CORP: Practical Tools for Improving Experimental Design and Reporting of Laboratory Studies of Cardiovascular Physiology and Metabolism

Kate L. Weeks[1, 2], Darren C. Henstridge[1], Agus Salim[1,3], Jonathan E. Shaw[1], Thomas H. Marwick[1], Julie R. McMullen[1,2]

[1] Baker Heart and Diabetes Institute, Melbourne, VIC 3004, Australia

[2] Department of Diabetes, Central Clinical School, Monash University, Clayton, VIC 3800, Australia

[3] Department of Mathematics and Statistics, PS2 Building, La Trobe University VIC 3086, Australia

Running Head: Experimental design: Survey, training & practical tools

Address for Correspondence:

Julie R. McMullen

Baker Heart and Diabetes Institute

Melbourne, VIC 3004

Australia

Tel: +61 (0) 3 8532 1194

Email: julie.mcmullen@baker.edu.au

Abstract

An exercise was undertaken to improve the quality of animal cardiac and metabolic studies at the Baker Heart and Diabetes Institute, Australia. The exercise consisted of: i) a short survey to acquire baseline data on current practices regarding the conduct of animal studies; ii) a series of presentations for promoting awareness, and providing advice and practical tools for improving experimental design; iii) a follow-up survey 12 months later to assess whether practices had changed. The surveys were compulsory for responsible investigators (N=16; paired data presented). Other investigators named on animal ethics applications were encouraged to participate (2017-total of 36 investigators; 2018-37 investigators). The major findings to come from the exercise included: 1) a willingness of investigators to make changes when provided with knowledge/tools and solutions which were relatively simple to implement (e.g. proportion of responsible investigators showing improved practices using a structured method for randomization was 0.44, 95% CI (0.19; 0.70), P=0.003, and de-identifying drugs/interventions was 0.40, 95% CI (0.12; 0.68), P=0.010); 2) resistance to change if this involved more personnel and time (e.g. as required for allocation concealment); and 3) evidence that changes to long term practices ("habits") require time and follow-up. Improved practices could be verified based on changes in reporting within publications, or documented evidence provided during laboratory visits. In summary, this exercise resulted in changed attitudes, practices, and reporting but continued follow-up, monitoring, and incentives are required. Efforts to improve experimental rigor will reduce bias and will lead to findings with the greatest translational potential.

Keywords: Experimental design, preclinical, randomization, allocation concealment, blinding

New & Noteworthy

The goal of this exercise was to encourage preclinical researchers to improve the quality of their cardiac and metabolic animal studies by: i) increasing awareness of concerns which can arise from sub-optimal experimental designs, ii) providing knowledge, tools and templates to overcome bias, and iii) conducting two short surveys over 12 months to monitor change. Improved practices were identified for the uptake of structured methods for randomization, and de-identifying interventions/drugs.

Introduction

Over the last decade there have been increasing concerns around the reproducibility of preclinical animal research and the lack of detail provided in publications in relation to experimental design (7-9, 29, 48). Leaders in the field have highlighted that there is a natural tendency for scientists to "see" results and report data in a manner that confirms their original hypothesis, and that failure to control for bias, and rationalizing behaviors such as P-hacking, can lead to results and conclusions which are less likely to be replicated (19, 35, 42, 45). The ARRIVE guidelines (Animal Research: Reporting In Vivo Experiments) have been developed to encourage improvements in experimental design and reporting (28). However, while these guidelines have received endorsement from numerous journals, universities, research institutes and funding agencies, there has been concern that the guidelines have not been adequately implemented in practice (6). This is not helped by a system in which individuals are judged on the numbers of publications with novel and positive results published in high impact factor journals rather than on the number and quality of publications with robust experimental design (44).

Investigators undertaking animal research are not typically provided with formal training in regard to experimental design. Furthermore, until relatively recently, there have been few mandatory requirements for reporting details such as methods used for randomization, allocation concealment, and blinding. It is now well recognised that studies without formal procedures for limiting bias have a high potential of leading to results and conclusions that are exaggerated (7-9, 29, 48), particularly those claiming that a particular drug/intervention provided significant benefit in a disease animal model (36).

Here we describe an exercise in which we sought to improve the quality of animal cardiac and metabolic studies at a research institute in Australia. The exercise consisted of: i) a short survey to acquire baseline data on current practices regarding the conduct of animal studies; ii) a series of presentations for promoting awareness, and providing advice and practical tools for improving experimental design; and iii) a follow-up survey 12 months later to assess whether practices had changed. Throughout the process, investigators were assured that this was not an exercise with repercussions on performance, but rather a process to obtain honest answers from investigators to understand any barriers/challenges hindering laboratories from undertaking more formal

procedures for reducing bias. It was emphasized that honest answers would provide an opportunity to identify solutions for improving the quality of preclinical animal research.

Materials and Methods

Study overview

The Baker Heart and Diabetes Institute (Melbourne, Australia) undertook an exercise to evaluate and improve practices encompassing the performance of preclinical research. An overview of the entire process is presented in Table 1 and included 1) presentations, 2) two short surveys rolled out approximately 12 months apart, 3) practical tools for encouraging improved practices, and 4) follow-up from senior members of the Institute, i.e. laboratory visits from the Institute Director (Prof Thomas Marwick) and the Head of the Basic Science Domain/Division (A/Prof Julie McMullen).

Surveys

Two short surveys (designed to be completed in 10-12 min or less) were developed and delivered via an online form on the Institute intranet. The surveys were provided to research staff and graduate students undertaking animal research in June 2017 and June/July 2018.

Inclusion criteria: The surveys were compulsory for responsible investigators from the Baker Institute with an active animal ethics application. A list of responsible investigators was provided by the Animal Ethics Officer (Alfred Health and Education Precinct Animal Ethics Committee).

The 2017 survey collected baseline data on existing practices within the institute in respect to the conduct of animal studies, with a focus on experimental design (e.g. blinding, randomization and allocation concealment). The aim of the 2018 survey was to determine whether practices had improved over a period of 12 months. Questions from both surveys are presented within Source Data (Appendix I and II, available at https://figshare.com/s/1788601bf6606df6f5a5).

In this paper, we present the responses to questions within the 2017 and 2018 surveys which were specifically related to: a) Pre-specified criteria for inclusions/exclusions; b) declaration of primary and secondary endpoints; c) sample size calculations; d) randomization of animals and method used; e) allocation concealment, de-identifying interventions, and blinding; and f) attitudes towards the quality and rigor of animal studies being undertaken. The main goal was to understand whether practices and behaviors had changed over a period of 12 months.

Pre-specified criteria for inclusions/exclusions. Within the research fields of cardiac biology, diabetes and obesity, it is not uncommon for investigators to exclude animals based on: i) unsuccessful cardiac surgery; ii) mice not becoming diabetic when administered streptozotocin (STZ) based on blood glucose levels; iii) inadequate weight gain on a high fat diet and/or skin lesions (prior to an intervention); iv) mice not weight-matched prior to an intervention assessing weight gain/metabolism. These criteria should be considered prior to experimentation and reported in publications. Depending on the study design, animals may be excluded: i) prior to entering a study e.g. a runt in a litter with a particularly low body weight that could not be weight-matched for a metabolic study; ii) after an intervention (e.g. cardiac surgery/STZ injection) but prior to a treatment e.g. due to surgical complication/low blood glucose not meeting the definition of diabetes; iii) after a treatment but prior to study endpoint due to animal welfare issues; or iv) after study endpoint once molecular analyses of tissue have been performed e.g. a study designed to assess gene therapy in the heart but transduction of the heart with gene therapy was unsuccessful/sub-optimal to test the hypothesis. To encourage investigators to track and report exclusions of animals in future work, investigators were asked if they would consider using a template/flowchart (an example was provided within the survey) which had been proposed by Drucker (18), and which he designated the Consolidated Standards of Animal Experiment ReporTing (CONSAERT) flow diagram.

Primary and secondary endpoints. In clinical research and journals publishing results of clinical trials it is mandatory/compulsory to define and report primary endpoints, which should be finalized before study completion and data analysis. This is typically not a requirement for preclinical research, and reporting of this information has been far less common, particularly for discovery research. The purpose of incorporating questions related to defined endpoints was to encourage preclinical researchers to consider stating primary endpoints for interventional studies in future, e.g. testing of a drug or genetic intervention in a mouse disease model on a cardiac function parameter such as fractional shortening by echocardiography. This provides differentiation between more robust findings relating to primary outcomes and hypothesis-generating outcomes when the primary outcome is not proven.

Sample size calculations. Power calculations are routinely part of clinical, but not laboratory, research culture. Nonetheless, their use is important to minimize the risk of type 2 error, as well as ensuring the ethical use of animals.

Randomization of animals and method used. The importance of a structured or systematic approach for randomization is well recognized in the clinic (55), and the need for randomization in animal studies has gained attention more recently (25). Within preclinical research, it is not uncommon for investigators to use unstructured methods or “haphazard selection” when assigning animals to different experimental groups (29).

Allocation concealment, de-identifying interventions, and blinding. It is recognised that when researchers are aware of experimental groups and treatments, unconscious bias may result in researchers caring for animals differently, and subsequently in larger effect sizes (37, 53). The differences and importance of allocation concealment and blinding for animal studies have been well described on the NC3Rs Experimental Design Assistant website (https://eda.nc3rs.org.uk/experimental-design-blinding). “Allocation concealment refers to concealing the allocation sequence (the treatment to be allocated to each individual animal) from the people assigning the animals to intervention groups, until the moment of assignment. Together with randomization, allocation concealment helps minimise selection bias, which introduces systematic differences in the characteristics of animals allocated to treatment groups. Allocation concealment enables blinding; for adequate blinding, the allocation sequence should not be revealed to the people conducting the experiment until the data has been analysed”.

A number of our preclinical researchers perform intervention studies in disease and/or genetic animal models (e.g. surgical for heart disease models, STZ-induced diabetes, high fat diet for obesity), with drugs, gene therapies (e.g. adeno-associated virus; AAV), and different dietary approaches. To gain further insight into what components of the experimental design were being concealed and blinded, and to determine the potential willingness of investigators to incorporate more formal strategies for allocation concealment, follow-up questions were asked. For example: “For drug/AAV/diet studies, are interventions labelled ‘A’ and ‘B’ or equivalent? If no, would you consider this for future studies?”

Presentations and tools

One hour educational sessions were presented to animal users throughout the exercise as outlined in Table 1. Sessions were delivered by the Domain Head of Basic Science and senior researchers (Dr Kate Weeks and Dr Darren Henstridge) who were actively conducting animal studies.

Flow charts/templates were provided to investigators based and adapted from the CONSAERT flow chart proposed by Drucker (18); see Source Data (Appendix III available at https://figshare.com/s/1788601bf6606df6f5a5).

Data collection

The surveys were provided via a web interface on the institute's intranet. Correspondence about the 2017 survey and the importance and rationale for conducting the survey was sent to all research staff and students via email from the Institute Director. Research staff listed as the responsible investigator on at least one active animal ethics application were informed that the survey was compulsory and received a follow-up email from the Head of the Basic Science Domain. To encourage investigators to respond openly and without reservations, examples of responses from the Head of the Basic Science Domain's laboratory were provided to highlight that the goal was to obtain honest responses, not 'perfect' responses. The 2018 survey was compulsory for all responsible investigators who completed the survey in 2017. Paired responses from 6 responsible investigators could not be obtained (2 overseas during at least 1 survey, 3 investigators leaving the institute in 2018, 1 non-responder). Other scientists undertaking animal research were encouraged to complete the survey. Two reminder emails were sent by the Head of the Basic Science Domain.

Data from both surveys were exported into Excel and the collated results graphed in GraphPad Prism 7.03. The data for the responsible investigators are presented as bar graphs and line graphs to demonstrate how practices of individuals changed. The raw data were checked by more than one author to ensure reliability.

Statistics

The two areas in which changes in practice were considered most likely to occur within 12 months were: 1) the uptake of using a structured method for randomization; and 2) de-identifying drugs/treatments by labelling them “A” and “B” or an equivalent. The Wilcoxon Sign-Rank Test for paired data was performed on the paired data (before and after) for these primary outcomes with a type I error set at 5%. The method of Agresti and Min (1) was used to calculate the proportion of investigators showing improved practices. This method takes into consideration bias when sample size is small (1). Other outcomes were not subjected to statistical testing, because of the combination of small sample size and to avoid the risk of increasing type 1 error due to multiplicity of outcomes. Those results should therefore be seen only as hypothesis generating.

Results

Study population

In total, 36 investigators completed the animal user survey in 2017 and 37 investigators completed the survey in 2018. The main analysis presents paired results from the 16 researchers listed as responsible/lead investigators on animal ethics applications. This included eight PIs/Lab Heads and eight Group Leaders/Research Officers. The results from all investigators (36 in 2017, 37 in 2018) were comparable with the findings from the 16 responsible investigators (see Data Supplement for results from all investigators: available at https://figshare.com/s/7d252ccaa26110170985).

All investigators were undertaking studies with genetic mouse models and/or mouse studies involving interventions e.g. drugs, gene therapy, diet interventions.

Pre-specified criteria for inclusions/exclusions of animals

In 2017, only 5 out of 16 responsible investigators were using pre-specified criteria for inclusions or exclusions of animals (e.g. unsuccessful cardiac surgery, mice not diabetic based on blood glucose levels) in 100% of their experiments (Fig 1A). In 2018, 9 out of 16 investigators reported using pre-specified criteria in all experiments, with the remaining 7 investigators reporting such use

in at least 50% of experiments (Fig 1A). Investigators were also asked whether pre-specified criteria were reported in publications and how often excluded animals were reported in publications. For both questions there was a spread of responses and no clear positive change over the 12 month period (Fig 1B and C). The most common reason for not including pre-specified criteria in publications was because this was not a requirement of journals. This also appeared to explain why some investigators reported pre-specified criteria and excluded animals in publications in 2017 but not 2018 i.e. this was the requirement of a journal when submitting in 2017 but not the requirement of another journal when submitting in 2018. One investigator provided answers of 100% in 2017 and 0% in 2018 (Fig 1B and C). In this case, the investigator had not published an animal study during the 12 month period. The inclusion criterion for responsible investigators to participate in the survey was to be listed as the responsible investigator on an active animal ethics application. There was not a requirement to publish a paper over the 12 month interval. The response from this one investigator (illustrated by a dotted line, Fig 1B and C) highlights a limitation of the survey questions. In this case, the reported measure of 0% provided by the investigator does not represent 0%. Future surveys should allow for a response of “Not applicable” and include a follow-up question to explain why “Not applicable” was selected. It should be noted that all other investigators who reported 0% had published papers over the 12 month period.

An example of a flowchart template was provided to assist and encourage researchers to track and report exclusions of animals (Fig 2A, modified version of CONSAERT (18)). In 2017, the majority of responsible investigators indicated they would use the template (Fig 2B). In 2018, 7 of 16 responsible investigators had begun using the template (Fig 2C). Reasons provided for not using the template included: i) new studies had not yet begun but the templates will be used for new animal studies; ii) the laboratory had created other templates or spreadsheets.

Primary and secondary endpoints

The number of responsible investigators using primary and secondary endpoints 100% of the time (i.e. having a key parameter for judging study outcome e.g. a change in cardiac function by echocardiography) increased over the 12 month period from 6 to 11 of 16 (Fig 3A). There may be situations (e.g. exploratory studies) in which primary endpoints and outcomes are specified after

the study is underway. However, a similar trend was observed for investigators specifying endpoints prior to commencement of experimentation (Fig 3B). There was minimal change in the number of investigators stating pre-specified endpoints in publications (Fig 3C). Reasons for not stating specific endpoints included: 1) journals had not requested this information; 2) for discovery research it is uncertain what parameters/endpoints might change, so there are multiple endpoints (i.e. exploratory endpoints; N.B. potential driver for false positive results given 1:20 endpoints would be positive by chance; conditional on the null hypothesis being true); 3) for new studies it can be difficult to predict effect sizes. A number of investigators noted they were planning to include this information in future publications, but they had not published any new studies since the recommendations were made in 2017. Two investigators indicated that pre-specified endpoints were reported in publications 100% of the time in 2017 but 0% in 2018 (Fig 3C). One investigator had not published an animal study during the 12 month period (response illustrated by a dotted line). The other investigator published in journals in which this information was not requested.

Sample size calculations

In 2017, 9 of 16 investigators were using power calculations 100% of the time to determine sample size. This increased to 11 of 16 investigators in 2018 (Fig 4). Though, of note, some of the investigators who indicated they were using power calculations 100% of the time in the 2017 survey, were performing power calculations only 50% or 75% of the time in 2018. The most common reason investigators gave for not using a power calculation was that basic science studies are often exploratory in nature. A pilot study is often required before effect sizes and variability in the measure of a parameter can be determined.

Randomization of animals and method used

In the 2017 survey, the majority of investigators claimed to be randomizing mice to a specific treatment or intervention (Fig 5A). However, upon asking what method of randomization was being used in a follow-up question, it became apparent that all of the 16 responsible investigators were using no structured method for randomization (Fig 5B, white bar). Within the 2017 survey we further asked: “If you are using “no structured method” for randomization, would you consider using

a structured method?". The majority of investigators (15 of 16) responded "yes" to this question. Given that there is the potential for bias to influence the allocation of animals into a specific group, information sessions to describe the pitfalls of not using a structured method were delivered to researchers, together with tools for performing randomization using computer-based methods e.g. NC3Rs Experimental Design Assistant, GraphPad QuickCalcs, RAND function in Excel; see Source Data (Appendix IV available at https://figshare.com/s/1788601bf6606df6f5a5). The uptake of a formal method for randomization was an area in which we considered changes in practice could occur relatively quickly (i.e. within 12 months) because it was a strategy which was relatively quick and easy. In 2018, all 16 responsible investigators were randomizing mice to a treatment/intervention 100% of the time (Fig 5A). Half of the investigators were using a computer-generated method or blind sealed envelope/other structured method (pulling numbers out of a hat) in 2018 (Fig 5B). The proportion of investigators demonstrating an improved practice (i.e. going from an unstructured method to structured: computer-generated/sealed envelope/numbers from a hat) was estimated to be 0.44 (95% CI: 0.19; 0.70; P=0.003). For the other half of investigators still randomizing mice with no structured method, reasons for this were requested. For most investigators, it was noted that studies/experiments had begun prior to the information session on new tools, and that for upcoming studies a structured method would be used. In some cases there were misconceptions about difficulties randomizing mice from small batches and study designs requiring mice to be body weight matched. In a follow-up session, examples of how structured randomization could be used in these situations were provided, e.g. stratified randomization for body weight matching (see Source Data, Appendix IV available at https://figshare.com/s/1788601bf6606df6f5a5).

Allocation concealment, de-identifying interventions, and blinding

Allocation concealment: The purpose of the first survey question was to assess the overall potential of performance bias within a study e.g. investigators providing different degrees of care to animals based on treatment or genotype. In 2017, the majority of investigators (13 of 16) indicated they were performing allocation concealment 50% or less of the time (Fig 6A). By contrast, in 2018,

311 the majority of investigators (11 of 16) were performing allocation concealment 50-100% of the
312 time (Fig 6A).
313 De-identifying interventions: Performance bias can be eliminated or reduced by labelling
314 drugs/interventions 'A' and 'B' or an equivalent. In 2017, most researchers were not blinded to the
315 intervention assigned to an animal group (Fig 6B; 13 of 16; white bars). However, 15 of 16
316 investigators indicated a willingness to consider de-identifying interventions or drugs in future
317 studies. Thus, this was another area in which significant improvements in practice were
318 considered possible within a 12 month period. At the time of the 2018 survey, 50% of investigators
319 were de-identifying interventions (Fig 6B, black bars). The proportion of responsible scientists
320 showing an improved practice was 0.40 (95% CI: 0.12; 0.68; P=0.010). Researchers were also
321 asked if they incorporated blinding into other components of their studies. In most cases this
322 included blinding of measurements and analysis of tissues post-animal experimentation (e.g.
323 molecular and histological analyses). All investigators incorporated blinding into some aspect of
324 their studies in 2017 (Fig 6C; 25-100%). In 2018, 8 of 16 responsible investigators were including
325 blinding 100% of the time (Fig 6C). Reported challenges associated with allocation concealment,
326 de-identifying drugs/interventions, and blinding included:
327 a) It can be difficult to conceal diets in diet intervention studies (e.g. high fat chow is a different
328 color to standard chow). However, genotypes/treatments would still be blinded within these
329 studies. If interventions are unable to be blinded, then animal identification numbers are
330 reassigned at the conclusion of the study to 're-blind' for analysis.
331 b) Lack of personnel- In some studies, drugs need to be mixed with vehicle just prior to
332 administration. Many labs only have one technician so the same technician will mix the
333 drug followed by immediate delivery to the mice.
334 c) Different monitoring/care requirements can be required for different interventions e.g.
335 diabetic animals require extra/daily care and have to have additional labelling on cages;
336 animals on a high fat diet require additional monitoring for skin lesions.
337 Practical solutions were provided to researchers in an information session. Examples are provided
338 in Table 2.
339

Quality and rigor of animal studies: Changes in attitudes, practices and uptake of tools

In 2017, 15 of 16 responsible investigators indicated that the quality and rigor of their animal studies could be improved (Fig 7A). In the 2018 survey, respondents were asked whether they perceived an improvement in their animal studies over the past 12 months. Fifteen of 16 investigators indicated that the quality and rigor of their animal studies had improved (Fig 7B), and 13 of 16 indicated they had changed practices regarding randomization, blinding and/or animal reporting (Fig 7C). Participants were asked to provide specific examples. These included: using a structured method for randomization, additional blinding, allocation concealment (e.g. labelling drugs “A” and “B”), more sample size calculations, and providing additional information in publications.

Discussion

The primary goal of this exercise was to encourage preclinical researchers to improve the quality of their cardiac and metabolic animal studies by incorporating strategies and protocols for the removal or reduction of bias. This was achieved by: i) increasing awareness of concerns which can arise from experimental designs that do not incorporate measures to exclude/limit bias; ii) providing knowledge, tools and templates to overcome bias/promote accurate reporting; and iii) rolling out two short surveys approximately 12 months apart to monitor any changes in practice and behavior. The major findings to come from the exercise included: 1) a willingness of investigators to make change when provided with knowledge and tools/solutions which were relatively simple to implement; 2) resistance to change if this involved more personnel and time; and 3) evidence that changes to long term practices (“habits” acquired over years) require time, follow-up, and incentives/mandatory requirements.

The most significant finding and change in practice identified by the survey was in regard to the randomization of animals. In 2017, most investigators were using unstructured methods for randomization. After highlighting the potential of unconscious bias to have an impact on the randomization process and providing information and tools for using formal methods, the number of people using a structured method in 2018 had increased to 50%. The options provided to

researchers included computer-generated approaches (highly encouraged: NC3Rs Experimental Design Assistant, GraphPad QuickCalcs, the RAND function in Excel), blind-sealed envelope, numbers from a hat, and flipping a coin or rolling a dice (though not encouraged: only suitable for large sample sizes). Uptake of these methods advanced relatively quickly and most investigators not currently using a structured method were planning to do so for future studies.

In general, the majority of investigators were willing to change practices when provided with the relevant information and tools, and when changes were easy to implement. Strategies requiring more personnel, e.g. for allocation concealment, were more challenging, particularly for small laboratories, to adopt. It is recognised that clinical trials and animal studies with inadequate or inappropriate methods for allocation concealment can overestimate treatment effects (37, 53). Thus, it is important allocation concealment is encouraged. Within an information session it was highlighted that some smaller laboratories were managing the problem of limited staff for allocation concealment and blinding by involving personnel from other laboratories. This was viewed as a practical solution for laboratories that work well together, but can be more challenging for others.

We and others have recognised that significant changes to experimental design and practices are likely to require monitoring, incentives and pressure from multiple groups, including journals, funding agencies, research institutes, universities, and animal ethics committees (3, 8, 13, 22, 23, 27). It has been reported that while researchers may endorse or subscribe to practices or changes in behavior, this may not correlate with actual behaviors in practice (5). Granting agencies within Australia (e.g. National Health and Medical Research Council) and overseas (e.g. National Institutes of Health, Wellcome Trust) have recognized the concern (14) and are encouraging enhanced reproducibility by incorporating relevant elements of experimental design (randomization, allocation concealment, blinding) within the assessment criteria of research proposals (16). Our training sessions also highlighted the more stringent experimental design and reporting requirements of journals and granting agencies as an incentive to encourage change. However, a requirement for explicit reporting from journals of specific experimental design details (e.g. method of randomization, de-identification of drugs) is likely to be necessary to drive

significant changes in practice. In undertaking a general review of recent publications from Am J Physiology- Heart and Circulation Physiology (using the search term “random”; Sept 2018-May 2019), a number of investigators noted that animals had been randomized (12, 20, 21, 26, 30, 38, 40, 41, 43, 50, 51, 54, 58). However, in the majority of cases, the method of randomization was not specifically stated. The exception to this was some examples of studies in large animals and humans (2, 31, 47, 49, 52). In these studies, structured methods of randomization had been included (e.g. computer-generated, random number table, stratified block randomization).

Within our exercise, follow-up and monitoring progress after and between surveys was considered very important. This consisted of information sessions, distribution of relatively simple tools (e.g. for randomization and tracking animal exclusions), and laboratory visits from the Institute Director and Domain Head of Basic Science. The online provision of tools, templates, and guidelines for many aspects of preclinical research is becoming increasingly common and considered a valuable resource. Many journals including *Am J Physiol Heart Circ Physiol* have been leading the way in providing guidelines, e.g. on data visualization, reporting statistics, experimental animal models of cardiac disease, measuring cardiac physiology in mice, formalized training for subjective measures including echocardiography, validation and correct use of reagents e.g. antibodies (11, 15, 17, 22, 24, 32-34, 56, 57).

Providing a balanced perspective on the concerns of irreproducible research was considered another key aspect of this exercise. As highlighted by others in editorials and reviews, discovery science is exploratory in nature (22). For such research, pre-specifying endpoints is not always possible, and tentative conclusions can still be beneficial to the scientific community and lead to important follow-up studies (22). Throughout the process, we were transparent about the potential deterrents of more rigid experimental designs, but also highlighted the benefits (Table 3). One issue is the acknowledgement that carefully conducted studies typically produce smaller effect sizes on average (4, 37). Although these smaller effect sizes are more accurate, they can make papers more difficult to publish in high impact journals because of the bias towards "breakthrough" and “exciting” results. Paradoxically, investigators who pursue this more careful approach could be

out-competed by less thorough researchers who win more funding on the back of their high impact (but sometimes less rigorous) publications. Though, of note, many high impact journals are now ensuring authors meet more rigorous checklists. Regardless, widespread and significant changes in practice are likely to require funders being made aware of the need to reward research quality regardless of the study outcomes.

Limitations of the study: This was a self-reported study which makes verification of data challenging. To check whether practices had indeed changed or were changing, the Institute Director and Domain Head of Basic Science visited laboratories during and after the exercise. In some cases, responses could also be verified by independently checking publications from the laboratory before 2017 and after 2018, e.g. use of animal templates, reporting of animal exclusions (10, 39, 46). However, in other cases this was not possible because the period of time between initial manuscript submission and acceptance/publication can often exceed 12 months. The scenario of an investigator not publishing an animal study over 12 months had not been considered. In future, an additional option of “Not applicable” should be available for selection. Another limitation is that this was a single site study with a relatively small cohort of responsible investigators. Whether this same exercise rolled out in an environment with larger numbers of animal users (larger institutes and universities) would identify changes in practice is currently unclear.

Future directions: When rolling out an exercise to improve practices and culture, it is important not to lose momentum, to reinforce desired practices among early adopters of change, and to promote change to more reluctant investigators. This could be achieved by a number of means:

a) Repeating the survey in future years to monitor the long term effects

b) Incorporating additional aspects of design and reproducibility into Institutional Animal Care and Use Committee (IACUC)/Animal ethics committee (AEC) applications, i.e. if defined randomization protocols and blinded treatments become a requirement in animal protocols, it is likely to accelerate implementation of better practices for improvement of better animal study design and reporting.

c) Inducting new staff/students on aspects of good experimental design and practice so they are indoctrinated into that culture from the start.

Summary: The approach we undertook to encourage improvements in preclinical cardiac and metabolic animal research included: 1) making researchers aware of the concerns; 2) providing knowledge, tools, skills and training to address the concerns; and 3) follow-up to monitor and encourage changed practices. This exercise resulted in changed attitudes, practices, and reporting. However, further improvements are needed and this will require continued follow-up, monitoring, and incentives.

Acknowledgements

The authors wish to acknowledge Leonie Cullen and Eliana Stanziano for assistance uploading and trialling surveys on the intranet and collating data from the survey, Eliana Stanziano and Celeste MK Tai for assistance with graphs and literature searches, Leia Demtschyna for providing a list of responsible investigators, and Prof David Howells (Uni of Tasmania) for delivering a seminar during the exercise and for helpful discussions. Current address for DCH: College of Health and Medicine, School of Health Sciences, University of Tasmania, Launceston, TAS 7250.

Grants

This work was supported by the Victorian Government's Operational Infrastructure Support Program. JRM and JES are National Health and Medical Research Council Fellows (IDs APP1079438, APP1078985). KLW and DCH are supported by an Emerging Leader Fellowship and Baker Fellowship, respectively, from The Shine On Foundation and the Baker Heart and Diabetes Institute.

Disclosures

No conflicts of interests, financial or otherwise, are declared by the authors.

Endnote

At the request of the author(s), readers are herein alerted to the fact that additional materials related to this manuscript may be found at [https://figshare.com/s/1788601bf6606df6f5a5]. These materials are not a part of this manuscript and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.

References

1. **Agresti A, and Min Y**. Simple improved confidence intervals for comparing matched proportions. *Stat Med* 24: 729-740, 2005.
2. **Akerman AP, Thomas KN, van Rij AM, Body ED, Alfadhel M, and Cotter JD**. Heat therapy vs. supervised exercise therapy for peripheral arterial disease: a 12-week randomized, controlled trial. *Am J Physiol Heart Circ Physiol* 2019.
3. **Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, and Chalmers I**. Increasing value and reducing waste in biomedical research regulation and management. *Lancet* 383: 176-185, 2014.
4. **Allen C, and Mehler DMA**. Open science challenges, benefits and tips in early career and beyond. *PLoS Biol* 17: e3000246, 2019.
5. **Anderson MS, Martinson BC, and De Vries R**. Normative dissonance in science: results from a national survey of u.s. Scientists. *Journal of empirical research on human research ethics : JERHRE* 2: 3-14, 2007.
6. **Baker D, Lidster K, Sottomayor A, and Amor S**. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol* 12: e1001756, 2014.
7. **Begley CG, and Ellis LM**. Raise standards for preclinical cancer research. *Nature* 483: 531, 2012.
8. **Begley CG, and Ioannidis JP**. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res* 116: 116-126, 2015.
9. **Begley CG, and Ioannidis JPA**. Reproducibility in Science. *Improving the Standard for Basic and Preclinical Research* 116: 116-126, 2015.
10. **Bernardo BC, Weeks KL, Pongsukwechkul T, Gao X, Kiriazis H, Cemerlang N, Boey EJH, Tham YK, Johnson CJ, Qian H, Du XJ, Gregorevic P, and McMullen JR**. Gene delivery of medium chain acyl-coenzyme A dehydrogenase induces physiological cardiac hypertrophy and protects against pathological remodelling. *Clin Sci (Lond)* 132: 381-397, 2018.
11. **Brooks HL, and Lindsey ML**. Guidelines for authors and reviewers on antibody use in physiology studies. *American Journal of Physiology-Heart and Circulatory Physiology* 314: H724-H732, 2018.
12. **Cahill LS, Zhou YQ, Hoggarth J, Yu LX, Rahman A, Stortz G, Whitehead CL, Baschat A, Kingdom JC, Macgowan CK, Serghides L, and Sled JG**. Placental vascular abnormalities in the mouse alter umbilical artery wave reflections. *Am J Physiol Heart Circ Physiol* 316: H664-h672, 2019.
13. **Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, Howells DW, Ioannidis JP, and Oliver S**. How to increase value and reduce waste when research priorities are set. *Lancet* 383: 156-165, 2014.
14. **Collins FS, and Tabak LA**. Policy: NIH plans to enhance reproducibility. *Nature* 505: 612-613, 2014.
15. **Curran-Everett D, and Benos DJ**. Guidelines for reporting statistics in journals published by the American Physiological Society. *American Journal of Physiology-Heart and Circulatory Physiology* 287: H447-H449, 2004.
16. **DeSoto KA**. NIH-Wide Policy Doubles Down on Scientific Rigor and Reproducibility. 2016.
17. **Donner DG, Kiriazis H, Du XJ, Marwick TH, and McMullen JR**. Improving the quality of preclinical research echocardiography: observations, training, and guidelines for measurement. *Am J Physiol Heart Circ Physiol* 315: H58-h70, 2018.
18. **Drucker DJ**. Never Waste a Good Crisis: Confronting Reproducibility in Translational Research. *Cell Metab* 24: 348-360, 2016.
19. **Editorial**. Let's think about cognitive bias. *Nature* 526: 163, 2015.
20. **Eldahshan W, Ishrat T, Pillai B, Sayed MA, Alwhaibi A, Fouda AY, Ergul A, and Fagan SC**. Angiotensin II type 2 receptor stimulation with compound 21 improves neurological function after stroke in female rats: a pilot study. *American Journal of Physiology-Heart and Circulatory Physiology* 316: H1192-H1201, 2019.
21. **Ferey JLA, Boudoures AL, Reid M, Drury A, Scheaffer S, Modi Z, Kovacs A, Pietka T, DeBosch BJ, Thompson MD, Diwan A, and Moley KH**. A maternal high-fat, high-sucrose diet induces transgenerational cardiac mitochondrial dysfunction independently of maternal mitochondrial inheritance. *Am J Physiol Heart Circ Physiol* 316: H1202-h1210, 2019.

22. **Flier JS**. Irreproducibility of published bioscience research: Diagnosis, pathogenesis and therapy. *Mol Metab* 6: 2-9, 2017.
23. **Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, and Wager E**. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 383: 267-276, 2014.
24. **Hart EC, Head GA, Carter JR, Wallin BG, May CN, Hamza SM, Hall JE, Charkoudian N, and Osborn JW**. Recording sympathetic nerve activity in conscious humans and other mammals: guidelines and the road to standardization. *American Journal of Physiology-Heart and Circulatory Physiology* 312: H1031-H1051, 2017.
25. **Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, and Heneghan C**. The need for randomization in animal trials: an overview of systematic reviews. *PLoS ONE* 9: e98856, 2014.
26. **Huang X, Liu S, Wu D, Cheng Y, Han H, Wang K, Zhang G, and Hu S**. Facilitated Ca(2+) homeostasis and attenuated myocardial autophagy contribute to alleviation of diabetic cardiomyopathy after bariatric surgery. *Am J Physiol Heart Circ Physiol* 315: H1258-h1268, 2018.
27. **Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, and Tibshirani R**. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 383: 166-175, 2014.
28. **Kilkenny C BW, Cuthill IC, Emerson M, Altman DG**. The ARRIVE Guidelines Checklist: animal research: reporting in vivo experiments. *Available: wwwnc3rsorguk/sites/default/files/documents/Guidelines*.
29. **Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, and Altman DG**. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE* 4: e7824, 2009.
30. **Kramer B, Franca LM, Zhang Y, Paes AMA, Gerdes AM, and Carrillo-Sepulveda MA**. Western diet triggers Toll-like receptor 4 signaling-induced endothelial dysfunction in female Wistar rats. *Am J Physiol Heart Circ Physiol* 315: H1735-h1747, 2018.
31. **Lee JB, Notay K, Klingel SL, Chabowski A, Mutch DM, and Millar PJ**. Docosahexaenoic acid reduces resting blood pressure but increases muscle sympathetic outflow compared with eicosapentaenoic acid in healthy men and women. *Am J Physiol Heart Circ Physiol* 316: H873-h881, 2019.
32. **Lindsey ML, Bolli R, Canty JM, Jr., Du XJ, Frangogiannis NG, Frantz S, Gourdie RG, Holmes JW, Jones SP, Kloner RA, Lefer DJ, Liao R, Murphy E, Ping P, Przyklenk K, Recchia FA, Schwartz Longacre L, Ripplinger CM, Van Eyk JE, and Heusch G**. Guidelines for experimental models of myocardial ischemia and infarction. *Am J Physiol Heart Circ Physiol* 314: H812-h838, 2018.
33. **Lindsey ML, Gray GA, Wood SK, and Curran-Everett D**. Statistical considerations in reporting cardiovascular research. *American Journal of Physiology-Heart and Circulatory Physiology* 315: H303-H313, 2018.
34. **Lindsey ML, Kassiri Z, Virag JAI, de Castro Bras LE, and Scherrer-Crosbie M**. Guidelines for measuring cardiac physiology in mice. *Am J Physiol Heart Circ Physiol* 314: H733-h752, 2018.
35. **MacCoun R, and Perlmutter S**. Blind analysis: Hide results to seek the truth. *Nature* 526: 187-189, 2015.
36. **Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman RA-S, Chan A-W, and Glasziou P**. Biomedical research: increasing value, reducing waste. *The Lancet* 383: 101-104, 2014.
37. **Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, and Donnan GA**. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* 39: 2824-2829, 2008.
38. **Mannoji H, Saku K, Nishikawa T, Tohyama T, Kamada K, Abe K, Sunagawa G, Kishi T, Sunagawa K, and Tsutsui H**. Estimation of the baroreflex total loop gain by the power spectral analysis of continuous arterial pressure recordings. *Am J Physiol Heart Circ Physiol* 316: H828-h839, 2019.
39. **Marshall JPS, Estevez E, Kammoun HL, King EJ, Bruce CR, Drew BG, Qian H, Iliades P, Gregorevic P, Febbraio MA, and Henstridge DC**. Skeletal muscle-specific overexpression of heat shock protein 72 improves skeletal muscle insulin-stimulated glucose uptake but does not alter whole body metabolism. *Diabetes Obes Metab* 20: 1928-1936, 2018.
40. **Maslov MY, Foianini S, Mayer D, Orlov MV, and Lovich MA**. Synergy between sacubitril and valsartan leads to hemodynamic, antifibrotic, and exercise tolerance benefits in rats with preexisting heart failure. *Am J Physiol Heart Circ Physiol* 316: H289-h297, 2019.

41. **Mozolevska V, Schwartz A, Cheung D, Goyal V, Shaikh B, Dingman B, Kim E, Mittal I, Asselin CY, Edel A, Ravandi A, Thliveris J, Singal PK, Czaykowski P, and Jassal DS**. Role of renin-angiotensin system antagonists in the prevention of bevacizumab- and sunitinib-mediated cardiac dysfunction. *Am J Physiol Heart Circ Physiol* 316: H446-h458, 2019.
42. **Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers E-J, Ware JJ, and Ioannidis JPA**. A manifesto for reproducible science. *Nature Human Behaviour* 1: 0021, 2017.
43. **Nguyen MN, Ziemann M, Kiriazis H, Su Y, Thomas Z, Lu Q, Donner DG, Zhao WB, Rafehi H, Sadoshima J, McMullen JR, El-Osta A, and Du XJ**. Galectin-3 deficiency ameliorates fibrosis and remodeling in dilated cardiomyopathy mice with enhanced Mst1 signaling. *Am J Physiol Heart Circ Physiol* 316: H45-h60, 2019.
44. **Nosek BA, Spies JR, and Motyl M**. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on psychological science : a journal of the Association for Psychological Science* 7: 615-631, 2012.
45. **Nuzzo R**. How scientists fool themselves - and how they can stop. *Nature* 526: 182-185, 2015.
46. **Parker BL, Calkin AC, Seldin MM, Keating MF, Tarling EJ, Yang P, Moody SC, Liu Y, Zerenturk EJ, Needham EJ, Miller ML, Clifford BL, Morand P, Watt MJ, Meex RCR, Peng KY, Lee R, Jayawardana K, Pan C, Mellett NA, Weir JM, Lazarus R, Lusis AJ, Meikle PJ, James DE, de Aguiar Vallim TQ, and Drew BG**. An integrative systems genetic analysis of mammalian lipid metabolism. *Nature* 2019.
47. **Patik JC, Curtis BM, Nasirian A, Vranish JR, Fadel PJ, and Brothers RM**. Sex differences in the mechanisms mediating blunted cutaneous microvascular function in young black men and women. *Am J Physiol Heart Circ Physiol* 315: H1063-h1071, 2018.
48. **Prinz F, Schlange T, and Asadullah K**. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10: 712, 2011.
49. **Purcell BP, Barlow SC, Perreault PE, Freeburg L, Doviak H, Jacobs J, Hoenes A, Zellars KN, Khakoo AY, Lee T, Burdick JA, and Spinale FG**. Delivery of a matrix metalloproteinase-responsive hydrogel releasing TIMP-3 after myocardial infarction: effects on left ventricular remodeling. *Am J Physiol Heart Circ Physiol* 315: H814-h825, 2018.
50. **Reynolds CA, O'Leary DS, Ly C, Smith SA, and Minic Z**. Development of a decerebrate model for investigating mechanisms mediating viscero-sympathetic reflexes in the spinalized rat. *Am J Physiol Heart Circ Physiol* 316: H1332-h1340, 2019.
51. **Schlaak RA, Frei A, Schottstaedt AM, Tsaih SW, Fish BL, Harmann L, Liu Q, Gasperetti T, Medhora M, North PE, Strande JL, Sun Y, Rui H, Flister MJ, and Bergom C**. Mapping genetic modifiers of radiation-induced cardiotoxicity to rat chromosome 3. *Am J Physiol Heart Circ Physiol* 316: H1267-h1280, 2019.
52. **Schmitz B, Niehues H, Lenders M, Thorwesten L, Klose A, Kruger M, Brand E, and Brand SM**. Effects of High-Intensity Interval Training on Microvascular Glycocalyx and Associated microRNAs. *Am J Physiol Heart Circ Physiol* 2019.
53. **Schulz KF, and Grimes DA**. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 359: 614-618, 2002.
54. **Shah A, Cooke CM, Kirschenman RD, Quon AL, Morton JS, Care AS, and Davidge ST**. Sex-specific effects of advanced maternal age on cardiovascular function in aged adult rat offspring. *Am J Physiol Heart Circ Physiol* 315: H1724-h1734, 2018.
55. **Suresh K**. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci* 4: 8-11, 2011.
56. **Weissgerber TL, Milic NM, Winham SJ, and Garovic VD**. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol* 13: e1002128, 2015.
57. **Weissgerber TL, Savic M, Winham SJ, Stanisavljevic D, Garovic VD, and Milic NM**. Data visualization, bar naked: A free tool for creating interactive graphics. *J Biol Chem* 292: 20592-20598, 2017.
58. **Wollborn J, Siemering S, Steiger C, Buerkle H, Goebel U, and Schick MA**. Phosphodiesterase-4 inhibition reduces ECLS-induced vascular permeability and improves microcirculation in a rodent model of extracorporeal resuscitation. *Am J Physiol Heart Circ Physiol* 316: H751-h761, 2019.

669 **Legends**

670 **Figure 1. A:** Responses from responsible investigators to a survey question related to
671 incorporation of pre-specified criteria for inclusions/exclusions of animals within experimental
672 designs from the original survey (2017) and follow-up survey (2018). **B:** Reporting of pre-specified
673 criteria in publications. **C**: Reporting of excluded animals in publications. Right panels: lines were
674 manually off-set within a band of responses (e.g. 0, 25, 50, 75, 100%) for easier visualization of
675 individual responses (N=16). Dotted lines illustrate the responses from an investigator who did not
676 publish a paper over the 12 month period. In this case, the reported measure of 0% provided by
677 the investigator does not represent 0%.

678

679 **Figure 2.** Responses from responsible investigators (N=16) regarding a willingness to use a formal
680 procedure/flow chart to track animal exclusions. **A:** Example of a flowchart/template to track
681 animals. **B:** Willingness of investigators to track animals with a flowchart/template. **C:** Uptake of
682 using the template over a 12 month period. N.B. In panel B one investigator provided no response.

683

684 **Figure 3. A and B:** Responses from responsible investigators (N=16) to survey questions related
685 to pre-specified primary and secondary endpoints from the original survey (2017) and follow-up
686 survey (2018). **C:** Reporting pre-specified endpoints/outcomes in publications. Right panels: lines
687 were manually off-set within a band of responses (e.g. 0, 25, 50, 75, 100%) for easier visualization
688 of individual responses (N=16). The dotted line illustrates the response from an investigator who
689 did not publish a paper over the 12 month period. In this case, the reported measure of 0%
690 provided by the investigator does not represent 0%.

691

692

693 **Figure 4.** Responses from responsible investigators (N=16) to a survey question related to sample
694 size calculations from the original survey (2017) and follow-up survey (2018). Right panel: lines
695 were manually off-set within a band of responses (e.g. 0, 25, 50, 75, 100%) for easier visualization
696 of individual responses (N=16).

697

698

699 **Figure 5. A and B:** Responses from responsible investigators (N=16) to survey questions related
700 to randomization of animals and the method of randomization used from the original survey (2017)
701 and follow-up survey (2018). **A:** Right panel; lines were manually off-set within a band of
702 responses (e.g. 0, 25, 50, 75, 100%) for easier visualization of individual responses (N=16). For

Fig 5B; P=0.005, Wilcoxon Sign-Rank Test when comparing a structured (desirable) method [computer-generated/sealed envelope/numbers from a hat] with an unstructured (undesirable) method.

Figure 6. A-C: Responses from responsible investigators (N=16) to survey questions related to allocation concealment and blinding from the original survey (2017) and follow-up survey (2018). Panel A assessed responses to allocation concealment in regard to genotype or treatment. Panel B refers to intervention studies in which a test drug, gene therapy (e.g. adeno-associated virus, AAV) or modified diet (e.g. high fat diet) is administered to mice, and whether these interventions are blinded by labelling one intervention “A” and one “B” or some equivalent. Panel C refers to blinding post-animal experimentation. For panel 6B, P=0.010, Wilcoxon Sign-Rank Test; comparison of 2017 with 2018. Right panels: lines were manually off-set within a band of responses (e.g. 0, 25, 50, 75, 100%) for easier visualization of individual responses (N=16).

Figure 7. A-C: Responses from responsible investigators (N=16) to survey questions related to their perception and practices in regard to the quality and rigor of animal studies from the original survey (2017, panel A) and follow-up survey (2018, panels B and C).

734

735 **Table 1**- Overview of exercise to improve the quality of preclinical animal research

Activity	Description
Presentation-Background and awareness (mid-April, 2017). One hour session including questions and discussion.	-Concerns regarding the quality of animal research -ARRIVE Guidelines -Types of bias -Practical examples and considerations when blinding -Feedback and consultation
Presentation-External speaker (late-April, 2017).	Seminar: Improving disease modeling and candidate drug evaluation
Survey preparation and testing (May 2017).	-Survey to assess baseline practices (May 2017) -Consultation on survey design and implementation -Trial testing of the survey with a scientist undertaking animal studies
Survey rollout (June 2017). Designed to be completed within 10-12 min.	-Email from Institute Director with instructions to complete the survey by week's end. -Emphasis on acquiring honest responses to understand obstacles and barriers; example response provided from Head of Basic Science Domain to highlight we were looking for honest responses -Follow-up reminders during the week
Communication of survey results (July 2017).	-A summary of the survey results and next steps was communicated to Science Faculty at a meeting July 21, 2017 -A report of the survey results was distributed to Lab Heads and Scientists by email July 23, 2017 -It was noted that animal templates and practical tools would be made available to researchers over the coming months
Presentation-Survey results and practical tools (Sept 2017).	-Areas for improvement -Methods for randomization -Flow charts for tracking animals and exclusions
Follow-up (March 2018 to present).	Institute Director and Head of Basic Science Domain visit labs on a monthly basis to: 1) Assess uptake of new tools 2) Identify potential issues/concerns
Follow-up survey-June 2018.	-Email from Head of Basic Science Domain with instructions to complete a short compulsory follow-up survey within a week (expected to take 6-8 min). The survey was sent to an investigator to trial before sending it to all investigators. -Re-emphasis- acquiring honest responses to understand obstacles and barriers -Follow-up reminders during the week
Communication and presentation of survey results-Sept 2018.	-A summary of survey results in 2018 compared to 2017 -A reminder about randomization tools and flow charts -Example of inserting information into publications

736

737

738

739

Table 2- Challenges and solutions for reducing bias

Challenges	Solutions
Diabetic and non-diabetic animals look physically different, e.g. diabetic mice have unkempt coats and require daily cage changes due to increased urination.	Re-blind collected tissues/samples at study end.
Drugs administered via a syringe are different in colour to the vehicle/control.	Wrap the syringes in foil.
Dietary intervention (e.g. high fat diet) is different in colour to the chow diet.	Re-blind collected tissues/samples at study end. In some cases, it is possible to request the supplier to make the intervention and chow/control diet a different colour.
Lack of personnel to label drugs “A” and “B” or to have different staff members blinding different analyses.	Assistance from personnel in other laboratories.

765

766 **Table 3.** Implications of the introduction of more structured randomization & blinding
767

Potential deterrents	Benefits
Increased time planning studies.	Improved quality of research.
Effect sizes are likely to be smaller. This may require larger animal numbers. Less dramatic results?	Negative results from well designed studies are more likely to be published in good journals than negative results from poorly designed studies.
More staff required for ensuring studies are blinded.	Fewer researchers following up on dead ends.
Additional costs associated with more animals and lab staff time.	More money, researchers time directed into research and targets which have the greatest opportunity for success. New drugs for patients.
Fewer positive results.	Lower risk of translational failure.
Potential implications for high impact factor papers, productivity & career progression (with current reward structures)	Less negative press. Greater confidence from the media and public. More philanthropic support.

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

Figure 1

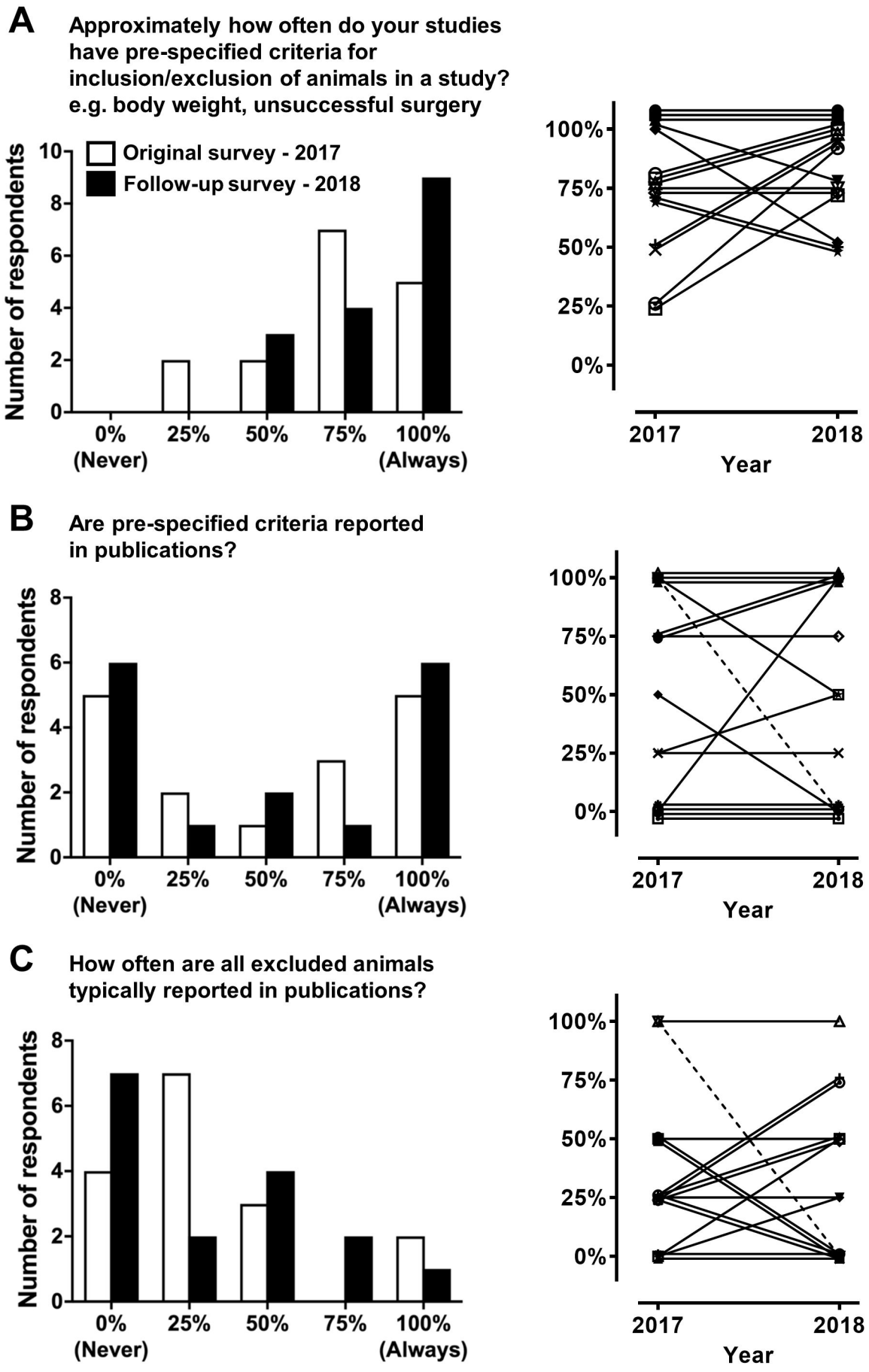
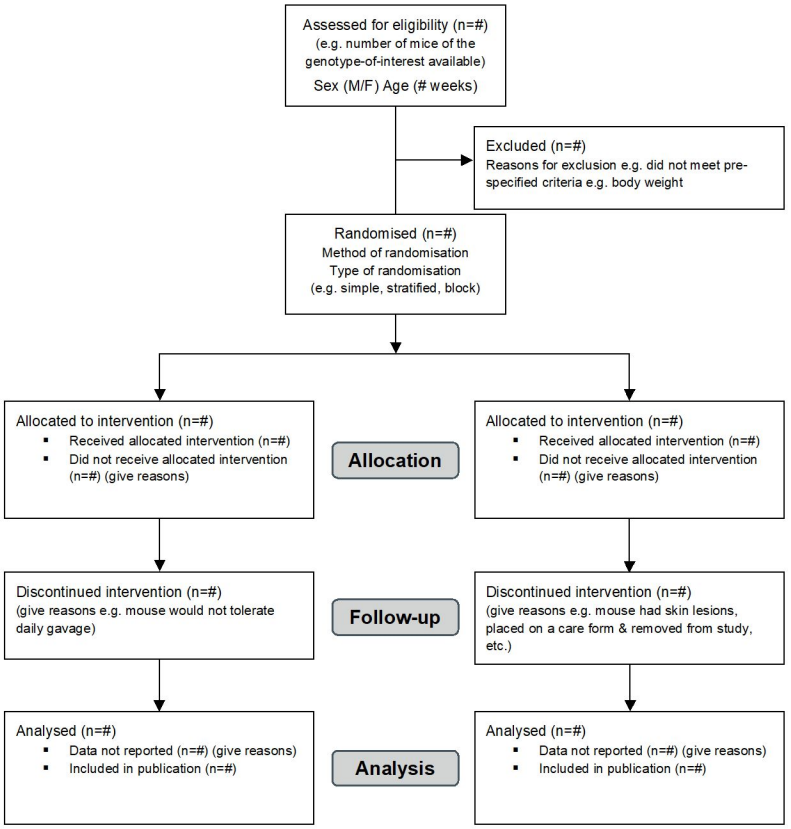

A
Approximately how often do your studies have pre-specified criteria for inclusion/exclusion of animals in a study? e.g. body weight, unsuccessful surgery
Original survey - 2017
Follow-up survey - 2018
Number of respondents
0%
(Never)
25%
50%
75%
100%
(Always)
2017
2018
Year
B
Are pre-specified criteria reported in publications?
C
How often are all excluded animals typically reported in publications?

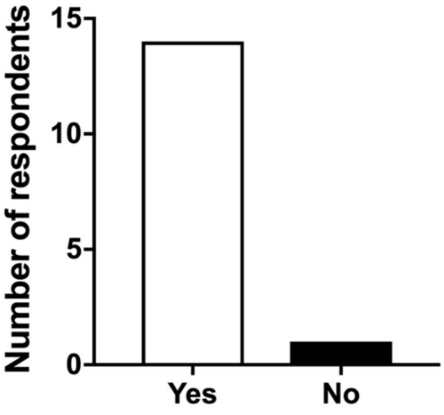
Figure 2

A **Consolidated Standards of Animal Experiment ReporTing (CONSAERT): Template for intervention studies in genetically modified mouse strains**



B

<u>Original survey (2017)</u>: Would you use this template in practice?



C

<u>Follow-up survey (2018)</u>: Have you started using any of the provided animal flow charts / templates or an equivalent to track animals in your studies?

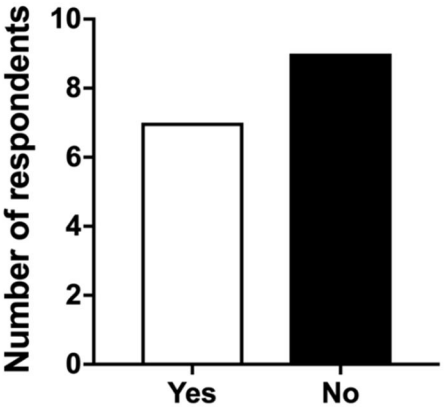


Figure 3

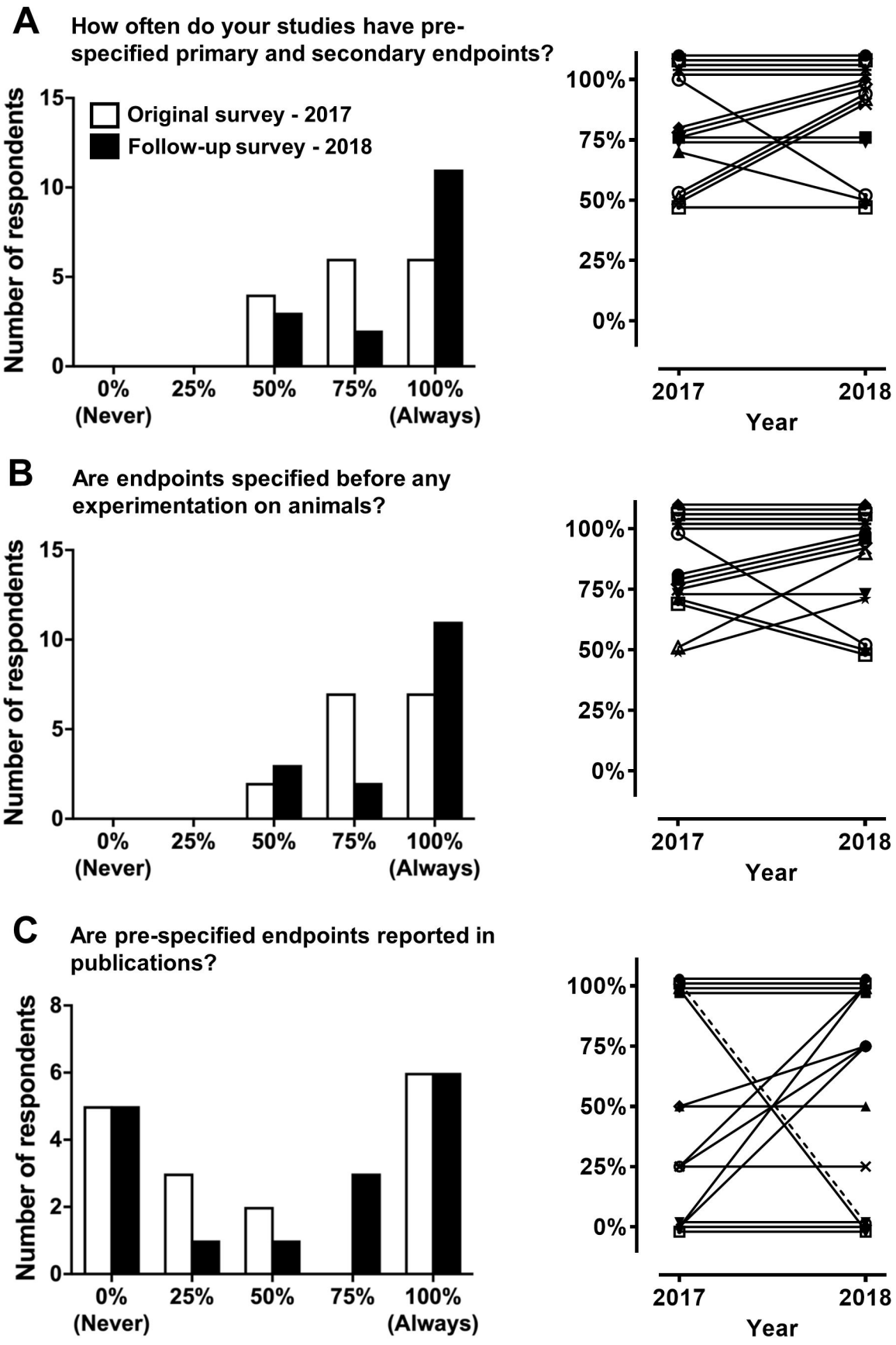
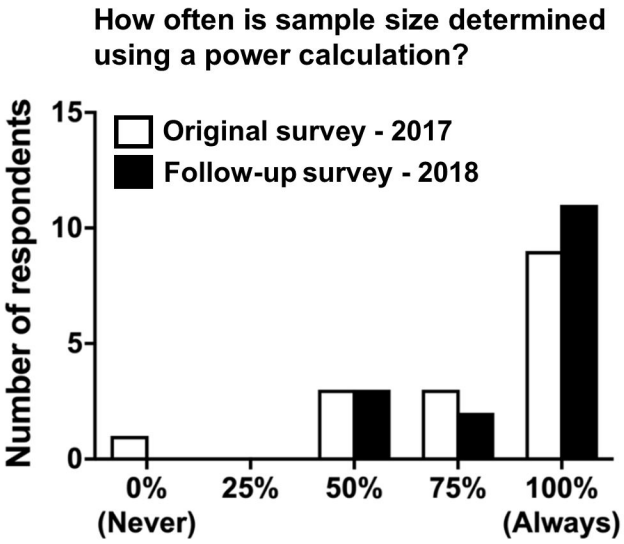

A
How often do your studies have pre-specified primary and secondary endpoints?
Original survey - 2017
Follow-up survey - 2018
Number of respondents
0%
(Never)
25%
50%
75%
100%
(Always)
2017
2018
Year
B
Are endpoints specified before any experimentation on animals?
C
Are pre-specified endpoints reported in publications?

Figure 4


How often is sample size determined using a power calculation?
Original survey - 2017
Follow-up survey - 2018
Number of respondents
15
10
5
0
0% (Never)
25%
50%
75%
100% (Always)

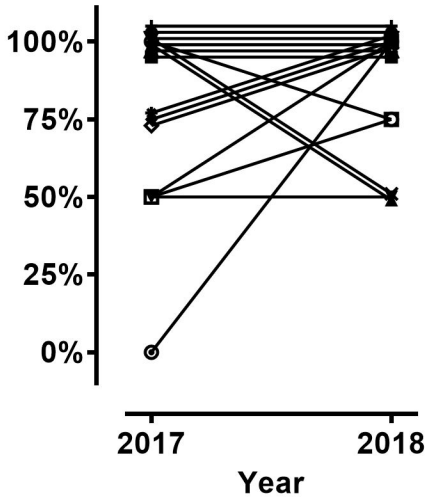
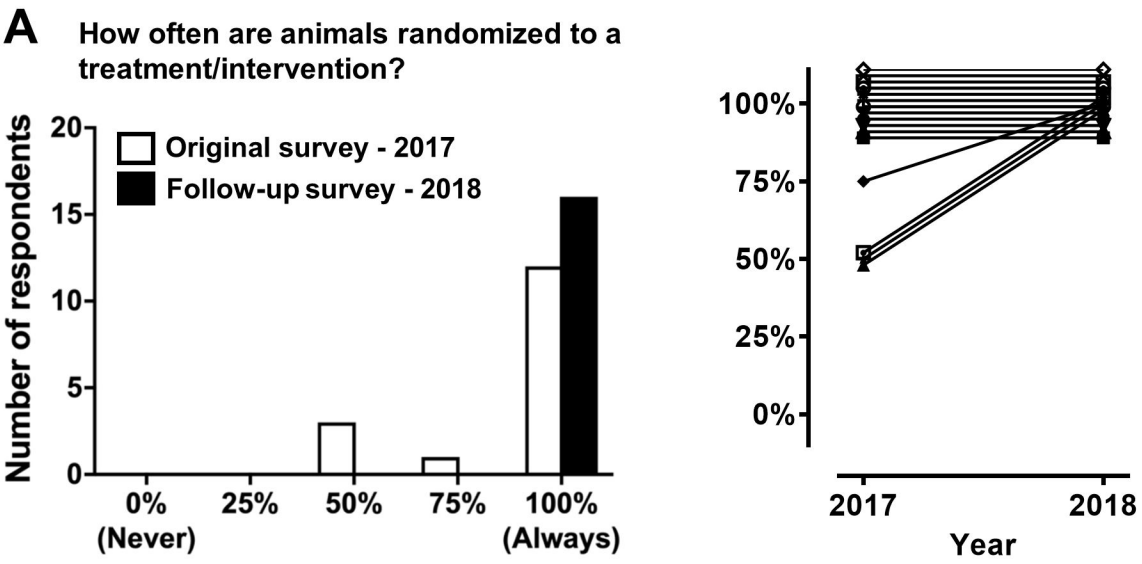

100%
75%
50%
25%
0%
2017
2018
Year

Figure 5


A
How often are animals randomized to a treatment/intervention?
Number of respondents
20
15
10
5
0
Original survey - 2017
Follow-up survey - 2018
0% (Never)
25%
50%
75%
100% (Always)
100%
75%
50%
25%
0%
2017
2018
Year

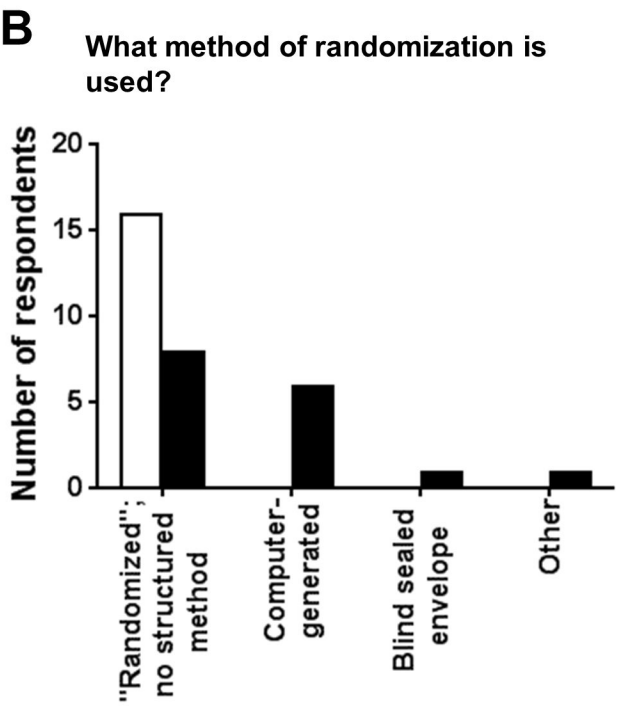
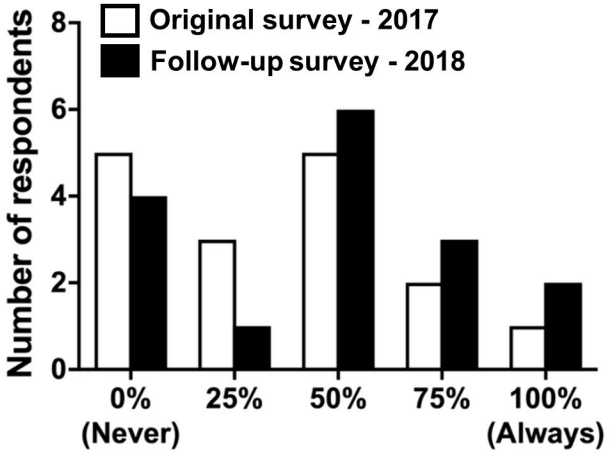
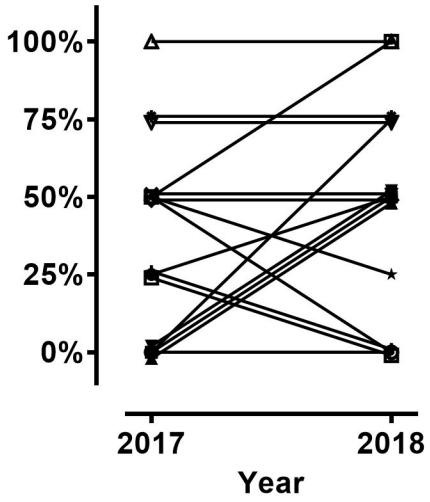

B
What method of randomization is used?
Number of respondents
20
15
10
5
0
"Randomized"; no structured method
Computer-generated
Blind sealed envelope
Other

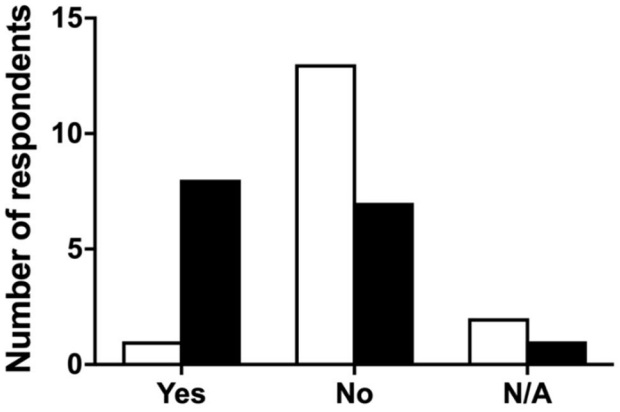
Figure 6

A Does experimental design include allocation concealment?

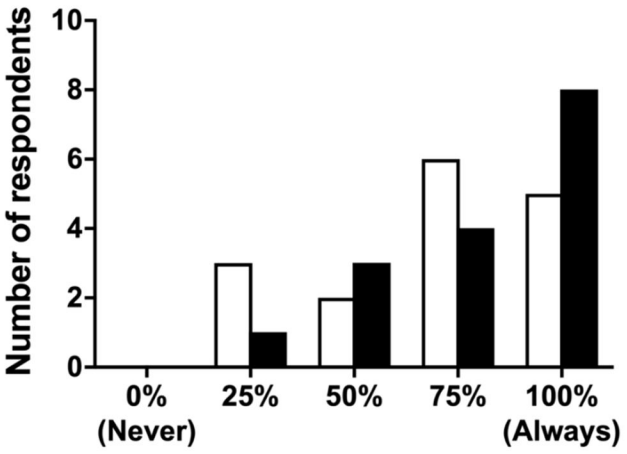




B For drug/AAV/diet studies, are interventions labelled ‘A’ and ‘B’ or equivalent?



C How often are components of the study blinded?



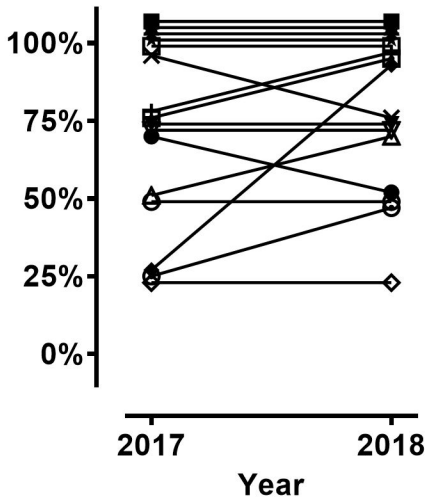
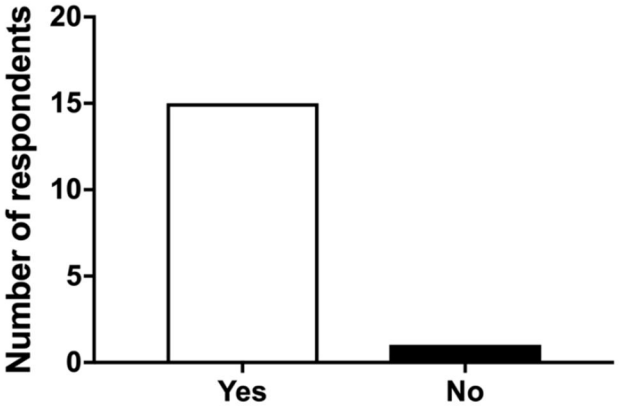


Figure 7

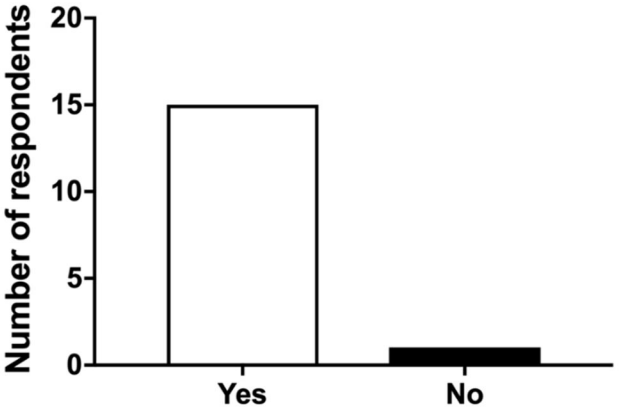
A **Original survey (2017):**

Do you think the quality/rigor of your animal studies could be improved?



B **Follow-up survey (2018):**

Do you think the quality/rigor of your animal studies has improved over the past 12 months?



C **Follow-up survey (2018):**

Has your lab changed practices regarding randomization, blinding and/or animal reporting since completing this survey 12 months ago?