did not. In our secondary analyses we found that randomization, allocation concealment, and blinding reduced effect sizes, especially where outcomes were subjective.

Conclusions: Our study demonstrates the need for randomization, allocation concealment, and blind outcome assessment in animal research across a wide range of outcomes and disease areas. Since human studies are often justified based on results from animal studies, our results suggest that unduly biased animal studies should not be allowed to constitute part of the rationale for human trials.

Introduction

Bias in Animal Studies

Clinical epidemiologists and proponents of evidence-based medicine (EBM) have been using methods to reduce bias in human studies for over four decades. [1–5] Random allocation of participants to treatment groups, concealing the allocation sequence from those assigning participants to intervention groups (allocation concealment), and blinding of investigators assessing outcomes are now viewed as fundamental ways of ensuring quality and minimizing bias in clinical trials. [6] This is because concealed random allocation reduces selection bias and blinding outcome assessors reduces detection bias. [5] Armed with these methods, researchers have exposed several common medical practices as ineffective. For example, observational studies led us to believe

that sodium fluoride reduced vertebral fractures, [7] that vitamin E reduced major coronary events, [8] and that high-dose aspirin was more effective than low-dose aspirin. [9] But subsequent randomized trials exposed all these treatments as useless or harmful. [10,11] Benefits of randomization, allocation concealment, and blinding have been confirmed in larger meta-epidemiological studies. In the earliest of these, Schulz et al. (1995) found that odds ratios were exaggerated by 30% in trials lacking allocation concealment and by 17% in studies that lacked blind outcome assessment. [12] Subsequent larger investigations have confirmed these results and also shown that adequate randomization reduces bias in human studies. [13,14]

A growing body of evidence is beginning to suggest that randomization, allocation concealment, and blinding outcome assessment can also reduce the risk of bias of animal studies. [15–

25] Some researchers hypothesize that avoidable biases in animal studies contribute to the failure to translate much experimental work for human benefit. [26,27] For example, while 503 of 835 candidate drugs for use in the management of stroke appeared effective in animal models, only one (tissue plasminogen activator) has proved sufficiently efficacious in humans. [28]

Much research into the empirical dimensions of bias in animal studies has been conducted by investigators from the Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies (CAMARADES) group. [29] CAMARADES researchers recently conducted an overview of systematic reviews of animal studies researching treatments for experimental stroke, and showed that failure to conceal allocation (but not failure to randomize or blind) exaggerated apparent treatment benefits in animal studies. [30] Despite this research, evidence-based principles have not yet been widely adopted in animal research; a recent study showed that only one in six controlled animal studies use randomization and only one in five use blind outcome assessment [31]. We therefore aimed to replicate the CAMARADES study independently and to expand its scope to include all conditions.

Methods

We conducted an overview of systematic reviews. The protocol (unpublished) was finalized by JH, CH, RP, and JA in October 2012. We modified the protocol once to add the secondary analysis (testing the "unpredictability paradox"; see below). We searched MEDLINE and Embase databases (19 April 2012) and scanned reference lists for systematic reviews of animal studies that measured effects of randomization, allocation concealment, or blinding of outcome assessment. We included reviews in any disease area, using any intervention, any control group, any outcome measure and any animal model. We limited our search to the last 20 years and excluded human studies (search strategy in Appendix S1). We also excluded conference papers, studies not reported in English, ecological studies, and epidemiological studies.

Two reviewers (JH and JAH) independently extracted data on numbers of studies, numbers of animals, disease/condition, outcomes, effect measures, and effect sizes with confidence intervals, using piloted data extraction forms. Disagreements were resolved by discussion with other authors. Authors were contacted to request data which were not reported. To enable inclusion of one review [32] we estimated the number of animals in randomized and non-randomized groups by calculating the mean number of animals per study. To test whether this estimation affected our results we carried out a sensitivity analysis by removing the study from the meta-analysis. We assessed the risk of bias of included systematic reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) criteria. [33]

We pooled results using the DerSimonian and Laird random effects model. [34] We reported outcomes for which differences between randomization/no randomization, allocation concealment/no allocation concealment, and blinding/no blinding were reported. We combined different outcomes and measurement units using standardized mean differences (SMDs), and quantified heterogeneity using the I-squared statistic. [35] We used meta-regression in a post-hoc analysis to examine whether various features influenced outcomes. Specifically, we investigated whether study size, disease state (stroke versus all other outcomes), or outcome measure were significantly associated with the effect size or could explain some of the heterogeneity.

For our secondary analysis we investigated the "unpredictability paradox", which was proposed in a similar study involving human subjects. [13] The paradox states that the difference between inadequately randomized and randomized studies, although real, is unpredictable in terms of direction. This is plausible, given that the direction of bias may relate to differences in expected results. To investigate the paradox we ignored direction to see whether there was an absolute difference between results in randomized and non-randomized studies. We used the same method to investigate the unpredictability paradox for adequate allocation concealment and blinding. This approach is useful only as a guide, since with a large enough sample some absolute difference is likely to arise by chance alone.

Results

We identified 238 articles from our electronic search, and a further 24 articles by hand searching references and contacting CAMARADES authors. Two authors (JH, JAH) excluded 199 articles after reading titles and abstracts. We assessed the full text of the remaining 63 articles and excluded a further 32 for not including outcome data. CAMARADES authors generously shared data from 19 reviews in which data were not included in the published reports. We were left with 31 systematic reviews involving 7339 comparisons (estimated 123,437 animals) to include in the meta-analysis (see Figure 1). Characteristics of the 31 included reviews are shown in Table 1, and our data are available freely from the authors.

Twenty systematic reviews investigated treatments for experimental stroke, [17–20,24,28,32,36–47] four reviews investigated treatments for spinal cord diseases, [48–51] one review each investigated treatments for bone cancer, [52] intracerebral hemorrhage, [39] glioma, [53] multiple sclerosis, [54] Parkinson's disease, [55] and any treatments used in emergency medicine. Animal types included baboons, cats, dogs, ewes, gerbils, guinea pigs, lambs, marmosets, mice, monkeys, pigs, rabbits, rats, and sheep. In our sample 29% of studies reported randomization, 15% reported allocation concealment, and 35% reported blinded outcome assessment.

1. Randomization

Thirty reviews with 7249 comparisons (121,784 animals) reported the effects of randomization. Randomized trials reduced effect sizes by a moderate and statistically significant amount (SMD = -0.07, 95% CI -0.12 to -0.02, $I^2 = 89.1\%$, $P = 0.008$) (Figure 2). In a subgroup analysis examining the effect of randomization by disease (stroke versus other), we found that randomization resulted in a lower effect size in areas other than stroke (SMD -0.18, 95% CI -0.30 to -0.06) but not stroke itself (SMD -0.03 95% CI -0.08 to 0.02). However, using meta-regression we found no significant difference between stroke and non-stroke on outcome measures ($P = 0.08$); additionally, meta-regression could not explain more than 3% of the heterogeneity. A sensitivity analysis excluding the single review [32] in which we had to estimate the number of animals, did not alter the overall result (SMD = -0.08 95% CI -0.13 to -0.03). In our secondary analysis (where we ignored direction of effect) we found a larger difference between randomized and non-randomized studies (SMD -0.16, 95% CI -0.21 to -0.11, $I^2 = 86.6\%$, $P<0.0001$) compared with the effect size in which we took direction into consideration.
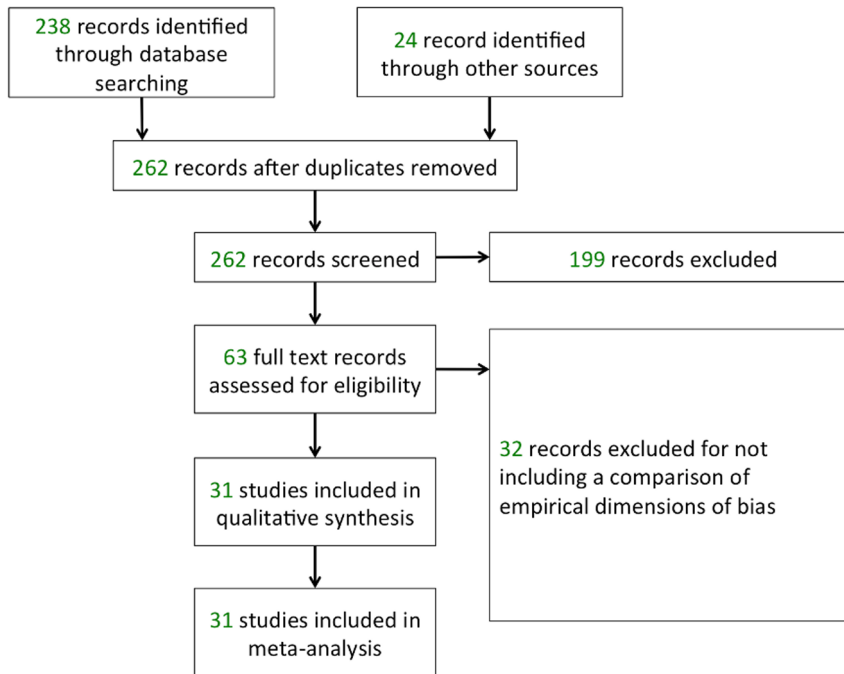
Figure 1. Flowchart of identified and included studies.
doi:10.1371/journal.pone.0098856.g001

2. Allocation concealment

Eighteen reviews with 2696 comparisons (39,405 animals) reported the effect of allocation concealment. Studies in which allocation concealment was used resulted in slightly decreased effect sizes, but this was not statistically significant (SMD = −0.04, 95% CI −0.09 to 0.00, $I^2 = 51.6\%$, P = 0.059) (Figure 3). Subgroup analysis examining different diseases (stroke and non-stroke) showed that allocation concealment in studies of stroke resulted in significantly lower effect sizes (SMD = −0.07, 95% CI −0.12 to −0.02, $I^2 = 48.5\%$, $P = 0.009$), whereas allocation concealment in other disease areas resulted in higher effect sizes (SMD 0.05, 95% CI −0.01 to 0.11, $I^2 = 0\%$, $P = 0.128$) but the difference between these groups was not found to be significant using meta-regression ($P = 0.073$). Meta-regression of the combination of disease and outcome measure was did not explain more than 9% of the heterogeneity. In our secondary analysis (where we ignored direction of effect) we found a larger difference between concealed and non-concealed studies (SMD −0.08, 95% CI −0.11 to −0.05, $I^2 = 13.8\%$, $P<0.0001$) compared with the effect size in which we took direction into consideration.

3. Blinding

Twenty-eight reviews involving 7140 comparisons (119,597 animals) reported the effects of blinding of outcome assessment. Effect sizes in studies that involved blind outcome assessment were not significantly different from studies that did not (SMD = −0.01, 95% CI −0.04 to 0.03; $I^2 = 68.3\%$; $P = 0.667$) (Figure 4). A sensitivity analysis excluding one study in which some estimates were made did not change results. [16] We did not find any differences in effect sizes when we sub-divided studies into stroke and non-stroke groups. In a post-hoc subgroup analysis, we showed that blinding in studies reporting infarct volume did not significantly change effect size (SMD = 0.03, 95% CI −0.02 to 0.08, $P = 0.187$)), whereas blinding in those reporting neurobehavioral outcomes did (SMD = −0.06, 95% CI −0.10 to −0.02,

$P = 0.003$) and this difference was significant when tested using meta-regression ($P = 0.014$). In our secondary analysis (in which effect direction was ignored) we found a larger difference between blinded and non-blinded studies (SMD = −0.08; 95% CI −0.11, −0.06; $I^2 = 49.5\%$; $P < 0.001$) compared with the effect size in which we took direction into consideration.

4. Risk of bias

Using AMSTAR (Table 2), we found a moderate risk of bias. It was encouraging that all 31 reviews assessed the quality of included studies, all but two reviews used clearly used appropriate methods, and all but two reviews performed comprehensive literature searches. Yet only 9 studies provided a protocol, and only 17 studies searched the grey literature.

Discussion

In this overview of systematic reviews we found that failure to randomize is likely to result in overestimation of the apparent treatment benefits of interventions across a range of disease areas and outcome measures. We also found a borderline effect of allocation concealment but no overall effect of blinding in our primary analysis. We hypothesize that the reason for an effect of randomization but not allocation concealment or blinding is that subjective judgments are less likely to influence outcomes in trials of (relatively homogeneous) animal models compared with (relatively heterogeneous) humans. While animal heart rates [56], blood flow [57], and behavior can be conditioned by human handling so that placebo controls are sometimes also used in animal studies, [58] there are no 'patient-reported' (subjective) outcomes in animal studies. This may make some measures of expectancy effects (for which blinding is useful [5]) smaller in animal studies. Our hypothesis is supported by our post hoc analyses, which showed that blinding reduced effect sizes for (more subjective) neurobehavioral scores, but not for (more objective) infarct volume. It may also be relevant that the comparison of

Table 1. Outcome measures, interventions, diseases, and effect sizes in included studies.

Author	Intervention	Disease	Outcome measure	Number of comparisons	Number of animals	Effect size (95% CI)			Animals used
						Randomized	Allocation concealed	Blinded	
Antonic (2013)	Stem cells Transplantation	Spinal Cord Diseases	Neurobehaviour score	315	5781	0.07 (0.01, 0.12)	0.08 (−0.01, 0.16)	−0.07 (−0.12, −0.02)	Rats, mice
Banwell (2009)	IL1 RA	Stroke	Infarct volume	44	784	−0.14 (−0.30, 0.02)	−0.22 (−0.38, −0.06)	0.05 (−0.10, 0.20)	Rats, mice
Batchelor (2013a)	Decompression	Spinal Cord Diseases	Neurobehaviour score	79	874	−0.09 (−0.23, 0.05)	0.15 (−0.06, 0.35)	−0.19 (−0.32, 0.05)	Dogs, mice, rats, sheep
Batchelor (2013b)	Hypothermia	Spinal Cord Diseases	Neurobehaviour score	25	448	−0.02 (−0.21, 0.17)	0.00 (−0.21, 0.22)	−0.04 (−0.23, 0.14)	Dogs, monkeys, rats
Bath (2009)	NXY-059	Stroke	Infarct volume	13	275	−0.25 (−0.49, −0.01)	−0.11 (−0.29, 0.08)	−0.14 (−0.33, 0.05)	Rats, mice, marmosets
Bebarta (2003)	Emergency medicine (all)	Any	Any outcome	290	*	−0.68 (−1.01, −0.29)	—	−0.66 (−1.13, −0.29)	Any (not specified)
Currie (2013)	Any	Bone cancer	Behavioural	202	4272	0.02 (−0.06, 0.10)	—	0.06 (−0.00, 0.12)	Rats, mice
			Histology, biochemistry	197	3228	−0.84 (−1.00, −0.68)	—	0.21 (0.13, 0.29)	
			Anatomical	27	470	—	—	0.08 (−0.10, 0.26)	
Egan (2014)	Exercise	Stroke	Neurobehaviour score	42	771	0.04 (−0.11, 0.18)	−0.02 (−0.17, 0.13)	−0.12 (−0.27, 0.02)	Rats, mice
			Infarct volume	65	987	0.13 (−0.01, 0.28)	−0.03 (−0.16, 0.10)	0.21 (0.02, 0.39)	
Frantzias (2011)	All drugs	Intracerebral hemorrhage	Neurobehaviour score	223	3932	−0.11 (−0.18, −0.05)	−0.09 (−0.18, −0.01)	0.02 (−0.05, 0.08)	Rats, mice, cats, rabbits, non-human primates
Gibson (2006)	Estrogen	Stroke	Infarct volume	22	372	0.49 (0.19, 0.79)	—	—	Rats, mice
Hirst (2013)	Temozolomide	Glioma	Median survival	123	2242	−0.49 (−0.57, −0.40)	—	—	Rats, mice
			Tumour volume	26	409	−0.01 (−0.25, 0.22)	—	0.16 (−0.17, 0.49)	
Horn (2001)	Nimodipine	Stroke	Infarct volume	7	121	—	—	0.16 (−0.23, 0.55)	Rats, rabbits, cats
Janssen (2010)	Enriched environment	Stroke	Learning	8	130	0.47 (0.11, 0.83)	—	—	Rats, mice
Jerndal (2010)	Erythropoietin	Stroke	Neurobehaviour score	29	489	−0.17 (−0.35, 0.01)	−0.21 (−0.46, 0.03)	−0.33 (−0.51, −0.15)	Rats, mice, gerbils
			Infarct volume	23	336	−0.27 (−0.48, −0.05)	−0.07 (−0.31, 0.16)	−0.17 (−0.38, 0.04)	
Lees (2012)	Stem cells	Stroke	Neurobehaviour score	233	3288	0.00 (−0.07, 0.07)	−0.01 (−0.09 to 0.07)	−0.03 (−0.10, 0.04)	Not specified
			Infarct volume	227	2804	−0.13 (−0.20, −0.05)	0.02 (−0.08, 0.11)	0.03 (−0.05, 0.11)	
Macleod (2004)	Nicotinamide	Stroke	Neurobehaviour score	52	711	−0.05 (−0.29, 0.19)	—	−0.05 (−0.29, 0.19)	Rats, mice
			Infarct volume	57	719	0.08 (−0.10, 0.27)	—	−0.01 (−0.17, 0.15)	
Macleod (2005a)	Melatonin	Stroke	Neurobehaviour score	6	47	—	—	−0.10 (−0.47, 0.28)	Rats, mice
			Infarct volume	27	419	0.20 (−0.02, 0.42)	0.31 (0.04, 0.58)	0.00 (−0.26, 0.27)	
Macleod (2005b)	FK 506 (tacrolimus)	Stroke	Neurobehaviour score	8	82	−0.89 (−1.35, −0.43)	—	—	Rats, mice, monkeys, gerbils
			Infarct volume	95	1569	0.10 (−0.06, 0.26)	—	0.12 (−0.01, 0.24)	

Table 1. Cont.

Author	Intervention	Disease	Outcome measure	Number of comparisons	Number of animals	Effect size (95% CI)			Animals used
						Randomized	Allocation concealed	Blinded	
Macleod (2008)	NXY-059	Stroke	Infarct volume	9	725	−0.44 (−0.65, −0.24)	−0.35 (−0.54, −0.17)	–	Mice, rats, rabbits, marmosets
Pedder (2014)	Any intervention	Lacunar stroke	Infarct volume	36	563	−0.01 (−0.17, 0.16)	−0.19 (−0.47, 0.09)	−0.00 (−0.17, 0.17)	Rats, rabbits, mice
Rooke (2011)	Dopamine	Parkinson's	Neurobehaviour score	601	5800	−0.03 (−0.09, 0.04)	−0.08 (−0.24, 0.08)	−0.08 (−0.14, −0.01)	Mice, rats, monkeys, guinea pigs
Sena (2007)	Tirlazad	Stroke	Neurobehaviour score	34	527	–	0.12 (−0.13, 0.36)	–	Rats, rabbits, cats
			Infarct volume	43	651	0.21 (0.03, 0.39)	−0.11 (−0.30, 0.07)	0.15 (−0.06, 0.36)	
Sena (2010)	Thrombotic occlusion	Stroke	Neurobehaviour score	69	1284	−0.04 (−0.15, 0.07)	−0.06 (−0.21, 0.08)	0.10 (−0.00, 0.22)	Monkeys, rats
			Infarct volume	231	3695	−0.01 (−0.07, 0.05)	−0.03 (−0.11, 0.05)	0.09 (0.02, 0.16)	
Van der Worp (2007)	Hypothermia	Stroke	Neurobehaviour score	55	870	−0.16 (−0.30, −0.01)	–	−0.05 (−0.21, 0.10)	Baboons, mice, rabbits, rats
			Infarct volume	222	3256	−0.10 (−0.17, −0.03)	−0.22 (−0.40, −0.03)	−0.06 (−0.13, 0.00)	
Vesterinen (2010)	Several interventions	Multiple sclerosis	Neurobehavioural outcomes	3190	64769	−0.01 (−0.03, 0.01)	–	−0.01 (−0.03, 0.01)	Mice, rats, guinea pigs, marmosets, monkeys, ewes
Vesterinen (2013)	Rho inhibitors	Stroke	Neurobehaviour score	30	502	0.14 (−0.04, 0.32)	0.09 (−0.20, 0.38)	−0.05 (−0.27, 0.16)	Rats, mice, dogs, gerbils
			Infarct volume	41	654	−0.04 (−0.20, 0.12)	0.00 (−0.29, 0.29)	−0.05 (−0.25, 0.14)	
Watzlawick (2014)	RhoA/ROCK-Blockade	Spinal Cord Diseases	Neurobehaviour score	30	655	−0.18 (−0.35, −0.00)	0.03 (−0.14, 0.20)	−0.19 (−0.37, −0.00)	Rats, mice
Wheble (2007)	Piracetam	Stroke	Infarct volume	14	197	0.44 (0.16, 0.72)	–	0.34 (0.05, 0.63)	Rats
Wilmot (2005a)	NOS inhibitors	Stroke	Neurobehaviour score	16	226	–	–	−0.03 (−0.29, 0.23)	Mice, gerbils, piglets, lambs, cats, rats
			Infarct volume	148	1998	−0.00 (−0.13, 0.12)	–	0.05 (−0.09, 0.20)	
Wilmot (2005b)	NOS Donors	Stroke	Infarct volume	40	483	0.09 (−0.10, 0.28)	–	0.01 (−0.34, 0.36)	Rats, rabbits
Wu (2014)	Edaravone	Stroke	Neurobehaviour score	30	519	−0.25 (−0.43, −0.08)	–	−0.15 (−0.32, 0.03)	Rats, mice
			Infarct volume	35	503	−0.16 (−0.33, 0.02)	–	−0.01 (−0.22, 0.21)	

* number of animals not reported and not required for analysis.
doi:10.1371/journal.pone.0098856.t001

5

Author	Year	Outcome	SMD (95% CI)	% Weight
Macleod	2008	Infarct volume	-0.44 (-0.65, -0.24)	2.17
Jerndal	2009	Infarct volume	-0.27 (-0.48, -0.05)	2.08
Bath	2009	Infarct volume	-0.25 (-0.49, -0.01)	1.92
Wu	2014	Infarct volume	-0.16 (-0.33, 0.02)	2.37
Banwell	2009	Infarct volume	-0.14 (-0.30, 0.02)	2.52
Lees	2012	Infarct volume	-0.13 (-0.20, -0.05)	3.10
Van de worp	2007	Infarct volume	-0.10 (-0.17, -0.03)	3.13
Vesterinen	2013	Infarct volume	-0.04 (-0.20, 0.12)	2.50
Sena	2010	Infarct volume	-0.01 (-0.07, 0.05)	3.16
Pedder	2014	Infarct volume	-0.01 (-0.17, 0.16)	2.47
Willmot	2004	Infarct volume	-0.00 (-0.13, 0.12)	2.76
Macleod	2004	Infarct volume	0.08 (-0.10, 0.27)	2.29
Willmot	2005	Infarct volume	0.09 (-0.10, 0.28)	2.26
Macleod	2005	Infarct volume	0.10 (-0.06, 0.26)	2.52
Egan	2014	Infarct volume	0.13 (-0.01, 0.28)	2.64
Macleod	2005	Infarct volume	0.20 (-0.02, 0.42)	2.06
Sena	2007	Infarct volume	0.21 (0.03, 0.39)	2.35
Wheble	2007	Infarct volume	0.44 (0.16, 0.72)	1.65
Gibson	2006	Infarct volume	0.49 (0.19, 0.79)	1.56
Currie	2013	Mixed	-0.84 (-1.00, -0.68)	2.50
Bebarta	2003	Mixed	-0.68 (-1.01, -0.29)	1.27
Hirst	2013	Mixed	-0.49 (-0.57, -0.40)	3.05
Hirst	2013	Mixed	-0.01 (-0.25, 0.22)	1.96
Currie	2013	Mixed	0.02 (-0.06, 0.10)	3.09
Janssen	2010	Mixed	0.47 (0.11, 0.83)	1.26
Macleod	2005	Neurobehaviour score	-0.89 (-1.35, -0.43)	0.91
Wu	2014	Neurobehaviour score	-0.25 (-0.43, -0.08)	2.37
Watzlawick	2014	Neurobehaviour score	-0.18 (-0.35, -0.00)	2.39
Jerndal	2009	Neurobehaviour score	-0.17 (-0.35, 0.01)	2.36
Van de worp	2007	Neurobehaviour score	-0.16 (-0.30, -0.01)	2.61
Frantzias	2011	Neurobehaviour score	-0.11 (-0.18, -0.05)	3.16
Batchelor	2013a	Neurobehaviour score	-0.09 (-0.23, 0.05)	2.64
Macleod	2004	Neurobehaviour score	-0.05 (-0.29, 0.19)	1.90
Sena	2010	Neurobehaviour score	-0.04 (-0.15, 0.07)	2.89
Rooke	2011	Neurobehaviour score	-0.03 (-0.09, 0.04)	3.16
Batchelor	2013b	Neurobehaviour score	-0.02 (-0.21, 0.17)	2.31
Vesterinen	2010	Neurobehaviour score	-0.01 (-0.03, 0.01)	3.30
Lees	2012	Neurobehaviour score	0.00 (-0.07, 0.07)	3.14
Egan	2014	Neurobehaviour score	0.04 (-0.11, 0.18)	2.65
Antonic	2013	Neurobehaviour score	0.07 (0.01, 0.12)	3.21
Vesterinen	2013	Neurobehaviour score	0.14 (-0.04, 0.32)	2.35
Overall (I-squared = 89.1%, p = 0.000)			-0.07 (-0.12, -0.02)	100.00

NOTE: Weights are from random effects analysis

-1.2   0   1

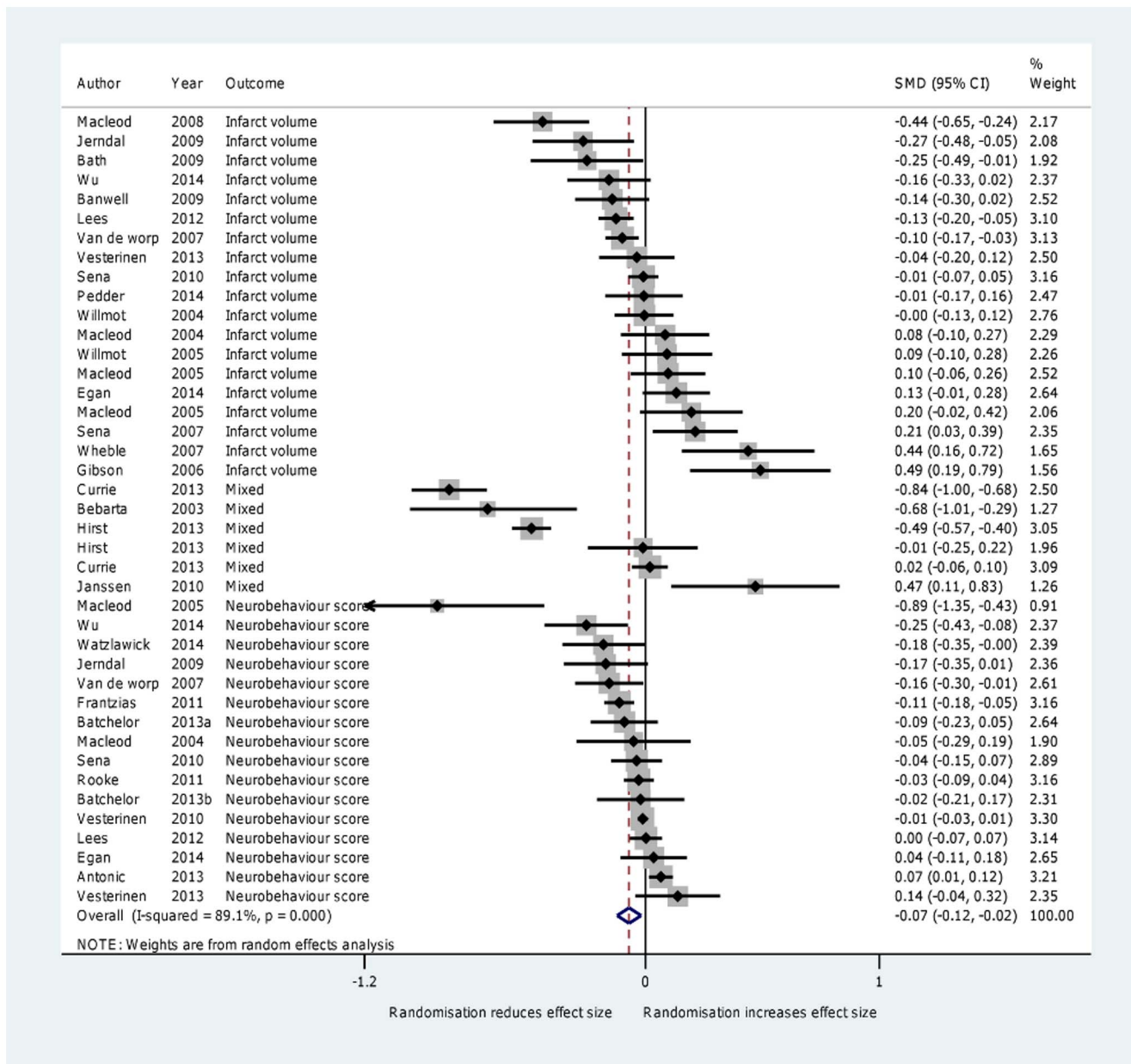Randomisation reduces effect size   Randomisation increases effect size

Figure 2. Forest plot showing the effect of random allocation on effect size.
doi:10.1371/journal.pone.0098856.g002

allocation concealment versus non-allocation concealment was reported far less frequently (about half as) as the other comparisons, so the failure to find an effect of allocation concealment could be due to insufficient power. A future individual major study of individual trials is now warranted to investigate the direction, magnitude, and conditions that must hold for randomization, allocation concealment, and blinding to reduce bias in animal studies.

Our results corroborate those of the CAMARADES study, in the sense that we also identified significant bias in animal studies. However, whereas they found a borderline effect of allocation concealment, but no effect for blinding or randomization, we found an effect of randomization, a borderline effect for allocation concealment, and no effect for blinding. The differences between the two reviews could be because our review covered all disease areas, whereas theirs was limited to experimental stroke. In addition, our methods were different; we calculated standardized mean differences rather than (the less widely used and more difficult to replicate) normalized mean differences used by the CAMARADES researchers.

Our study had several potential limitations. First, outcomes, animal models, and disease types were heterogeneous. The high levels of between-study heterogeneity of our overview could not be explained using meta-regression but may result from heterogeneity of the included reviews (and it was beyond the scope of our study to examine the sources of heterogeneity within our included reviews). Secondly, we relied on reports of systematic reviews; these, in turn, relied on reports of individual trials. Some trials may have failed to report randomization, allocation concealment, and blinding when in fact these were used, and vice versa. Evidence
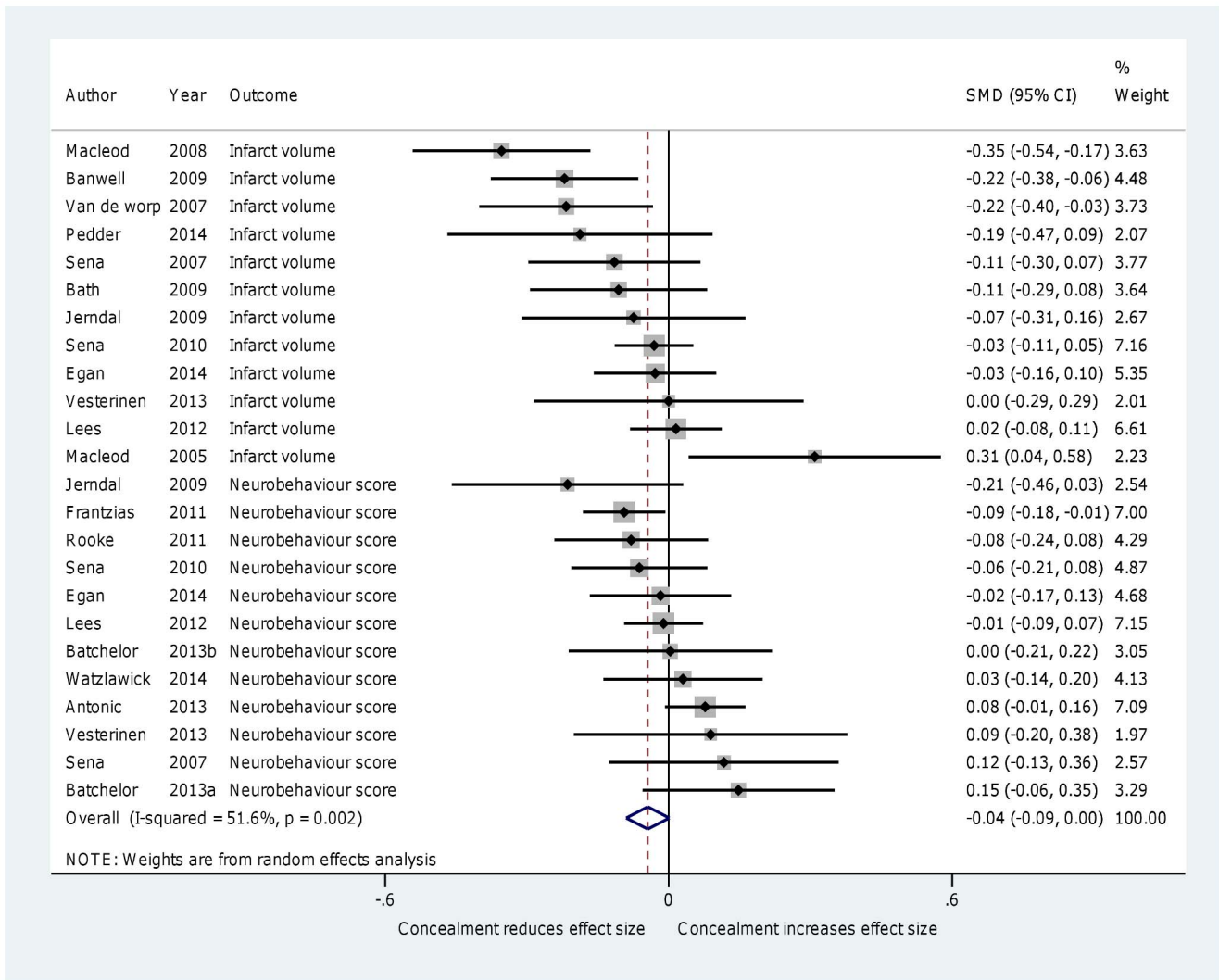
Figure 3. Forest plot showing the effect of allocation concealment on effect size.
doi:10.1371/journal.pone.0098856.g003

from clinical trials suggests that reporting quality is a good surrogate for actual risk of bias. If a similar relationship between reporting quality and study quality in animal studies holds, incomplete reporting may not have affected our results [59]. Based on reporting standards for clinical studies (that require, among other things, descriptions of how randomization, concealment, and blinding were achieved [60]) reporting standards for animal studies have been are emerging. [61] The Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines, developed in 2010, [62] arguably constitute the leading candidate for becoming a requirement, although development work in this area continues [63]. More recently, it has been suggested that until formal reporting guidelines become required: "at a minimum, authors of grant applications and scientific publications should report on randomization, blinding, sample-size estimation, and the handling of all data". [61]

Thirdly, it is unclear whether publication bias may have affected our results. It has been estimated that 1 in 6 animal trials remain unpublished, [64] so publication bias may have affected our results. If we assume that unpublished studies were equally likely to be randomized, allocation concealed, and blinded as they were to be non-randomized, not adequately concealed, and unblinded,

then publication bias may not have affected the direction of our results. As with human studies, [65] compulsory registration of preclinical studies [66] would reduce publication bias and allow more precise estimates of the empirical dimensions of bias in animal studies.

Fourthly, many of the individual trials included in the systematic reviews applied randomization, allocation concealment, and blinding together, whereas we examined these features independently. Of the 31 included reviews, 19 investigated experimental stroke. If stroke studies tend to be different from other types of studies this might have influenced the results, although we explored this using sub-group analysis and meta-regression. Fifthly, there were a disproportionate number of stroke studies included in out overview of systematic reviews. This was due to the fact that stroke researchers have spearheaded empirical investigations of bias in animal research. Finally, this study was restricted to an investigation of the effects of randomization, allocation concealment, and blinding. Other features, such as lack of power, publication bias, choice of animal models, choice of sex of animals, and choice of outcome may also contribute to the internal and external validity of animal studies. [22,31,54,67] A future
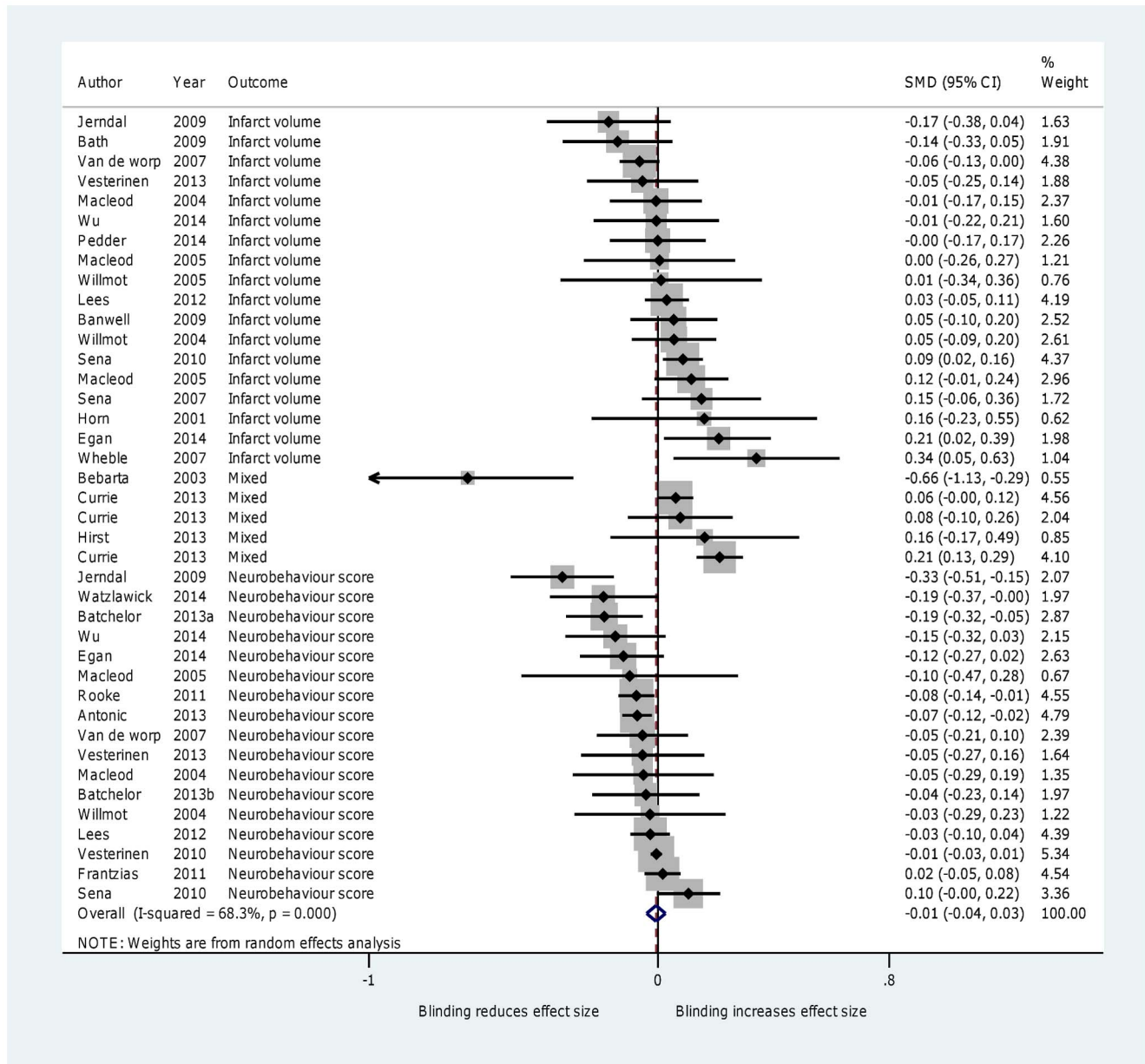
Author	Year	Outcome		SMD (95% CI)	% Weight
Jerndal	2009	Infarct volume		-0.17 (-0.38, 0.04)	1.63
Bath	2009	Infarct volume		-0.14 (-0.33, 0.05)	1.91
Van de worp	2007	Infarct volume		-0.06 (-0.13, 0.00)	4.38
Vesterinen	2013	Infarct volume		-0.05 (-0.25, 0.14)	1.88
Macleod	2004	Infarct volume		-0.01 (-0.17, 0.15)	2.37
Wu	2014	Infarct volume		-0.01 (-0.22, 0.21)	1.60
Pedder	2014	Infarct volume		-0.00 (-0.17, 0.17)	2.26
Macleod	2005	Infarct volume		0.00 (-0.26, 0.27)	1.21
Willmot	2005	Infarct volume		0.01 (-0.34, 0.36)	0.76
Lees	2012	Infarct volume		0.03 (-0.05, 0.11)	4.19
Banwell	2009	Infarct volume		0.05 (-0.10, 0.20)	2.52
Willmot	2004	Infarct volume		0.05 (-0.09, 0.20)	2.61
Sena	2010	Infarct volume		0.09 (0.02, 0.16)	4.37
Macleod	2005	Infarct volume		0.12 (-0.01, 0.24)	2.96
Sena	2007	Infarct volume		0.15 (-0.06, 0.36)	1.72
Horn	2001	Infarct volume		0.16 (-0.23, 0.55)	0.62
Egan	2014	Infarct volume		0.21 (0.02, 0.39)	1.98
Wheble	2007	Infarct volume		0.34 (0.05, 0.63)	1.04
Bebarta	2003	Mixed		-0.66 (-1.13, -0.29)	0.55
Currie	2013	Mixed		0.06 (-0.00, 0.12)	4.56
Currie	2013	Mixed		0.08 (-0.10, 0.26)	2.04
Hirst	2013	Mixed		0.16 (-0.17, 0.49)	0.85
Currie	2013	Mixed		0.21 (0.13, 0.29)	4.10
Jerndal	2009	Neurobehaviour score		-0.33 (-0.51, -0.15)	2.07
Watzlawick	2014	Neurobehaviour score		-0.19 (-0.37, -0.00)	1.97
Batchelor	2013a	Neurobehaviour score		-0.19 (-0.32, -0.05)	2.87
Wu	2014	Neurobehaviour score		-0.15 (-0.32, 0.03)	2.15
Egan	2014	Neurobehaviour score		-0.12 (-0.27, 0.02)	2.63
Macleod	2005	Neurobehaviour score		-0.10 (-0.47, 0.28)	0.67
Rooke	2011	Neurobehaviour score		-0.08 (-0.14, -0.01)	4.55
Antonic	2013	Neurobehaviour score		-0.07 (-0.12, -0.02)	4.79
Van de worp	2007	Neurobehaviour score		-0.05 (-0.21, 0.10)	2.39
Vesterinen	2013	Neurobehaviour score		-0.05 (-0.27, 0.16)	1.64
Macleod	2004	Neurobehaviour score		-0.05 (-0.29, 0.19)	1.35
Batchelor	2013b	Neurobehaviour score		-0.04 (-0.23, 0.14)	1.97
Willmot	2004	Neurobehaviour score		-0.03 (-0.29, 0.23)	1.22
Lees	2012	Neurobehaviour score		-0.03 (-0.10, 0.04)	4.39
Vesterinen	2010	Neurobehaviour score		-0.01 (-0.03, 0.01)	5.34
Frantzias	2011	Neurobehaviour score		0.02 (-0.05, 0.08)	4.54
Sena	2010	Neurobehaviour score		0.10 (-0.00, 0.22)	3.36
Overall (I-squared = 68.3%, p = 0.000)				-0.01 (-0.04, 0.03)	100.00

NOTE: Weights are from random effects analysis

```
-1                    0                    .8
   Blinding reduces effect size    Blinding increases effect size
```

Figure 4. Forest plot showing the effect of blinding of outcome assessment on effect size.
doi:10.1371/journal.pone.0098856.g004

individual study systematic review and meta-analysis is now warranted to address these potential limitations.

Our study has implications that extend beyond the conduct of animal studies. Only animal studies that do not suffer from avoidable bias should be accepted as justification for human studies. For this reason, the United States Food and Drug Administration (FDA), [68] the Medical Research Council (MRC) in the United Kingdom, [69] and the World Health Organization (WHO) [70] insist on fair tests, often involving systematic reviews of high quality randomized trials. Our study therefore supports the requirement for adequate conduct and reporting of animal studies, including those being promoted by CAMARADES, and SABRE Research UK. [71]

Conclusions

Our overview of systematic reviews and meta-analyses revealed that failure to randomize leads to exaggerated effect sizes in animal studies across a wide range of disease areas. In our secondary analysis we found that failure to conceal allocation or employ blind outcome assessment exaggerates effect sizes in animal studies. Biased animal research is less likely to provide trustworthy results, is less likely to provide a rationale for research that will benefit humans, and wastes scarce resources. Requiring compulsory study registration and adherence to emerging evidence-based standards for the conduct and reporting of animal research is likely to reduce the risk of bias in animal studies and improve translatability of animal research.

Table 2. AMSTAR Criteria for included studies*.

	1. Was an 'a priori' design provided?	2. Was there duplicate study selection and data extraction?	3. Was a comprehensive literature search performed?	4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	5. Was a list of studies (included and excluded) provided?	6. Were the characteristics of the included studies provided?	7. Was the scientific quality of the included studies assessed and documented?	8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	9. Were the methods used to combine the findings of the studies appropriate?	10. Was the likelihood of publication bias assessed?	11. Were conflicts of interest stated?
Antonic (2013)	1	1	1	1	1	1	1	1	1	1	1
Banwell (2009)	1	2	1	2	1	1	1	1	1	1	2
Batchelor (2013a)	1	1	1	2	2	1	1	1	1	1	1
Batchelor (2013b)	1	3	1	2	2	1	1	1	1	1	1
Bath (2009)	2	2	1	1	1	1	1	1	3	1	1
Bebarta (2003)	2	2	2	2	2	1	1	1	1	3	2
Currie (2013)	2	1	1	1	1	1	1	1	1	2	1
Egan (2014)	2	1	1	2	1	1	1	1	1	1	1
Frantzias (2011)	2	1	1	1	2	1	1	1	1	3	1
Gibson (2006)	2	1	1	1	1	1	1	1	1	1	2
Hirst (2013)	2	1	1	2	2	1	1	1	1	1	1
Horn (2001)	2	2	1	1	1	1	1	1	1	1	2
Janssen (2010)	2	1	1	2	2	1	1	1	1	2	1
Jerndal (2010)	2	1	1	2	1	1	1	1	1	3	1
Lees (2012)	2	3	1	2	1	1	1	1	1	1	1
Macleod (2004)	2	3	1	1	2	1	1	1	1	1	2
Macleod (2005a)	2	3	1	1	2	1	1	1	1	2	2
Macleod (2005b)	2	3	1	1	2	1	1	1	1	1	1
Macleod (2008)	2	2	1	2	1	1	1	1	3	2	2
Pedder (2014)	1	3	1	2	2	1	1	1	1	1	1
Rooke (2011)	2	1	1	1	2	2	1	1	1	1	1
Sena (2007)	2	1	1	1	2	1	1	1	1	1	2
Sena (2010)	2	1	1	1	2	2	1	1	1	1	1
Van der Worp (2007)	2	3	1	1	2	2	1	2	1	1	2
Vesterinen (2010)	2	1	2	3	2	2	1	1	1	3	1
Vesterinen (2013)	1	1	1	2	2	1	1	1	1	1	1
Watzlawick (2014)	1	1	1	1	2	2	1	1	1	1	1
Wheble (2007)	2	1	1	1	2	1	1	1	1	2	2
Wilmot (2005a)	1	1	1	1	2	2	1	1	1	1	2
Wilmot (2005b)	2	1	1	1	2	1	1	1	1	1	2
Wu (2014)	1	1	1	2	2	2	1	1	1	1	1

*1 = yes, 2 = no, 3 = can't answer, 4 = not applicable.
doi:10.1371/journal.pone.0098856.t002

Supporting Information

Appendix S1   Search Strategy.
(DOCX)

Checklist S1   PRISMA Checklist.
(DOC)

Author Contributions

Conceived and designed the experiments: JAH JH JKA RP CK CH. Performed the experiments: JAH JH NR CK. Analyzed the data: JAH JH CK. Contributed reagents/materials/analysis tools: RP. Wrote the paper: JAH JH JKA CK CH.

References

1. Sackett DL (1969) Clinical epidemiology. Am J Epidemiol 89: 125–128.
2. Sackett DL (1986) Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 89: 2S–3S.
3. Sackett DL, Richardson WS, Rosenberg W, Haynes B (1997) Evidence-based medicine: How to Practice & Teach EBM. London: Churchill Livingstone.
4. Chalmers I (2007) The lethal consequences of failing to make full use of all relevant evidence about the effects of medical treatments: the importance of systematic reviews. In: Rothwell PM, editor. Treating individuals: from randomised trials to personalized medicine. London: The Lancet.
5. Howick J (2011) The Philosophy of Evidence-Based Medicine. Oxford: Wiley-Blackwell.
6. Juni P, Altman DG, Egger M (2001) Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 323: 42–46.
7. Farley SM, Libanati CR, Odvina CV, Smith L, Eliel L, et al. (1989) Efficacy of long-term fluoride and calcium therapy in correcting the deficit of spinal bone density in osteoporosis. J Clin Epidemiol 42: 1067–1074.
8. Knekt P, Reunanen A, Jarvinen R, Seppanen R, Heliovaara M, et al. (1994) Antioxidant vitamin intake and coronary mortality in a longitudinal population study. Am J Epidemiol 139: 1180–1189.
9. Decousus H, Leizorovicz A, Parent F, Page Y, Tardy B, et al. (1998) A clinical trial of vena caval filters in the prevention of pulmonary embolism in patients with proximal deep-vein thrombosis. Prevention du Risque d'Embolie Pulmonaire par Interruption Cave Study Group. N Engl J Med 338: 409–415.
10. Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, et al. (1990) Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med 322: 802–809.
11. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P (2000) Vitamin E supplementation and cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342: 154–160.
12. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273: 408–412.
13. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, et al. (2011) Randomisation to protect against selection bias in healthcare trials. Cochrane database of systematic reviews: MR000012.
14. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, et al. (2012) Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials. Annals of internal medicine.
15. Bath PM, Macleod MR, Green AR (2009) Emulating multicentre clinical stroke trials: a new paradigm for studying novel interventions in experimental models of stroke. International journal of stroke: official journal of the International Stroke Society 4: 471–479.
16. Bebarta V, Luyten D, Heard K (2003) Emergency medicine animal research: does use of randomization and blinding affect the results? Academic emergency medicine: official journal of the Society for Academic Emergency Medicine 10: 684–687.
17. Jerndal M, Forsberg K, Sena ES, Macleod MR, O'Collins VE, et al. (2010) A systematic review and meta-analysis of erythropoietin in experimental stroke. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism 30: 961–968.
18. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA (2005) Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism 25: 713–721.
19. Macleod MR, O'Collins T, Howells DW, Donnan GA (2004) Pooling of animal experimental data reveals influence of study design and publication bias. Stroke; a journal of cerebral circulation 35: 1203–1208.
20. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, et al. (2008) Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke; a journal of cerebral circulation 39: 2824–2829.
21. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, et al. (2007) Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 334: 197.
22. Sena E, Wheble P, Sandercock P, Macleod M (2007) Systematic review and meta-analysis of the efficacy of tirilazad in experimental stroke. Stroke 38: 388–394.
23. van der Worp HB, de Haan P, Morrema E, Kalkman CJ (2005) Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. Journal of neurology 252: 1108–1114.
24. van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR (2007) Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain: a journal of neurology 130: 3063–3074.
25. Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, et al. (2011) Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the Journal of Cerebral Blood Flow and Metabolism. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism 31: 1064–1072.
26. MacLeod M (2010) How to avoid bumping into the translational roadblock. Rodent Models of Stroke Neuromethods. 47: 7–15.
27. Ioannidis JP (2006) Evolution and translation of research findings: from bench to where? PLoS Clin Trials 1: e36.
28. Sena ES, Briscoe CL, Howells DW, Donnan GA, Sandercock PA, et al. (2010) Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism 30: 1905–1913.
29. Macleod M (2011) CAMARADES. Edinburgh: CAMARADES.
30. Crossley NA, Sena E, Goehler J, Horn J, Van Der Worp B, et al. (2008) Empirical evidence of bias in the design of experimental stroke studies: A metaepidemiologic approach. Stroke 39: 929–934.
31. Macleod M, van der Worp HB (2010) Animal models of neurological disease: are there any babies in the bathwater? Practical neurology 10: 312–314.
32. Janssen H, Bernhardt J, Collier JM, Sena ES, McElduff P, et al. (2010) An enriched environment improves sensorimotor function post-ischemic stroke. Neurorehabilitation and neural repair 24: 802–813.
33. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, et al. (2007) External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2: e1350.
34. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clinical Trials 7: 177–188.
35. Higgins JJ, Green S (2008) The Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration.
36. Banwell V, Sena ES, Macleod MR (2009) Systematic review and stratified meta-analysis of the efficacy of interleukin-1 receptor antagonist in animal models of stroke. Journal of stroke and cerebrovascular diseases: the official journal of National Stroke Association 18: 269–276.
37. Bath PMW, Gray LJ, Bath AJG, Buchan A, Miyata T, et al. (2009) Effects of NXY-059 in experimental stroke: an individual animal meta-analysis. British Journal of Pharmacology 157: 1157–1171.
38. Egan KJ, Janssen H, Sena ES, Longley L, Speare S, et al. (2014) Exercise Reduces Infarct Volume and Facilitates Neurobehavioral Recovery: Results From a Systematic Review and Meta-analysis of Exercise in Experimental Models of Focal Ischemia. Neurorehabil Neural Repair.
39. Frantzias J, Sena ES, Macleod MR, Al-Shahi Salman R (2011) Treatment of intracerebral hemorrhage in animal models: meta-analysis. Annals of neurology 69: 389–399.
40. Gibson CL, Gray LJ, Murphy SP, Bath PM (2006) Estrogens and experimental ischemic stroke: a systematic review. J Cereb Blood Flow Metab 26: 1103–1113.
41. Horn J, De Haan RJ, Vermeulen M, Luiten PGM, Limburg M (2001) Nimodipine in animal model experiments of focal cerebral ischemia: A systematic review. Stroke 32: 2433–2438.

42. Lees JS, Sena ES, Egan KJ, Antonic A, Koblar SA, et al. (2012) Stem cell-based therapy for experimental stroke: a systematic review and meta-analysis. Int J Stroke 7: 582–588.

43. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA (2005) Systematic review and meta-analysis of the efficacy of melatonin in experimental stroke. Journal of pineal research 38: 35–41.

44. Pedder H, Vesterinen HM, Macleod MR, Wardlaw JM (2014) Systematic review and meta-analysis of interventions tested in animal models of lacunar stroke. Stroke 45: 563–570.

45. Sena E, van der Worp HB, Howells D, Macleod M (2007) How can we improve the pre-clinical development of drugs for stroke? Trends in neurosciences 30: 433–439.

46. Wheble PCR, Sena ES, Macleod MR (2008) A systematic review and meta-analysis of the efficacy of piracetam and piracetam-like compounds in experimental stroke. Cerebrovascular Diseases 25: 5–11.

47. Vesterinen HM, Currie GL, Carter S, Mee S, Watzlawick R, et al. (2013) Systematic review and stratified meta-analysis of the efficacy of RhoA and Rho kinase inhibitors in animal models of ischaemic stroke. Syst Rev 2: 33.

48. Antonic A, Sena ES, Lees JS, Wills TE, Skeers P, et al. (2013) Stem cell transplantation in traumatic spinal cord injury: a systematic review and meta-analysis of animal studies. PLoS Biol 11: e1001738.

49. Batchelor PE, Skeers P, Antonic A, Wills TE, Howells DW, et al. (2013) Systematic review and meta-analysis of therapeutic hypothermia in animal models of spinal cord injury. PLoS One 8: e71317.

50. Batchelor PE, Wills TE, Skeers P, Battistuzzo CR, Macleod MR, et al. (2013) Meta-analysis of pre-clinical studies of early decompression in acute spinal cord injury: a battle of time and pressure. PLoS One 8: e72659.

51. Watzlawick R, Sena ES, Dirnagl U, Brommer B, Kopp MA, et al. (2014) Effect and reporting bias of RhoA/ROCK-blockade intervention on locomotor recovery after spinal cord injury: a systematic review and meta-analysis. JAMA Neurol 71: 91–99.

52. Currie GL, Delaney A, Bennett MI, Dickenson AH, Egan KJ, et al. (2013) Animal models of bone cancer pain: systematic review and meta-analyses. Pain 154: 917–926.

53. Hirst TC, Vesterinen HM, Sena ES, Egan KJ, Macleod MR, et al. (2013) Systematic review and meta-analysis of temozolomide in animal models of glioma: was clinical efficacy predicted? Br J Cancer 108: 64–71.

54. Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, et al. (2010) Improving the translational hit of experimental treatments in multiple sclerosis. Multiple sclerosis 16: 1044–1055.

55. Rooke ED, Vesterinen HM, Sena ES, Egan KJ, Macleod MR (2011) Dopamine agonists in animal models of Parkinson's disease: A systematic review and meta-analysis. Parkinsonism & related disorders 17: 313–320.

56. Lynch JJ, Fregin GF, Mackie JB, Monroe RR, Jr. (1974) Heart rate changes in the horse to human contact. Psychophysiology 11: 472–478.

57. Newton JE, Ehrlich W (1993) Coronary blood flow in dogs: effect of person. Integrative physiological and behavioral science: the official journal of the Pavlovian Society 28: 280–286.

58. Breuer K, Hemsworth PH, Barnett JL, Matthews LR, Coleman GJ (2000) Behavioural response to humans and the productivity of commercial dairy cows. Applied animal behaviour science 66: 273–288.

59. Liberati A, Himel HN, Chalmers TC (1986) A quality assessment of randomized control trials of primary treatment of breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 4: 942–951.

60. Simera I (2008) EQUATOR Network collates resources for good research. BMJ 337: a2471.

61. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, et al. (2012) A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490: 187–191.

62. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. Journal of pharmacology & pharmacotherapeutics 1: 94–99.

63. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG (2013) Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. PLoS Med 10: e1001489.

64. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS biology 8: e1000344.

65. International Committee of Medical Journal Editors (2009) Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

66. Kimmelman J, Anderson JA (2012) Should preclinical studies be registered? Nat Biotechnol 30: 488–489.

67. Beery AK, Zucker I (2011) Sex bias in neuroscience and biomedical research. Neurosci Biobehav Rev 35: 565–572.

68. FDA (2005) CFR 314.126: Applications for FDA Approval to Market a New Drug. United States Food and Drug Administration.

69. Medical Research Council (2013) The MRC and Clinical Trials. London: MRC.

70. Vilar J, Duley L (2003) The need for large and simple randomized trials in reproductive health. Geneva: The World Health Organization Library.

71. SABRE (2012) SABRE Research UK. In: SABRE, editor.