Want to comment? Your input is important. OLAW welcomes questions and comments from viewers of this recording. OLAW will post the comments, questions, and answers on the OLAW website. Please go to the OLAW Webinars and Podcasts page and click on the seminar title for further information.

Note: Text has been edited for clarity.

Improving Experimental Design: Ethical Implications and How the Experimental Design Assistant (EDA) Can Help

Speaker:

- Esther Pearl, PhD, Programme Manager for Experimental Design, National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs)

Broadcast Date: September 14, 2023

Slide 1: Improving Experimental Design: Ethical Implications and How the Experimental Design Assistant (EDA) Can Help

>>Nicolette Petervary: Good afternoon. I'm Dr. Nicolette Petervary part of the NIH Office of Laboratory Animal Welfare. And today is Thursday, September 14th, 2023. And I'm pleased to welcome you and our speaker to our webinar today entitled, "Improving Experimental Design: Ethical Implications and How the Experimental Design Assistant can help." There are just a few housekeeping details before we get started.

If you have questions throughout the webinar, please enter them in the Q & A box. The Q & A box does allow questions to be submitted anonymously, and the chat will also be enabled for this webinar. Dr. Pearl will be taking questions at the end of the webinar, and if the question is a little more nuanced or context-specific, or if there are any that we have no time to answer, we'll forward the questions to her after the webinar and then we'll pin the question and answer to the end of the transcript. We'll monitor the chat as best we can, and we encourage you to use it to interact with us and with other participants. We do ask people to abide by the OLAW Code of Conduct found on our webinars and podcast page.

The slides, transcript, and webinar recording will be available after the webinar on our website, but they do need to be processed for 508 compliance compatibility before posting and this can take a few weeks, so please do bear with us. There will be a participant survey at the end of the webinar. The survey will pop up once you exit. We appreciate your feedback in answering these because it helps us improve and identify topics of interest for future webinars. And now let's get started with an introduction for Dr. Pearl.

Esther Pearl is the Programme Manager for Experimental Design at the UK's National Center for the Replacement, Refinement, and Reduction of Animals in Research (the NC3Rs). Esther works to develop tools and resources to help researchers design experiments this includes the Experimental Design Assistant, or EDA, an online tool to guide researchers through the design of animal experiments, and the ARRIVE guidelines, which encourage improved design and reporting of animal research. Esther completed a PhD in Biochemistry at the University of Otago, New Zealand and worked with Xenopus as a postdoctoral researcher in the field of developmental biology at the Clinical Research Institute of Montreal, Canada; the National Xenopus Resource, USA; and King's College London, UK. Welcome, Dr. Pearl. The floor is yours.

Slide 2: Improving experimental design: Ethical implications and how the Experimental Design Assistant (EDA) can help (2)

>>Dr. Pearl: Thank you, Nicolette. Thank you very much for the introduction. So as Nicolette said, today I'm going to be talking to you about "Improving Experimental Design: Ethical Implications and How the Experimental Design Assistant Can Help."

Slide 3: www.nc3rs.org.uk

But I wanted to start with a very quick introduction to the NC3Rs where I work. So we are the UK's National Center for the 3Rs. So we collaborate with the research community to create and disseminate resources to help implement the 3Rs in practice. So this screenshot is from our website where we have free-to-use resources that cover all of the Rs, but today we'll be focusing on reduction. So we'll be discussing how we can improve experimental design to reduce the number of animals wasted in experiments that give unreliable results.

Slide 4: Learning objectives

So let's take a look at our learning objectives for today. So by the end of this talk you'll understand why we need to improve experimental design. You'll be able to identify some key ways we can improve biomedical research. You will know what the Experimental Design Assistant is and the benefits it has for researchers. And you'll also understand how the EDA can be used as part of the ethical review process.

Slide 5: Poll

So I wanted to begin by seeing what kind of roles you have when it comes to planning experiments, assessing experimental plans, and actually conducting experiments, and also what kind of support you have access to. So we're going to launch a poll. And I'd like you to click all of the things that apply. So the questions are:

- Are you involved in designing experiments?
- Have you had any formal training in experimental design?
- Do you have access to statistical support either for designing your own experiments or to help you understand other plans?
- Do you sit on an IACUC?
- And do you conduct your own research whether that's animal, clinical, cell-based or in silico?

So if we go ahead and launch the poll and then just click all of the ones that apply to you. You can click them all if they all apply. And we'll just give you a couple of minutes or a minute or so to answer. And I thought this would be a nice way of just seeing where everyone sits in terms of assessing experiments and designing them themselves.

>>Nicolette Petervary: It looks like most of the answers are slowing down.

>>Dr. Pearl: Yes. So should we go ahead and close the poll and see what kind of audience we have. Excellent. So we've got 70% of our audience sits on IACUC, and about half are involved in designing experiments. And about half have had formal training in experimental design, which I say is a very nice high proportion because in the UK we find lots of researchers don't actually unfortunately have experimental design training. And about a third of people conduct their own research. And half have access to statistical support, which is also excellent.

Slide 6: Why we need to improve experimental design

Let's go ahead and think about why we need to improve experimental design.

Slide 7: Research waste

So we've all heard about issues with reproducibility of animal and preclinical research, and we are all concerned about it because it contributes to research waste. So the Lancet published a landmark paper about 15 years ago and they estimated in that paper that about 85% of the investment in biomedical research was wasted. But a more recent study in ecology research came to similar conclusions. They reviewed 33 meta-studies, and together those 33 meta-studies combined over 10,000 studies. They found that actually 82% to 89% of research was being wasted, and they found this was mostly down to issues with experimental design, incomplete reporting, or studies that were just not published at all. So if we flip that round that turns into about 10% to 15% of research that's funded and generated is actually usable. So that's really bad news.

Slide 8: Research waste in *in vivo* studies: Ethical implications

And this has additional implications when we are talking about animal research. The authorization we have for animal experiments is given based on a harm-benefit analysis. The harms to the animals— these are things that are inherent to their transport, their housing or consequence of the scientific procedures they receive— these are weighed against the likely benefits to science and society such as new scientific knowledge or improvements in human or veterinary health. But if the findings of our research aren't reliable or that research is not reported in enough detail, then these benefits can't be realized and the research is simply unethical. So these studies are a waste of animals, but that also has problems when it comes to the implications of entire programs of clinical work based on the findings of unreliable animal research.

Irreproducible findings published in the scientific literature can trigger clinical studies and that can subject patients to treatments unlikely to be effective, and that can also delay the discovery of new treatments while we chase things that end up being a dead end. So improving the rigor and transparency of animal research will improve the entire drug development process. And irreproducible in vitro research can also contribute to waste in animal use. In vitro experiments often use animal cells or tissues, and also spurious findings in in vitro experiments can be used to justify a future animal experiment. Poor quality in vitro research can damage its credibility and that can delay the implementation of replacement technologies for people that are considering moving from animals to in vitro.

Slide 9: Contemporary definitions of the 3Rs

So reducing research waste by improving how experiments are designed and reported is really consistent with the 3Rs. So this slide shows the standard and contemporary definitions of the 3Rs. And I just want us to focus on the contemporary definition of reduction. This definition is: “Appropriately designed and analyzed animal experiments that are robust and reproducible, and truly add to the knowledge base.” So this shows that reducing animal waste is really a key part of improving experimental design.

Slide 10: Resources from the NC3Rs

Over recent years the NC3Rs has focused on improving reproducibility as well as rigor of animal experiments as part of our work on reduction. We've primarily done this by launching an online tool to help researchers with in vivo experimental design— that's the Experimental Design Assistant— and by publishing guidelines on the reporting of animal experiments, so that's the ARRIVE guidelines. So as part of my talk today I will tell you about these resources, most especially the EDA, and how they can support the IACUC, particularly as both EDA and ARRIVE are recognized by major US organizations.

So for example, the NIH recommends using the EDA in preparation of grant applications. And the Working Group established by the NIH Director to advise on Enhancing Rigor, Transparency, and Translatability in Animal Research published its findings in 2021 and that included the recommendation that applicant's planning vertebrate studies should use the Experimental Design Assistant when designing their studies. And that all publications of NIH funded preclinical animal research should include information outlined by the ARRIVE guidelines Essential 10. ARRIVE has also been highlighted by the National Academy of Sciences. So I will be describing both resources later in my talk but obviously focusing on the EDA.

Slide 11: Image of Academy of Medical Sciences report cover page

So there are many causes of reliability and reproducibility issues with biomedical research and lots of groups have identified them and they basically have identified the same set. This slide is from the Academy of Medical Sciences. They produced this report in collaboration with major UK biomedical research funders, and I'll briefly go over these causes of reliability problems.

- So one of the causes is omitting null results, and this is also known as publication bias: when scientists or journals decide not to publish studies unless the results are statistically significant. This obviously biases our literature.
- There's data dredging which is also known as p-hacking: this involves repeatedly searching a dataset or trying alternative statistical analysis until a magical statistically significant result is found.
- Another issue is weak experimental design: a study may have one or more methodological flaws that mean it's unlikely to ever produce reliable or valid results.
- We have underpowered studies. Statistical power is a probability that the analysis will detect an effect if that effect exists, and an underpowered study is too small to reliably indicate whether or not the effect exists. So essentially, all of the animals in an underpowered study are wasted.
- And then we have errors. Technical errors may exist within a study such as a misidentified reagent or computational error.
- Then we have underspecified methods. So a study may be really, really robust, but if its method is not shared with other scientists in enough detail then others can't precisely replicate it and we are not kind of passing that knowledge of a good rigorous experiment on.

Slide 12: Key ways to improve biomedical research

So let's move on to the key ways that we can improve biomedical research. We know it's not as rigorous as it should be, so what can we do to both improve research and simultaneously reduce the number of animals wasted on unreliable research?

Slide 13: Key aspects of good experimental design

So the next few slides we'll go over key aspects of good experimental design, and this will help us ensure results are more likely to be reliable. And these are some of the things that you can look out for when you're reviewing experimental protocols. [Some] aspects of good experimental design you can look for are randomization, masking also known as blinding, has the study included both sexes, is the sample size appropriate, and is there a pre-planned statistical analysis method.

Slide 14: Reporting of rigour criteria in animal research publications

So we have made some progress in improving our experimental design over the last few years, but it has been a little slow. This study really illustrates it. It was published in 2020, and researchers used the

software SciScore to analyze 1.6 million open access papers from across PubMed Central. So the graph on the left was specifically looking at papers reporting animal research in 2018, and we can see that the percentage of papers mentioning measures to reduce subjective bias is quite low. So randomization, which is kind of the yellowy bar, was only mentioned in 37% of publications; blinding in only 12%; and power calculation to justify sample size in only 7%.

But what I really wanted to highlight on this slide was the graph on the right. So that shows our progress when it comes to reporting these measures over the last 25 years. This graph is across all papers, so it's clinical and preclinical. And we can see that there has been some improvement over the years, especially for the reporting of randomization. So this was at about 10% in papers in the year 2000 and it got up to 30% by 2018, but we can see that the prevalence is still low. It's still only 30%, for example, for randomization and our progress seems to have plateaued a bit over the last few years.

Slide 15: Randomisation

So to dig a bit deeper into some of these let's start with randomization. How the randomization sequence is generated is really important. Haphazard, which is what happens when a human tries to create what they think is a random sequence, is not actually random. True random sequences can be generated in different ways including computer generated sequences, rolling a die, flipping a coin, etc. And to demonstrate why we need to use valid randomization methods, I'd like you all to pick a number between one and 10 and just pick the first number that pops into your mind. And we are going to start a Zoom poll and just get you to choose that number, that first number between one and 10 that popped into your head. And we'll close the poll once the responses start slowing down a bit.

All right. So as you can see in our lovely straw poll about third of us chose the number seven. So this is kind of illustrating that as humans we are not very good at being random. And the reason I asked you to pick a number between one and 10 is that someone did something very similar online a few years ago but they used numbers between one and 20. Now, Zoom won't let us give you 20 options in a multiple choice so we had to narrow it down to 10 for today's talk. But what this online poll did basically was it got people to choose a number between one and 20 and it got a computer to pick numbers between one and 20.

And you can see that there's a dashed horizontal line is at 5%. You can see that the pale bars, which are the computer's choice, they oscillate around the 5% mark and that's what you'd expect for a truly random choice. And the choices of the humans, which is the kind of darker blue bars, were not very evenly distributed. So in this case there were about 18 [% of] people chose number 17 and about 12% chose number seven and you'd expect those numbers to be closer to 5%. So this is why when we generate a randomization sequence, we really need to take the person out of the equation. We can't be trusted to do it.

Slide 16: Randomisation (2)

But why is it so important to randomize? Random allocation to interventions is really crucial for a couple of reasons. One, it helps us minimize selection bias, so that helps reduce systematic differences and the characteristics of animals allocated to different groups so [that] we don't start off with groups that are already different. One example of this is if I did haphazard selection when allocating mice to groups, I may accidentally end up with all the slowest mice in the same group. Randomization also helps meet one of the key assumptions of the statistical analysis methods which is that different groups should be drawn from the same background population using random sampling. So that means if researchers have not randomized animals into groups, then they shouldn't be running statistical analysis on the data.

Slide 17: Randomisation (3)

We can see the impact of randomization in the published literature. These data are from a systematic review looking at the efficacy of interventions in in vivo models of multiple sclerosis. So they combined the results from 126 different studies testing the efficacy of 36 different drugs and then compared studies that were randomized to studies that didn't randomize. And this is at the point of allocating to groups.

As you can see in this graph, on the y-axis we have the percent improvement in neurobehavioral score, so that is how much better did the drugs make things? The bar on the left is where random allocation to groups was not used and the average effect size, the average improvement from the drugs in those cases was just over 40%. But when random allocations to groups was used, the effect size was only just over 20%. So we can see that if you're not randomizing it really overestimates the treatment efficacy and that really contributes to reproducibility issues.

Slide 18: Masking

So let's move on to masking and why this is really important to help reduce bias. Masking is also known as blinding, and this means being unaware of which intervention an animal has received. So if possible, being unaware of which animals have received the same intervention as each other and when analyzing the data being unaware of which experimental group each group of data come from.

Masking can be used at different steps in an experiment:

- During the allocation and intervention, so when the animals are being assigned to experimental groups and steps where they receive the experimental treatments. Masking during allocation minimizes selection bias, reducing the chances of us subconsciously putting more similar animals together, for example, in one group.
- We also need to mask, if possible, during the conduct of the experiment [that is] during housing and welfare management of the animals to make sure they're all treated the same way.
- If possible, we should mask when assessing the outcome. So this [includes] steps where an outcome is measured or a sample is processed in preparation for a measurement. And this really helps us minimize confirmation bias when taking measurements or processing samples.
- And when assessing results [making sure that we're] blinding there. When we are doing data processing and statistical analysis, [this] minimizes our confirmation bias because there are often some subjective decisions about the analysis pipeline.

But sometimes it's not possible for a researcher to mask every stage of an experiment. For example, if there are visible differences between their groups, say they're doing an intervention looking at different genotypes and one has a different coat color (mouse with a different coat color), they won't be able to be blinded to which group is which when they're assessing the outcome. So in that case they just need to be careful to conduct some of the other stages masked. So they've just got to be really careful. If you know you can't mask one stage, be really careful about the others.

Slide 19: Reducing bias

So this experiment shows the importance of masking. In this experiment, 157 veterinary medicine students watched footage of pigs and they had to identify positive behavior between pigs and negative behaviors. So they watched two five-minute video clips of pigs, and they scored how frequently the different groups of pigs showed positive behaviors and how frequently they showed negative behaviors. One group of pigs were control pigs and the other group were 'High Social Breeding Value' pigs, so this

means pigs that were basically bred to be better behaved. And the results are at the bottom of the slide. And you can see in the observations that the 'High Social Breeding Value' pigs (that's the purple bars) they showed more positive social behaviors than the control pigs and fewer negative social behaviors than the control pigs. You can see that the differences between control and 'High Social Breeding Value' pigs were statistically significant in both cases.

Slide 20: Reducing bias (2)

But what's really interesting is, it turns out that the footage of the pigs was actually identical, and they'd just been slightly visually altered to mislead the students into thinking they were actually watching two different videos. So just like in the cartoon at the top of the slide, the footage of the pigs was mirrored, the brightness changed, the dates and pen numbers written on the videos were altered to make the students think they were looking at different videos. So half of the group was told they saw the 'High Social Breeding Value' pigs first and half were told that they saw the better-behaved pigs second. And so all of the differences they observed in the graphs were actually a result of their own expectations. So this really highlights the importance of masking when we are measuring the outcome of an experiment and that helps remove our expectations and subjective bias from our experiments.

Slide 21: Using masking/blinding in in vivo experiments

So if you do have researchers approaching you asking about strategies for implementing masking in real-life experiments, we have a resource page on the NC3Rs website based around our recent paper. And this paper was written in collaboration with AstraZeneca. The link to both the webpage and the paper at the bottom of the slide, and as Nicolette said, the slides will be available on the OLAW website. So the information here includes strategies for masking different types of interventions, strategies for masking data analysis, and then some ideas for how we as a whole community can make it easier for masking to be implemented more regularly.

Slide 22: Using both sexes

So another way to improve our experiments is to reduce sex bias. So in animal experiments most research is conducted on males still rather than females or both sexes, and this is a problem because we can't necessarily extrapolate findings from males to females. So clinical trials did use to have this problem as well, but about 30 years ago there was a real push to include both men and women in trials and nowadays your typical clinical trials would expect to be half and half. But this hasn't happened for preclinical research, so cell and animal research yet, and sex is often not considered as an important factor in the design and analysis of preclinical research. And this is a real missed opportunity for us to collect really crucial information that could inform the design of clinical trials or provide warnings on adverse effects that are sex specific.

A really famous example of this is thalidomide. So this was a drug developed in the 50s to treat nausea in pregnant women, but it did lead to the death of approximately 2,000 children and serious birth defects in 10,000 children. Evidence found in the 1980s actually showed that when they did in vitro testing on human female tissue they could have predicted that there'd be problems. And there are many other examples of drugs on the market that had to be withdrawn because of side effects that were either specific to women or specifically worse in women. So this really demonstrates the importance of studying females in preclinical research. Most biomedical research funders actually require a full justification for the sex of animals to be used in grant applications. For example, the NIH specifically states that using both sexes should be the default for preclinical experiments and that sex should be taken into account in the design, analysis, and reporting of the research.

Slide 23: Sex bias in animal experiments

So in animal research we do have a clear bias towards using males. The situation has been improving over the years, but we still have a higher proportion of male only studies. So this graph is from an article in the journal Pain, and it was an analysis of all the primary research articles published in that journal. The areas in blue are the overall proportion of studies that used males only. And we can see that over time that is reducing, but by 2019 about half of this papers reported were still using only males and only about 30% were using both sexes. And this is in Pain research this is an area where we know there are massive differences between male and female animals and clinically most patients with chronic pain are women. So we know there are sex differences here, but we're still not using both sexes very often.

Slide 24: Sex bias in animal experiments (2)

So let's just focus in on the studies where they did use both sexes and they looked for a difference, that's the ones highlighted on this slide. We can see that about half of the studies that used both sexes and looked for a difference discovered something. So this is information that would be missed in studies that only use one sex or in studies that use both sexes but aren't actually looking for sex differences.

Slide 25: Using both male and female animals

So in most cases researchers should be including both sexes in their protocol, and it should be clear when you are reviewing a protocol what the researchers are trying to do. Are they including both sexes to ensure the results are generalizable or are they directly comparing the sexes? And if they are directly comparing sexes, do they have adequate numbers to make this comparison? Will they need to include more animals? If the researcher is not directly comparing the sexes, they may not need to increase animal numbers, they could just do a power calculation for their different groups and just make sure half of each group is male and half female. And if the researcher is not sure about this then you can suggest they consult a statistician and that will help come up with a number that will give a nice reliable result.

Slide 26: Sample size

Using an appropriate sample size is also essential for reliability of results. So the number of animals in an experiment needs to be explained, and if the data will be compared with a statistical test (for example, tests that give a p-value or an F statistic), the sample size should be determined with a formal justification such as a power calculation or a simulation study. There are a couple of problems when sample sizes are just guessed.

Some studies will use too many animals and they tend to generate statistically significant results for things that aren't actually biologically relevant. So the example on this slide is a graph from a study looking at the effect of a drug on body temperature in ferrets. The difference between the vehicle and drug groups was only 0.25 degrees Celsius—so not a biologically relevant difference. But, because the study was overpowered, the difference was statistically significant. So this study used more animals than it needed to. But on the other hand, we need to make sure that studies use enough animals to give reliable results. Underpowered studies waste animals because the unreliable results mean we just can't tell if the effect we're looking for is there or not. And also underpowered studies are much more likely to give us false positives, [which] tell us there's a statistically significant difference when there really isn't.

For the conclusion of a study to be scientifically valid the sample size needs to be chosen correctly so that biological relevance and statistical significance complement each other, and this is what power calculations help researchers to do.

Slide 27: Sample size (2)

But if the data in a study won't be compared with statistical tests, researchers still need to explain how they decided on the number of animals needed for the study. So for example, if a study aims to establish if a new surgical technique works in mice the number of animals needed to determine this is basically how many would you need to ascertain if the technique would work and be practical for the future application; so what further studies they want to use this technique in. And things like that obviously need to be assessed on a case-by-case basis.

Slide 28: Analysis plan

And finally, we get to analysis plans. So it's really important that researchers have made their statistical analysis plans before they start experiments. Analysis plans are an integral part of experimental design, and having a clear analysis plan is really important. It promotes careful consideration of variables and outcome measures— so it makes sure researchers have really carefully thought about the interventions they're testing, the outcomes they're measuring, and do these answer their specific research question. Do they need to include any covariates or blocking factors or other things to take account of some other variability that could affect their results?

It also helps researchers identify the appropriate analysis path. Not all studies require a statistical analysis; in some cases descriptive statistics like means and standard deviations, confidence interval, median and quartiles, sometimes they are enough. And [in] other cases, inferential statistics are needed. That is making a statistical comparison between groups. Deciding your analysis plan ahead of time also helps prevent p-hacking. This is where researchers try different statistical analysis methods until they get a statistically significant result and this just produces false positives.

But another really key part about having an analysis plan ahead of time is that researchers can identify if they need help with experimental design before they start the experiment. Do they need help with the design? Do they need help with the analysis? If they realize their study is going to be more complex than they initially thought, then before they start, they can contact statisticians. Proper statistical support can be the difference between a study that is worth doing and a study that's a waste of animals. And I'm not sure how easy it is to get statistical support in the US, from the survey at the beginning it seems it's a bit easier than in the UK, but some institutions in the UK lack biostatistical support and this means researchers often have to include a statistician in their grant. So they'll collaborate with a statistician and add them into their grant as a member of their study group. And so obviously if you're going to add a statistician into your grant, you need to know that you need them early enough to write them into the grant. So that's why it's really important to think about this early so that researchers can do this if they need to.

Slide 29: The EDA: benefits for researchers and IACUCs

Now let's move on to the EDA and the benefits that it has for researchers and IACUCs.

Slide 30: Experimental Design Assistant (EDA)

So the Experimental Design Assistant is an online tool to help researchers design in vivo experiments. We are aware that many researchers don't have access to statistical support, so we have tried to replicate that support as much as possible in the EDA. So the EDA is free to use, it's secure, and it

focuses on internal validity: so rigorous study design, things that will lead to reliable results, things that can help us reduce bias. And it was developed with the help of a team of experts in experimental design and statistics.

Slide 31: Features of the EDA

So the EDA is a bit like having a personal statistical assistant. It gives researchers bespoke advice and feedback on experimental plans. It gives them recommendations on statistical tests appropriate for their specific design. It helps researchers determine the appropriate sample size. It has dedicated support for randomization and masking. It can improve transparency of experimental design and analysis plans. So it allows researchers to share and discuss their plans in detail with colleagues and collaborators. And it can also be used to help teach experimental design.

Slide 32: The EDA Diagram

So in the system experiments are represented as diagrams. And this is an example of what an EDA diagram looks like. So researchers build their experiment using these different shapes. Each shape denotes a stage or a part of their experiment, and the arrows show how these are linked together.

This diagram represents a simple two-group comparison. So it starts with a pool of animals and they are allocated by complete randomization. In this case to group one who will receive an injection of vehicle, group two who will receive an injection of drug. And then the plasma glucose in all animals is measured and then this data is fed into the analysis.

Each of these shapes has more information about that stage in the experiment, and in many cases there are dropdown lists to help researchers decide what's appropriate. So for example, we are looking here inside the allocation node and one of the dropdown lists is about the randomization strategy. A complete randomization: are they going to randomize within factors, within blocks, etc.? Another dropdown list is about the randomization procedure, so how that random sequence is generated. And this really helps researchers figure out an appropriate method or which valid randomization method will work for them.

And so the EDA provides feedback on experimental plans as well. So for example, helping identify possible sources of bias and mitigation strategies.

Slide 33: Images showing informational popups from the EDA program

So the example I've chosen here is this diagram does not include any nuisance variables. The prompt from the EDA gives an explanation of what a nuisance variable is: [it's] another source of variability or a condition which may influence the outcome and so it's not the thing you're specifically studying in your experiment. It goes on to explain common nuisance variables in animal experiments, and then it goes through different options for mitigating against the effects of this in the experiment.

Another example of a common problem which is explaining whether or not a step is masked or blinded. The prompt explains why this is important, [and] goes through different ways that researchers could implement masking in their experiments. And this helps them figure out what's most appropriate for their experiment and what's actually feasible in their lab or facility. Using this feedback researchers can amend their diagram; they can improve their experimental plan. And once this has been done, they can share it with others in the lab who'll be conducting the work. And actually, in my experience it's the researcher who is doing the experiment that tends to use EDA. They design the experiment and then they go to their boss or their PI and they go over it and just make sure everyone's on the same page.

Slide 34: Key features of rigorous research

The EDA has features to help researchers use rigorous methods.

- [For] randomization sequence, the EDA can do three different types of randomization: complete randomization, block randomization, and randomization within factors.
- The workflow helps enable blinding or masking. The random[ization] sequence is emailed directly to the person who's helping you with the masking. So if you are the researcher conducting the study, you can remain unaware of which of the group each animal has been allocated to for the whole experiment.
- And we also have some sample size calculators in the EDA for paired and unpaired t-tests, along with the decision tree to help decide what calculator is appropriate, [as well as] full guidance to help identify the parameters for the calculation ([for example], help with identifying biologically relevant effect sizes or estimating variability).

Slide 35: EDA report

The EDA can also help researchers communicate experimental plans once their experiment has been finalized in the EDA. So there are a couple of ways of doing this: there's a PDF and a read-only diagram. I'll start with a PDF. So the EDA report gives researchers [and] ethical review committees the confidence that the experimental design has been reviewed and had feedback on it. And we have found in our experience, researchers that submit this PDF report to ethical review boards in the UK can use that as a really good starting point for a constructive dialogue between the researcher and the ethical review board.

So the PDF report contains [information] really clearly laid out: the objectives and the hypothesis of the experiment; the animal numbers and a justification for that sample size;

Slide 36: EDA report (2)

any steps taken to minimize the effect of bias such as randomization and blinding; what the primary and any secondary outcome measures are in the experiment and the plan statistical analysis methods;

Slide 37: EDA report (3)

information on the characteristics of animals used in this study;

Slide 38: EDA report (4)

and it has a summary of any feedback from the EDA that has not been addressed. So if researchers are creating this, say ahead of submitting it for ethical review for example, they can have a look at it and they can go, "Oh, I forgot to take care of that." And they can go back and adjust their diagram and improve the experiment and then produce a new report or, if they want some help from the IACUC, for example, from this they can submit it as is and that could be a nice starting point for a conversation between the IACUC and the researcher. And it can also be a way to help members of the IACUC identify when there's missing information or identify things that they can start a conversation about. Like for example, in this report it says, “other sources of variability are not accounted for in the experiment.” So you could ask the researcher, "Okay. Are there other sources of variability that could influence your results? And are there ways you can mitigate around them?" And you can also point researchers to the EDA website pages that outline information about nuisance variables or if you do have statistical support in your institution, you can signpost how the researcher can access that.

The report does also include the statistical analysis method recommended by the EDA.

Slide 39: EDA report (5)

And at the end there's a visual overview, so an image of the diagram. And that's a really nice visual overview for people that prefer kind of visual information. You can easily see that in this case it's two groups being compared and there's just the one variable, etc. The information in that image is also in the table, so you don't miss out any information but you can get a nice visual overview.

Slide 40: Read-only EDA diagram

So the other way researchers can share their experiment is [with] our new read-only diagram. They can generate a static online version of the EDA diagram that they can then share with anyone regardless of whether or not they have an EDA account. The read-only diagram has that same table of key experimental design information, and that's shown on the left in this slide. But instead of a static image of the EDA diagram at the end, if we hide that table of information, there's a clickable version. So others can click on a part of the experiment they want more information on.

The read-only diagram is shared using an unguessable URL, so just a really long link with lots of numbers and letters, and it also has an access code so that if the researcher wanted to share this with you they'd have to send you both the link and the access code. And that means the researcher can control when they want that to be generally publicly available. For example, when they publish a study, they can then remove the access code.

Slide 41: The EDA: Benefits for researchers

So I just wanted to highlight the benefits for researchers and then the benefits for ethical review boards of the EDA. So for researchers: The EDA gives bespoke advice on experimental plans, leads to better experimental design; analysis recommendations that are appropriate for the specific design; help determine the appropriate sample size; and support for randomization and masking.

Slide 42: The EDA: Benefits for IACUCs

For the ethical review committee: your IACUC can consider, for example, requesting an EDA report or a link to the read-only diagram from researchers that have used the EDA to plan their experiments. So by asking for the EDA PDF report or a link to the diagram as part of an ethical review application, you can know that researchers have received feedback on their experimental plans. And the PDF report lays out the key experimental design information in the standardized format, it means the key information is in the same place in each report. So if you are looking at a lot of protocols then the key information is in the same position and missing information is really clearly marked. So you can follow up for more information if that's something that's important for the experiment.

And you can also use the EDA report to help you identify information or experimental design issues that you may not have confidence finding yourself if you're not as confident with experimental design. And also if you do have any researchers that approach you asking for experimental design advice or resources, you can point them to the EDA software and the supporting website.

Slide 43: Live: inside the EDA

So I thought now I'd take you inside the EDA, just have a little look about what it's about inside. So I'll just quickly switch my screen share. So we'll do a new share and we'll share the other screen.

So this is the EDA homepage, and I thought I'd just show you what the EDA looks like from a user perspective in case you're not actually going to use the EDA yourself. The EDA homepage has a little description about what the EDA's about and a quick one-minute intro video. So if you ever need to explain the EDA to anyone you could send them there. But let's just log in and have a quick look.

So I'll just log in here. Here we go. So when someone first logs into the EDA, they arrive at the start page and it gives kind of tips on how to get started using the software. On the right there's a flow diagram, how to get the most out of the support in the EDA. We have templates and examples to help get people started.

But what I really wanted to show you today was how the EDA can provide feedback to researchers. And to do this, I've designed an experiment in the EDA that has some flaws. So I'll just go ahead and open it. So the flaws that are in this experiment mean that the experiment's not as rigorous as it could be, and we are not maximizing the information possible from the experiment. We want to get as much information as possible from an experiment and make it as rigorous as possible because that's all part of the harm-benefit analysis, right? So a poorly designed experiment won't give reliable results and that's a waste of animals.

So this is the experiment; it's the simple two-group comparison like I showed on the slides. Each of the shapes here represents part of the experiment. And I just thought… inside each shape there's more information. So let's have a wee look. *[The presenter clicks on a shape.]* I started filling in this one. [Here] we've got male mice. I could also add really important information like the strain, the age, the weight, have they had any previous procedures, the genetic background, etc. You can put all the really key information about the animals you're using there. You can also add information about different parts of the experiment in all of the shapes. So let's just go ahead and get some feedback on the experiment.

So if I go up to “tools” here, these are things that can help design the experiment. And if you go over to the question mark, it opens the help center so you can figure out what each of those tools are. In this case, “critique” gives targeted feedback on the experimental design and that's what we want. So let's go ahead and run the critique. What’s happening here is [that] behind the scenes, the EDA is looking at the experimental plan and it's looking for some different types of issues. It's looking for clarity: Is it clear what's happening in the experiment? [It] checks for missing information: Is there anything that's really important to know that's not being put in? And it checks for good experimental design practices: Is there some advice that EDA could give that will improve the experiment?

[A popup appears onscreen as the critique tool completes the review] So basically now that the critique is finished, you can see that there are a few icons on some of the different shapes and these are different types of feedback. And we have a legend in the top right. So I'm just going to quickly show you maybe one of the warnings or maybe two of them.

[The presenter clicks on a warning flag on a shape.] So in here on our experiment there's a couple of warnings: one is that we're only using one sex in this experiment. So this goes on to say that basically if you only use one sex it really limits how generalizable the results are. So that gives you a warning about that. And then the other one is saying, other sources of variability are not accounted for in the design of the experiment. So this is the one about nuisance variables and how if we don't mitigate against those we could end up in trouble.

So with nuisance variables we can have a couple of problems. One is, it could confound the experiment entirely, so you may have no way of telling what's an effect of this other source of variability and what's

the effect of the thing that you are actually testing. Or it could just mean that your experiment is not as sensitive to testing the intervention that you want to test and that you'd need to use more animals to be able to see a difference. So taking account of these other sources of variability can actually mean you can use fewer animals in your experiment. So both of those pieces of guidance I showed you can help researchers improve their experiment and make it more likely to give reliable and generalizable results and hopefully help them reduce the number of animals they need as well.

So I mentioned in the slides that the EDA can be used to communicate experimental plans via making a PDF or this read-only diagram. I just want to show you… if the researcher has used the EDA to design their experiment, I want to show you how easy it is for them to create these outputs. So we go "tools," "experimental design report," and what that does is it opens a new window. And what the EDA is doing in the background now is it's pulling out this key information from the experiment and it's putting it in a PDF…so it's building that PDF. It takes a couple of minutes. So while it's building the PDF, I'm just going to show you how you can create that read-only diagram as well.

So if we go back into our diagram. We just go "tools", "read-only diagram." We've got to make sure we save it first, and then it gives us this window that just tells us a little bit about the read-only diagram. For example, it tells us it's a static point in time, so if we made changes to the diagram it won't change the read-only diagram, things like that. So all the researcher needs to do, in this case, is click "create new version" and it will start generating this diagram as well. These are two ways that experimental plans can be shared with people that don't have an EDA account.

So the report should be finished shortly and then we'll go and show you basically how we can have a look for missing information or things that could be done to improve the experiment. So this is how the researcher can create these and then I'll show you how you as an IACUC can look through them and have a look for key information. So the amount of time [the reports take to generate] depends a little bit on how many people are using the EDA at once. I did just give an EDA demonstration about an hour and a half ago, so I'm not sure if we've got some new users or not. But it looks like the report is finished.

So it gives us the title of the diagram and when we created the report. And then it's got information from inside our nodes. So this is information about the objectives, group and sample size, what we are doing to mitigate against the effects of bias. You can see here under characteristics of the animal, all I put in that diagram was the sex and the species. And you can see that "information not provided" really clearly flags where I've got missing information. So you can see that, and you can also see here that I haven't included any blocking factors or covariates in the analysis. And in the feedback [section] provided by the EDA, it's saying that I hadn't actually specified the method of analysis or statistical assumptions, so there's some missing information there, and it's also spelling out that I've only used one sex in this experiment and [that] I haven't taken account of other sources of variability. So you can see how that's pulling up things that you could discuss with researchers as part of the IACUC process.

And if we go back to looking at the read-only diagram. So if I wanted to share that with someone else, I'd have to copy the URL [and] paste that in an email or a message, however you want to share, and then I'd also have to copy and paste the access code. But since we are in the account that created this read-only diagram, I can just go ahead and click "open." And you can see that we've got the same table of key experimental design information on the left here. If we scroll down, it's got the same flagging where there's information missing and the same list of things that we could do to improve the experiment. But if I go ahead and close that by clicking on these double arrows here, we've got a clickable version of my EDA diagram and there's a legend down the right. So if you share it with someone who's never used the EDA before they have an idea of what all the different shapes mean.

But the thing is that you can click, if someone shares this with you, you can click on the bit you are most interested in. For example, if you want to look at the pharmacological interventions, you can click on that node and you can see, okay: the drug, the dose of the drug, its subcutaneous, and they're not anticipating any possible adverse effects. Now, if this study included surgical interventions, then the surgical intervention node would have information about the analgesia, the anesthesia, pre- and post-op, and the monitoring and things like that. So you can click on the different nodes to get more information about this experiment. So that means that you get lots of details in the read-only diagram. So that's given you a quick overview of the EDA and how it can help make communicating experiments easier.

So I'm just going to go back to the slideshow. So, PowerPoint slideshow. There we go. And we will continue.

Slide 44: EDA demonstration

So if you would like to encourage researchers to use the EDA or you'd like to use it yourself, we do run regular live demonstrations about every other month. And we do also have a recording of a live demonstration online. So for example, this link at the bottom has a recording of a demonstration of how to use the EDA to build an experimental design, get feedback and improve the experiment, generate a randomization sequence, and generate both the PDF report and the read-only diagram, as well as how to find help that's in the system. And this link at the top links to the page advertising our next couple of live demos which are done via Zoom.

Slide 45: ARRIVE guidelines

So before I wrap up, I just wanted to talk about a complimentary resource to the EDA that also helps us with reproducibility of animal research and that's the ARRIVE guidelines. So they were developed initially to improve the quality of reporting of in vivo experiments and thus improve reproducibility of animal research. And they are recommended by the NIH and the National Academy of Sciences. So they are a checklist containing key information necessary to describe a study comprehensively and transparently. And we saw from the Academy of Medical Sciences report earlier in my talk that inadequate reporting really contributes to irreproducible biomedical research and therefore wasted animal lives. But the ARRIVE Guidelines can also be used as a framework for planning research studies.

Slide 46: Skipped

Slide 47: www.arriveguidelines.org

So when we revised the ARRIVE Guidelines recently, we wanted to address issues around researchers not understanding some of the reporting requirements or not understanding the concepts behind them. So we published an extensive explanation and elaboration document alongside the guidelines. This is a screenshot from the ARRIVE Guidelines website and it has things like why it's important to report specific information. But also really key, clarifying key concepts particularly about experimental design statistics and this is really useful when you're planning studies. We do recommend that people look at this alongside using the EDA to really make the most out of the experimental design advice. And there are also examples from the published literature showing how to report this information. If you do want more information about the ARRIVE Guidelines specifically, the NC3Rs website has an ARRIVE webinar. https://arriveguidelines.org/resources/webinar And OLAW itself has recorded a couple of webinars on the ARRIVE Guidelines by our ARRIVE Guidelines working group member Dr. Penny Reynolds. So they are available on the OLAW website.

Slide 48: The RIVER recommendations

Many animal experiments also involve an *in vitro* or *ex vivo* component, and we have developed the RIVER recommendations to help describe information to report so that methodological rigor and reliability of results of these studies can be evaluated. They're currently out in a preprint and we are about to begin road testing them with researchers. So if you are interested or you know someone else that is interested, please scan the QR code to find out some more about that.

Slide 49: Experimental design resources links

And finally, I have compiled some links to resources that you might find useful either for yourself or to direct others to, and a PDF of all of my slides will be available on the OLAW website, so you can access all of these links that I've used throughout the session once that goes up.

Slide 50: Thank you!

And I'd like to thank you very much for your attention. I'm happy to take any questions. And I'll hand back over to you, Nicolette.

>>Nicolette Petervary: Thank you, Dr. Pearl. We've got a lot of chat in the chat, and we've got a few really good questions in the Q & A box. So I will start with the question-and-answer box and then if we have time, we'll highlight some of the chat comments and questions.

So the first question is: What would you recommend are the key components to robust experimental design training?

>>Dr. Pearl: I guess other than identifying the concepts that are important in experimental design, it's also if you want people to be able to make changes then we need to give them practical things that they can actually implement. So one of the things we are trying to do is talk to researchers that have implemented things well such as masking, find out what the barriers were and then try and make resources that can help people get around those barriers. In some cases it's misconceptions, in other cases it's things that we need to change as a community. And I think one important part of experimental design training is we need to make sure it's all very well to say “this, this, and this are ideal” but it's not always possible in a real situation. So to make it more likely to make a difference I guess, you want to make sure there's practical tips that people can come away and do something about.

>>Nicolette Petervary: Great. And then we have a follow-up question. What are the main methods of randomization? That's a huge topic I'm sure, but if you can just provide an overview.

>>Dr. Pearl: So I guess one of the easiest ways to randomize is to use a random number generator. So you can list your say animals by their animal ID in Excel and use the =RAND function and create a random number and then sort them based on the lower to higher of the random number. The EDA [also] generates the random numbers for the random sequence for you. There's a random.org that has a list randomizer, so you can enter all the names of your animals or identifiers of your animals and have that randomized. So I would really recommend computer generated randomization.

But I'd also say that complete randomization isn't always appropriate. So in animal studies that have smaller groups, if you did a method of complete randomization, you may still end up with say more heavy animals in one group rather than another. That's where things like block randomization can come into play. So you could say, “Okay. We're going to split the animals into groups of high weight, medium weight, and low weight” based on the bottom third, the middle third, the top third, and then make sure that that top third are evenly spread amongst your experimental groups, etc. So that's a way to make sure that you're spreading out the variability.

Another way that's really well used in things like cancer studies is minimization, and that is where the randomization... So you have the characteristic of the animal you're worried about, which could be tumor size, in a spreadsheet and you use minimization to make sure that they are evenly spread amongst the groups. We actually have a collaborator who is working with someone on a Shiny app, a freely available online software to do minimization, and they are going to let us host it on I think the NC3Rs website in the future. So I'm really hoping we'll be able to offer people a freely available minimizing app at some stage because I don't think there's anything out there if you can't code it yourself. So that's a really good way of randomizing, or distributing the characteristics you're worried about evenly. But there isn't anything freely available for people that can't code it themselves at the moment. So hopefully that's coming.

>>Nicolette Petervary: Well, we'll be looking forward to that when it comes out for sure. Another question we have is: Is the EDA tool updated when new designs or stats are available?

>>Dr. Pearl: Yes. The EDA is kind of iteratively developed and improved, and this year we added extra advice and guidance for covariates. So for example, if you want to take account of any baseline differences in animals. The read-only diagram that I mentioned came out this year as well and we've kind of improved the usability of our randomization sequence function. So those are all things that have happened this year. And basically, it's iteratively improved and developed.

The next thing we want to focus on in terms of stats and advice is equivalence testing. So let's say you've got a refined method for something and you need to prove to the community that uses the gold standard method that your refined method gives equivalent results, then you'd use an equivalence test and test your refined method against the gold standard. So we don't actually have any support for equivalence testing at the moment, at the moment it's all difference testing. So that's the next thing we are working with. And we've got some statisticians who very kindly volunteering their time to make sure that we're giving the appropriate advice at the appropriate time. And we make sure all of the statisticians agree before we put it in the EDA.

>>Nicolette Petervary: And that's a great point that statistical experts are weighing in on the EDA.

>>Dr. Pearl: Yes.

>>Nicolette Petervary: And it's on an ongoing basis. So the next question that we have. This is a great question and I'll take this one. It says: I'm not aware that OLAW requires the IACUC to evaluate if the statistical analysis of data is appropriate to the study. Our IACUC has historically relied on the fact that the grant was awarded and that scientific review of the proposal has okayed all of the procedures as well as the statistical analyses. Would you like to comment on that because truth be told we're tempted but we do not have policies to back us up. And that's a great question and it's very gray, right?

So OLAW FAQ D12 talks about whether the IACUC is responsible for judging the merit of scientific proposals. And there are certainly overlaps with the scientific review group, but it's not the same thing. We do look for the IACUC to evaluate if the study is sound and if it's scientifically valuable research because presumably a study that can't meet those criteria would be inherently unnecessary and wasteful of the animals. So there is that. But I realize that especially with the varied expertise among IACUCs that can be a really big ask. The reasons we're featuring the Experimental Design Assistant is because this can be a resource to educate investigators and it can also be a resource for finding really big red flags.

We don't expect every member of the IACUC to have extensive experience at every statistical method. But if you see some outrageous number of animals, and for USDA species they require a justification for animal numbers I believe, this would be something that you could just say, "Hey, we have this tool available to help with this." Or just open up a conversation with a PI and say, "Can you explain the animal numbers a little better because we don't quite understand." So I realize that that is hugely challenging. But I think just knowing what tools are available to help PIs make it easier for them, especially if you don't have a lot of statistical support at your institution. And framing things as a question and an opportunity to become educated on the PI's research is a really good approach. I hope that helps.

>>Nicolette Petervary: And let's see. Okay. So we have other questions. We don't have too much time left so I'll just quickly run through these. Everyone who attends the webinar live should receive an email after the fact and that serves as proof of participation. So that will come after. And you will also receive a survey link afterwards. So if you don't get that for any reason let us know.

And I think we already answered this question: Will the NC3Rs continue to develop and evaluate this tool over time? And it sounds like they've got a lot of exciting projects in the works.

And here's a good question that I think we should answer: What does the EDA have in place to protect information?

>>Dr. Pearl: That is a very good question. So the NC3Rs is a research funder as well as a developer of resources and tools, so we've basically made the EDA as secure as we make our grant management systems. So if you log into the EDA, you are the only person that can access your EDA account; you are the only person that can access your diagrams. We have I think one person who works with the developers who can do it in extreme emergencies, but there are distinctly set out rules they need to follow in order to access an account. And so if you came to me for help with the EDA, you would have to explicitly share your diagram with me.

If you are worried, we also have multifactor authentication which you can add to your account if you would like and that can use either an app or an email. So it sends you a verification code to make sure

that it's you when you're logging in. We do regular penetration testing to make sure that we are covering bases in terms of being hacked and things like that. Basically, we are treating it as if this is your data for a grant application, so we are being as careful with experimental design data in the EDA as we would be if that was a grant application from you. Yeah. We take that very seriously.

>>Nicolette Petervary: Perfect. And we have one last question I quickly want to get to and then I think we'll have to go to the final slides. But would you recommend the EDA tool for large animal studies as well and for multi-variable studies?

>>Dr. Pearl: Yes. So the idea behind the EDA was [that] it is modular so you can put things in whatever order, [basically] like Lego bricks, in a way that represents your experiment. The Experimental Design Assistant is looking primarily on principles of experimental design that are fairly universal. So it shouldn't matter whether you are using large or small animals the principles are the same. Some of the practicalities might be a little bit different in terms of masking [and] randomizing. And I guess in some cases with very large animals it's more expensive to have appropriate sample sizes, but that's more about thinking about the research questions you can realistically answer. But the principles of experimental design apply throughout. Well, the other bit was about multi-variate studies, was it?

>>Nicolette Petervary: Correct.

>>Dr. Pearl: So yes. So you can put multifactorial studies in there. You just put more independent variable of interest nodes for each independent variable of interest you have. So yes, it can give advice on factorial studies as well.

>>Nicolette Petervary: Okay. And if we can proceed to the closeout slides. There's one more slide after this.

Slide 51: In case you missed it!

In case you missed it, we've been focusing a lot on experimental design this year. So we have two other OLAW online seminars recorded: one, Foundations for Evaluating Study Design; and the other was on the ARRIVE 2.0 Essential 10. So please feel free to take a look at that. This recording, as I mentioned before, will take a couple of weeks but we'll get that out as soon as possible.

Slide 52: Next Webinar: Winter 2023

And our next OLAW webinar will be in the winter with a topic to be determined but we hope to see you all then. And we will try and take these questions and any that are unanswered we will append to the transcript. Thank you all for attending, and a big thank you to Dr. Pearl. We'll see you next time.

Questions

These questions were collected from the chat, Q&A, and email after the session and provided to the speaker. The responses represent the speaker's comments and opinions.

1. Regarding the use of both sexes in experiments: when housed in the same cage, male mice fight; female mice don't. How do we overcome that practical barrier?

   We are aware that aggression amongst male mice can be a problem, but there are ways to reduce this. The NC3Rs conducted a study in collaboration with several universities to monitor aggression in mice and came up with some guidance on the subject. The guidance can be found here: https://www.nc3rs.org.uk/minimising-aggression-group-housed-male-mice.

   The paper this is based on is here: https://link.springer.com/content/pdf/10.1038/s41598-019-51674-z.pdf

   Other papers on the topic include https://doi.org/10.3390/ani7120088 and https://doi.org/10.1038/laban.1219

   The NC3Rs also jointly hosted a webinar, with the UK's Medical Research Council, on using both sexes in animal research. One of the experts discussed practical challenges when using both sexes, including housing and husbandry issues. The link to the webinar is here: https://nc3rs.org.uk/events/nc3rs-mrc-joint-webinar-using-both-sexes-animal-experiments

2. Regarding the use of both sexes in experiments: When using males and females, shouldn't groups of females in different stages of estrus also be used to provide more generalizable results for the human female population?

   Unless the stage of oestrus cycle is biologically relevant to your research question, you do not have to explore the effect of oestrus stage in your experiments. There is a common misconception that the reproductive hormone cycle means that female animals are inherently more variable than males. This is not always the case. Meta-research in both rats and mice have found no evidence that randomly cycling females had greater variability than male animals in the outcomes they measured (https://doi.org/10.1186/s13293-016-0087-5 and https://doi.org/10.1016/j.neubiorev.2014.01.001).

3. How is randomization of cage placement beneficial? For example, does it address issues with light exposure (cages in the middle vs edge)? Or is it with how they're handled by researchers (order of the treatments, etc.)?

   Randomization of cage placement can be beneficial as a way of minimizing potential confounding introduced by cage location. This could be in relation to light levels, or distance from the door (cages near the door will be disturbed more often than cages further from the door). If you have a small number of cages, you could use cage location as a blocking factor (e.g. if you had eight cages in total, you could randomize two control and two experimental cages to the four positions nearest the door, then randomize two control and two experimental cages to the four positions further from the door). Another option is to arrange cages in a staggered pattern (one control cage, one experimental cage, one control cage, one experimental cage etc.) as this spreads any variability from the different cage positions amongst your groups and is better than having all control cages together and all experimental cages together. The EDA website has more information about blocking factors here https://eda.nc3rs.org.uk/experimental-design-variables#blocking.

Randomizing the order of treatments and/or measurements is another way to reduce potential confounding your experiment. For example, if all control animals are treated first and then all experimental animals afterwards, any fatigue in the experimenter that could affect how the treatment was administered would predominantly affect the experimental animals. Using masking (aka blinding) as well as randomization reduces bias further. For example, with a subcutaneous injection, an experimenter may be less careful to inject the full volume if they know the syringe only contains saline. This also applies when taking measurements. An experimenter looking for clinical signs in animals may subconsciously look more carefully at the animals that have received the active intervention compared to control animals making them more likely to notice clinical signs in those animals (even if the signs are there in animals from both groups).