

SIRF: Sex Inclusive Research Framework

An evaluation framework to assess whether an *in vivo* research proposal follows the sex-inclusive research philosophy.

Natasha A. Karp¹, Manuel Berdoy², Lilian Hunt³, Maggy Jennings⁴, Angela Kerton⁵, Matt Leach⁶, Jordi L. Tremoleda⁷, Esther J. Pearl⁸, Nathalie Percie du Sert⁸, Benjamin Phillips¹, Penny S Reynolds⁹, Kathy Ryder¹⁰, S Clare Stanford¹¹, Sara Wells¹², Lucy Whitfield¹³.

1: Data Sciences & Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK

2: BMS, University of Oxford, UK

3: Wellcome Trust, London, UK

4: RSPCA, Animals in Science Department, UK

5: The Learning Curve (Development) Ltd, Ware, UK

6: Comparative Biology Centre, Newcastle University, UK

7: Queen Mary University of London, UK

8: The NC3Rs, London, UK

9: Department of Health, Stormont Estate, Belfast, UK

10: University of Florida, USA

11: University College London, UK

12: The Mary Lyon Centre at MRC Harwell, Harwell Science and Innovation Campus, UK

13: OWL Vets Ltd, UK

Table of Contents

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| SIRF: Sex Inclusive Research Framework | 1 |
| An evaluation framework to assess whether an <i>in vivo</i> research proposal follows the sex-inclusive research philosophy..... | 1 |
| Why is there a need for a decision framework to assess the sex inclusion position of project proposals? | 3 |
| The evaluation framework | 3 |
| Limitations of the framework..... | 3 |
| Supporting material | 5 |
| Q1: Does the experiment set include identifiable male and female study samples throughout the research project? | 5 |
| Q2: Can the sex of the study sample be determined? | 5 |
| Q3: Is the experiment an acceptable exception?..... | 6 |
| Q4: Is the justification a statement that the disease model can only be induced in one sex? | 7 |
| Q5: Is the justification a generic statement around variability? | 7 |
| Q6: Is the justification a misunderstanding around statistical power? | 8 |
| Q7: Is the justification fear of/avoiding change? | 9 |
| Q8: Is the justification a generic statement around welfare management? | 9 |
| Q9: Does the explanation for the model/species provides a harm- &/or cost-benefit justification sufficient to justify the use of one sex? | 10 |
| Q10: Does the experiment set include groups that will be mathematically compared? | 10 |
| Q11: Does the analysis plan adequately consider sex-related variation? | 10 |
| Q12: Will the design have a balanced inclusion of female and male samples? | 11 |
| Glossary..... | 12 |
| References | 12 |




Why is there a need for a decision framework to assess the sex inclusion position of project proposals?

Within preclinical research, an endemic and persistent sex bias exists where research is predominately conducted with a single sex, typically male animals or male cell lines [1-3]. This would explain the finding that our fundamental biological knowledge base may be biased [4]. Translatability issues compromise the ethical use of animals in research [5]. To improve translation and back translation of results between humans and other animals, numerous funding bodies have released inclusion mandates [6, 7] that require males and females to be included as standard unless strong justification is provided. These increasingly lead to a need for many organisations (i.e., funders, national regulatory bodies, ethical review bodies) to assess whether a proposal is sex inclusive.

Research has shown that scientists are supportive and believe sex matters in early research but there may be barriers to implementing sex inclusive designs [8, 9]. Frequently, the blockers mentioned are culturally embedded misconceptions. Here we present an evaluation framework to rapidly assess an *in vivo* research proposal to determine whether the proposed strategy is sex inclusive and appropriately planned where possible. When a decision has been made to study only a single sex, the framework evaluates whether the justification is scientifically appropriate and is not based on common misconceptions. By addressing the misconceptions, justifications will become considered and as a community we can truly understand when sex inclusive research is a possibility. The clarity of the framework will also provide transparency in the assessment process for both researchers and those evaluating the proposals.

The evaluation framework

The framework consists of a decision tree comprising of twelve questions and associated supporting information for each question. The decision tree (**Figure 1**) leads to one or more “traffic light” outcome classifications, indicating whether a proposal is appropriate, carries some risks, or is insufficient with regards to sex inclusion. The classifications can be grouped into three types:

-  **Green:** Proposal is appropriate
-  **Amber:** Caution is required (i.e., the proposed design/analysis carries some risk)
-  **Red:** Justification for single sex study in the proposal is not sufficient

A [glossary](#) of technical terms has been provided to support accessibility.

Limitations of the framework

- The framework covers only *in vivo* and *ex vivo* studies.
- If a proposal contains multiple sets of experiments, then the framework would need to be independently applied to each.
- Many of the questions require a subjective evaluation, which could lead to variation in the judgement reached. The provision of supporting information for each question should mitigate that risk.
- Decisions may shift in time as science/culture evolves (e.g., ability to determine the sex).

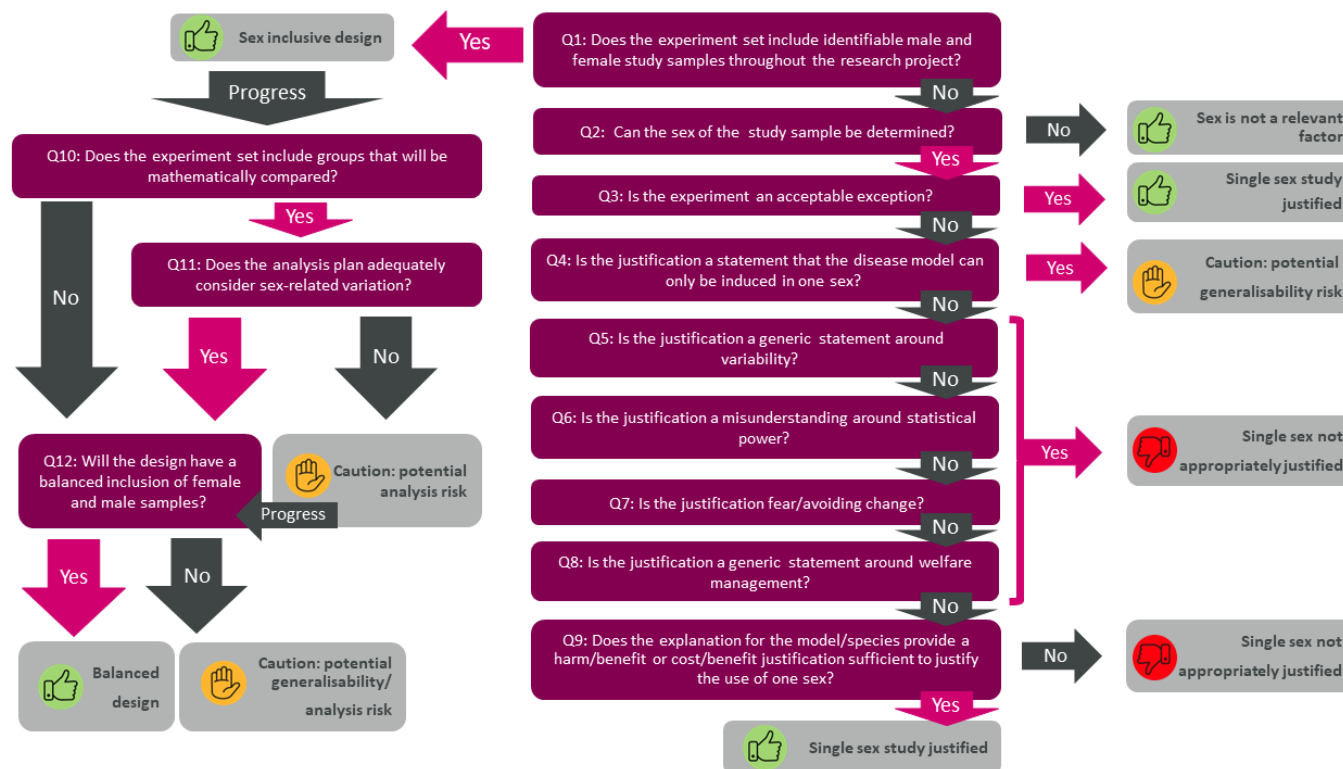


Figure 1: A decision tree to assess a research proposal justification around sex inclusion.


Supporting material can be found for each question by following the links below:

| | | | |
|--------------------|---------------------|---------------------|---------------------|
| Q1 | Q2 | Q3 | Q4 |
| Q5 | Q6 | Q7 | Q8 |
| Q9 | Q10 | Q11 | Q12 |

Supporting material

Q1: Does the experiment set include identifiable male and female study samples throughout the research project?

Assessment advice:

- Males and females should be included in all studies and should be studied simultaneously.
- If the researchers' plan to include females and males in all experiments, then the proposal is classed as being " **Sex inclusive design**" but has the potential to accumulate other classifications based on questions around the design and analysis; consequently, proceed to [Q10](#). However, if any of the studies will be conducted on one sex, the assessment progresses to [Q2](#).

Why is this question included?


- Males and females need to be studied simultaneously within an experiment to allow the sex inclusive research philosophy to be realised. The inclusive strategy will enable an estimation of a generalisable effect (average effect across females and males) and an assessment of whether any intervention effect observed depended on sex.
- Studying one sex first and then later the second sex, typically male then female, is a structurally biased research strategy. When this strategy is implemented, science is being explored through a male lens and "true crux of the research question for females can be missed" [10].
- Furthermore, the reality is that often, researchers never get round to studying the second sex.

Supporting resources

- [Canadian Institutes of Health Research - online training module on 'Sex and Gender in Biomedical Research'](#).

Q2: Can the sex of the study sample be determined?

Assessment advice:

- If the sex of the sample cannot be determined (see Table 1 for examples) then the outcome classification " **Sex is not a relevant factor**" would be applied (i.e., the design is appropriate).
- Examples [11] of appropriate justification where the sex cannot be determined include:
 - "All animals will be P3 or younger, thus no sex determination is possible. We anticipate that animals of both sexes will be present at approximately equal proportions."
 - "The experimental unit is the litter, and male and female littermates will be pooled together and analysed as one group"
 - "We will be unable to consistently distinguish between male and female prairie dogs and therefore will pool both sexes in analyses."
- In situations where the sex could be determined, Q2 should be answered 'yes' and the assessment progresses to [Q3](#).

Why is this question included?

- Unless the sex of the subject of the research cannot be assessed, the default position is that sex should be determined, operationalised, reported, and considered in the design and subsequent data analysis [6, 7, 12, 13].
- There are multiple traits, such as genetic, endocrinological and anatomical features, that can be used to categorise an individual's sex [14]. However, there are no universally agreed guidelines for defining sex [15]. Consequently, to improve reproducibility and provide context, it is recommended that sex is operationalized by defining and reporting the concrete and measurable variables that will be used to identify the sex [13].

Table 1: Example scenarios where the ability to assess sex is considered

| Situation | Ability to determine the sex |
|--------------------------------------|------------------------------|
| Early embryos (dependant on species) | No |
| Certain juvenile fish species | No |
| Certain invertebrate species | No |
| Animal tissue | Yes |
| Organoids | Yes |
| Model with animal and donor cells | Yes, and Yes |

In the examples above, the sex could often be determined from an assessment of physical features. When physical features are insufficient, the sex could potentially be determined with a genetic screen though the cost might be prohibitive.

Note

- The terms sex and gender are often used interchangeably but in fact have different meanings (Table 2)[12]. In the context of *in vivo* research, we are only able to consider sex and hence this is the terminology used throughout this framework.

Table 2: Definitions for sex and gender

| Term | Definition |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sex | Is a classification system that refers to a set of attributes that defines an organism as male, female, intersexed or hermaphrodite. A variety of characteristics can be used. For example chromosomes, hormones or reproductive organs. |
| Gender | Refers to the socially constructed roles, behaviours, expressions and identities of girls, women, boys, men, and gender diverse people [12]. |

Q3: Is the experiment an acceptable exception?

Assessment advice:

- Two acceptable exceptions have been identified ('acutely scarce resources' and 'single sex mechanism or disease'). In these situations, the proposals would be classified as "📋 **Single sex study justified**" (see Table 3 for examples).
- A disease being rare in a sex (e.g., haemophilia A in females) does not necessarily indicate a single sex mechanism and therefore is not an acceptable justification under this question. Consequently, Q3 should be answered 'no' and the assessment progresses to [Q4](#). There is a later question that deals with this scenario ([Q9: Does the explanation for the model/species provides a harm- &/or cost-benefit justification sufficient to justify the use of one sex?](#)).
- A difference in effect size between the sexes following an intervention (e.g., on model induction) does not, in isolation, indicate a single sex mechanism and is therefore not an appropriate justification. Consequently, Q3 should be answered 'no' and the assessment progresses to [Q4](#).

Table 3: Example research scenarios where the studies may fulfil the criteria as an acceptable exception.

| Scenario | Example |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Acutely scarce resources | Human tissue samples of rare diseases Conservation studies where access to one sex may be challenging (e.g., aggression) |
| Single sex mechanism or disease | Ovarian/Prostate cancer Mechanism of action that is unique to a sex Implantation of embryos Sperm quality Pregnancy |

Q4: Is the justification a statement that the disease model can only be induced in one sex?

Assessment advice:

- If the justification given is that the disease model can only be induced in one sex. In this situation, an outcome classification of “🚫 **Caution potential Generalisability risk**” would arise.
- This is an amber classification, as a decision to proceed is dependent on whether other disease model constructs are available for this disease and a reflection on the risk arising when proceeding with only one sex.
- If the answer is no, the assessment progresses to [Q5](#).

Why is this question included?

- This is a common justification given when working with induced models. The preceding question has already accounted for a disease that is unique to a single sex. As the disease occurs in females and males, an induced model that is only recapitulated in one sex raises questions about generalisability and whether the model is a good construct to study the disease of interest.

Q5: Is the justification a generic statement around variability?

Assessment advice:


- Assess the justification to determine whether it is a generic and unsubstantiated statement suggesting that one sex shows greater variation based on the perspective that outcome measures are inherently more variable in females, or that sex differences will introduce overall variability in the data.
- In this situation, an outcome classification of “🚫 **Single sex is not appropriately justified**” (i.e., the design is not appropriate) would arise.
- Examples [11] where the misconceptions have resulted in inappropriate justifications include:
 1. “Because collecting vaginal smears to control oestrous cycle in females would add an extra layer of stress to our experiments”.
 2. “We focused upon a single gender [sic], as there are sex differences in glucose metabolism”.
 3. “To avoid the influence of hormonal changes, only male rats were included in this study”.
 4. “Using male mice ensured reduced gender [sic]-dependent variability in plaque deposition.”
- If the variability in one sex has a molecular basis relevant to the research question (e.g., molecular interaction with oestrous cycle) or there is data demonstrating that the variance is significantly higher in one sex than the other that results in a prohibitive increase in animal use, Q5 should be answered ‘no’ and the assessment should proceed to [Q6](#). There is a later question that deals with these scenarios ([Q9: Does the explanation for the model/species provides a harm- &/or cost-benefit justification sufficient to justify the use of one sex?](#)).
- If the answer is no, the assessment progresses to [Q6](#).

Why is this question included?

- A meta research article found that 27% of the justifications for single sex studies focused on variability (N=30) [11].
- There are two misconceptions frequently seen around variability that have led to researchers using one sex:
 1. The reproductive hormone cycle means that females are inherently more variable [16, 17].
 - This misconception has led to a belief that, for a specific effect size of interest, a higher number of randomly cycling female animals would be needed to achieve the same statistical power compared to a study with male animals. A further common interpretation is that the researchers would therefore need to control for the stage of the oestrous cycle, thereby adding a burdensome procedure with potential welfare costs.
 - However, meta-analysis research [16-20] looking at experimental data from mice and rats found no evidence that randomly cycling females had greater variability than male animals.
 - Furthermore, it is not necessary to design the experiment(s) to explore the effect of the intervention at each stage of the oestrous cycle unless this is biologically relevant to the research question.
 2. Differences between the sexes introduces variation:
 - This misconception arises from concerns that sex-dependent variation, either because of a baseline difference between the males and females (sex effect) or because the effect of the intervention is different between females and males (an interaction effect), would reduce statistical power [21].
 - Baseline differences between males and females are common [22]. However, if data are appropriately analysed then the variance explained by sex is accounted for and does not decrease sensitivity to detect an intervention effect [21].
 - Furthermore, if sex explains variation in the intervention effect, power is only lost when the effect of the intervention has an opposite effect between females and males or only occurs in one sex [21]. This is biologically rare [22], and, in this situation, power is passed to the interaction term highlighting that the biology is highly dependent on sex and should be explored further [21].

Q6: Is the justification a misunderstanding around statistical power?

Assessment advice:

- Assess the justification to determine whether there is a generic statement suggesting that the inclusion of males and females will increase animal usage.
- In these situations, an outcome classification of “ **Single sex is not appropriately justified**” would arise.
- Examples where the misconception have resulted in an inappropriate justification include:
 - “Studying both sexes will double the number of animals I will need to use.”
 - “Using one sex will reduce the cost of the experiments as I can use fewer animals”.
- If the answer is no, the assessment progresses to [Q7](#).

Why is this question included?

- There is an embedded misconception that studying females and males will increase the number of animals being used [8, 9].
- The mandates [6, 7] implementing a sex inclusive research philosophy do not require researchers to directly study sex differences. Rather, the goal is to ensure that the intervention assessments will be a generalisable average estimate of the treatment effect and if large differences exist between males and females then this could be detected.
- This introduces sex as an additional factor of interest and typically results in the implementation of a factorial designs [23]. These designs are inherently more powerful as males and females are used to estimate the intervention effect whilst, at the same time, taking into account the sex-related variation.

- The advice is to design your experiments as you would normally, but then change half the animals to include the other sex [21, 24].

Supporting resources:

- [Phillips et al](#) 'Statistical simulations show that scientists need not increase overall sample size by default when including both sexes in in vivo studies' (PLOS Biology 2023).
- [Miller et al](#) 'Considering sex as a biological variable in preclinical research' (2017 FASEB)
- [Buch et al](#) 'Benefits of a factorial design focusing on inclusion of female and male animals in one experiment' (2019 Journal of Molecular Medicine).

Q7: Is the justification fear of/avoiding change?

Assessment advice:

- Assess the justification as to whether it is simply arguing against change.
- In these situations, an outcome classification of “**Single sex is not appropriately justified**” would arise.
- Examples of justifications that would be classified as researchers trying to avoid change include:
 - “To date, sex hasn’t explained variation in my model”.
 - “My prior work has only been considered in only one sex”.
- If the answer is no, the assessment progresses to [Q8](#).

Why is this question included?

- Change can be difficult, and our first emotional reaction is resistance by looking for arguments that support our current position [8].
- It is easy to delay embracing change: however, continued procrastination [2] needs to be challenged. Consider the quote: “To change is difficult. Not to change is fatal” William Pollard.
- The lack of data regarding sex explaining variation does not indicate there are none and is not considered sufficient justification [6, 25].
- Meta-analysis has demonstrated that data analysis has often been poorly conducted, and hence historic conclusions can be misleading [26].

Q8: Is the justification a generic statement around welfare management?

Assessment advice:

- Assess the justification to determine whether it is a generic and unsubstantiated statement claiming that welfare issues prevent inclusion of males and females.
- In these situations, an outcome classification of “**Single sex is not appropriately justified**” would arise.
- Examples where the justification is a generic statement around welfare include:
 - “Male mice fight and cannot be co-housed”.
 - “Cannot house male and female mice in the same room as it will trigger aggression”.
 - “Cannot test male and female mice in the same apparatus without the males fighting”.
- If the welfare issues are explored at the model/species level with mitigation discussed leading to a harm- and/or cost-benefit analysis that is sufficient to justify the use of one sex, Q8 should be answered ‘no’. There is a later question that deals with these scenarios ([Q9: Does the explanation for the model/species provides a harm- &/or cost-benefit justification sufficient to justify the use of one sex?](#)).
- If the answer is no, the assessment progresses to [Q9](#).

Why is this question included?

Considerable progress has been made in reducing aggression in male mice through amendments to housing, husbandry, and the handling of mice, all of which consider the triggers of male mouse aggression (ethological theory of aggression) [27, 28]. Whilst these strategies have reduced the incidence of severe aggression, there is no one

solution that will work for all strains/protocols nor completely avoid male mouse aggression [27, 29]. To avoid potentially confounding the experiment it is important to apply equivalent housing and husbandry for all sexes studied within the same experiment.

Supporting resources:

- [Lidster et al](#) 'Cage aggression in group-housed laboratory male mice: an international data crowdsourcing project' (2019 Nature Scientific Reports).
- [Kappel et al](#) 'To Group or Not to Group? Good Practice for Housing Male Laboratory Mice' (2017 Animals).
- [Weber et al](#) 'Aggression in group-housed laboratory mice: why can't we solve the problem?' (2017 Lab Animals).

Q9: Does the explanation for the model/species provides a harm- &/or cost-benefit justification sufficient to justify the use of one sex?

Assessment advice:

- The justification could explore logistical, ethical, or cost implications relative to the benefit of using male and female animals in a research proposal.
- The strength of the justification may vary, and assessment might require expert evaluation/specialist knowledge.
- If the justification is sufficient then the outcome classification would be “👍 Single sex study justified”. If, however, the justification is considered weak the assessor would classify the outcome as “👎 Single sex is not appropriately justified”.

Why is this question included?

- Whilst the mandates [6, 7] are moving us towards a framework where sex should be included as the default, single-sex studies remain “valid and warranted, provided there is evidence-based rationale for the case” [11].

Q10: Does the experiment set include groups that will be mathematically compared?

Assessment advice:

- Some research studies are purely descriptive; for example, a study to establish the neuron atlas of the mouse brain. In such studies, the data obtained will not be compared relative to a control group nor to the other sex.
- In these situations, where the answer is no, proceeds to [Q12](#).
- If the answer is yes, the assessment progresses to [Q11](#).

Why is this question included?

- In purely descriptive studies, studying females and males is important for generalisability. However, the question about analysis in the framework is not relevant.

Q11: Does the analysis plan adequately consider sex-related variation?

Assessment advice:

- Proposal should look not only to include males and females, but also explore the effect of sex on the intervention effect through visualisation and analysis of the resulting data.
- Where possible, and appropriate, a factorial analysis with a statistical test of whether the intervention depends on sex (a statistical interaction) should be applied. An interaction term should be included even when the statistical power for an interaction is low. This is to support the inclusion philosophy where females and males are included to allow the estimate of a generalisable effect and detection of a large differences in intervention effect if it arises.

- For some research questions, a complex statistical pipeline precludes a direct statistical assessment of an interaction. For example, a principal component analysis of metabolomic data.
- If the justification explains how the analysis and / or visualisation will take sex into account as a potential source of variation, then the assessment progresses to [Q12](#). If, however, the analysis and or visualisation does not directly explore sex as a potential source of variation the proposal gains a classification of “👉 **Caution potential analysis risk**” and the assessment progresses to [Q12](#).
- The decision to proceed/fund etc will depend on a risk-benefit analysis of the assessing body and the ability to provide feedback to the researchers.

Why is this question included?

- A ten-year follow-up meta-analysis study looking at papers published in 2019 from 34 journals across nine biological disciplines, found there was a significant increase in the proportion of studies that included females and males, but there was no improvement in the proportion of studies that considered sex in the data analysis [11].
- A subsequent paper reviewed the analysis plans of the publications where males and females were studied and considered in the data analysis and found that data analysis errors (such as pooling, comparison of p values) were common and appropriate factorial analysis rarely conducted [26].
- Ideally a factorial analysis with a test of interaction should be implemented, to:
 - Estimate an average intervention (e.g., drug treatment) effect across the sexes.
 - Determine if the intervention effect depends on sex.
 - Maintain sensitivity of the study by accounting for the sex-related variation.

However, in situations where the sample size is low (e.g. 2) per intervention group per sex the power for an interaction effect will be low and could reduce the power for the intervention effect. Fundamentally, these experiments have limited power and could be classed as exploratory. In these scenarios, the researchers could fit a model where baseline sex differences are accounted for as a blocking factor with a visual inspection of the data to determine if further research is needed to assess whether the intervention effect depended on sex.

Supporting resources:

- [Phillips et al](#) ‘Statistical simulations show that scientists need not increase overall sample size by default when including both sexes in in vivo studies’ (PLOS Biology 2023).
- [Garcia-Sifuentes and Maney](#) ‘Reporting and misreporting of sex differences in the biological sciences’ (eLife 2021).

Q12: Will the design have a balanced inclusion of female and male samples?

Assessment advice:

- Where possible, proposals should ensure a balanced design.
- Definition of a balanced design: “any research design in which the number of observations or measurements obtained in each experimental condition is equal”, (i.e., include equal numbers of female and male samples evenly distributed across interventions).
- Example concerning imbalance: 20% Male and 80% female.
- Sampling restrictions (for example, aggression in males limit sampling opportunities in the field) could lead to an imbalance in the design. If this is an issue, this should be discussed, and sampling amended to minimise risk as far as possible.

- As you proceed through the decision tree, several additional classifications can accumulate. If the answer is yes then the “👍 **Balanced design**” classification is assigned if no, then the warning “👎 **Caution potential generalisability / analysis risk**” is assigned.
- The decision to proceed/fund etc will depend on a risk-benefit analysis of the assessing body and the ability to provide feedback to the researchers.

Why is this question included?

- Meta-research has identified several published articles where the studies were not balanced for sex of the samples at the start of the experiment (for example, one study had a ratio of 4:1 (male: female)) [30].
- Having a balanced design is important to ensure conclusions represent both populations (generalisable) and allows the potential for a sex difference to be observed. From a statistical perspective, balanced designs have higher power and more reliable test statistics.

Glossary

| Word | Definition |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Blocking | Blocking is where you manage sources of variation in the experiment by creating homogeneous groups (blocks). Within a block, experimental units are randomly assigned to the treatment groups. |
| Effect size | Quantitative measure of differences between groups, or strength of relationships between variables. |
| Experiment set | A series of related experiments where a decision on the design regarding sex inclusion will be common to all individual experiments. |
| Factor | Factors are independent categorical variables that the experimenter controls during an experiment to determine their effect on the outcome variable. Example factors include sex or drug dose. |
| Factorial design | An experiment that consists of two or more factors, with each factor having multiple discrete possible values (levels). |
| Generalisability | Also called external validity. The extent to which the results of a given study enable application or generalisation to other studies, study conditions, animal strains/species, or humans. |
| Interaction effect | When the effect of one independent variable (factor) depends on the level of another. For example, the observed intervention effect depends on the sex of the animals. |
| Harm- &/or cost- benefit | In this context, refers to the need to weigh the scientific benefits of using males and females against the likely harm (in terms of pain, suffering and/or distress) to animals and the economic or practical costs of this approach. |
| Intervention effect | Sometimes called treatment effect. This is the change in outcome that is attributed to the intervention applied. |
| Sex inclusive research | The research philosophy that emphasises the importance of including females and males in <i>in vivo</i> studies in such a way that a generalisable treatment effect is detectable. Critically, sex should be treated as a variable of primary biological interest. There is no requirement to power to detect a baseline difference between the sexes or treatment by sex interaction, but studies will detect large differences where they exist. |
| Levels | The values that a factor can take. E.g., For the factor “sex” the levels are male and female. |
| Statistical power | For a predefined effect size, the probability that the statistical test will detect the effect if it exists (i.e., the null hypothesis is rejected correctly). |

References

1. Beery, A.K. and I. Zucker, *Sex bias in neuroscience and biomedical research*. Neurosci Biobehav Rev, 2011. **35**(3): p. 565-72.

2. Mazure, C.M. and D.P. Jones, *Twenty years and still counting: including women as participants and studying sex and gender in biomedical research*. BMC women's health, 2015. **15**(1): p. 94.
3. Taylor, K.E., et al., *Reporting of sex as a variable in cardiovascular studies using cultured cells*. Biology of sex differences, 2011. **2**(1): p. 11.
4. Mogil, J.S., *Qualitative sex differences in pain processing: emerging evidence of a biased literature*. Nature Reviews Neuroscience, 2020. **21**(7): p. 353-365.
5. Graham, M.L. and M.J. Prescott, *The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease*. European Journal of Pharmacology, 2015. **759**: p. 19-29.
6. NIH. *NIH Policy on Sex as a Biological Variable*. 2016.
7. MRC. *Sex in experimental design - Guidance on new requirements*. 2022 [cited 2022].
8. Karp, N.A. and N. Reavey, *Sex bias in preclinical research and an exploration of how to change the status quo*. Br J Pharmacol, 2019. **176**(21): p. 4107-4118.
9. Group, M.W. *Working Group on Sex in Experimental Design of Animal Research* 2022.
10. Shansky, R.M., *Are hormones a "female problem" for animal research?* Science, 2019. **364**(6443): p. 825-826.
11. Woitowich, N.C., A. Beery, and T. Woodruff, *A 10-year follow-up study of sex inclusion in the biological sciences*. Elife, 2020. **9**.
12. Heidari, S., et al., *Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use*. Res Integr Peer Rev, 2016. **1**: p. 2.
13. Pape, M., et al., *Sex contextualism in laboratory research: enhancing rigor and precision in the study of sex-related variables*. Cell, 2024. **187**(6): p. 1316-1326.
14. Kuhn, T.S., *The structure of scientific revolutions Fourth Edition*. 2012, Chicago: The University of Chicago Press.
15. Velocci, B., *The history of sex research: Is "sex" a useful category?* Cell, 2024. **187**(6): p. 1343-1346.
16. Becker, J.B., B.J. Prendergast, and J.W. Liang, *Female rats are not more variable than male rats: a meta-analysis of neuroscience studies*. Biology of sex differences, 2016. **7**(1): p. 1-7.
17. Prendergast, B.J., K.G. Onishi, and I. Zucker, *Female mice liberated for inclusion in neuroscience and biomedical research*. Neuroscience & Biobehavioral Reviews, 2014. **40**: p. 1-5.
18. Kaluve, A.M., J.T. Le, and B.M. Graham, *Female rodents are not more variable than male rodents: A meta-analysis of preclinical studies of fear and anxiety*. Neuroscience & Biobehavioral Reviews, 2022: p. 104962.
19. Levy, D.R., et al., *Mouse spontaneous behavior reflects individual variation rather than estrous state*. Current Biology, 2023. **33**(7): p. 1358-1364. e4.
20. Zajitschek, S.R., et al., *Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications*. elife, 2020. **9**: p. e63170.
21. Phillips, B., T. Haschler, and N. Karp, *Including both sexes in in vivo research does not necessitate an increase in sample size: a key role for factorial analysis methods*. bioRxiv, 2022: p. 2022.09. 29.510061.
22. Karp, N.A., et al., *Prevalence of sexual dimorphism in mammalian phenotypic traits*. Nat Commun, 2017. **8**: p. 15475.
23. Miller, L.R., et al., *Considering sex as a biological variable in preclinical research*. The FASEB Journal, 2017. **31**(1): p. 29.
24. McCarthy, M.M., *Incorporating sex as a variable in preclinical neuropsychiatric research*. Schizophrenia bulletin, 2015. **41**(5): p. 1016-1020.
25. Prager, E.M., *Addressing sex as a biological variable*. 2017, Wiley Online Library. p. 11-11.
26. Garcia-Sifuentes, Y. and D.L. Maney, *Reporting and misreporting of sex differences in the biological sciences*. Elife, 2021. **10**.
27. Kappel, S., P. Hawkins, and M.T. Mendl, *To Group or Not to Group? Good Practice for Housing Male Laboratory Mice*. Animals (Basel), 2017. **7**(12).
28. Weber, E.M., et al., *Aggression in group-housed laboratory mice: why can't we solve the problem?* Lab Anim (NY), 2017. **46**(4): p. 157-161.
29. Weber, E.M., et al., *Aggression in group-housed male mice: a systematic review*. Animals, 2022. **13**(1): p. 143.
30. Stanford, S.C., et al., *Considering and reporting sex as an experimental variable II: An update on progress in the British Journal of Pharmacology*. British Journal of Pharmacology, 2023.