# Evaluating Study Design Rigor in Preclinical Cardiovascular Research: A Replication Study

Isaiah C. Jimenez*[1, 2], Gabrielle C. Montenegro*[1,3], Keyana Zahiri*[1,4], Damini Patel[1,5], Adrienne Mueller[1]

[1] Stanford Cardiovascular Institute, Stanford, CA
[2] Saint Mary's College of California, Moraga, CA
[3] Macalester College, Saint Paul, MN
[4] Warren Alpert Medical School of Brown University, Providence, RI
[5] Central Michigan University, Mt Pleasant, MI

*ICJ, GCM, and KZ contributed equally to this work and are co-first authors.

## ABSTRACT

### Background

Methodological rigor is a major priority in preclinical cardiovascular research to ensure experimental reproducibility and high quality research. Lack of reproducibility results in diminished translation of preclinical discoveries into medical practice and wastes resources. In addition, lack of reproducibility fosters uncertainty in the public's acceptance of reported research results.

### Methods

We evaluate the reporting of rigorous methodological practices in preclinical cardiovascular research studies published in leading scientific journals by screening articles for the inclusion of the following key study design elements (SDEs): considering sex as a biological variable, randomization, blinding, and sample size power estimation. We have specifically chosen to screen for these SDEs across articles pertaining to preclinical cardiovascular research studies published between 2011 and 2021. Our study replicates and extends a study published in 2017 by Ramirez et al. We hypothesized that there would be higher SDE inclusion across preclinical studies over time, that preclinical studies that also include human and animal substudies within the same study will exhibit greater SDE inclusion than animal-only preclinical studies, and that there will be a difference in SDE usage between large and small animal models.

### Results

Overall, inclusion of SDEs was low. 15.2% of animal only studies included both sexes as a biological variable, 30.4% included randomization, 32.1% included blinding, and 8.2% included sample size estimation. Incorporation of SDE in preclinical studies did not significantly increase over the ten year time period in the articles we assessed. Although the inclusion of sex as a biological variable increased over the 10 year time frame, that change was not significant ($p=0.411$, corrected $p=8.22$). These trends were consistent across journals. Reporting of randomization and sample size estimation differs significantly between animal and human substudies (corrected $p=3.690e-06$ and corrected $p=7.252e-08$, respectively.) Large animal studies had a significantly greater percentage of blinding reported when compared to small

animal studies (corrected p=0.01.) Additionally, overall, large animal studies tended to have higher SDE usage.

**Conclusions**

In summary, evidence of methodological rigor varies substantially depending on the study type and model organisms used. Over the time period of 2011-2021, the reporting of SDEs within preclinical cardiovascular studies has not improved and suggests extensive evaluation of other SDEs used in cardiovascular research. Limited incorporation of SDEs within research hinders experimental reproducibility that is critical to future research.

## INTRODUCTION

Preclinical studies using animal models play an important role in developing new treatments and evaluating the safety and efficacy of novel therapies. Preclinical cardiovascular research has greatly contributed to our understanding of heart disease (Houser et al., 2012; Bacmeister et al,, 2019), yet we often find failed translations from “bench-to-bedside” (Justice & Dhillon, 2016; Seok et al., 2012). Methodological rigor in these studies remains a major priority to establish a level of consistent reproducibility across preclinical research.

One means to enhance reproducibility of preclinical research is to increase the frequency of use of study design elements (SDEs) that enhance rigor, such as inclusion of sex as a biological variable, randomization of samples or subjects, blinding, and sample size estimation. Inclusion of both biological sexes in experimental design removes sex as a potential confounding variable in establishing causal relationships between variables of interest. Implementation of randomization in experimental design is an additional means to control for *all* potential confounders between variables of interest. Moreover, randomization helps reduce bias. The use of blinding in experimental design limits potential bias in assessment of experimental outcomes from study participants and researchers themselves. Finally, sample size estimation limits impractical significance in experimental results and false-negative results. Ultimately, each of these four SDEs influence the reproducibility of preclinical experimental outcomes.

This study is a replication and extension of a study performed by Ramirez et al. (2017), which investigated the prevalence of these four SDEs in preclinical studies published in five leading cardiovascular journals of the American Heart Association (*Circulation*; *Circulation Research*; *Hypertension*; *Stroke*; and *Arteriosclerosis, Thrombosis, and Vascular Biology* (*ATVB*)) between July 2006 and June 2016. Their study found low prevalence of SDEs across screened studies, reflecting low methodological rigor in preclinical cardiovascular research in that decade (Ramirez et al., 2017). This study investigates inclusion of these four SDEs in randomly selected preclinical cardiovascular studies published between 2011-2021 in nine different leading biomedical and scientific journals outside of American Heart Association publications: *Science*, *Nature*, *European Heart Journal*, *Journal of the American College of Cardiology*, *New England Journal of Medicine*, *Cell*, *Lancet*, *Journal of the American Medical Association*, and *Proceedings of the National Academy of Sciences of the United States of America*. This study also examines the use of rigorous SDEs by comparing animal only studies with studies that have both animal and human substudies (human/animal studies) over the ten year period. The

decade 2011-2021 was selected for analysis to provide a more recent evaluation of methodological rigor in preclinical cardiovascular research, following the work of Ramirez et al., 2017.

By identifying trends in sex of study subjects used, randomization, blinding, and sample size estimations, we assessed the methodological rigor of scientific practices carried out in preclinical cardiovascular research.

**METHODS**

We reviewed preclinical cardiovascular articles published between 2011-2021 in leading biomedical and scientific journals. Studies were included from nine leading journals: *Science*, *Nature*, *European Heart Journal*, *Journal of the American College of Cardiology*, *New England Journal of Medicine*, *Cell*, *Lancet*, *Journal of the American Medical Association*, and *Proceedings of the National Academy of Sciences of the United States of America*. These journals were selected to complement those used in a previous study: *Circulation*; *Circulation Research*; *Hypertension*; *Stroke*; and *Arteriosclerosis, Thrombosis, and Vascular Biology* (*ATVB*) (Ramirez et al., 2017). Using a search string in a Pubmed query, we identified primary research articles (excluding editorials or comments) describing cardiovascular experiments using animal models. The complete search string used was:

((cardi*[Title]) OR (heart[Title]) OR (arteri*[Title]) OR (hypertensi*[Title]) OR (atherosclero*[Title]) OR (arrhythm*[Title])) AND ((pig[Title/Abstract]) OR (rat[Title/Abstract]) OR (mouse[Title/Abstract]) OR (guinea pig[Title/Abstract]) OR (gerbil[Title/Abstract]) OR (hamster[Title/Abstract]) OR (monkey[Title/Abstract]) OR (rabbit[Title/Abstract]) OR (dog[Title/Abstract])) NOT ((review[Publication Type]) OR (systematic review[Publication Type]) OR (editorial[Publication Type]) OR (comment[Publication Type])) AND (("2011/01/01"[Date - Publication] : "2021/12/31"[Date - Publication])) AND (("Lancet (London, England)"[Journal]) OR ("Nature"[Journal]) OR ("Science (New York, N.Y.)"[Journal]) OR ("JAMA"[Journal]) OR ("The New England journal of medicine"[Journal]) OR ("Proceedings of the National Academy of Sciences of the United States of America"[Journal]) OR ("Cell"[Journal]) OR ("European heart journal"[Journal]) OR ("Journal of the American College of Cardiology"[Journal])).

309 articles were returned by the PubMed Search query. No stopping rule was utilized, as sample size was predetermined prior to data collection. Studies were included if they were published manuscripts and used animal subjects. Articles were excluded from data analysis if they did not include animal-model experiments, if they were not related to a cardiovascular research topic, or were published as an abstract, editorial, or any form other than a published full manuscript. Thus, although the search string yielded 309 studies for screening, 11 studies were excluded for not meeting inclusion criteria. A total of 298 studies were ultimately included in our data analyses.

Articles were screened on the basis of four study design elements (SDEs): use of both biological sexes in study subjects, randomization, blinding, and sample size estimation. The screening database included animal only studies, as well as animal studies that included human

substudies (human/animal studies). For human/animal studies, the same SDEs were used to evaluate methodological rigor across human subject populations. We evaluated the four SDEs separately for studies that only performed animal experiments and studies that performed human/animal experiments, if applicable. Screening definitions were predefined (Table I). Articles were also screened for what cardiovascular research topic they were investigating, as well as which animal species were used in the study (Table II). We initially used cardiovascular research topics that were defined in Ramirez et al. (2017) and also expanded the topic list to also include 2 additional topics that occurred more frequently in our dataset: congenital heart disease and heart development/repair/regeneration.

Articles were distributed equally among members of the research team for screening. Articles were initially screened in the order they were returned from the PubMed search. Investigators did not select which articles from the database of 298 articles to screen based on any specific criteria. To ensure accuracy and consistency in screening, each article was independently screened by two investigators. Screeners were randomly assigned for the second screening of an article. Discrepancies in screening were resolved by consensus.

**Statistical Methods**

Chi-squared tests, *t*-tests, and multiple regressions were performed to evaluate statistical significance of comparisons using RStudio and Microsoft Excel software.

The collected data was analyzed to evaluate the prevalence of the use of SDEs in preclinical cardiovascular research between 2011-2021. Categorical variables are reported as a number (%) and were compared via chi-square tests. A threshold of $p < 0.05$ was considered statistically significant, however we used Bonferroni correction to correct for multiple comparisons.

**Pre-registration and Data Availability**

This study was pre-registered in the Open Science Framework (OSF) registry. The preregistration can be accessed via the following link: https://doi.org/10.17605/OSF.IO/F4NH9 (Patel et al., 2022). We adhered to the methodology detailed in the pre registration for this analysis.

The data and analytical methods used for this study are available at https://osf.io/52Q6W/ (Zahiri et al., 2022.)

## RESULTS

A total of 298 preclinical research studies published between 2011 and 2021 were included in our analyses. 61.7% (N=184) of these studies were animal only and 38.3% (N=114) were human/animal studies (Figure 1A). Approximately the same number of cardiovascular preclinical research studies were published for each of the ten years in our sample (Figure 1B). The majority of the studies in our analysis were published in Proceedings of the National Academy of Sciences of the United States of America, 45% (N=135) (Figure 1C). In addition, a wide range

of species were used in the preclinical studies we analyzed (Figure 1D). Mice were the most commonly studied species and were used in 77.2% (N=230) of studies. This was followed by rats in 20.1% (N=60) of studies. Additionally, a wide range of topics relating to cardiovascular disease were investigated in the studies we assessed (Figure 1E). We categorized the topics as follows: cardiomyopathy or heart failure (28.2%), atherosclerosis or vascular homeostasis (16.1%), myocardial infarction (14.8%), cardiac arrhythmia (9.1%), heart development/repair/regeneration (8%), hypertension (4.4%), metabolic or endocrine disease (4%), congenital heart disease (2.7%), valvular disease (2%), cardiac transplantation (1.7%), hematological disorder (0.3%), and other (8.7%), based on topics identified in the original Ramirez et al. (2017) study.

*Overall SDE Inclusion*

Table III shows the proportion of studies that included each of the four SDEs, stratified by animal-only studies, as well as animal and human substudies in human/animal studies. In animal only studies, both sexes were used in 15.2% (N=28) studies, single sex was used in 48.4% (N=89) studies, and there was no reporting on sex of study subjects used in 36.4% (N=67) studies. In animal substudies in human/animal studies, both sexes were used in 17.5% (N=20) studies, single sex was used in 53.5% (N=61) studies, and there was no reporting on sex of study subjects used in 29% (N=33) studies. In human substudies in human/animal studies, both sexes were used in 36% (N=41) of studies, single sex was used in 8.8% (N=10) of studies, and there was no reporting on sex of study subjects used in 55.2% (N=63) studies.

In terms of randomization, 30.4% (N=56) of animal only studies, 36% (N=41) of animal substudies in human/animal studies, and 6% (N=7) of human substudies in human/animal studies used randomization to any degree in their experiments. In terms of blinding, 32.1% (N=59) of animal only studies, 24.5% (N=28) of animal substudies in human/animal studies, and 11.4% (N=13) of human substudies in human/animal studies used blinding to any degree in their experiments.

In terms of sample size estimations for animal only studies, 8.2% (N=15) used statistical analysis to determine sample size for experiments, 4.9% (N=9) provided other justification for the sample size they selected, 2.7% (N=5) indicated that no sample size estimation was done, and 84.2% (N=155) did not report sample size justification. Examples of ‘other justification for sample size selection’, include estimating sample size based on pilot studies or determining sample size based on previous standards in the field. In terms of sample size estimations for animal substudies in human/animal studies, 7% (N=8) used statistical analysis to determine sample size for experiments, 2.6% (N=3) provided other justification for the sample size they selected, 5.3% (N=6) indicated that no sample size estimation was done, and 85.1% (N=97) did not report any sample size justification. Lastly, in terms of sample size estimations for human substudies in human/animal studies, 1.8% (N=2) used statistical analysis to determine sample size for experiments, 12.3% (N=14) provided other justification for the sample size they

selected, 2.6% (N=3) indicated that no sample size estimation was done, and 83.3% (N=95) did not report any sample size justification.

*Changes in SDEs Over Time*

A regression analysis of SDE inclusion between 2011-2021 revealed that the proportion of studies including animals of both biological sexes generally increased between 2011 and 2021, though not significantly ($R^2$= 0.0762, F(1,9)= 0.742, p= 0.411 (corrected p=8.22)) (Figure 2). Interestingly, of the four SDEs analyzed, sex as a biological variable was the only variable observed to increase in prevalence across preclinical cardiovascular studies over the decade examined, although the increase was not determined to be statistically significant (p>0.05). Of studies screened from 2011, 10% (N=3 of 29) included animals of both biological sexes. Of studies screened from 2017, 27% (N=9 of 33) included animals of both biological sexes. This proportion, however, decreased to 10% (N=2 of 21) of studies in 2021 (Figure 2).

Conversely, regression analysis revealed that the proportion of studies implementing blinding generally decreased between 2011-2021, though not determined statistically significant ($R^2$= 0.1357, F(1,9)= 1.41, p= 0.265 (corrected p= 5.3)). Of studies screened from 2011, 24% (N=7 of 29) implemented blinding. This proportion decreased to 19% (N=4 of 21) in 2021 (Figure 2). Similar trends were found for randomization– $R^2$=0.0698, F(1,9)= 0.675, p=0.433 (corrected p=8.66)– and sample size estimation– $R^2$=0.0466, F(1,9)= 0.439, p=0.524 (corrected p=10.48). Approximately half of preclinical cardiovascular studies implemented randomization in 2011, as 48% (N=14 of 29) 2011 studies mentioned randomization in their methods. This proportion declined to 29% (N=6 of 21) of studies in 2021 (Figure 2). Likewise, of 2011 studies, 14% (N=4 of 29) justified their sample size using size estimation (justification or statistical estimation). This proportion dropped to 0% (N=0 of 22) studies in 2018. However, this proportion jumped back up to 14% (N=3 of 21) in 2021 (Figure 2). Although general decreasing trends were found for blinding, randomization, and sample size estimation, these decreases were not found to be statistically significant (p>0.05).

*Differences in SDE Reporting Across Journals*

Of all 298 journal articles screened in this study, the four journals with the highest numbers of cardiovascular preclinical studies were: *Proceedings of the National Academy of Sciences of the United States of America (Proc Natl Acad Sci)* 45% (N=135), *European Heart Journal (Eur Heart J)* 20% (N=61)*, Journal of the American College of Cardiology (J Am Coll Cardiol)* 14% (N=42)*,* and *Nature* 10% (N=31). A general difference in the reporting of the four SDEs was observed across all four journals, but differences were not determined to be statistically significant by *t*-tests: both sexes: t(22)=4.804, p=0.0064 (corrected p=0.128); randomization: t(41)=3.179, p=0.0107 (corrected p=0.214); blinding: t(38)=1.489, p=0.2077 (corrected p=4.154), sample size estimation: t(6)=14.853, p=0.0264 (corrected p=0.528) (Figure 3).

Articles screened across the four most prevalent journals in this study varied in frequency across the ten year period (2011-2021). Among studies screened from *Eur Heart J,* (N=1 of 11) 11% included animals of both biological sexes in 2011 while (N=2 of 8) 25% included both biological sexes in 2019. Across studies screened from the *J Am Coll Cardiol*, (N=1 of 6) 17% included animals from both biological sexes in 2011 while 50% (N=2 of 4) included both biological sexes in 2019 (Figure 3). A difference of +14% in the *Eur Heart J* and +33% in the *J Am Coll Cardiol* between the journals was not found to be statistically significant (p=0.0064, corrected p=0.128).

In studies screened from *Eur Heart J*, 56% (N=5 of 9) reported using randomization in 2011 and 43% (N=3 of 7) reported using randomization in 2021. Meanwhile, in studies screened from the *Proc Natl Acad Sci*, 44% (N=4 of 9) reported using randomization in 2011 and 22% (N=2 of 9) reported using randomization in 2021 (Figure 3). Similarly, although a difference of -13% over time was observed in *Eur Heart J* versus a difference of -22% in the *Proc Natl Acad Sci*, no statistically significant difference between these decreasing rates was found across the ten year period (p=0.0107, corrected p=0.214).

With regards to reporting of blinding, in studies screened from the *Proc Natl Acad Sci,* 22% (N=2 of 9) reported implementing blinding in 2011 and 33% (N=3 of 9) reported implementing blinding in 2021. In studies screened from the *J Am Coll Cardiol*, 33% (N=2 of 6) used blinding in 2011 while 100% (N=1 of 1) used blinding in 2021 (Figure 3). A change of +11% overtime was observed across studies screened from the *Proc Natl Acad Sci* versus a change of +77% across studies screened from the *J Am Coll Cardiol*, but no statistically significant difference was found in comparing these differences as well (p=0.2077, corrected p=4.154).

Finally, in assessing differences in reporting statistical estimations for study sample size, a difference of +3% (N=2 of 9 to N=1 of 4) was found across studies screened from the *Eur Heart J* between 2011 and 2020 (Figure 3). A difference of -10% (N=3 of 18 to N=1 of 15) was observed between studies screened from *Proc Natl Acad Sci* from 2012-2017 (Figure 3). No studies prior to 2012 nor beyond 2017 from the *Proc Natl Acad Sci* reported using sample size estimations. Observed differences were not determined to be statistically significant (p=0.0264, corrected p=0.528).

*Differences in SDE Reporting Across Experimental Models*

There were slight differences in SDE reporting for animal substudies in human/animal studies vs animal only studies (Figure 4). In human/animal studies, 18% (N=20) used subjects of both sexes, 53% (N=61) used only one sex of study subjects, and 33% (N= 29) did not report sex of study subjects used, as opposed to 16% (N=28), 47% (N=89), and 37% (N=67), respectively, for animal only studies. Randomization was only used in 36% (N=41) of human/animal studies and in 30% (N=56) of animal only studies. Blinding was only used in 25% (N=28) of human/animal studies and in 32% (N=59) of animal only studies. In terms of sample size estimations for

human/animal studies, 7% (N=8) performed sample size estimations, 3% (N=3) justified not using sample size estimations, 5% (N=6) did not perform sample size estimation, and 85% (N=97) did not report any information on sample size estimation. For animal only studies, these values were 8% (N=15), 5% (N=9), 3% (N=5), and 84% (N=155), respectively.

However, there were no statistically significant differences in SDE reporting for in human/animal or animal only studies in terms of sex of study subjects used ($X^2$=1.775, df=2, p=0.4117, corrected p=1.6468), randomization ($X^2$=0.7448, df=1, p=0.388, corrected p=1.5524), blinding ($X^2$=1.5715, df=1, p=0.21, corrected p=0.84), or sample size estimation ($X^2$=2.2518, df=3, p=0.5218, corrected p=2.0872).

*Differences in SDE Reporting For Animals vs Humans within the Same Study*

Within human/animal studies, there were variations in SDE reporting for human vs animal substudies (Figure 5). For human substudies in human/animal studies, 36% (N=41) used subjects of both sexes, 53% (N=61) used only one sex of study subjects, and 33% (N= 29) did not report sex of study subjects used, as opposed to 16% (N=28), 47% (N=89), and 37% (N=67), respectively, for animal only studies. Randomization was used in 6% (N=7) of human substudies and 36% (N=41) of animal substudies, and blinding was used in 11% (N=13) of human substudies and 25% (N=28) of animal substudies in human/animal studies. In terms of sample size estimations for human substudies, 2% (N=2) performed sample size estimations, 12% (N=14) justified not using sample size estimations, 3% (N=3) did not perform sample size estimation, and 83% (N=95) did not report any information on sample size estimation. For animal substudies, these values were 7% (N=8), 3% (N=3), 5% (N=6), and 85% (N=97), respectively.

There was a statistically significant difference in use of randomization ($X^2$=24.083, df=1, p=9.226e-07, corrected p=3.690e-06) and sample size estimation inclusion ($X^2$=38.911, df=3, p=1.813e-08, corrected p=7.252e-08) of animal vs human substudies in human/animal studies. However, there was no statistically significant difference for blinding ($X^2$=5.4878, df=1, p=0.01915, corrected p=0.0766) and sex of study subjects used in animal vs human substudies in human/animal studies ($X^2$=3.123, df=2, p=0.2098, corrected p=0.8392).

*Differences in SDE Reporting in Different Animal Models*

Reporting of biological sex SDE is greater in large animal studies whether including both sexes or stating only one sex was used when compared to small animal studies (Figure 6). Note that the number of small animal studies was far greater than the number of large animal studies that were used for this analysis. For small animals, both sexes were used in 16% (N=44) studies, a single sex was used in 50% (N=135) studies, and sex was not reported for 34% (N=93) studies. Large animal studies exhibited a similar trend in that both sexes were used in 15% (N=4) studies, a single sex was used in 58% (N=15), and sex was not reported in 26% (N=7) studies.

There was no significant difference in proportions of studies reporting sex between the small and large animals (chi square test of interdependence, $X^2$=0.689, df=2, p=0.709, corrected p=2.83). In small animal studies, randomization was used in 30% (N=82) whereas for large animal studies, 58% (N=15) studies used randomization. There was a significant difference in proportions of studies using randomization between the large and small animals (chi square test of interdependence, $X^2$=6.995, df=1, p=0.008, corrected p=0.017). Small animal studies had 27% (N=74) of studies that used blinding where large animals had 50% (N=13) of studies use blinding. There was no significant difference in proportion of studies using blinding between the two variables (chi square test of interdependence, $X^2$=4.913, df=1, p=0.0267, corrected p=0.107). Lastly, 86% (N=234) of small animal studies did not report any information regarding sample size estimation. For large animal studies, 69% (N=18) did not report any information regarding sample size estimation. There was no significant difference in proportion of studies using sample size estimation between the two variables (chi square test of interdependence, $X^2$= 7.7154, df=3, p=0.052, corrected p=0.210). This data is summarized in Figure 6.

The proportion of single sex studies was found to be greater for other animal models in comparison to rodents (Figure 7). Of the studies in which rodents were the primary animal model, 48.1% (N=128) were single sex studies. This is less than other animal model studies where 68.8% (N=22) were single sex studies. There is a statistically significant difference between single sex inclusion in rodent studies and other animal studies ($X^2$=4.073, df=1, p=0.0436).

**Discussion**

This study is a replication and extension of a study performed by Ramirez et al. (2017), which quantifies the use of methodologically rigorous practices across preclinical cardiovascular research by measuring the inclusion of specific study design elements that promote rigor. We assessed the reporting of four study design elements - inclusion of both sexes, randomization, blinding, and sample size analysis - in preclinical cardiovascular studies with either animal only studies, or studies with both human and animal substudies, over a 10 year period from 2011 to 2021. Overall, inclusion of SDEs was low. 15.2% of animal only studies included both sexes as a biological variable, 30.4% included randomization, 32.1% included blinding, and 8.2% included sample size estimation. Also, incorporation of SDE in preclinical studies did not significantly increase over the ten year time period in the articles we assessed. Among human and animal substudies of human/animal studies, a significantly larger proportion of animal substudies reported usage of randomization and sample size estimation. We also discovered that a significantly greater proportion of large animal substudies report usage of randomization when compared to small animal substudies. Our conclusions serve as an informative checkpoint for the impacts implementing ARRIVE (Animal Research: Reporting in In Vivo Experiments) and other protocols for enforcing methodological rigor (Lapchak et al., 2013; Ramirez et al., 2020; Williams et al., 2022).

Preclinical research is foundational for clinical treatment, and rigorous methodological practice is critical for safely translating research from bench to bedside. Unfortunately, low reproducibility of preclinical studies may partially be attributed to the cultural pressure to publish in high ranked journals, which is more easily achieved if a study has positive results (Mlinaric et al., 2017). Researchers are also incentivized to share positive results to secure funding or highly competitive academic jobs. Studies with limited SDE incorporation also make it harder to perform systematic reviews (O'Connor, 2018). Failure to include the necessary SDEs may also undermine the credibility of a study's results. Though the reproducibility crisis has been acknowledged by the NIH (Collins & Tabak, 2014) and mitigating efforts have been employed, the progress and success of these efforts continue to warrant assessment. In addition to Ramirez et al. (2017)'s finding that preclinical cardiovascular research studies had low SDE inclusion, Williams et al. (2022) determined that preclinical research articles do not adequately adhere to many ARRIVE guidelines, including having sufficient power for t-tests, and use of randomization and blinding. Our results support the findings of both these previous studies.

*Study Design Elements*

Although some SDEs can be relevant only to specific domains of study, (Provencher, 2018), several SDEs are relevant to preclinical research in general. Use of a single sex in a study means the results are not generalizable. Omission of sex as a variable may hinder later applications or reproducibility of results in translational or clinical settings (Ramirez et al., 2017). Animal studies have exhibited a significant lack of incorporation of both sexes. In addition, animal model types used in studies exhibit a significant difference in reporting of both sexes when comparing rodent studies to non-rodent studies (Figure 7).

Randomization diminishes confounding variables by assigning cohorts of animals or human subjects to an experimental or control group at random. Failure to apply randomization can overexaggerate results thereby leading to unreliable findings (Hirst et al., 2014). The importance of the SDE is critical to preventing selection bias and is also the assumption upon which many statistical tests are based. In our study, the proportion of studies reporting the usage of randomization was low for both large and small animals. However, a greater proportion of large animal model studies used randomization compared to small animal studies (Figure 6).

The use of blinding in research is advantageous since it limits selection and procedural bias that could influence experimental results. Articles were screened for any reporting of blinding used whether it was single, double, or triple blinding, however, no significant differences were found.

Sample size estimation seeks to predict the necessary number of subjects for conclusions to be drawn and applied to the general population. This consists of predetermining a set number of animals needed for random group assignment. Sample size estimation was the lowest reported SDE among the four SDE's observed - only 8% of animal-only studies provided any justification, statistical or otherwise, for their sample size.

*Study Design Element Inclusion Over Time in the Past Decade*

Although we hypothesized that the level of SDE inclusion would increase over time, the general prevalence of SDEs, whether they be applied to animal or human substudies, was remarkably low across the 298 articles screened (Figure 2). This contrasts with Ramirez et al. (2017), who found a positive trend in reporting for the journals they observed. One other study has also found a positive trend in temporal SDE inclusion in elements such as randomization and blinding, however, sample size estimation and inclusion of both sexes remained low (Jung et al., 2021). In contrast, our findings suggest, if anything, decreasing SDE inclusion rates for randomization, blinding, and sample size estimation.

*The Influence of Animal and Human Subject Models on Study Design Element Inclusion*

We evaluated SDE usage in animal-only studies compared to human/animal studies and also evaluated SDE usage in animal substudies vs. human substudies within the same overall study. No large contrast in SDE incorporation existed between studies that were animal only and those that had both animal and human substudies, though the proportion of studies that reported SDEs was low in both. Sample size estimation remained the lowest reported SDE for both study types.

Evaluating differences between SDE usage for animal or human subjects within the same study showed differences in inclusion of randomization and sample size estimation. For these studies, a larger proportion of studies exhibit greater reporting of randomization and sample size estimation in animal substudies compared to human substudies.

*Study Limitations*

Although all articles were subjected to randomized double screenings, we cannot be completely confident that an article did not use an SDE. In some instances, certain SDEs are not able to be incorporated due to the conditions of the study, however, studies where SDE inclusion was not possible were still considered as lacking the SDE. Also, although the SDEs evaluated in this study are critical to conducting rigorous experimentation, other SDEs discipline-specific may ultimately be more relevant to the reproducibility of a paper.

*Future Directions*

Future studies should expand the scope of SDEs reviewed in preclinical cardiovascular publications to include more domain-specific SDEs such as a use of comparison group, units of concern, or other forms of subject allocation (O'Connor, 2018). A more detailed investigation of

reasons publications do not include sample size estimation would also be extremely valuable, given the complexity of justification for including or not including this SDE. A further extension of this work would be to survey the authors of these studies to determine the underlying reason for either omission or inclusion of different SDES. This would allow both verification of this study's screening, as well as shed light on the reasons why investigators do or do not use SDEs in their studies.

*Conclusion*

Inclusion of SDEs improves reproducibility, which is important for translating preclinical findings into clinical outcomes. Contrary to our hypothesis, we found that over the past decade there has been no increase in SDE incorporation in the preclinical cardiovascular publications we studied. Sample size estimation remains the least reported study design element. These trends all indicate the need for further efforts to increase the incorporation of rigorous study design elements in research projects to the point that they become routine. Future research efforts should evaluate a wider range of SDEs and investigate the reasons why SDEs have such low incorporation in preclinical cardiovascular research.

**Funding**
This work was supported by an NIH NHLBI R25 training award "Stanford Undergraduate URM Summer Cardiovascular Research Program" (R25HL147666), an AHA institutional training award "AHA - Stanford Cardiovascular Institute Undergraduate Fellowship Program," (18UFEL33960207,) and an NIH NHLBI T35 training award "Stanford Cardiovascular Summer Research Training Program for Medical Students" (T35HL160496.)

**Disclosures**

The authors have no financial conflicts of interest to disclose.

**Acknowledgements**

The authors acknowledge support from the Stanford Program on Research Rigor & Reproducibility (SPORR). GCM acknowledges support from Kelsey Grinde, PhD from Macalester College for her guidance on the statistical analysis for this project. The authors would also like to acknowledge the BioInfograph web application for figure generation: https://baohongz.github.io/bioInfograph/**.**

**References**

Bacmeister, L., Schwarzl, M., Warnke, S., Stoffers, B., Blankenberg, S., Westermann, D., & Lindner, D. (2019). Inflammation and fibrosis in murine models of heart failure. *Basic Research in Cardiology*, *114*(3). https://doi.org/10.1007/s00395-019-0722-5

Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature*, *505*(7485), 612–613. https://doi.org/10.1038/505612a

Gosselin, R.-D. (2021). Insufficient transparency of statistical reporting in Preclinical Research: A scoping review. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-83006-5

Hirst, J. A., Howick, J., Aronson, J. K., Roberts, N., Perera, R., Koshiaris, C., & Heneghan, C. (2014). The need for randomization in animal trials: An overview of Systematic Reviews. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0098856

Houser, S. R., Margulies, K. B., Murphy, A. M., Spinale, F. G., Francis, G. S., Prabhu, S. D., Rockman, H. A., Kass, D. A., Molkentin, J. D., Sussman, M. A., & Koch, W. J. (2012). Animal models of heart failure. *Circulation Research*, *111*(1), 131–150. https://doi.org/10.1161/res.0b013e3182582523

Jung, R. G., Stotts, C., Makwana, D., Motazedian, P., Di Santo, P., Goh, C.-Y., Verreault-Julien, L., Simard, T., Ramirez, F. D., & Hibbert, B. (2021). Methodological rigor in preclinical cardiovascular research: Contemporary Performance of AHA Scientific Publications. *Circulation Research*, *129*(9), 887–889. https://doi.org/10.1161/circresaha.121.319921

Justice, M. J., & Dhillon, P. (2016). Using the Mouse to model human disease: Increasing validity and reproducibility. *Disease Models & Mechanisms*, *9*(2), 101–103. https://doi.org/10.1242/dmm.024547

Lapchak, P. A., Zhang, J. H., & Noble-Haeusslein, L. J. (2012). Rigor guidelines: Escalating stair and steps for effective Translational Research. *Translational Stroke Research*, *4*(3), 279–285. https://doi.org/10.1007/s12975-012-0209-2

Mlinarić, A., Horvat, M., & Šupak Smolčić, V. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia Medica*, *27*(3). https://doi.org/10.11613/bm.2017.030201

O'Connor, A. M., Totton, S. C., Cullen, J. N., Ramezani, M., Kalivarapu, V., Yuan, C., & Gilbert, S. B. (2018). The study design elements employed by researchers in preclinical animal experiments from two research domains and implications for automation of Systematic Reviews. *PLOS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0199441

Patel, D. N., Zahiri, K., Jimenez, J. C., Montenegro, G., & Mueller, A. (2022). Evaluating Study Design Rigor in Preclinical Cardiovascular Research: A Replication Study. https://doi.org/10.17605/OSF.IO/F4NH9

Provencher, S., Archer, S. L., Ramirez, F. D., Hibbert, B., Paulin, R., Boucherat, O., Lacasse, Y., & Bonnet, S. (2018). Standards and methodological rigor in Pulmonary Arterial Hypertension Preclinical and translational research. *Circulation Research*, *122*(7), 1021–1032. https://doi.org/10.1161/circresaha.117.312579

Ramirez, F. D., Motazedian, P., Jung, R. G., Di Santo, P., MacDonald, Z. D., Moreland, R., Simard, T., Clancy, A. A., Russo, J. J., Welch, V. A., Wells, G. A., & Hibbert, B. (2017). Methodological rigor in preclinical cardiovascular studies. *Circulation Research*, *120*(12), 1916–1926. https://doi.org/10.1161/circresaha.117.310628

Ramirez, F. D., Jung, R. G., Motazedian, P., Perry-Nguyen, D., Di Santo, P., MacDonald, Z., Clancy, A. A., Labinaz, A., Promislow, S., Simard, T., Provencher, S., Bonnet, S., Graham, I. D., Wells, G. A., & Hibbert, B. (2019). Journal initiatives to enhance preclinical research: Analyses of S*troke, Nature Medicine, Science Translational Medicine*. *Stroke*, *51*(1), 291–299. https://doi.org/10.1161/strokeaha.119.026564

Seok, J., Warren, H. S., Cuenca, A. G., Mindrinos, M. N., Baker, H. V., Xu, W., Richards, D. R., McDonald-Smith, G. P., Gao, H., Hennessy, L., Finnerty, C. C., López, C. M., Honari, S., Moore, E. E., Minei, J. P., Cuschieri, J., Bankey, P. E., Johnson, J. L., Sperry, J., … Wong, W. H. (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences*, *110*(9), 3507–3512. https://doi.org/10.1073/pnas.1222878110

Williams, J. L., Chu, H. (Cindy), Lown, M. K., Daniel, J., Meckl, R. D., Patel, D., & Ibrahim, R. (2022). Weaknesses in experimental design and reporting decrease the likelihood of reproducibility and generalization of recent cardiovascular research. *Cureus*. https://doi.org/10.7759/cureus.21086

Zahiri, K., Jimenez, I. C., Montenegro, G., Patel, D. N. & Mueller, A. (2022). Methodological Rigor in Cardiovascular Publications. https://osf.io/52Q6W/ 10.17605/OSF.IO/52Q6W

**Table I:** This table depicts the predetermined screening numerical codes and descriptions used for study analysis of methodological rigor across all studies.

| | Numerical Code | Definition | Chi Square True Or False |
|---|---|---|---|
| Study Type | 1 | Human-Animal Study | TRUE |
| | 0 | Animal-only Study | FALSE |
| Sex As A Biological Variable | 2 | Both sexes used for any experiment in the study | TRUE |
| | 1 | Both sexes not used for any experiment in the study | FALSE |
| | 0 | No reporting of sex of study subjects | FALSE |
| Randomization | 1 | Randomization (to any degree) used in any experiment in the study, or if randomization used in data analysis | TRUE |
| | 0 | No reporting of randomization | FALSE |
| Blinding | 1 | Blinding (to any degree) used in any experiment in the study or, or if blinding used in data analysis | TRUE |
| | 0 | No reporting of blinding | FALSE |
| Sample Size Estimation | 3 | Statistical analysis used to determine sample size for any experiment in the study | TRUE |
| | 2 | Other justification used to determine sample size for any experiment in the study | FALSE |
| | 1 | No sample size estimation done for any experiment in the study | FALSE |
| | 0 | No reporting of sample size estimation | FALSE |

**Table II:** This table lists the information collected during screening across all studies.

| | Definition |
|---|---|
| Topic | Short version of cardiovascular disease under study, including 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, hematologic disorders (including thrombosis), congenital heart disease, and heart development/repair/regeneration. Used "Other - [Details]" if none of these topics were applicable. First listed topic considered most relevant. |
| Species | List of non-human species used in study. If a combination of non-human species were used, then all species used were listed and separated by commas. |

**Table III:** These tables depict the descriptive statistics for SDEs analyzed for animal only studies vs animal or human substudies in human/animal studies.

**Both Sex Inclusion**

| | *Both Sexes (N, %)* | *Single Sex (N, %)* | *Not Reported (N, %)* |
|---|---|---|---|
| **Animal Only** | 28 (15.2) | 89 (48.4) | 67 (36.4) |
| **Human/Animal - Animal Substudy** | 20 (17.5) | 61 (53.5) | 33 (29) |
| **Human/Animal - Human Substudy** | 41 (36) | 10 (8.8) | 63 (55.2) |

**Randomization Inclusion**

| | *Randomization Used (N, %)* | *Randomization Not Used (N, %)* |
|---|---|---|
| **Animal Only** | 56 (30.4) | 128 (69.6) |
| **Human/Animal - Animal Substudy** | 41 (36) | 73 (64) |
| **Human/Animal - Human Substudy** | 7 (6) | 107 (94) |

**Blinding Inclusion**

| | *Blinding Used (N, %)* | *Blinding Not Used (N, %)* |
|---|---|---|
| **Animal Only** | 59 (32.1) | 125 (67.9) |
| **Human/Animal - Animal Substudy** | 28 (24.5) | 86 (75.5) |
| **Human/Animal - Human Substudy** | 13 (11.4) | 101 (88.6) |

**Sample Size Estimation Inclusion**

| | *Used Size Estimation (N, %)* | *Iustified Not Using Size Estimation (N, %)* | *Size Estimation Not Used (N, %)* | *Not Reported (N, %)* |
|---|---|---|---|---|
| **Animal Only** | 15 (8.2) | 9 (4.9) | 5 (2.7) | 155 (84.2) |
| **Human/Animal - Animal Substudy** | 8 (7) | 3 (2.6) | 6 (5.3) | 97 (85.1) |
| **Human/Animal - Human Substudy** | 2 (1.8) | 14 (12.3) | 3 (2.6) | 95 (83.3) |

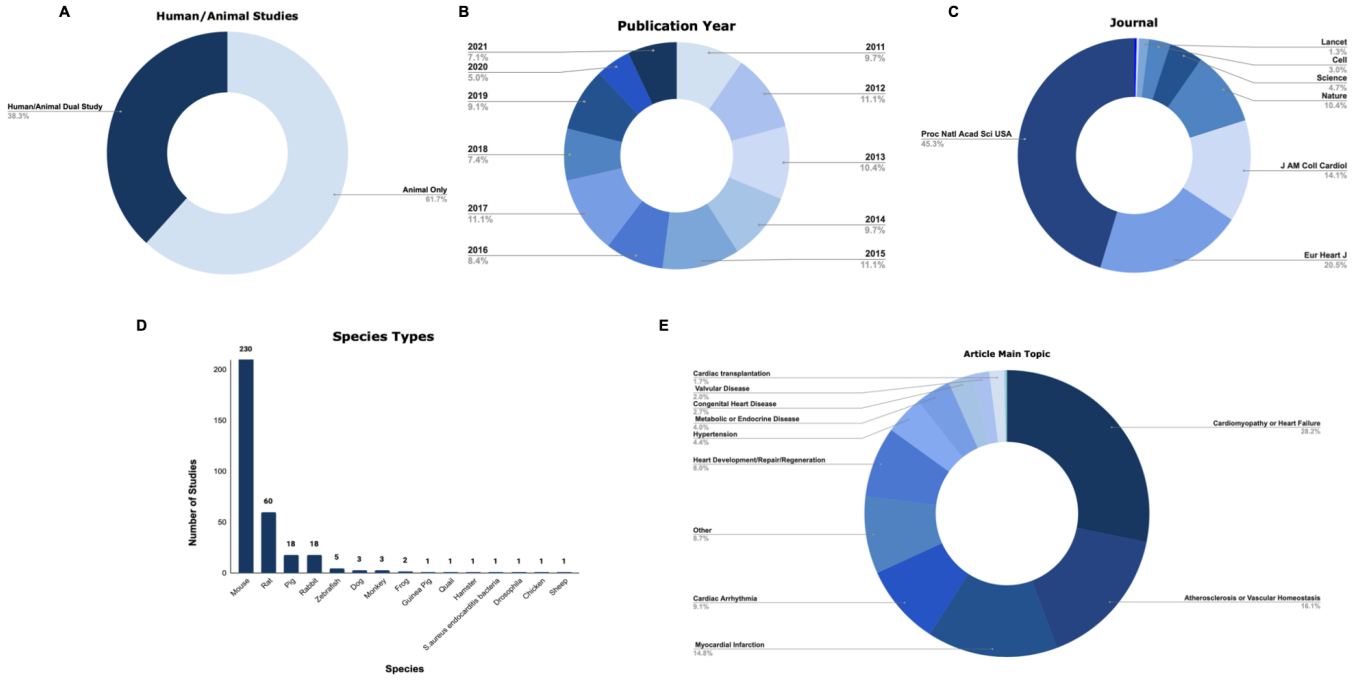


**Figure 1: There is a diverse range of preclinical cardiovascular research studies published over the last decade.**
**(A)** This figure depicts the proportion of studies included in our analysis were animal only or human/animal studies. (**B)** This figure depicts the proportions of studies included in our analysis from various publication years between 2011-2021. (**C)** This figure depicts the proportions of studies included in our analysis from each of nine high impact scientific journals. (**D)** This figure depicts the range of species used in the sample populations of studies included in our analysis. Mice and rats were the most commonly studied animal models across all studies. (**E)** This figure depicts the main range of topics the articles used in this study primarily investigated.

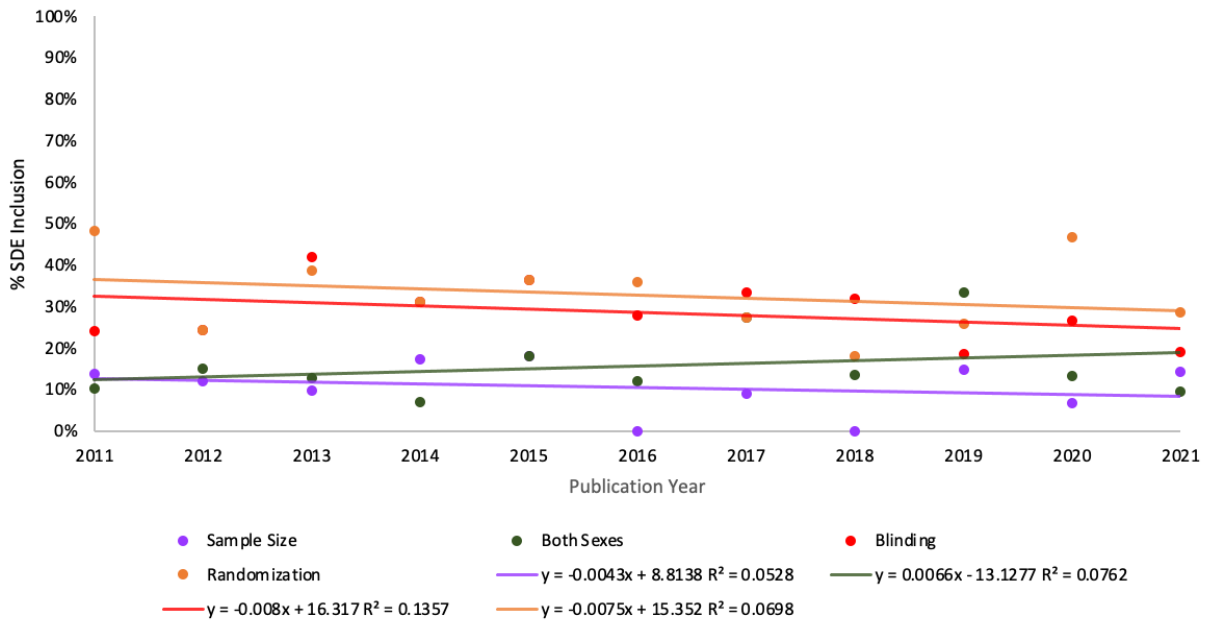


**Figure 2: Percentage of studies with SDE inclusion in preclinical cardiovascular research between 2011-2021 differed across the four SDEs screened.**

Percentages are reported for animal models used in all 298 eligible studies, regardless of whether the experimental design included human substudies. The percentage of studies including both biological sexes increased between 2011-2021, while the percentage of studies implementing blinding, randomization, and sample size estimation decreased between 2011-2021.

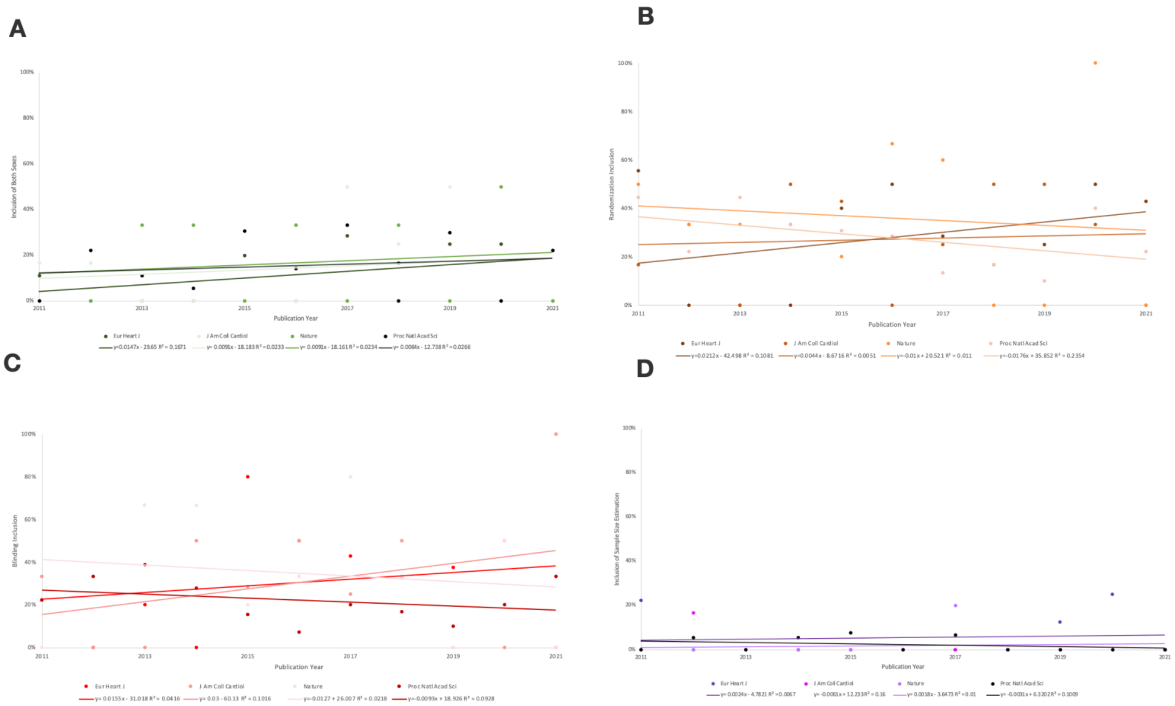


**Figure 3: The percentage of SDE inclusion in preclinical cardiovascular research between 2011-2021 generally varied across the four most prevalent journals.**
**(A) Both Sexes Analysis:** The percentage of studies including animals of both biological sexes generally increased across all four most journals. (**B) Randomization Analysis:** The percentage of studies using randomization generally increased in studies from *Eur Heart J* and *J Am Coll Cardiol*, but decreased across studies from *Proc Natl Acad Sci* and *Nature*. **(C) Blinding Analysis:** Similar results for randomization inclusion apply to blinding inclusion. (**D) Sample Size Analysis:** The percentage of studies implementing statistical sample size estimations increased in studies screened from *Eur Heart J* and *Nature*, and decreased across those screened from *Proc Natl Acad Sci* and *J Am Coll Cardiol.*

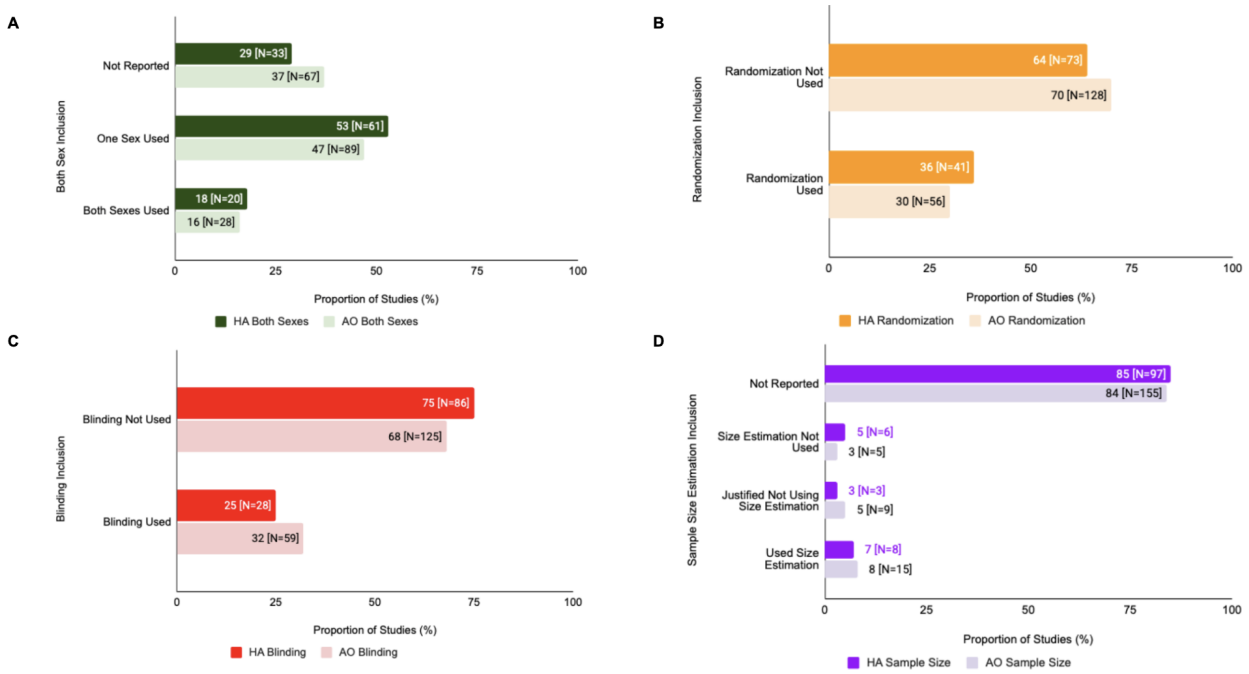


**Figure 4: SDE inclusion varied in Human/Animal (HA) vs Animal Only (AO) studies.**
**(A) Both Sexes Analysis:** This figure depicts the inclusion of both sexes vs single sex in human/animal studies vs animal only studies. **(B) Randomization Analysis:** This figure depicts the inclusion of randomization in human/animal studies vs animal only studies. **(C) Blinding Analysis:** This figure depicts the inclusion of blinding in human/animal studies vs animal only studies. (**D) Sample Size Analysis:** This figure depicts the inclusion of sample size estimation in human/animal studies vs animal only studies.

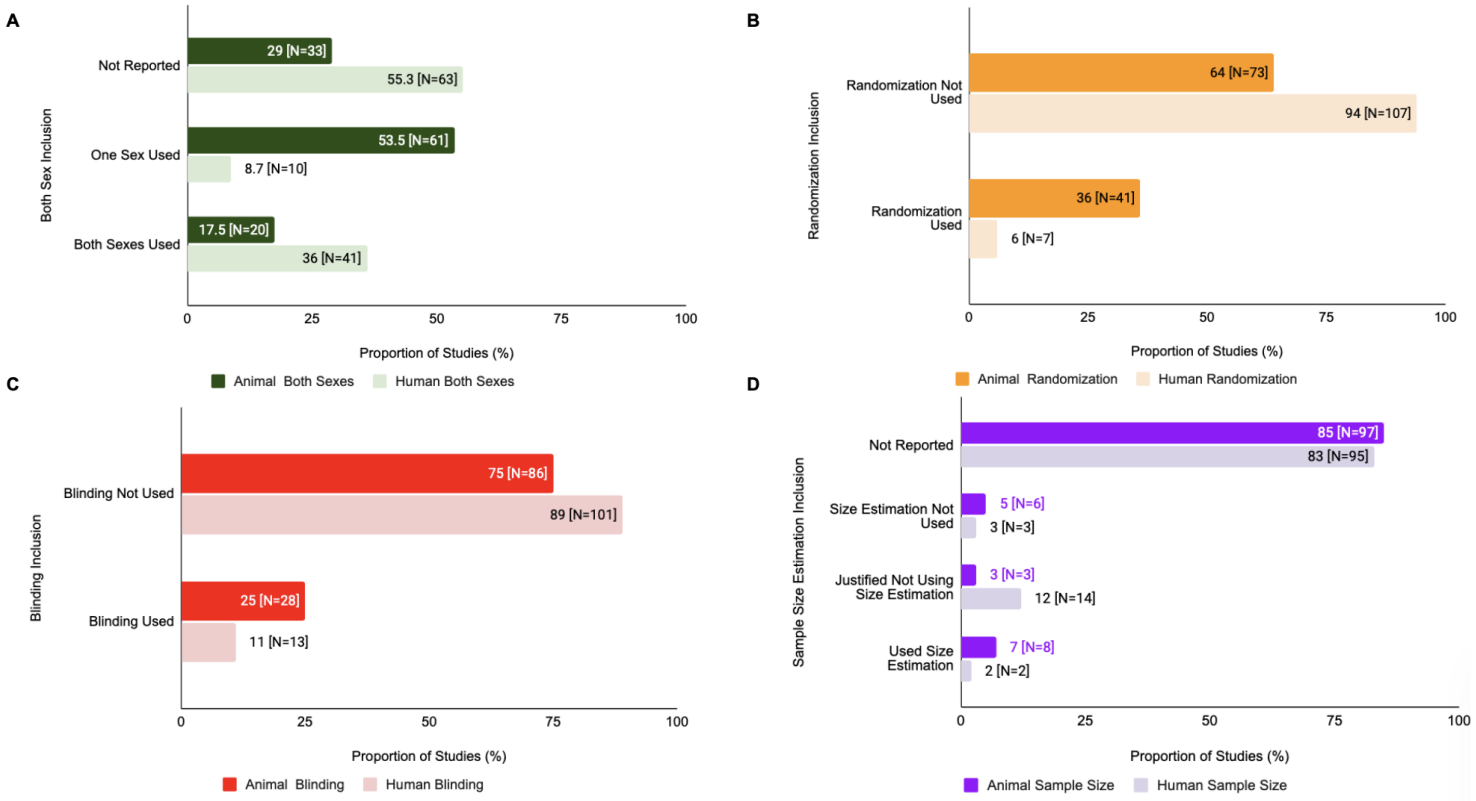


**Figure 5: SDE inclusion for animal vs human substudies also varied within human/animal studies.**
**(A) Human/Animal Studies: Both Sexes Analysis.** This figure depicts the inclusion of both sexes vs single sex in human vs animal substudies in human/animal studies. **(B) Human/Animal Studies: Randomization Analysis.** This figure depicts the inclusion of randomization in human vs animal substudies in human/animal studies.**(C) Human/Animal Studies: Blinding Analysis.** This figure depicts the inclusion of blinding in human vs animal substudies in human/animal studies. **(D) Human/Animal Studies: Sample Size Analysis** This figure depicts the inclusion of sample size estimation in human vs animal substudies in human/animal studies.

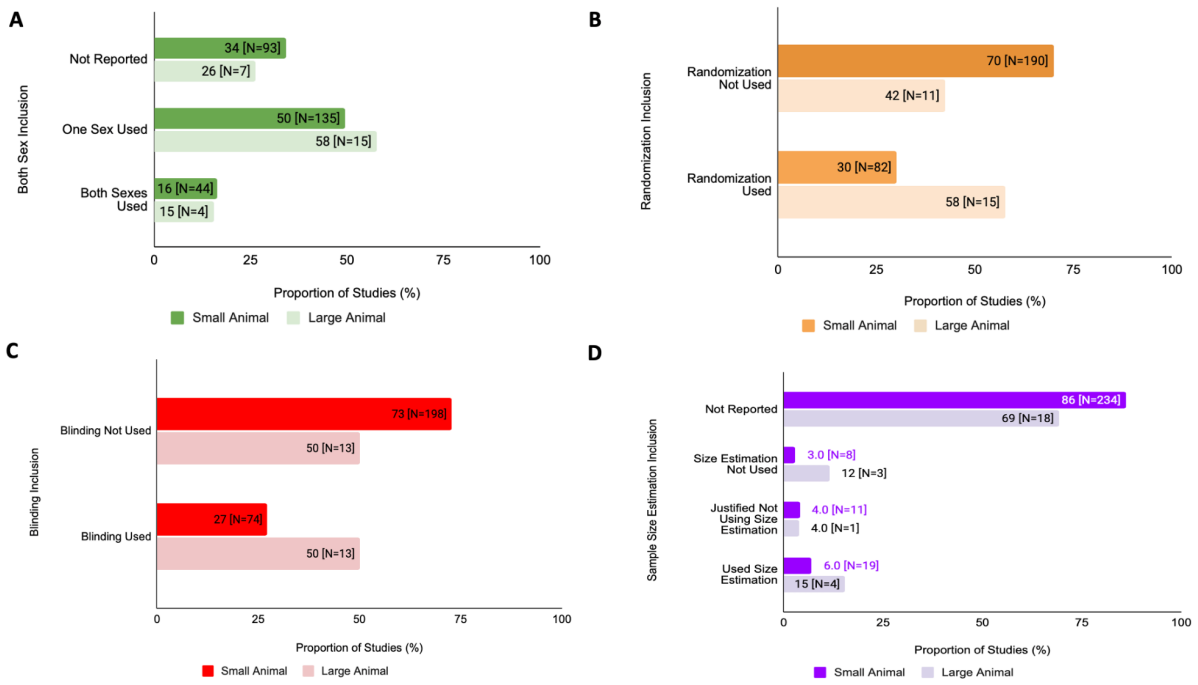


**Figure 6: Proportion of SDE inclusion in preclinical cardiovascular studies differs between studies consisting of either small or large animal substudies from 2011-2021.**
**(A) Sex as a Biological Variable Analysis:** This figure compares the proportion of articles that reported incorporation of both sexes, a single sex, or omission of reporting for small or large animals. (**B) Randomization Analysis:** This figure compares the proportion of articles that reported randomization for small or large animals. (**C) Blinding Analysis:** This figure compares the proportion of articles that reported randomization for small or large animals. **(D) Sample Size Estimation Analysis:** This figure compares the proportion of studies that reported use of sample size estimation for small or large animals.

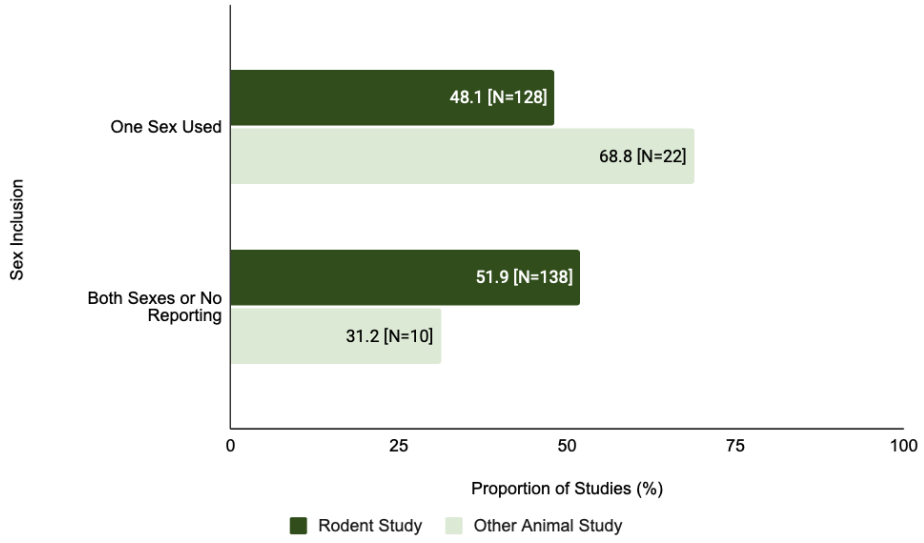


**Figure 7: Sex inclusion in the study population varied in rodent vs other animal studies.**
This figure compares the proportion of rodent studies and non-rodent animal studies that were single sex studies.