OPEN

Methodological Rigor in Preclinical Cardiovascular Studies
Targets to Enhance Reproducibility and Promote Research Translation

F. Daniel Ramirez, Pouya Motazedian, Richard G. Jung, Pietro Di Santo,
Zachary D. MacDonald, Robert Moreland, Trevor Simard, Aisling A. Clancy, Juan J. Russo,
Vivian A. Welch, George A. Wells, Benjamin Hibbert

Rationale: Methodological sources of bias and suboptimal reporting contribute to irreproducibility in preclinical science and may negatively affect research translation. Randomization, blinding, sample size estimation, and considering sex as a biological variable are deemed crucial study design elements to maximize the quality and predictive value of preclinical experiments.

Objective: To examine the prevalence and temporal patterns of recommended study design element implementation in preclinical cardiovascular research.

Methods and Results: All articles published over a 10-year period in 5 leading cardiovascular journals were reviewed. Reports of in vivo experiments in nonhuman mammals describing pathophysiology, genetics, or therapeutic interventions relevant to specific cardiovascular disorders were identified. Data on study design and animal model use were collected. Citations at 60 months were additionally examined as a surrogate measure of research impact in a prespecified subset of studies, stratified by individual and cumulative study design elements. Of 28 636 articles screened, 3396 met inclusion criteria. Randomization was reported in 21.8%, blinding in 32.7%, and sample size estimation in 2.3%. Temporal and disease-specific analyses show that the implementation of these study design elements has overall not appreciably increased over the past decade, except in preclinical stroke research, which has uniquely demonstrated significant improvements in methodological rigor. In a subset of 1681 preclinical studies, randomization, blinding, sample size estimation, and inclusion of both sexes were not associated with increased citations at 60 months.

Conclusions: Methodological shortcomings are prevalent in preclinical cardiovascular research, have not substantially improved over the past 10 years, and may be overlooked when basing subsequent studies. Resultant risks of bias and threats to study validity have the potential to hinder progress in cardiovascular medicine as preclinical research often precedes and informs clinical trials. Stroke research quality has uniquely improved in recent years, warranting a closer examination for interventions to model in other cardiovascular fields.   (*Circ Res*. 2017;120:1916-1926. DOI: 10.1161/CIRCRESAHA.117.310628.)

Key Words: animal models ∎ bias ∎ biomedical research ∎ cardiovascular diseases ∎ reproducibility of results

Preclinical research and clinical therapy development exist in a symbiotic continuum yet are guided by different reporting standards, regulatory forces, and reward systems.[1,2] Consequent discrepancies in study designs and analytical methods can compromise scientific validity,[3] result in irreproducible results (particularly from preclinical experiments using animal models[2]), and may contribute to the high rate of attrition seen in early stages of clinical development.[4,5] Improved methodological rigor

and transparent reporting practices have therefore been advocated to improve the predictive value of animal model data.[2,6–10]

Editorial, see p 1852
In This Issue, see p 1843
Meet the First Author, see p 1844

The issue of irreproducibility in preclinical research and its impact are well-recognized in academic circles[2,11–14] and

Novelty and Significance

What Is Known?

- Preclinical research often precedes and informs clinical trials.
- Randomization, blinding, sample size estimation, and considering sex as a biological variable are considered crucial study design elements to maximize the predictive value of preclinical experiments.

What New Information Does This Article Contribute?

- Key study design elements are rarely implemented in preclinical cardiovascular research.
- The implementation of these elements has not appreciably improved over the past decade, except in stroke research.
- Methodological shortcomings in preclinical experiments may be overlooked when researchers design subsequent studies.

Preclinical and clinical research are guided by different methodological standards. Resultant discrepancies in study designs can compromise scientific validity, give rise to experimental irreproducibility, and may contribute to failures in research translation. Randomization, blinding, sample size estimation, and considering sex as a biological variable have been identified as key study design elements to promote rigor and reproducibility of animal experiments and research relevance to both men and women. However, there are limited data on the extent of suboptimal methodological rigor in preclinical research. We identified and analyzed 3396 preclinical studies in leading cardiovascular journals over a 10-year period, including citation counts at 60 months as a surrogate measure of research impact for papers published in the first 5 years. We show that these design elements were rarely implemented—a situation that has not substantially improved over the past decade, except in stroke research. The individual and cumulative implementation of these design elements was not associated with increased citations. Together, these data indicate that methodological shortcomings are prevalent in preclinical cardiovascular research and suggest that they may be overlooked when researchers design subsequent studies.

Nonstandard Abbreviations and Acronyms	
CVD	cardiovascular disorder
NIH	National Institutes of Health

increasingly so among the lay public[15–17] and the biopharmaceutical industry.[13,18] In 2011, Bayer HealthCare published an analysis of 67 in-house target identification and validation projects in the fields of oncology, women's health, and cardiovascular disease over a 4-year period, reporting that the published data were reproducible in less than one third of cases and that in nearly two thirds inconsistencies either prolonged the validation process or led to project termination.[11] In 2012, Amgen similarly published the results of their attempt to reproduce 53 "landmark" studies in the fields of hematology and oncology, succeeding in only 6 cases even after contacting the original authors for guidance, exchanging reagents, and in certain instances repeating experiments in the laboratory of the original investigators.[12] Some findings could not be reproduced even by the original investigators in their own laboratories when the experiments were repeated in a blinded fashion.[19] Disconcertingly, some of the irreproducible research were deemed to have prompted clinical studies. Others have likewise described candidate therapies that have advanced to testing in humans despite irreproducible or inconsistent results in animals.[8,20,21]

Since the mid-1990s, pharmaceutical research and development productivity has declined with decreasing numbers of medications approved, increasing attrition rates at all stages of research, and lengthening development times despite rising expenditures.[4] This decline has been attributed, in part, to flawed preclinical research methodology, highlighting the importance of preclinical research for advancing clinical care, but equally portending a waning confidence in animal model data to identify promising therapeutic targets.[5,7,11,12,22,23] Cardiovascular research and development, in particular, has fared poorly in recent years relative to other disease categories, exhibiting the greatest decline as a percentage of total projects[4] and among the lowest success rates and likelihood of approval at all phases of clinical trials.[24] Moreover, research and development investment patterns are increasingly deterring incremental innovation (for instance, improving upon available effective therapies). Instead, increasing focus is being placed on novel disease targets with higher revenue potential but at higher risk of failure (eg, specific cancers).[4] Poor reproducibility in preclinical research coupled with this increasingly unfavorable landscape for successful cardiovascular therapy development could, therefore, substantially hinder progress in cardiovascular care.

In response to the above systematic issues and in collaboration with editors from major journals, funding agencies, and scientific leaders, the National Institutes of Health (NIH) proposed a set of reporting guidelines and funding policies to improve the reproducibility and rigor of preclinical research.[25] A core set of standards first proposed by the NIH's National Institute of Neurological Disorders and Stroke were adopted, which identified randomization, blinding, sample size estimation, and data handling as minimum reporting requirements to promote transparency in animal studies.[7] In addition, the NIH announced that it would require that sex be considered as a biological variable in applications for preclinical research funding.[26] These principles and guidelines have been endorsed by prominent academic societies, associations, and journals, including all American Heart Association (AHA) journals,[25] with evidence of editorial commitment to complying.[27,28] However, there are limited data on the extent of suboptimal methodological rigor or incomplete reporting in preclinical cardiovascular science and therefore no baseline from which to gauge progress. A commitment to improving the quality and impact of preclinical research and to maintaining the trust of public and private stakeholders requires transparency on current and future states of scientific practice.[13] We, therefore, reviewed all preclinical cardiovascular studies published in leading AHA journals over the past decade to determine the prevalence of randomization, blinding, sample size estimation, and inclusion of both sexes and trends in these practices over time.

Methods

As previously described,[29] all preclinical cardiovascular studies published in AHA journals with archives spanning at least 10 years were reviewed. Five journals met these criteria: *Circulation*; *Circulation Research*; *Hypertension*; *Stroke*; and *Arteriosclerosis, Thrombosis, and Vascular Biology* (*ATVB*). All reports published during a 10-year period (July 2006 to June 2016) were screened. Studies were included if they represented original research published as full manuscripts; reported results of in vivo experiments in nonhuman mammals; and described pathophysiology, genetics, or therapeutic interventions that were stated to be directly relevant to a specific cardiovascular disorder (CVD) in humans (see prespecified list of CVDs below). Studies on physiological and genetic characteristics were included if potential therapeutic applications or implications of the study findings were proposed in the article. These criteria are consistent with previously proposed definitions of preclinical ("confirmatory" or "proof-of-concept") experiments.[7,30,31] CVDs of interest included atherosclerosis or vascular homeostasis, arterial aneurysms or dissections, myocardial infarction, valvular disease, cardiomyopathy or heart failure, cardiac transplantation, pulmonary hypertension, cardiac arrhythmia, stroke, resuscitative medicine, hypertension, metabolic or endocrine diseases, and hematologic disorders (including thrombosis). Studies deemed to report on clinically relevant cardiovascular conditions but not falling into one of the prespecified categories could be included at the journal reviewers' discretion (category "other"). Each journal article was independently reviewed and data extracted using standardized case report forms by 2 reviewers, allowing for the assessments of inter-rater agreement. To permit our team to review a large volume of articles, we allowed for studies to be excluded as soon as it was clear that a manuscript violated any of our inclusion criteria, which was most often because of the model used (eg, zebrafish or humans) or because they were published in formats other than full manuscripts (eg, conference abstracts). The specific inclusion criterion/criteria that was/were deemed to have been violated for each excluded article, therefore, varied according to journal reviewer and were often multiple (eg, a nonmammalian animal model was used and no specific reference to therapeutic implications/applications was made). Discrepancies were resolved by consensus or by an independent adjudicator before building a final locked database for analysis.

Pre-specified data, including the date of publication, CVD investigated, animal model(s) used and their sex, whether animals were randomized to treatment groups, whether any blinding was implemented (concealed allocation or blinded outcome assessment), and whether a priori sample size/power estimations were performed were collected. Subgroup analyses restricted to studies of therapeutic interventions and by CVD studied were performed as were post hoc comparisons

of these practices before and after the publication of NIH guidelines and policies for reporting preclinical research and the implementation of a "Basic Science checklist" by *Stroke*,[28] which is purported to have improved the quality and designs of preclinical studies published in that journal.[32]

Finally, the number of citations has been used as a surrogate measure of research impact and influence.[33–37] Therefore, Scopus (Elsevier) was queried to identify original research articles that cited preclinical cardiovascular studies published between July 2006 and June 2011. This 5-year period was selected as it ensured that each article had ≥5 years of citation data available in June 2016. Reviews, conference papers, editorials, letters, books/book chapters, and errata were excluded. Scopus was selected because it is the largest database available for citation analysis and retrieves a greater number of citations when compared with others.[38] Prespecified analyses of citations at 60 months after the index publication were performed, stratified by individual and cumulative study design elements and adjusted for journal of publication, year of publication, and CVD studied. Given the possibility that certain publications may have garnered attention in the short-term but could have ultimately been disproven or deemed irreproducible, a sensitivity analysis of citation counts at 36 months was also undertaken as this time point approximates contemporary mean time-to-retraction.[39]

Categorical variables are reported as number (%) and were compared via χ^2 tests. Continuous variables are reported as median (interquartile range) and were compared using Wilcoxon signed-rank or Kruskal–Wallis tests. Inter-rater agreement was calculated using Cohen κ statistic and percent agreement. Temporal patterns in the proportions of studies reporting randomization, blinding, and sample size estimations were evaluated via Cochrane–Armitage trend tests and journal-specific logistic regression models adjusting for CVD studied and animal model used when the number of events per predictor variable was adequate.[40–42] For the latter adjusted analyses, backward elimination was used for model building using a criterion of $P<0.20$ for specific CVD and animal model predictor variable inclusion. Associations between study design elements and citation counts were examined via stratification and multivariable linear regression. Non-normally distributed variables were log-transformed, when required. All analyses were performed using SAS 9.4 (SAS Institute Inc, Cary, NC) using a 2-tailed α level of 0.05 to define statistical significance.

Results

Study Selection and Characteristics

As previously described,[29] of 28 636 articles screened, 3396 met inclusion criteria and were analyzed (Figure 1). Inter-rater
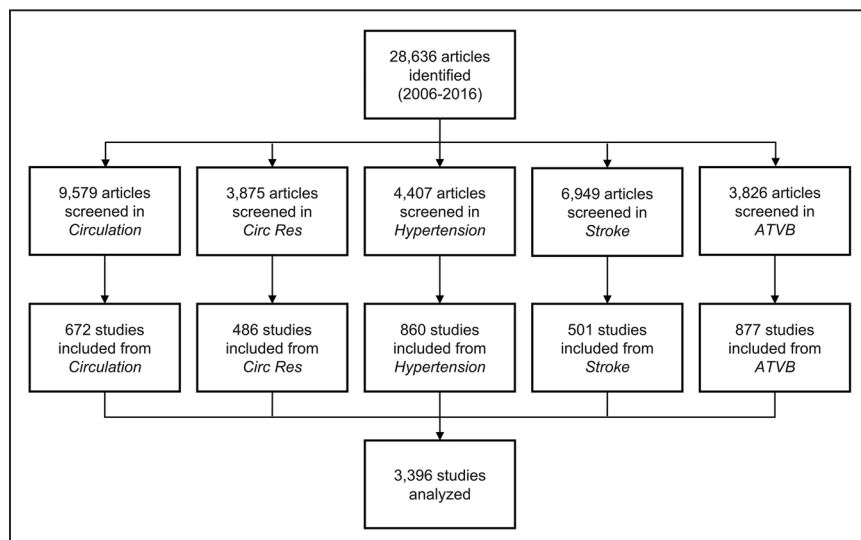
Figure 1. Literature search and results. *ATVB* indicates *Arteriosclerosis, Thrombosis and Vascular Biology*; and *Circ Res*, *Circulation Research*.

agreement for study inclusion before resolution was 94.5% (κ=0.72; 95% CI, 0.70–0.73). Atherosclerosis, hypertension, stroke, and cardiomyopathy/heart failure were the most commonly studied CVDs (range 14.2%–19.5%). Most other CVDs were the focus of <5.0% of studies. Ten studies on resuscitative medicine were identified and were included in the "other" category, which comprised 3.0% of studies. Nearly one third of studies examined a therapeutic intervention. Mice or rats were most often used by researchers (used in 89.8% of studies) whereas guinea pigs, gerbils, or hamsters were used in <0.2%. Multiple animal models were used in 2.8% of studies (Table).

Randomization

Randomization of animals was reported in 740 studies (21.8%) overall, but was more frequently noted in the subset examining therapeutic interventions (38.3% versus 15.0%, *P*<0.0001). When studies were stratified by CVD studied, significant differences in the proportions reporting randomization were noted (*P*<0.0001, range 5.6%–46.5%) with a lack of randomization predominating in all cases except stroke (Figure 2A). Significant differences in the proportions

of studies reporting randomization were also observed when stratified by animal model used (*P*<0.0001, range 14.6%–45.0%) with a lack of randomization predominating in all cases except when pigs or a combination of animal models were used (Figure 2B).

Blinding

Blinding of treatment allocation or outcome assessment was reported in 1110 studies (32.7%) overall, but was also more frequently noted in the subset examining therapeutic interventions (41.9% versus 28.9%, *P*<0.0001). When stratified by CVD studied, significant differences in the proportions of studies reporting any blinding were observed (*P*<0.0001, range 10.9%–62.6%). Studies on stroke ranked highest among CVDs although arterial aneurysms/dissections and transplantation medicine had comparable numbers of blinded and nonblinded studies. For all other CVDs, studies with blinding formed the minority (Figure 3A). No significant difference in the proportions of studies reporting blinding were seen when stratified by animal model used (*P*=0.369, range 20.9%–36.8%; Figure 3B).

Table.   Study Characteristics

	Circulation (n=672)	*Circ Res* (n=486)	*Hypertension* (n=860)	*Stroke* (n=501)	*ATVB* (n=877)	Total (n=3396)
Disease studied						
Atherosclerosis	119 (17.7)	113 (23.3)	14 (1.6)	2 (0.4)	385 (43.9)	633 (18.6)
Hypertension	34 (5.1)	35 (7.2)	577 (67.1)	1 (0.2)	14 (1.6)	661 (19.5)
Arterial aneurysm	16 (2.4)	21 (4.3)	15 (1.7)	24 (4.8)	62 (7.1)	138 (4.1)
Myocardial infarction	95 (14.1)	56 (11.5)	19 (2.2)	0 (0)	115 (13.1)	285 (8.4)
Cardiomyopathy/heart failure	173 (25.7)	154 (31.7)	135 (15.7)	0 (0)	21 (2.4)	483 (14.2)
Pulmonary hypertension	57 (8.5)	29 (6.0)	34 (4.0)	0 (0)	15 (1.7)	135 (4.0)
Arrhythmia	65 (9.7)	35 (7.2)	4 (0.5)	0 (0)	0 (0)	104 (3.1)
Valvular disease	9 (1.3)	4 (0.8)	0 (0)	0 (0)	14 (1.6)	27 (0.8)
Stroke	20 (3.0)	4 (0.8)	9 (1.1)	435 (86.8)	18 (2.1)	486 (14.3)
Transplantation medicine	28 (4.2)	15 (3.1)	0 (0)	31 (6.2)	32 (3.7)	106 (3.1)
Metabolic/endocrine disorders	19 (2.8)	10 (2.1)	23 (2.7)	2 (0.4)	75 (8.6)	129 (3.8)
Hematologic disorders	19 (2.8)	8 (1.7)	1 (0.1)	0 (0)	79 (9.0)	107 (3.2)
Other	18 (2.7)	2 (0.4)	4 (0.5)	6 (1.2)	47 (5.4)	102 (3.0)
Animal model(s) used						
Mouse	463 (68.9)	391 (80.5)	357 (41.5)	223 (44.5)	744 (84.8)	2178 (64.1)
Rat	84 (12.5)	44 (9.1)	443 (51.5)	237 (47.3)	63 (7.3)	872 (25.7)
Rabbit	10 (1.5)	15 (3.1)	16 (1.9)	18 (3.6)	23 (2.6)	82 (2.4)
Dog	37 (5.5)	12 (2.5)	13 (1.5)	5 (1.0)	0 (0)	67 (2.0)
Pig	84 (12.5)	4 (0.8)	9 (1.1)	3 (0.6)	19 (2.2)	60 (1.8)
Primate	3 (0.5)	1 (0.2)	2 (0.2)	3 (0.6)	8 (0.9)	17 (0.5)
Other	10 (1.5)	4 (0.8)	9 (1.1)	1 (0.2)	1 (0.1)	25 (0.7)
Combination	40 (6.0)	15 (3.1)	11 (1.3)	11 (2.2)	18 (2.1)	95 (2.8)
Therapeutic study	162 (24.1)	85 (17.5)	221 (25.7)	264 (52.7)	256 (29.2)	988 (29.1)

Values reported as number (%). *ATVB* indicates *Arteriosclerosis, Thrombosis, and Vascular Biology*; and *Circ Res, Circulation Research*.

Sample Size Estimation

A priori sample size estimations or power calculations were reported in 79 studies (2.3%). This aspect of study design was also more frequently reported in the subset examining therapeutic interventions (4.2% versus 1.6%, $P<0.0001$). Significant differences in the proportion of studies reporting sample size calculations were observed when stratified by CVD studied ($P<0.0001$, range 0%–9.5%) with stroke ranking highest (Figure 4A), but not when stratified by animal model used ($P=0.270$, range 0%–6.1%; Figure 4B).

Inclusion of Both Sexes

Sex bias prevalence and detailed temporal trends have been described previously.[29] After excluding studies that did not report the sex of the animals used, significant differences were noted in the proportions including animals of both sexes when stratified by CVD of interest ($P<0.0001$, range 2.4%–34.7%) and by animal model used ($P<0.0001$, range 6.7%–26.9%; Online Figure).

Temporal Patterns in Study Design Element Implementation

Over the past 10 years, there have been no significant changes in the proportions of studies reporting blinding or randomization (inclusion of both sexes has been reported previously[29]). There has been a significant increase in the proportion reporting a priori sample size estimations ($P_{trend}<0.0001$), although it remains below 7% (Figure 5). There was no substantial difference in the prevalence of any of these study design elements before and after the NIH principles and guidelines for reporting preclinical research were published in 2014 (range of differences 0.7%–3.6%). Post hoc calculations suggest that our sample sizes had ≥80% power to detect a difference of 5.0% for all study design elements.

CVD-specific temporal analyses suggest that preclinical studies on stroke are increasingly incorporating randomization, blinding, sample size estimations, and inclusion of both sexes—a pattern that is not seen in studies on atherosclerosis, hypertension, or cardiomyopathy/heart failure (the most
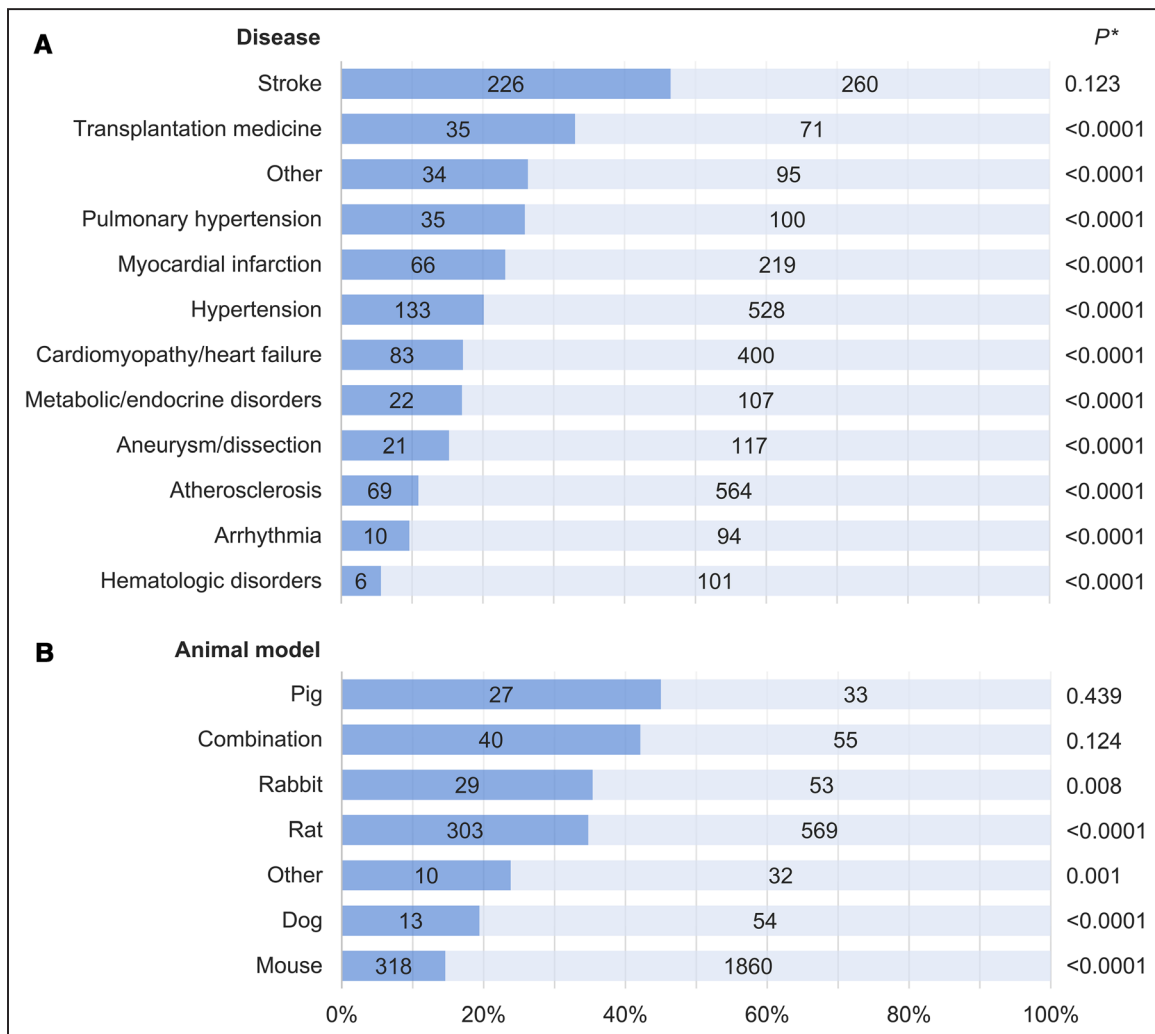
Figure 2. Randomization in preclinical cardiovascular research studies published over a 10-y period stratified by (A) disease studied and (B) species used. Dark blue corresponds to the proportion of studies implementing randomization; numbers in bars correspond to absolute numbers of studies. Valvular disease and resuscitative medicine studies were included in the "Other" category because of the small number of relevant publications. "Combination" refers to more than one animal species used within the same publication; animal models in the "Other" category included guinea pig, gerbil, and hamster. *For comparison of studies incorporating randomization vs not.
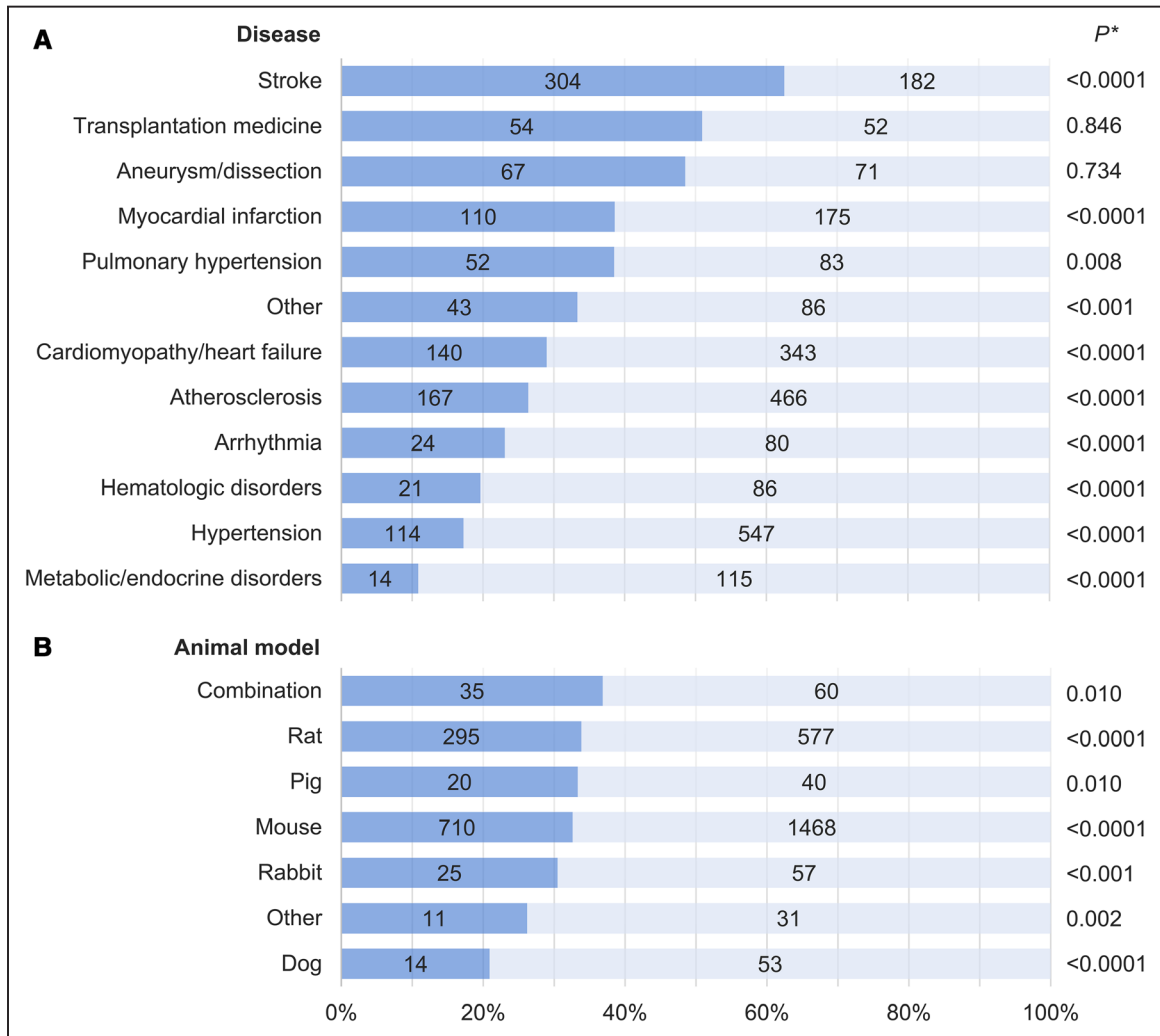
Figure 3. Blinding in preclinical cardiovascular research studies published over a 10-y period stratified by (A) disease studied and (B) species used. Dark blue corresponds to the proportion of studies implementing blinding; numbers in bars correspond to absolute numbers of studies. Valvular disease and resuscitative medicine studies were included in the "Other" category because of the small number of relevant publications. "Combination" refers to more than one animal species used within the same publication; animal models in the "Other" category included guinea pig, gerbil, and hamster. *For comparison of studies incorporating blinding vs not.

commonly studied CVDs). Studies on atherosclerosis have shown a modest increase in the proportion reporting sample size estimation in recent years, but an overall significant decrease in the proportion reporting blinding, whereas the proportion of studies on cardiomyopathy/heart failure that include both sexes has significantly decreased. No other significant changes were observed during the 10-year period (Figure 6).

Given the disproportionate amount of stroke research published in *Stroke*, which has reported improvements in the quality and design of published preclinical research following the introduction of their "Basic Science Checklist" in 2011,[32] CVD-adjusted and animal model-adjusted comparisons of study design implementation before and after this time point were performed for each journal. *Stroke* uniquely exhibited significant improvements in all measures of methodological quality (range of adjusted odds ratios 2.4–8.2, $P<0.0001$ for all study design elements; Online Table). These analyses also identified stroke as the CVD studied as an independent

positive predictor of one or more study design element in every journal.

Citations According to Index Study Methodology

At 60 months, 41 441 articles citing the 1681 preclinical studies that were published between July 2006 and June 2011 were identified. The median citation count per preclinical study was 20 (interquartile range 13–31, range 0–131). Studies that implemented randomization, blinding, or sample size estimation had similar numbers of citations as those that did not; however, studies that included both males and females were cited less frequently (median 18 versus 20, $P=0.023$). The cumulative number of these study design elements was not associated with citation counts (Figure 7). No study implemented all 4 of the above design elements and only 20 studies (1.2%) incorporated 3; therefore, studies with ≥2 were grouped together for this analysis. The above associations (or lack thereof) between study design element(s) and number of citations persisted after adjusting for journal of publication, year of publication,
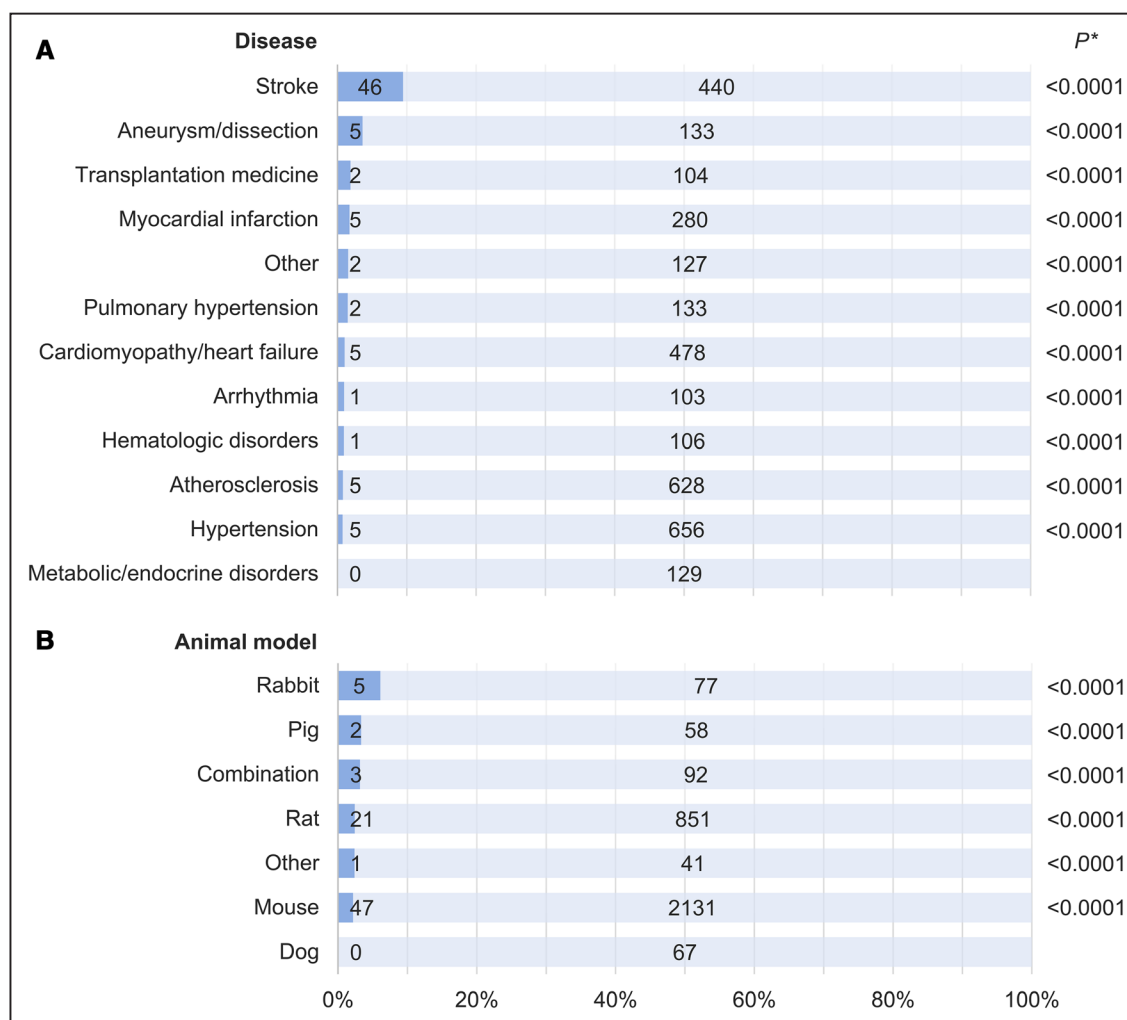
Figure 4. Sample size estimation in preclinical cardiovascular research studies published over a 10-y period stratified by (A) disease studied and (B) species used. Dark blue corresponds to the proportion of studies reporting sample size estimations/power calculations; numbers in bars correspond to absolute numbers of studies. Valvular disease and resuscitative medicine studies were included in the "Other" category because of the small number of relevant publications. "Combination" refers to more than one animal species used within the same publication; animal models in the "Other" category included guinea pig, gerbil, and hamster. *For comparison of studies incorporating sample size estimation vs not.

and CVD studied in multivariable regression models and all findings were comparable in sensitivity analyses of citation counts at 36 months.

Discussion

Preclinical cardiovascular research using animal models plays an integral role in advancing the care of patients afflicted with CVDs. However, its impact is contingent on its scientific validity, reproducibility, and relevance to human physiology and disease. It is understood that inherent limitations of using animals to model human diseases can undermine the predictive value of preclinical findings, rendering it difficult for even the most skilled scientists to make impactful discoveries.[12] However, this difficulty is compounded when methodological sources of bias are introduced. We systematically examined a continuous and large body of preclinical cardiovascular research to determine how often randomization, blinding, and a priori sample size estimation are incorporated and did so over a sufficiently long period to draw meaningful conclusions on

the temporal patterns of these practices. We report that, overall, these design elements are rarely implemented in studies published in leading peer-reviewed cardiovascular journals, paralleling the previously reported prevalence of sex bias.[29] Furthermore, apart from a modest increase in the proportion of studies reporting sample size calculations, there has been no improvement in these practices over the past decade, although stroke research may be a notable exception. Finally, analyses of citation counts suggest that crucial methodological aspects of preclinical studies may be overlooked by cardiovascular researchers, affording potentially biased research comparable influence on scientific research efforts as methodologically more robust studies.

Poorly designed preclinical studies not only contribute to experimental irreproducibility[2,7] and wasted resources[43,44] but also may result in erroneous conclusions regarding the treatment effects,[9,21,23,45,46] which can ultimately spur or deter clinical trials in humans with consequent risk of harm.[12,20,47] The experimental design elements evaluated in our study are
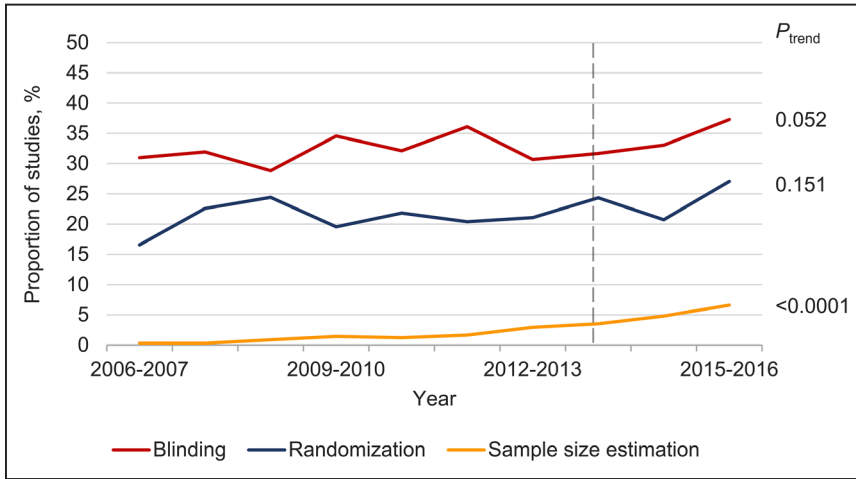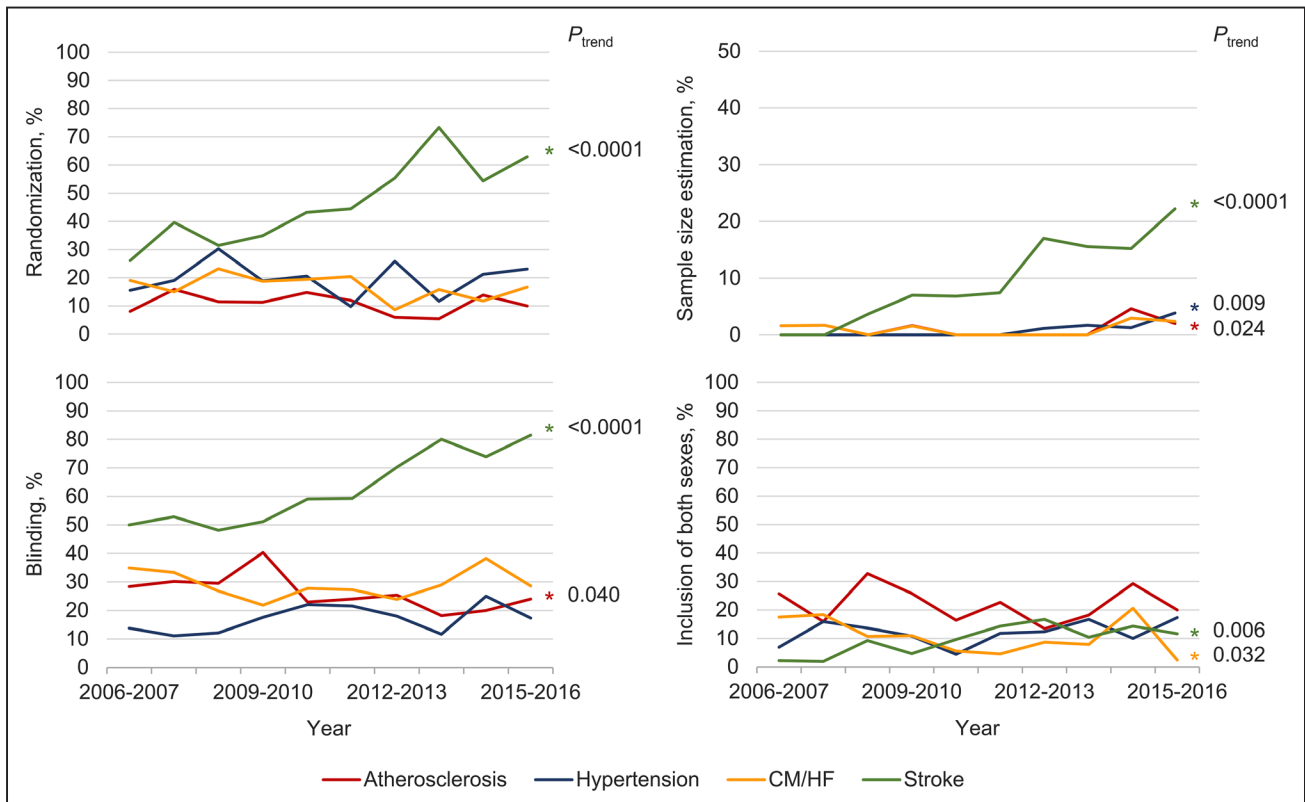
Figure 5. Temporal patterns in randomization, blinding, and sample size estimation in preclinical cardiovascular studies. Dashed line indicates the publication of the National Institutes of Health guidelines and policies for reporting preclinical research. Data for inclusion of both sexes have been reported previously.[29]

deemed crucial to improving the quality of preclinical research by addressing selection bias (randomization); minimizing performance, detection, and attrition bias (blinding); ensuring adequate statistical power and the ethical use of animals (sample size estimation); and promoting research relevance for both men and women (inclusion of both sexes).[1,6,7,10,13,26,48,49] Although these elements are routinely implemented in clinical trials as they have been shown to protect against bias and imprecision,[50–53] they have conspicuously not permeated preclinical experiments.[1,23,54–56] The high degree of experimental control that is possible in preclinical research (eg, via genetic and environmental homogeneity) can reduce variation and the sample sizes required relative to clinical trials; however,

variations in injury or disease induction and the potential for persistent and unrecognized confounders represent important sources of bias.[6,31,57]

Furthermore, despite a pervasive belief among scientists that there is a reproducibility "crisis" in research that is attributable to methodological and reporting shortcomings,[14] our citation analyses suggest that greater methodological rigor in preclinical cardiovascular research does not translate into greater scientific influence. A similar observation was noted by Amgen in a study of 53 preclinical cancer research publications: studies that they could not adequately reproduce were cited as often, if not more often, than those that they could successfully reproduce, irrespective of the journal of

Figure 6. Temporal patterns in randomization, blinding, sample size estimation, and inclusion of both sexes in preclinical research for the most commonly studied cardiovascular diseases. Note the different scale of *y*-axis for sample size estimation. CM/HF indicates cardiomyopathy/heart failure.
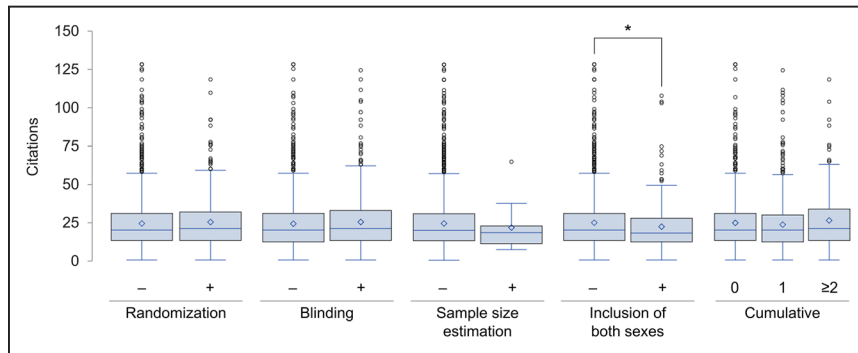
Figure 7. Box-whisker plot of the number of citations by methodological rigor of index preclinical study (n=1681). Diamond symbol plotted at mean. Outliers identified as beyond 1.5 interquartile range. Cumulative refers to sum of randomization, blinding, sample size estimation, and inclusion of both sexes (each contributing 1 point). *P<0.05.

publication's impact factor.[12] The suggestion that methodologically robust studies are given equal consideration as studies at greater risk of bias by cardiovascular researchers is troubling as it undermines the reputed "self-correcting" tenet of science and increases the risk of pursuing unfruitful avenues of research.[2,5,44,48]

Cogent analyses and arguably practical solutions to improve preclinical research quality (and secondarily to enhance research translation) have been proposed. However, evidence of changes in research practice and data on the impact of corrective actions are scarce.[13] Several guidelines and checklists have been developed to improve preclinical methodology[45,58,59] and reporting,[25,45,60–62] yet there is little indication that they are effecting change despite being widely endorsed.[43,58,63] Systematic reviews and meta-analyses have been advocated as safeguards to expose bias in preclinical research before embarking on clinical trials,[9,47,64,65] but they are limited by the internal validity of included studies[66] and are not routinely performed. Changes in research funding such as those undertaken by the NIH[2,26] may bolster these efforts; however, our analyses of animal studies in the cardiovascular sciences have not shown signs of substantial changes in research practices or reporting since they were announced.[29]

Journal editors and reviewers directly influence what is published thereby often serving as ultimate gatekeepers of research findings; yet, few journals effectively encourage the use of reporting guidelines.[67] At the time of writing, there exists considerable variation in author guidelines for reporting animal studies among leading cardiovascular journals with few requiring authors to report randomization, blinding, sample size estimation, or the sex of the animals used—requirements that could increase transparency and reinforce the importance of these study design elements.[13] Uniquely, however, *Stroke* implemented a "Basic Science Checklist" in 2011 (updated in 2016),[28] which includes all of these relevant elements, forms part of the manuscript submission process, and is evaluated by editors and reviewers. Over the 2.5 years after its introduction, Minnerup et al[32] noted improvements in randomization, blinding, and allocation concealment compared with the preceding 18 months. We noted significant improvements in the quality of preclinical stroke research, 90% of which was published in *Stroke*, raising the question of whether these improvements were due to journal editorial policies/culture or driven by changes in research practices among stroke researchers in general. Our post hoc analyses suggest that it is likely a combination of both: *Stroke* uniquely showed improvements in all

study design elements even after adjusting for CVD studied and animal model used, but stroke as the CVD studied was identified as an independent predictor for at least one study design element for every journal examined. Our findings, therefore, corroborate those of Minnerup et al[32] and expand upon them by demonstrating that these improvements have continued into 2016, extend to sample size estimation (and to a lesser extent inclusion of both sexes), contrast with an extended preceding period of suboptimal reporting, and have not been paralleled in other prominent cardiovascular journals. Therefore, although not conclusive, our data suggest that journal editors and reviewers can exert considerable influence on preclinical research practices and reporting. Yet, our findings also suggest that stroke research has independently improved in quality—a finding that may be attributable to its community's early appreciation of the importance of methodological rigor in preclinical science and extensive involvement in efforts to improve research translation.[7,20,21,46,58,68–71]

Our study is not without limitations. The study sample comprised articles from 5 cardiovascular journals over 10 years, which may not fully represent preclinical cardiovascular research publications or practices. However, we deliberately selected AHA journals given their prominence in the field of cardiovascular medicine, established reputation, collective focus on a broad range of CVDs, and unanimous endorsement of the NIH guidelines on rigor and reproducibility. Given that none of the American College of Cardiology or European Society of Cardiology journals have endorsed the guidelines so far,[25] our data may actually overestimate the methodological quality of most preclinical cardiovascular research. There is no accepted definition for preclinical studies therefore criteria were developed for this study. It is, therefore, possible that not all relevant studies were included in our analysis. However, the inclusion criteria used, the number of journals reviewed, and the substantial inter-rater agreement for study inclusion support the validity of our results and render our data the best available on methodological rigor in preclinical cardiovascular research. Furthermore, our inclusion criteria are in line with previously proposed distinctions between "exploratory" and "confirmatory" preclinical research.[7,30,31] Our analysis considered blinding as a single and dichotomous variable, which limits detailed assessments of this design element, and did not examine all relevant potential sources of bias. Data handling and publication bias, for instance, are important factors that may contribute to irreproducibility in science.[2,6,7,9–11,13,43] As well, factors other than experimental

methodology can influence the predictive value of preclinical research, including how well an animal model reflects human physiology and disease,[6,9,45,72] the appropriateness of statistical analyses,[73] and a lack of standardization of definitions and surrogate markers,[74,75] which were not examined. Although article citation is the most commonly used measure of research impact, it is an imperfect indicator.[33,34] Finally, study design elements may have been implemented but not reported, which could result in underestimates of their prevalence. However, underreporting of measures of methodological rigor is believed to be low[49,65] and therefore unlikely to have significantly influenced our results.

Our analysis of preclinical research published in leading cardiovascular journals over the past 10 years demonstrates that methodological sources of potential bias and imprecision are prevalent, have not appreciably improved over time, and may be overlooked by researchers when basing subsequent studies. Concerted efforts to address this problem are urgently needed. Stroke research has uniquely shown substantial improvement in several measures of quality in recent years, warranting a closer examination to identify drivers of its success.

Disclosures

None.

References

1. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166–175. doi: 10.1016/S0140-6736(13)62227-8.
2. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014;505:612–613.
3. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. *PLoS Med*. 2013;10:e1001489. doi: 10.1371/journal.pmed.1001489.
4. Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov*. 2011;10:428–438. doi: 10.1038/nrd3405.
5. Bunnage ME. Getting pharmaceutical R&D back on target. *Nat Chem Biol*. 2011;7:335–339. doi: 10.1038/nchembio.581.
6. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR. Can animal models of disease reliably inform human studies? *PLoS Med*. 2010;7:e1000245. doi: 10.1371/journal.pmed.1000245.
7. Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490:187–191. doi: 10.1038/nature11556.
8. Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, Kelly N, Bostrom A, Theodoss J, Al-Nakhala BM, Vieira FG, Ramasubbu J, Heywood JA. Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler*. 2008;9:4–15. doi: 10.1080/17482960701856300.
9. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ*. 2007;334:197. doi: 10.1136/bmj.39048.407928.BE.
10. Muhlhausler BS, Bloomfield FH, Gillman MW. Whole animal experiments should be more like human randomized controlled trials. *PLoS Biol*. 2013;11:e1001481. doi: 10.1371/journal.pbio.1001481.
11. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712. doi: 10.1038/nrd3439-c1.
12. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483:531–533. doi: 10.1038/483531a.
13. Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res*. 2015;116:116–126. doi: 10.1161/CIRCRESAHA.114.303819.
14. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533:452–454. doi: 10.1038/533452a.
15. Unreliable research: Trouble at the lab. *The Economist*. 2013.
16. Lehrer J. The truth wears off: Is there something wrong with the scientific method? *The New Yorker*. 2010.
17. Freedman DH. Lies, damned lies, and medical science. *The Atlantic*. 2010.
18. Osherovich L. Hedging against academic risk. *Science-Business eXchange*. 2011;4:doi:10.1038/scibx.2011.1416
19. Begley CG. Six red flags for suspect work. *Nature*. 2013;497:433–434. doi: 10.1038/497433a.
20. Horn J, de Haan RJ, Vermeulen M, Luiten PG, Limburg M. Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. *Stroke*. 2001;32:2433–2438.
21. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke*. 2008;39:2824–2829. doi: 10.1161/STROKEAHA.108.515957.
22. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124. doi: 10.1371/journal.pmed.0020124.
23. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. *JAMA*. 2006;296:1731–1732. doi: 10.1001/jama.296.14.1731.
24. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32:40–51. doi: 10.1038/nbt.2786.
25. National Institutes of Health. Principles and guidelines for reporting preclinical research. https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. Accessed December 2, 2016.
26. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. *Nature*. 2014;509:282–283.
27. Daugherty A, Hegele RA, Mackman N, Rader DJ, Schmidt AM, Weber C. Complying with the national institutes of health guidelines and principles for rigor and reproducibility: refutations. *Arterioscler Thromb Vasc Biol*. 2016;36:1303–1304. doi: 10.1161/ATVBAHA.116.307906.
28. Vahidy F, Schäbitz WR, Fisher M, Aronowski J. Reporting standards for preclinical studies of stroke therapy. *Stroke*. 2016;47:2435–2438. doi: 10.1161/STROKEAHA.116.013643.
29. Ramirez FD, Motazedian P, Jung RG, Di Santo P, MacDonald Z, Simard T, Clancy AA, Russo JJ, Welch V, Wells GA, Hibbert B. Sex bias is increasingly prevalent in preclinical cardiovascular research: implications for translational medicine and health equity for women: a systematic assessment of leading cardiovascular journals over a 10-year period. *Circulation*. 2017;135:625–626. doi: 10.1161/CIRCULATIONAHA.116.026668.
30. Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. *PLoS Biol*. 2014;12:e1001863. doi: 10.1371/journal.pbio.1001863.
31. Unger EF. All is not well in the world of translational research. *J Am Coll Cardiol*. 2007;50:738–740. doi: 10.1016/j.jacc.2007.04.067.
32. Minnerup J, Zentsch V, Schmidt A, Fisher M, Schäbitz WR. Methodological quality of experimental stroke studies published in the stroke journal: time trends and effect of the basic science checklist. *Stroke*. 2016;47:267–272. doi: 10.1161/STROKEAHA.115.011695.
33. Ranasinghe I, Shojaee A, Bikdeli B, Gupta A, Chen R, Ross JS, Masoudi FA, Spertus JA, Nallamothu BK, Krumholz HM. Poorly cited articles in peer-reviewed cardiovascular journals from 1997 to 2007: analysis of 5-year citation rates. *Circulation*. 2015;131:1755–1762. doi: 10.1161/CIRCULATIONAHA.114.015080.
34. Danthi N, Wu CO, Shi P, Lauer M. Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants. *Circ Res*. 2014;114:600–606. doi: 10.1161/CIRCRESAHA.114.302656.
35. Kaltman JR, Evans FJ, Danthi NS, Wu CO, DiMichele DM, Lauer MS. Prior publication productivity, grant percentile ranking, and topic-normalized citation impact of NHLBI cardiovascular R01 grants. *Circ Res*. 2014;115:617–624. doi: 10.1161/CIRCRESAHA.115.304766.
36. Lauer MS, Danthi NS, Kaltman J, Wu C. Predicting productivity returns on investment: thirty years of peer review, grant funding, and publication of highly cited papers at the National Heart, Lung, and Blood Institute. *Circ Res*. 2015;117:239–243. doi: 10.1161/CIRCRESAHA.115.306830.
37. Fabsitz RR, Papanicolaou GJ, Sholinsky P, Coady SA, Jaquish CE, Nelson CR, Olson JL, Puggal MA, Purkiser KL, Srinivas PR, Wei GS, Wolz M, Sorlie PD. Impact of National Heart, Lung, and Blood Institute-Supported Cardiovascular Epidemiology Research, 1998 to 2012. *Circulation*. 2015;132:2028–2033. doi: 10.1161/CIRCULATIONAHA.114.014147.

38. Kulkarni AV, Aziz B, Shams I, Busse JW. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA*. 2009;302:1092–1096. doi: 10.1001/jama.2009.1307.

39. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS One*. 2013;8:e68397. doi: 10.1371/journal.pone.0068397.

40. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol*. 1995;48:1495–1501.

41. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510.

42. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–1379.

43. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383:267–276. doi: 10.1016/S0140-6736(13)62228-X.

44. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol*. 2015;13:e1002165. doi: 10.1371/journal.pbio.1002165.

45. Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke*. 2004;35:1203–1208. doi: 10.1161/01.STR.0000125719.25853.20.

46. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. *J Cereb Blood Flow Metab*. 2005;25:713–721. doi: 10.1038/sj.jcbfm.9600064.

47. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I; Reviewing Animal Trials Systematically (RATS) Group. Where is the evidence that animal research benefits humans? *BMJ*. 2004;328:514–517. doi: 10.1136/bmj.328.7438.514.

48. Clayton JA. Studying both sexes: a guiding principle for biomedicine. *FASEB J*. 2016;30:519–524. doi: 10.1096/fj.15-279554.

49. Macleod MR, Lawson McLean A, Kyriakopoulou A, et al. Risk of bias in reports of in vivo research: a focus for improvement. *PLoS Biol*. 2015;13:e1002273. doi: 10.1371/journal.pbio.1002273.

50. Savović J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med*. 2012;157:429–438. doi: 10.7326/0003-4819-157-6-201209180-00537.

51. Tzoulaki I, Siontis KC, Ioannidis JP. Prognostic effect size of cardiovascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study. *BMJ*. 2011;343:d6829.

52. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061–1066.

53. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174:469–476. doi: 10.1503/cmaj.050090.

54. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One*. 2009;4:e7824. doi: 10.1371/journal.pone.0007824.

55. Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. *Neurosci Biobehav Rev*. 2011;35:565–572. doi: 10.1016/j.neubiorev.2010.07.002.

56. Yoon DY, Mansukhani NA, Stubbs VC, Helenowski IB, Woodruff TK, Kibbe MR. Sex bias exists in basic science and translational surgical research. *Surgery*. 2014;156:508–516. doi: 10.1016/j.surg.2014.07.001.

57. Reardon S. A mouse's house may ruin experiments. *Nature*. 2016;530:264. doi: 10.1038/nature.2016.19335.

58. Fisher M, Feuerstein G, Howells DW, Hurn PD, Kent TA, Savitz SI, Lo EH; STAIR Group. Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke*. 2009;40:2244–2250. doi: 10.1161/STROKEAHA.108.541128.

59. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol*. 2014;14:43. doi: 10.1186/1471-2288-14-43.

60. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010;8:e1000412. doi: 10.1371/journal.pbio.1000412.

61. Announcement: Reducing our irreproducibility. *Nature*. 2013;496

62. Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. *Altern Lab Anim*. 2010;38:167–182.

63. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for preclinical animal studies. *PLoS Biol*. 2014;12:e1001756. doi: 10.1371/journal.pbio.1001756.

64. Sandercock P, Roberts I. Systematic reviews of animal experiments. *Lancet*. 2002;360:586. doi: 10.1016/S0140-6736(02)09812-4.

65. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, Heneghan C. The need for randomization in animal trials: an overview of systematic reviews. *PLoS One*. 2014;9:e98856. doi: 10.1371/journal.pone.0098856.

66. van Luijk J, Bakker B, Rovers MM, Ritskes-Hoitinga M, de Vries RB, Leenaars M. Systematic reviews of animal studies; missing link in translational research? *PLoS One*. 2014;9:e89981. doi: 10.1371/journal.pone.0089981.

67. Hirst A, Altman DG. Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. *PLoS One*. 2012;7:e35621. doi: 10.1371/journal.pone.0035621.

68. Stroke Therapy Academic Industry Roundtable (STAIR). Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*. 1999;30:2752–2758

69. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. *Ann Neurol*. 2006;59:467–477. doi: 10.1002/ana.20741.

70. Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, Macleod M, Dirnagl U. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke*. 2008;39:929–934. doi: 10.1161/STROKEAHA.107.498725.

71. 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;30:433–439. doi: 10.1016/j.tins.2007.06.009.

72. Perrin S. Preclinical research: Make mouse studies work. *Nature*. 2014;507:423–425. doi: 10.1038/507423a.

73. García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. *BMC Med Res Methodol*. 2004;4:13. doi: 10.1186/1471-2288-4-13.

74. Johnson VE. Revised standards for statistical evidence. *Proc Natl Acad Sci U S A*. 2013;110:19313–19317. doi: 10.1073/pnas.1313476110.

75. Ioannidis JP. How to make more published research true. *PLoS Med*. 2014;11:e1001747. doi: 10.1371/journal.pmed.1001747.