OPEN

Journal Initiatives to Enhance Preclinical Research

Analyses of *Stroke, Nature Medicine, Science Translational Medicine*

F. Daniel Ramirez, MD, MSc; Richard G. Jung, BSc; Pouya Motazedian, MD; Dylan Perry-Nguyen, BMSc; Pietro Di Santo, MD; Zachary MacDonald, MD; Aisling A. Clancy, MD, MSc; Alisha Labinaz; Steven Promislow, MD; Trevor Simard, MD; Steeve Provencher, MD, MSc; Sébastien Bonnet, PhD; Ian D. Graham, PhD; George A. Wells PhD; Benjamin Hibbert, MD, PhD

Background and Purpose—Preclinical research using animals often informs clinical trials. However, its value is dependent on its scientific validity and reproducibility, which are, in turn, dependent on rigorous study design and reporting. In 2011, *Stroke* introduced a Basic Science Checklist to enhance the reporting and methodology of its preclinical studies. Except for *Nature* and *Science* journals, few others have implemented similar initiatives. We sought to estimate the impact of these journal interventions on the quality of their published reports.

Methods—All articles published in *Stroke*, *Nature Medicine*, and *Science Translational Medicine* over 9 to 18 years and in 2 control journals without analogous interventions over a corresponding 11.5 years were reviewed to identify reports of experiments in nonhuman mammals with proposed clinical relevance. The effect of journal interventions on the reporting and use of key study design elements was estimated via interrupted time-series analyses.

Results—Of 33 009 articles screened, 4162 studies met inclusion criteria. In the 3.5 to 12 years preceding each journal's intervention, the proportions of studies reporting and using key study design elements were stable except for blinding in *Stroke* and randomization in *Science Translational Medicine*, which were both increasing. Post-intervention, abrupt and often marked increases were seen in the reporting of randomization status (level change: +17% to +44%, $P \leq 0.005$), blinding (level change: +20% to +40%, $P \leq 0.008$), and sample size estimation (level change: 0% to +40%, $P \leq 0.002$ in 2 journals). Significant but more modest improvements in the use of these study design elements were also observed. These improvements were not seen in control journals.

Conclusions—Journal interventions such as *Stroke*'s author submission checklist can meaningfully improve the quality of published preclinical research and should be considered to enhance study transparency and design. However, such interventions are alone insufficient to fully address widespread shortcomings in preclinical research practices.

Visual Overview—An online visual overview is available for this article. **(*Stroke*. 2020;51:291-299. DOI: 10.1161/STROKEAHA.119.026564.)**

Key Words: animal model ■ bias ■ biomedical research ■ research design ■ sample size

See related article, p 6

Preclinical experiments using animal models are an integral part of biomedical research. However, their value is dependent on their scientific validity and reproducibility,[1,2] which are, in turn, dependent on rigorous study design and reporting.[3,4] Poorly designed or inadequately reported preclinical studies can, therefore, reduce the dividends from research investments,[5,6] result in animals being unjustifiably subjected to harm,[2,7,8] and inappropriately spur or deter clinical trials in humans.[3,9,10] Randomization, blinding, sample size estimation,

Received June 2, 2019; final revision received August 8, 2019; accepted August 28, 2019.

From the Division of Cardiology (F.D.R., P.D.S., S. Promislow, T.S., B.H.), CAPITAL Research Group (F.D.R., R.G.J., P.M., D.P.-N., P.D.S., A.L., S. Promislow, T.S., B.H.), Vascular Biology and Experimental Medicine Laboratory (R.G.J., P.M., D.P.-N., A.L., T.S., B.H.), and Cardiovascular Research Methods Centre (G.A.W.), University of Ottawa Heart Institute, ON, Canada; School of Epidemiology and Public Health (F.D.R., P.D.S., I.D.G., G.A.W.), Department of Cellular and Molecular Medicine (R.G.J., T.S., B.H.), Faculty of Medicine (R.G.J., P.M., D.P.-N., Z.M.), Department of Emergency Medicine (Z.M.), Department of Obstetrics and Gynecology (A.A.C.), and Faculty of Science (A.L.), University of Ottawa, ON, Canada; Electrophysiology and Ablation Unit, Hôpital cardiologique du Haut-Lévêque, Centre hospitalier universitaire (CHU) de Bordeaux, Bordeaux-Pessac, France (F.D.R.); L'Institut de rythmologie et modélisation cardiaque (LIRYC), Université de Bordeaux, Bordeaux-Pessac, France (F.D.R.); Department of Medicine, University of Calgary Cumming School of Medicine, AB, Canada (P.M.); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA (A.A.C.); Pulmonary Hypertension Research Group, Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec (S. Provencher, S.B.) and Department of Medicine (S. Provencher, S.B.), Université Laval, Québec City, Canada; and Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada (I.D.G.).

The online-only Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.119.026564.

Correspondence to Benjamin Hibbert, MD, PhD, University of Ottawa Heart Institute, 40 Ruskin St, H-4238, Ottawa, Ontario, Canada K1Y 4W7. Email bhibbert@ottawaheart.ca

and considering sex as a biological variable are considered essential study design elements (SDEs) to improve the reproducibility and predictive value of preclinical experiments.[1,8,11–14] Randomization protects against selection bias; blinding protects against performance, detection, and attrition biases; sample size estimation ensures adequate statistical power, which improves effect estimate precision and protects against spurious conclusions; and considering sex as a biological variable increases the likelihood that results will be relevant to both men and women.[2,5,7,14–16] These SDEs are routinely used in clinical trials but are rarely implemented at the preclinical stage, even though animal and human studies are susceptible to similar risks of bias and threats to study validity.[8,12,14,15]

Recently, the National Institutes of Health and prominent journals, including *Nature* and *Science* publishing groups, have focused considerable effort on improving the rigor and reproducibility of preclinical research, particularly through enhanced experimental conduct and reporting standards.[1,11,13,17] In May 2013, *Nature* journals introduced editorial measures that included a mandatory checklist to prompt authors to disclose information regarding crucial experimental SDEs[18]—a similar initiative to the Basic Science Checklist introduced by *Stroke* in 2011.[19] *Science Translational Medicine* (*Science TM*) similarly began requiring that authors describe and defend their study designs, including disclosing whether key SDEs were implemented.[20] Yet, robust data on the impact of such interventions are limited and at times contradictory.[12,19,21–23] We hypothesized that the introduction of targeted journal initiatives had improved the reporting and experimental design of published preclinical research.

Methods

The data that support the findings of this study are available from the corresponding author on reasonable request.

Study Screening and Selection

All studies published in *Stroke* between January 2000 and December 2017 were reviewed. The start date of January 2000 was selected as the first Stroke Academic Industry Roundtable guidelines on preclinical experiments were published exclusively in *Stroke* in December 1999,[24] arguably marking an editorial awareness of their goals and importance. The screening and study selection process for this journal has been described previously.[12] Briefly, studies were included if they described original research published as full articles; 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 disorder in humans. Studies on physiological or genetic characteristics were included if potential therapeutic applications or implications were proposed in the article. These criteria were used in an effort to enrich our study sample with confirmatory or proof-of-concept experiments[2,25,26] while minimizing the inclusion of reports of exploratory experiments with as yet unclear relevance to human disorders. All articles in *Stroke* were independently selected and data extracted using standardized case report forms by 2 reviewers, including via review of all articles' supplementary material. Discrepancies were resolved by consensus or by an independent adjudicator. Our group has reported 94.5% interrater agreement for study inclusion before resolution using this approach in a previous review of >28 000 articles.[12]

All studies published in *Nature Medicine* and *Science TM* between 2009 and 2017 were similarly reviewed, but the study selection process differed slightly for these journals as described in methods in the online-only Data Supplement. Data extraction was performed by 2 independent assessors as was done for *Stroke*. The starting year of 2009 for data collection was selected because the inaugural issue of *Science TM* was published in that year.

We previously examined the reported methodological quality of articles published in an additional 4 leading cardiovascular journals between July 2006 and June 2016 using before-after comparisons.[12] Two of these journals (*Circulation* and *Circulation Research*) publish on all aspects of cardiovascular medicine, allocate a substantial amount of space to preclinical research that is of similar reported methodological quality,[12] have among the highest journal impact factors in the cardiovascular field, and did not explicitly and publicly implement initiatives to improve the reporting or methodology of their articles until October 2017.[17] Limited journal- and disease-specific data from those before-after comparisons have been reported previously.[12,14,27] The data set from these journals was extended to December 2017, expanded in scope (see below), and combined into a single control group to increase statistical power. Study selection and data extraction for these control journals were performed as was done for *Stroke*.

Diseases or disease systems of interest in each journal are listed in methods in the online-only Data Supplement. Studies on sex-specific diseases (for instance, prostate or ovarian cancer) were prospectively identified during data collection.

Data Extraction

The date of publication, disease system studied, and animal model(s) used were collected for all articles. Disease systems studied and animal models used in <2% of reports in all journals were included in "other" categories. Data on randomization, blinding (concealed allocation or blinded outcome assessment), sample size/power estimations, and the sex of the animals used were also collected. These data have been identified as minimum reporting requirements by the National Institutes of Health.[2,13] A distinction was made between the reporting of any given SDE and its reported use such that cases in which studies reported not randomizing animals to treatment groups or *not* using blinding, for instance, were captured. Similarly, studies that reported using a specific SDE, but its description revealed that it had not been properly executed, were captured (eg, studies that reported being adequately powered because group comparisons yielded statistically significant results). The reporting and use of SDEs are, therefore, described separately. These elements were deemed to have been reported on or used if done for any experiment or animal model within an article. Because of the small number of *Science TM* studies published in 2009 (the first issue was published in October of that year), studies from 2009 to 2010 were combined for that journal for all analyses. Similarly, studies published in control journals between July and December 2006 were included with articles published in 2007.

Given the volume of articles screened, journal-specific stylistic features (page layouts, article organization, color schemes, and appearance of figures), and the variability and volume of supplementary materials, de-identification of articles was not performed and reviewers were not blinded to journal or date of publication during screening or data extraction, although they were blinded to scores from other reviewers. Articles were also reviewed in a predominantly chronological sequence rather than in a randomized fashion.

Statistical Analysis

Journal-specific sample size calculations are described in methods in the online-only Data Supplement. Categorical variables are reported as number (%) and were compared via χ^2 tests. Continuous variables are reported as mean±SD and were compared using *t* tests. Interrater agreement for study inclusion for *Stroke* and control journals was calculated using percent agreement and Cohen's κ statistic. Secular trends in the number of preclinical studies published over time were examined using Spearman rank correlation coefficients (ρ); trends in the proportions of studies reporting on or using SDEs were examined using Cochrane-Armitage trend tests. Randomization, blinding, and sample size estimation were specifically targeted in all initiatives studied; therefore, temporal changes in the proportions of studies

reporting on and using these SDEs were evaluated as journal-specific interrupted time series with editorial initiatives introduced as step functions in segmented regression models and are reported as absolute cumulative means or absolute percentages with 95% CIs. Data on sex reporting and sex of animals used were similarly but separately evaluated and excluded studies on sex-specific disorders. For temporal analyses, *Stroke* was examined in 12-month intervals whereas *Nature Medicine*, *Science TM*, and control journals were examined in 6-month intervals to render sample sizes similar per time point between journals. The influence of journal initiatives was also estimated by comparing reporting and methodological quality measures before versus after their implementation in each journal and via simple and multivariable modified Poisson regression models with robust error variance (sandwich estimation) to generate journal-specific crude and adjusted relative risks (RRs) with 95% CIs. Covariate category inclusion in multivariable models was prespecified and included relevant disease system studied and animal model used. Within these covariate categories, specific disease system and animal model covariates were selected via backward elimination using an inclusion criterion of $P<0.20$ to ensure adequate events per predictor variable. 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

Of 33 009 indexed articles, 4162 met inclusion criteria and were analyzed (Figure I in the online-only Data Supplement). Interrater agreement for study inclusion before resolution was 97.9% (κ=0.82). The number of relevant articles published yearly in *Nature Medicine* and in control journals remained stable over the period examined (105±10, ρ=0.234, P=0.544 and 118±20, ρ=−0.378, P=0.252), but increased and decreased in *Science TM* (108±28, ρ=0.976, $P<0.001$), and *Stroke* (55±13, ρ=−0.634, P=0.005), respectively (Figure II in the online-only Data Supplement).

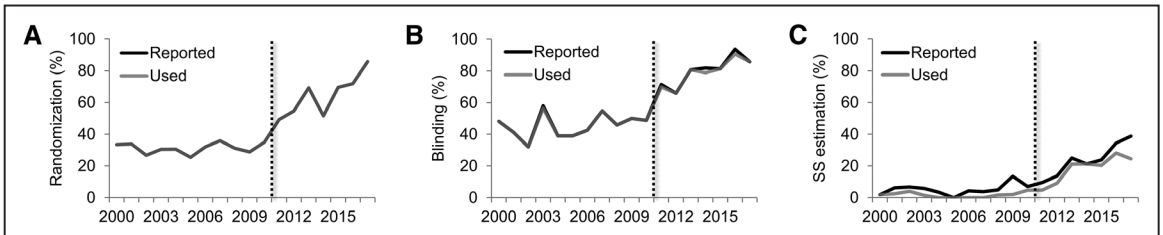
In *Stroke* and in control journals, ≥93.0% of studies focused on cardiovascular diseases and ≥85.5% used mice or rats. In *Nature Medicine* and *Science TM*, the most commonly studied disease systems were oncology, infectious diseases, neurology, endocrinology, cardiovascular, and "other," with 71.3% and 85.6% using mice in each journal, respectively. Psychiatry, anesthesia, obstetrics/gynecology, ophthalmology, and toxicology were each studied in <2% of studies in all journals and were therefore included in the "other" category (Table I in the online-only Data Supplement). One hundred eight articles were deemed to be sex-specific and were excluded from sex-based analyses.

Stroke

In the 12 years leading up to and including the year in which *Stroke* introduced its Basic Science Checklist, the proportions of studies that reported on and used blinding were increasing (P_{trend}=0.002 for both) whereas the reporting or use of randomization and sample size estimations were stable. Beginning in 2012, significant trend level or slope increases in the reporting and use of these SDEs were observed (Figure 1). No study reported not having randomized animals, and 2 studies reported not having used blinding after the checklist was introduced (Table). Among studies that reported having randomized animals, the proportion describing the method of randomization increased from 14/233 (6.0%) to 30/183 (16.4%; $P<0.001$). Temporal increases in the cumulative numbers of reported and used SDEs were observed (Figure III in the online-only Data Supplement).

Nature Medicine

In the 4.5 years preceding *Nature Medicine*'s author submission checklist, the reporting of randomization, blinding status, or sample size estimation remained stable. After the journal implemented its checklist, significant improvements in the trend lines for the reporting and use of randomization, blinding, and sample size estimation were observed. Large numbers of articles reported not having implemented various SDEs as shown by diverging trend lines of the reporting and use of individual SDEs after the initiative (Figure 2) and as shown in the Table. Among studies that reportedly randomized animals, the proportion describing the method used for randomization did not appreciably change following checklist implementation (1/50 [2.0%] versus 14/245 [5.7%], P=0.480). Corresponding temporal increases in the cumulative number



Panel	SDE	Δ Level (95% CI)	P	Δ Slope (95% CI)	P
A	Randomization (reported)	+17% (6, 28)	0.005	+5% (3, 8)	<0.001
	Randomization (used)	+17% (6, 28)	0.005	+5% (3, 8)	<0.001
B	Blinding (reported)	+20% (6, 34)	0.008	+3% (0, 6)	0.073
	Blinding (used)	+19% (6, 32)	0.008	+3% (0, 6)	0.063
C	SS estimation (reported)	+2% (-5, 9)	0.488	+4% (3, 6)	<0.001
	SS estimation (used)	+6% (1, 12)	0.034	+3% (2, 5)	<0.001

Figure 1. Reporting and use of individual study design elements (SDEs) in *Stroke*. **A**, Randomization, (**B**) blinding, and (**C**) sample size (SS) estimation. Vertical interrupted line denotes the journal intervention. Δ Level/Slope indicates change in level/slope.

Table. Comparison of Study Design Element Reporting and Use in Preclinical Studies Before and After the Implementation of Relevant Initiatives in *Stroke*, *Nature Medicine*, *Science Translational Medicine*, and Control Journals

	Before, n (%)	After, n (%)	Crude RR, (95% CI)	*P* Value	Adjusted RR, (95% CI)*	*P* Value*
Stroke	N=715	N=269				
Study design element reporting						
Randomization	233 (32.6)	183 (68.0)	2.1 (1.8–2.4)	<0.0001	2.2 (2.0–2.6)	<0.0001
Blinding	338 (47.3)	218 (81.0)	1.7 (1.6–1.9)	<0.0001	1.7 (1.6–1.9)	<0.0001
Sample size estimation	40 (5.6)	70 (26.0)	4.7 (3.2–6.7)	<0.0001	5.5 (3.8–8.1)	<0.0001
Sex of animals used	601 (84.1)	246 (91.5)	1.1 (1.0–1.1)	<0.001	1.1 (1.0–1.1)	0.008
Study design element use						
Randomization	233 (32.6)	183 (68.0)	2.1 (1.8–2.4)	<0.0001	2.2 (2.0–2.6)	<0.0001
Blinding	336 (47.0)	216 (80.3)	1.7 (1.5–1.9)	<0.0001	1.7 (1.6–1.9)	<0.0001
Sample size estimation	14 (2.0)	55 (20.5)	10.4 (5.9–18.5)	<0.0001	14.1 (7.8–25.5)	<0.0001
Inclusion of both sexes	50 (7.0)	26 (9.7)	1.1 (1.0–1.2)	0.044	1.1 (1.0–1.2)	0.090
Nature Medicine	N=476	N=473				
Study design element reporting						
Randomization	50 (10.5)	356 (75.3)	7.2 (5.5–9.4)	<0.0001	7.1 (5.4–9.3)	<0.0001
Blinding	112 (23.5)	390 (82.5)	3.5 (3.0–4.1)	<0.0001	3.5 (3.0–4.1)	<0.0001
Sample size estimation	6 (1.3)	334 (70.6)	56.0 (25.2–124.3)	<0.0001	NR	
Sex of animals used	311 (68.5)†	414 (91.2)†	1.3 (1.2–1.4)	<0.0001	1.3 (1.2–1.4)	<0.0001
Study design element use						
Randomization	50 (10.5)	245 (51.8)	4.9 (3.7–6.5)	<0.0001	4.8 (3.7–6.4)	<0.0001
Blinding	112 (23.5)	226 (47.8)	2.0 (1.7–2.4)	<0.0001	2.0 (1.7–2.4)	<0.0001
Sample size estimation	1 (0.2)	145 (30.7)	145.9 (20.5–1038.6)	<0.0001	NR	
Inclusion of both sexes	98 (21.6)†	186 (41.0)†	1.5 (1.4–1.6)	<0.0001	1.5 (1.3–1.6)	<0.0001
Science Translational Medicine	N=290	N=574				
Study design element reporting						
Randomization	54 (18.6)	375 (65.3)	3.5 (2.7–4.5)	<0.0001	3.5 (2.7–4.4)	<0.0001
Blinding	88 (30.3)	404 (70.4)	2.3 (1.9–2.8)	<0.0001	2.3 (2.0–2.8)	<0.0001
Sample size estimation	14 (4.8)	274 (47.7)	9.9 (5.9–16.6)	<0.0001	NR	
Sex of animals used	175 (64.8)‡	370 (69.0)‡	1.1 (1.0–1.2)	0.238	1.1 (1.0–1.2)	0.284
Study design element use						
Randomization	53 (18.3)	335 (58.4)	3.2 (2.5–4.1)	<0.0001	3.2 (2.5–4.1)	<0.0001
Blinding	85 (29.3)	280 (48.8)	1.7 (1.4–2.0)	<0.0001	1.7 (1.4–2.1)	<0.0001
Sample size estimation	4 (1.4)	146 (25.4)	18.4 (6.9–49.3)	<0.0001	NR	
Inclusion of both sexes	47 (17.4)‡	121 (22.6)‡	1.2 (1.0–1.3)	0.022	1.1 (1.0–1.3)	0.035
Control journals	*N*=889§	*N*=476§				
Study design element reporting						
Randomization	169 (19.0)	93 (19.5)	1.0 (0.8–1.3)	0.813	1.1 (0.8–1.3)	0.614
Blinding	308 (34.7)	161 (33.8)	1.0 (0.8–1.1)	0.761	1.0 (0.8–1.1)	0.839
Sample size estimation	31 (3.5)	42 (8.8)	2.5 (1.6–4.0)	<0.0001	2.7 (1.7–4.1)	<0.0001
Sex of animals used	626 (70.9)‖	361 (76.3)‖	1.1 (1.0–1.1)	0.028	1.1 (1.0–1.1)	0.041
Study design element use						
Randomization	164 (18.5)	91 (19.1)	1.0 (0.8–1.3)	0.762	1.1 (0.9–1.3)	0.575

(*Continued*)

Table. Continued

	Before, n (%)	After, n (%)	Crude RR, (95% CI)	*P* Value	Adjusted RR, (95% CI)*	*P* Value*
Blinding	307 (34.5)	159 (33.4)	1.0 (0.8–1.1)	0.676	1.0 (0.8–1.1)	0.765
Sample size estimation	8 (0.9)	20 (4.2)	4.7 (2.1–10.5)	<0.001	NR	
Inclusion of both sexes	143 (16.2)‖	76 (16.1)‖	1.0 (0.8–1.3)	0.952	0.9 (0.7–1.2)	0.613

NR indicates not reported due to small number of events per predictor variable; and RR, relative risk.

*Adjusted for most relevant disease system studied and animal model used.

†N=454 and 454 before and after, respectively (excluding studies on sex-specific diseases).

‡N=270 and 536 before and after, respectively (excluding studies on sex-specific diseases).

§Before and after the introduction *Nature Medicine*/*Science TM*'s initiatives. All effect estimates were similar when analyzed as before versus after the introduction of *Stroke*'s initiative (data not shown).

‖N=883 and 473 before and after, respectively (excluding studies on sex-specific diseases).

of reported and used SDEs per article were detected (Figure III in the online-only Data Supplement).

Science Translational Medicine

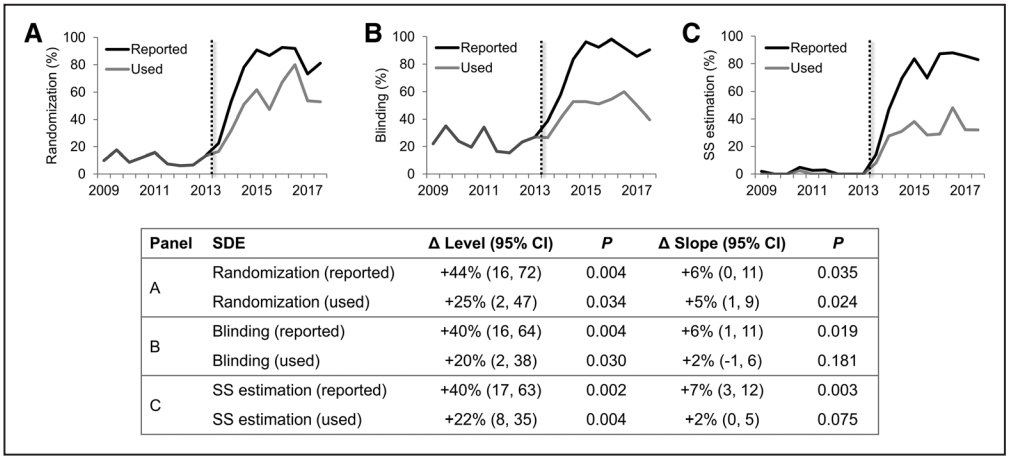
In the 3.5 years before *Science TM* revised its reporting requirements and policies, both the reporting and use of randomization were increasing (P_{trend}=0.004 and 0.007, respectively) whereas the reporting and use of blinding and sample size estimation were stable. After the introduction of this journal's initiative, significant improvements were also observed in the reporting of randomization, blinding status, and sample size estimations and in the use of randomization and sample size estimations. As was seen in *Nature Medicine*, a substantial number of studies reported not having used specific SDEs, particularly blinding and power calculations (Figure 3, Table), but among studies that reported using randomization there was no change in the proportion describing how it was performed (0/53 [0%] versus 16/335 [4.8%], *P*=0.143). Temporal increases in the cumulative numbers of reported and used SDEs were also detected (Figure III in the online-only Data Supplement).

Control Journals

In contrast, in control journals, there were no temporal improvements in the reporting or use of randomization, blinding, or sample size calculations over a corresponding 11.5 year period (Figure 4), excepting a modest improvement in the reporting of sample size calculations beginning in 2012 when analyzed using *Stroke*'s initiative as the time series interruption (1% absolute trend line slope increase [95% CI, 0%–2%], *P*=0.049; data not shown). No changes in the cumulative numbers of reported or implemented SDEs were detected in these control journals (Figure III in the online-only Data Supplement).

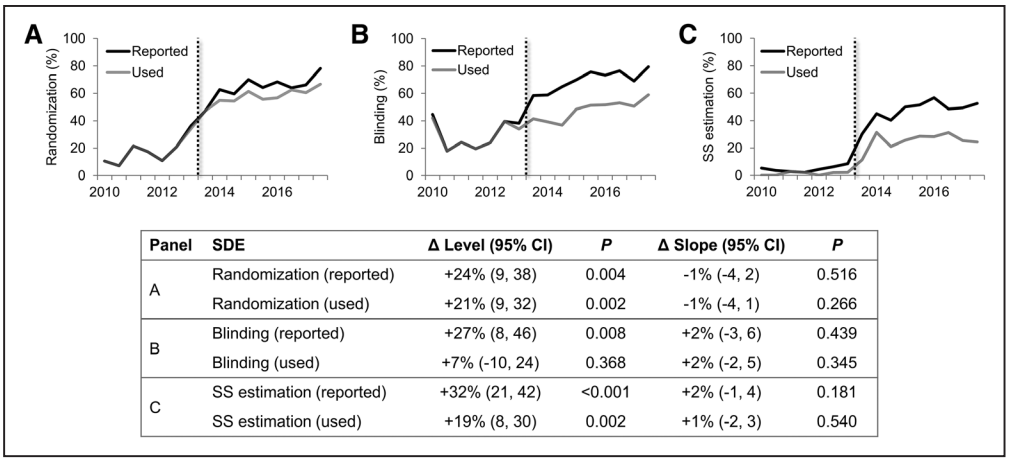
Sex Reporting and Inclusion

The proportions of articles published in *Stroke* that reported the sex of the animals used and included both males and females in experiments were stable before its checklist was introduced and did not improve afterwards. The proportions of studies in *Nature Medicine* and *Science TM* that reported the sex of the animals used and included animals of both sexes were increasing before their respective interventions (P_{trend} ≤0.040 for both in both journals), but also did not improve after the journals' initiatives. In control journals, before *Nature Medicine*, *Science TM*, or *Stroke*'s interventions, the proportions of articles reporting the sex of animals used were stable and the proportions reporting the use of both males and females were decreasing (P_{trend} <0.001). A trend level increase (+10%, [95% CI, 0%–20%] *P*=0.050) and slope increase (+2%, [95% CI, 0%–3%] *P*=0.014) in the use of both sexes were observed in these journals when analyzed using *Stroke*'s initiative as the time series interruption, but not when analyzed using *Nature Medicine* or *Science TM*'s initiatives. A marked and persistent preferential use of male animals was evident in articles in *Stroke* and control journals (Figure 5).



Panel	SDE	Δ Level (95% CI)	*P*	Δ Slope (95% CI)	*P*
A	Randomization (reported)	+44% (16, 72)	0.004	+6% (0, 11)	0.035
	Randomization (used)	+25% (2, 47)	0.034	+5% (1, 9)	0.024
B	Blinding (reported)	+40% (16, 64)	0.004	+6% (1, 11)	0.019
	Blinding (used)	+20% (2, 38)	0.030	+2% (-1, 6)	0.181
C	SS estimation (reported)	+40% (17, 63)	0.002	+7% (3, 12)	0.003
	SS estimation (used)	+22% (8, 35)	0.004	+2% (0, 5)	0.075

Figure 2. Reporting and use of individual study design elements (SDEs) in *Nature Medicine*. **A**, Randomization, (**B**) blinding, and (**C**) sample size (SS) estimation. Vertical interrupted line denotes the journal intervention. Δ Level/Slope indicates change in level/slope.



Panel	SDE	Δ Level (95% CI)	*P*	Δ Slope (95% CI)	*P*
A	Randomization (reported)	+24% (9, 38)	0.004	-1% (-4, 2)	0.516
	Randomization (used)	+21% (9, 32)	0.002	-1% (-4, 1)	0.266
B	Blinding (reported)	+27% (8, 46)	0.008	+2% (-3, 6)	0.439
	Blinding (used)	+7% (-10, 24)	0.368	+2% (-2, 5)	0.345
C	SS estimation (reported)	+32% (21, 42)	<0.001	+2% (-1, 4)	0.181
	SS estimation (used)	+19% (8, 30)	0.002	+1% (-2, 3)	0.540

Figure 3. Reporting and use of individual study design elements (SDEs) in *Science Translational Medicine*. **A**, Randomization, (**B**) blinding, and (**C**) sample size (SS) estimation. Vertical interrupted line denotes the journal intervention. Δ Level/Slope indicates change in level/slope.

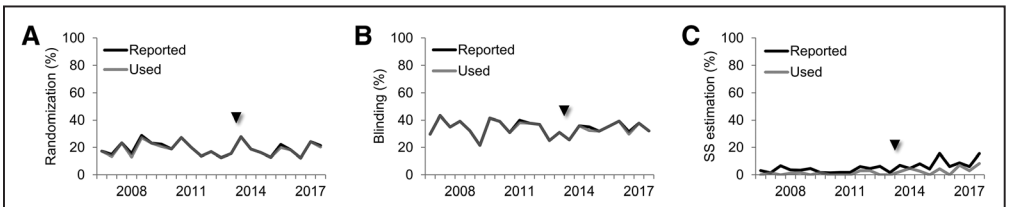
Adjusted Analyses

All temporal patterns persisted in before-after comparisons adjusting for the most relevant disease studied and animal model(s) used (Table).

Discussion

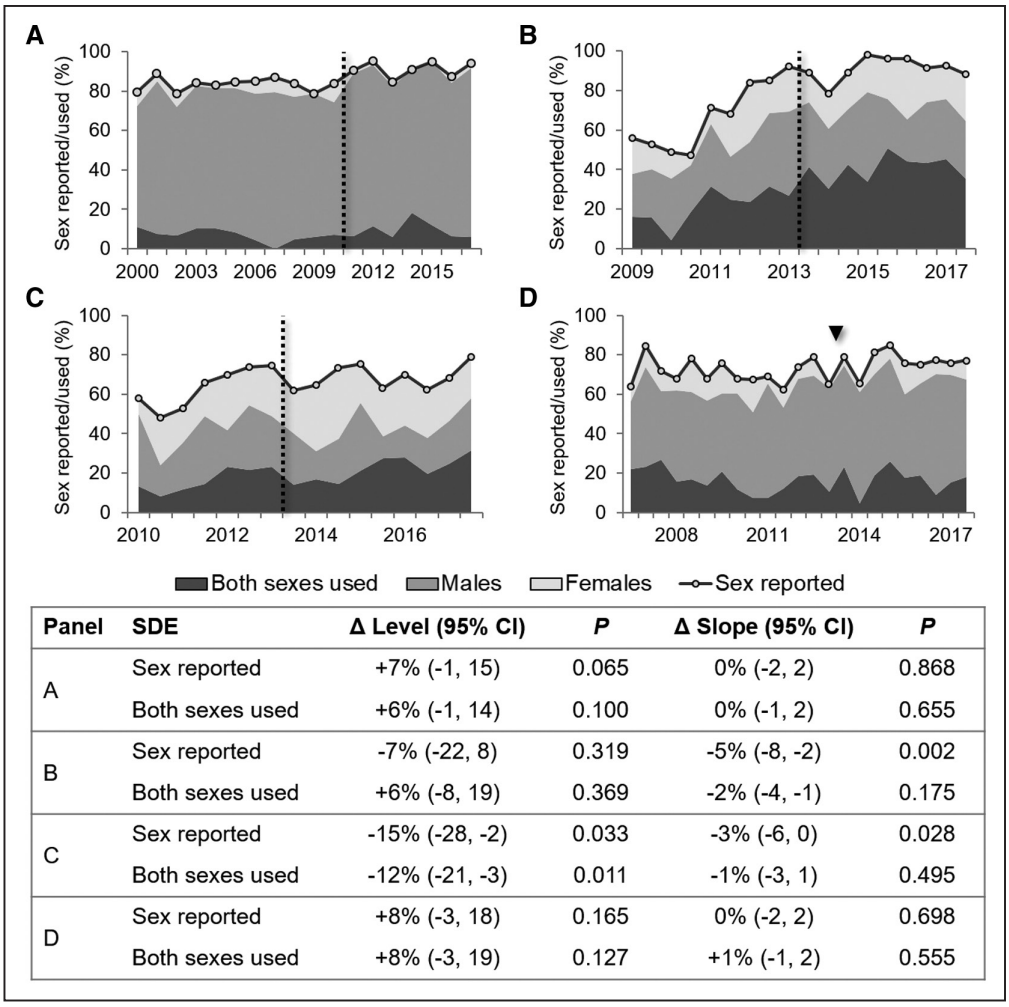
Time-series analyses identified significant and marked improvements in the quality of reporting of published preclinical studies coinciding with the introduction of relevant initiatives in three journals—findings that persisted after accounting for secular trends in reporting and methodological practices and that were not seen in influential journals without analogous interventions over the same period. These findings strongly support the notion that journal interventions can meaningfully influence the quality of published preclinical research.

Mechanistically, well-designed author submission checklists and analogous journal initiatives could improve the quality of published preclinical research in at least 4 ways: (1) by improving the reporting of study methods without affecting the methodology used (ie, improving transparency for readers)[28]; (2) by improving transparency for journal editors and peer reviewers, which could influence whether a manuscript is accepted for publication[29]; (3) by influencing authors' decision to submit to a particular journal (eg, dissuading authors from submitting a manuscript if their study is felt to have methodological shortcomings that will be highlighted)[17]; and (4) by influencing experimental study design practices among researchers who hope to publish in a particular journal.[28] However, few robust analyses of the impact of journal initiatives on published research quality exist. The NPQIP (Nature Publication Quality Improvement Project) Collaborative group examined the potential impact of *Nature* journals' checklist on the reporting quality of in vivo research in 448 articles using primarily a controlled before-after study design.[21] Our analysis of *Nature* journals' publications differs by focusing on *Nature Medicine* (considered to be focused on translational research) and by evaluating a larger sample size (n=949) over a longer period to increase our confidence that changes observed were due to the publishing group's intervention rather than secular trends in editorial practices or culture. Nevertheless, our data support their conclusions and expand on them by demonstrating improvements in study quality after instituting reporting and methodological standards at other journals. The recently reported IICARus



Panel	SDE	Δ Level (95% CI)	*P*	Δ Slope (95% CI)	*P*
A	Randomization (reported)	+3% (-6, 13)	0.446	0% (-1, 2)	0.855
	Randomization (used)	+3% (-6, 12)	0.479	0% (-2, 2)	0.964
B	Blinding (reported)	-1% (-11, 9)	0.847	+1% (-1, 2)	0.360
	Blinding (used)	-1% (-11, 9)	0.791	+1% (-1, 3)	0.326
C	SS estimation (reported)	+2% (-3, 7)	0.432	+1% (0, 2)	0.143
	SS estimation (used)	+1% (-3, 4)	0.679	0% (0, 1)	0.137

Figure 4. Reporting and use of individual study design elements (SDEs) in control journals. **A**, Randomization, (**B**) blinding, and (**C**) sample size (SS) estimation. Analyzed using *Nature Medicine* and *Science TM*'s initiatives as the time series interruption (arrowhead; similar results were obtained using *Stroke*'s initiative as interruption). Δ Level/Slope indicates change in level/slope.



Panel	SDE	Δ Level (95% CI)	*P*	Δ Slope (95% CI)	*P*
A	Sex reported	+7% (-1, 15)	0.065	0% (-2, 2)	0.868
	Both sexes used	+6% (-1, 14)	0.100	0% (-1, 2)	0.655
B	Sex reported	-7% (-22, 8)	0.319	-5% (-8, -2)	0.002
	Both sexes used	+6% (-8, 19)	0.369	-2% (-4, -1)	0.175
C	Sex reported	-15% (-28, -2)	0.033	-3% (-6, 0)	0.028
	Both sexes used	-12% (-21, -3)	0.011	-1% (-3, 1)	0.495
D	Sex reported	+8% (-3, 18)	0.165	0% (-2, 2)	0.698
	Both sexes used	+8% (-3, 19)	0.127	+1% (-1, 2)	0.555

Figure 5. Sex reporting and sex of animals used in preclinical studies. (**A**) *Stroke*, (**B**) *Nature Medicine*, (**C**) *Science Translational Medicine* (*Science TM*), and (**D**) control journals. Vertical interrupted lines denote journal interventions. Results shown for sex reporting and use of both sexes. Control journals were analyzed using *Nature Medicine* and *Science TM*'s initiatives as the time series interruption (arrowhead). Δ Level/Slope indicates change in level/slope.

(Intervention to Improve Compliance with the ARRIVE guidelines) study[22] randomized articles reporting in vivo animal research submitted to *PLoS One* to journal-requested completion of the ARRIVE (Animal Research: Reporting of In Vivo Experiments)[29] checklist versus standard practice (no checklist). Contrasting with our results and those of the NPQIP study, the study found no difference in the proportion of articles fully complying with the ARRIVE guidelines. However, the primary outcome in IICARus was reporting compliance with *all* 38 items in the checklist, which authors may have deemed onerous or judged as unlikely to be enforced.

The journal interventions included in the present analysis improved the reporting and use of SDEs to varying degrees. Differences in the nature of the initiatives may account for some of this. For instance, the checklist introduced by *Nature* journals requires authors to describe whether and how specific SDEs were used.[18] In contrast, *Stroke*'s Basic Science Checklist, when first introduced, was a simple and brief questionnaire consisting of dropdown menus with limited options[19] (although it has since been updated to a more comprehensive version[30]). *Science TM* does not have a readily accessible checklist for authors, but instead provides instructions in its Information for Authors section, which its Editor has stated are enforced.[20] Differences in how vigorously journals enforce their requirements and how carefully they screen for author adherence are also be expected to be influential, could fluctuate over time, and may be affected by how attuned reviewers and editors are to them.

The use of key SDEs overall remains low in preclinical research, even in recent years and despite the prominence of the journals examined—a fact that is more clearly brought to light with improved reporting. Based on reported methods, ≈40% to 50% did not use any randomization, >50% did not use any blinding, and up to 75% did not ensure adequate statistical power. Additionally, consistent across journals was a lack of improvement in the inclusion of animals of both sexes after accounting for baseline secular trends. Despite the National Institutes of Health requiring sex inclusion plans in preclinical research funding applications and the importance placed on sex considerations in clinical research, existing journal reporting requirements place little emphasis on this aspect of study design for preclinical experiments.

There are several limitations to our study. The journals examined are widely considered leading translational research journals in terms of impact metrics and reputation. It is possible that their initiatives are more likely to be accepted and adhered to by the research community. The control journals are focused on cardiovascular research. Whether improvements have occurred in noncardiovascular journals that have not implemented interventions analogous to those examined in our study is unknown. However, similar conclusions were reached by the NPQIP Collaborative group using a more heterogeneous control group.[21] Our interrupted time series study design suggests but does not prove causal effects of the interventions examined. However, *Nature Medicine* and *Science TM* are among the few translational research journals that have endorsed the National Institutes of Health guidelines on rigor and reproducibility[11] and communicated to the research community their initiatives with specific implementation dates,[18,20] rendering them ideal case study candidates. In addition, we show a lack of meaningful improvement in the reporting or conduct of preclinical research practices in high-ranking journals that did not implement similar initiatives over the same period despite having similar impact factors to *Science TM*, while confirming previously suggested improvements in *Stroke* after the implementation of analogous initiatives.[12,19] Furthermore, all initiatives addressed randomization, blinding, and sample size estimation, but minimally addressed considering sex as a biological variable, which corresponded to the patterns of improvements seen. Our analysis relied on published reports, therefore the risk of misclassification exists. For instance, the method of randomization used was rarely described, raising the possibility that certain articles that reported randomly assigning animals may have referred to haphazard assignment. Lastly, the study protocol was not registered.

As gatekeepers of what is published, journals' interventions can meaningfully influence the quality of preclinical research and are promising strategies to improve study transparency and design. However, although the studied initiatives were associated with substantial improvements in reporting, improvements in the use of key SDEs were relatively modest, suggesting that additional strategies focused on improving preclinical study design are required to address widespread methodological shortcomings, including sex bias.

Sources of Funding

Dr Ramirez was supported by a Canadian Institutes of Health Research (CIHR) Frederick Banting & Charles Best Canada Graduate Scholarship, a CIHR Banting Postdoctoral Fellowship, the University of Ottawa Heart Institute Foundation, the Ernest and Margaret Ford Research Endowed Fund, and the Cardiac Arrhythmia Network of Canada (CANet) as part of the Networks of Centres of Excellence. R.G. Jung was supported by a CIHR Vanier Canada Graduate Scholarship. Dr Graham was supported by a CIHR Foundation grant (FDN number 143237).

Disclosures

Dr Provencher reports having received grants and personal fees from Actelion Pharmaceuticals, grants from AstraZeneca, and grants from Resverlogix outside the submitted work. Dr Hibbert reports funding as a clinical trial investigator from Abbott and Edwards Lifesciences outside the submitted work. The other authors report no conflicts.

References

1. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014;505:612–613. doi: 10.1038/505612a
2. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490:187–191. doi: 10.1038/nature11556
3. 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
4. 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
5. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166–175. doi: 10.1016/S0140-6736(13)62227-8
6. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol*. 2015;13:e1002165. doi: 10.1371/journal.pbio.1002165
7. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? *PLoS Med*. 2010;7:e1000245. doi: 10.1371/journal.pmed.1000245
8. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. 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
9. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature*. 2012;483:531–533. doi: 10.1038/483531a
10. 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. doi: 10.1161/hs1001.096009
11. National Institutes of Health. Principles and guidelines for reporting preclinical research. Available at https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. Accessed July 29, 2019.
12. Ramirez FD, Motazedian P, Jung RG, Di Santo P, MacDonald ZD, Moreland R, et al. Methodological rigor in preclinical cardiovascular studies: targets to enhance reproducibility and promote research translation. *Circ Res*. 2017;120:1916–1926. doi: 10.1161/CIRCRESAHA.117.310628
13. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. *Nature*. 2014;509:282–283. doi: 10.1038/509282a
14. Ramirez FD, Motazedian P, Jung RG, Di Santo P, MacDonald Z, Simard T, et al. 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
15. 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
16. Provencher S, Archer SL, Ramirez FD, Hibbert B, Paulin R, Boucherat O, et al. Standards and methodological rigor in pulmonary arterial hypertension preclinical and translational research. *Circ Res*. 2018;122:1021–1032. doi: 10.1161/CIRCRESAHA.117.312579
17. Bolli R. New initiatives to improve the rigor and reproducibility of articles published in Circulation Research. *Circ Res*. 2017;121:472–479. doi: 10.1161/CIRCRESAHA.117.311678
18. Reducing our irreproducibility. *Nature*. 2013;496:398.
19. 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
20. Kelner KL. Playing our part. *Sci Transl Med*. 2013;5:190ed197
21. The NPQIP Collaborative Group. Did a change in Nature journals' editorial policy for life sciences research improve reporting? *BMJ Open Science*. 2019;3:e000035.
22. Hair K, Macleod MR, Sena ES; IICARus Collaboration. A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Res Integr Peer Rev*. 2019;4:12. doi: 10.1186/s41073-019-0069-3
23. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards

for pre-clinical animal studies. *PLoS Biol*. 2014;12:e1001756. doi: 10.1371/journal.pbio.1001756
24. Stroke Therapy Academic Industry Roundtable. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*. 1999;30:2752–2758. doi 10.1161/01.str.30.12.2752
25. 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
26. 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
27. Ramirez FD, Hibbert B. Letter by Ramirez and Hibbert regarding article, "Consideration of sex differences in design and reporting of experimental arterial pathology studies: a statement from the Arteriosclerosis, Thrombosis, and Vascular Biology Council." *Arterioscler Thromb Vasc Biol*. 2018;38:e99–e100. doi: 10.1161/ATVBAHA.118.310942
28. Checklists work to improve science. *Nature*. 2018;556:273–274.
29. 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
30. 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