**ILAR J**

Volume 55, Number 3, 2014

***Experimental Design and Statistics***

**Festing and Nevalainen. The Design and Statistical Analysis of Animal Experiments: Introduction to this Issue, pp. 379-382**

Domain 3: Research

SUMMARY: Scientists have emphasized the importance of animal welfare using the “3Rs” and the conditions in which laboratory animals are kept now are at a high standard. However, there remains a cause for concern: poorly designed or inadequately analyzed experiments. These experiments cause waste in animals and resources because they are found to be unrepeatable. This article is a review to address some of the issues to ensure that scientists using animals are well trained in both the biology and the use of animals in well-designed and properly reported experiments.

Animal research has made enormous contributions to human health and welfare. Examples include the rabies and anthrax vaccines, chemotherapy, insulin, development of organ transplantation, among others. Medical research still requires the use of some animals, so we must carefully look at the way they are used in order to develop new treatments to diseases.

The purpose of this issue of ILAR is to be critical of the current methods of animal research, in order to serve as a wake-up call to all those involved in animal research.

4 major themes in this issue:

1)  The quality of animal experiments – there is poor reproducibility and translational failure with many experiments. These problems may arise from disregard for the basic principles of experimental design such as randomization and blinding. They also may come from ignoring variation from other areas such as animal husbandry, season, reproductive cycles, diurnal rhythm, etc.

2)  Systematic reviews and meta-analysis of animal experiments – normally, critical appraisal is performed informally on published research, but in this issue a case is made for a more formal approach. Systematic reviews can also help improve the methodology and quality of animal experiments.

3)  Choice of animal model – in new drug trials, the failure rate is very high. This may be due, in part, to the fact that some animal models may not truly represent the human condition it is supposed to represent.

4)  What can be done to improve the quality of animal experiments? Scientists need better training in experimental design and basic statistics.  The NIH is currently developing a strategy that includes teaching relevant methods to investigators as well as taking steps to relieve pressure on investigators to publish papers in high-impact journals.

QUESTIONS

1. True or False? Randomization and blinding are not important concepts in experimental design

2. Which of the following can cause variation in a study, but is frequently overlooked by investigators in the study design?

a. Season/time of year

b. Cage change cycles

c. Reproductive cycles

d. Diurnal rhythm

e. All of the above

ANSWERS

1. False

2. e. All of the above

**Bailoo et al. Refinement of Experimental Design and Conduct in Laboratory Animal Research, pp. 383-391**

SUMMARY: A review of literature revealed that there are significant problems with the validity of biomedical research. Poor reproducibility of results and a failure to translate preclinical animal research into clinical trials highlight the concerns of researchers. The authors discuss many areas of bias and validity that routinely affect the reported outcomes. Lack of internal validity leads to bias while lack of external validity contributes to poor reproducibility and translational failure.

The four primary types of bias with respect to internal validity are selection bias, attrition bias, performance bias, and detection bias. Selection bias is caused by lack of randomization or when the criteria for inclusion and exclusion of animals are poorly defined. Performance bias occurs when there is a systematic difference in the interaction with the animals. Detection bias occurs when the outcome is measured differently between treatment groups. Blinding investigators and researchers will help prevent both performance and detection bias. Attrition bias occurs when there is an unequal distribution of dropouts among treatment groups.

Another significant source of bias is sample sizes that are too large or too small. Overpowered studies use more animals than needed to detect a significant effect of a given size. Underpowered studies are more prevalent and more problematic. Underpowered studies are unable to detect biologically relevant effects which mean the animals are wasted for inconclusive research.

A common approach to reducing poor experimental conduct has been the implementation of reporting guidelines. The Animals in Research: Reporting In Vivo Experiments (ARRIVE) guidelines have been endorsed by over 430 journals, funding agencies, and academic institutions in hopes of improving scientific reporting. The guidelines provide a useful tool for improving internal validity of animal research. Additional precautions such as splitting experiments into small batches of animals that are tested some time apart, using multiple experimenters for testing and data collection, and spreading test sessions across time of day may improve the external validity of the research.

QUESTIONS

1. Construct validity depends on the specific disease that is modeled, and there is no simple method for assessing construct validity.
2. ARRIVE guidelines are based on the CONSORT guidelines developed for human clinical trials.
3. Testing many hypotheses in parallel using small sample sizes is a valid way to produce reliable results.
4. Reproducibility is a cornerstone of the scientific method and poor reproducibility threatens the credibility of the entire field of animal research.
5. The gold standard of experimental design is to hold constant all factors except for the independent variables under investigation.  This standard compromises external validity and reproducibility in laboratory animal studies.
6. The principals of reproducibility and falsifiability are the reason that science is often viewed as self-correcting.

ANSWERS

1. True
2. True
3. False
4. True
5. True
6. True

**Nevalainen. Animal Husbandry and Experimental Design, pp. 392-398**

SUMMARY: This article discusses the importance of controlling all major variables and changes of variation in regard to laboratory animal husbandry practices for the experimental design. It recommends that all researchers should familiarize themselves with the current routine animal care of the facility serving them, including their capabilities for monitoring of biological and psychochemical environment.

Husbandry variables can impact the study design. Controlling all major variables as far as possible is the key issue when establishing an experimental design, the only difference between the study and control groups is the procedure and nothing else. Bias is any systematic difference between the groups in addition to the procedure in the study design. The other common mechanism affecting study results is a change in variation. While establishing an experimental design for an experiment, it is important to invest adequate time for the identification of such factors and building up effective strategies to cope with them. The following selected animal housing and care related items and their potential to interfere with the experimental design are discussed: husbandry-related cycles, in-house transport, caging, psychochemical environment, material-related environment, feeding and social aspects.

Animal life follows many different cycles. They are capable of sensing the season even when housed in tightly regulated environment with no windows. In rodents, female reproductive cycle is sensitive to the pheromones present in male urine. The Whitten effect is one example. Therefore controlling spread of odors within a room becomes an important issue. During the weekend, fewer people are present in the animal facility, and hence a reduced acoustic environment audible to animals in the animal room. It has shown that blood pressure in the workings days in rats are higher than during weekends. Cage changes and room sanitation represent major exposures for laboratory animals. Rats housed in their home cages display increased locomotive activity, bedding manipulation, and defecation in both the mornings and afternoons of cage change days for several hours., In mice, post cleaning activity also includes aggression. Therefore sensitive behavioral experiments should not be performed the cage change day or the following day. To prevent pheromonal interference and stress-induced pheromonal release in their research subjects, the lowest possible frequency of cage change out should be chosen to avoid unnecessary stress to the animals. Transfer of specific olfactory cues or of nesting material during cage change out can decrease aggression and stress in group-housed male mice.

Moving the animals to another space than the animal housing room for testing has shown that it takes at least one hour of acclimatization for the body temperature to be restored. It is recommended that the investigators allocate the animals at random into the study group, and then cages should be placed into the cage rack also at random. If cages are assigned to the cage racks in a systematic manner, then external factors can introduce bias into the statistical analyses. In mice, temperature and humidity variation can affect the age of puberty. Low temperatures and extremely low humidity levels have shown to delay sexual maturation. Laboratory animals have a much better developed sense of smell than humans. Any disruption to it carries wide range of consequences to the animals and hence to the study results.

The most common clear cage material, polycarbonate, has been shown to leach Bisphenol A, a monomer with estrogenic activity. This compound becomes a problem if the polycarbonate box is exposed to high temperatures and alkaline conditions, allowing an increase in the amount of leaching. Softwood bedding contains compounds with liver microsomal enzyme induction properties and can contain a measurable amount of endotoxins. Cage complexity becomes environmental enrichment once it has been verified to confer a welfare outcome on the animals. For the researcher, the most important aspect of complexity is consistency; the added items must be present all of the time and have to do the same for all animals. In group-housed animals, social hierarchies may influence the use of complexity items. Minor complexity changes may alter the physiology and development of the animals in an unpredictable way. Considerable variation between batches of the same diet brand has been observed. Keeping to one batch in an experiment is advisable. Good laboratory practice-guidelines emphasize that the study design should take account of kinship, since litter effects are large accounting for 61% of the variation in behavioral outcomes. Animal technicians should be considered as key individuals in the studies. They are in charge of daily routines, and the way they deal with the animal can make a difference. Lifting rats from the base of the tail to move them to a new cage can make them aggressive and they do not habituate to handling. Anxiety of rats decreases when transferred to a new cage in cupped hands or transparent tunnel.

Laboratory animal husbandry issues are an integral part of the experimental design, which if ignored can cause major interference with the results. While defining the experimental design for an experiment, it is important to invest adequate time in the identification of such issues and to devise effective strategies to deal with them.

QUESTIONS

1.  True or False? In mice, temperature and humidity variation can affect the age of puberty.

2.  Which of the following husbandry variables can impact the study?

a.   Husbandry-related cycles

b.  Reproductive cycles

c.  Weekend-Working days cycles

d.  Caging

e.  All of the above

3.  True or False? Cage complexity becomes environmental enrichment once it has been verified to confer a welfare outcome on the animals

ANSWERS

1.   True

2.  e

3.  True

**Festing. Evidence Should Trump Intuition by Preferring Inbred Strains to Outbred Stocks in Preclinical Research, pp. 399-404**

Domain 3 - Research; T3- Design and conduct research; K3- Animal models

SUMMARY: The majority of scientists performing research in mice use genetically defined (inbred) strains, commonly C57BL/6 and BALB/c, and focus on fundamental research. Mice are desirable for such studies because they are small and economical. In contrast, the majority of scientist performing research in rats use genetically undefined (outbred) strains, such as Wistar and Sprague Dawley, and focus on preclinical research. However, no scientific case for using outbred stocks in research has been made, and the question arises as to whether the quality of research involving outbred animals could be improved, and total numbers reduced, if inbred strains were used instead. Inbred animals. Advantages of inbred stains include the fact that they are homozygous and isogenic, have less phenotypic variability, they are for the most part genetically stable, and strains have unique biological characteristics that make them favorable models for studying specific conditions. Outbred strains are heterozygous, which some believe is more relevant to human populations, but there can be a significant amount of variability in heterozygosity depending on the history of the strain in a particular location. Some outbred strains may have so little heterozygosity as to be virtually inbred.  Outbred strains that are reproductively separate will differ genetically, even if they have the same strain designation, i.e., CD-1. Also, outbred stocks are more genetically labile and susceptible to genetic drift. Another consideration is that due to their heterozygosity, outbred stocks are not as phenotyically uniform as inbred stocks, which may necessitate the use of significantly more animals in experiments to see differences in response to an experimental manipulation and thus achieve statistically significant differences between groups. The use of inbred strains is not without its challenges, however, as environmental factors may contribute to undesirable effects within the strains. For example, the NTP carcinogenesis bioassay program which historically used F344/N inbred rats switched to using outbred Sprague-Dawley rats due to decreasing fecundity of the F344/N strain, as well as increased incidence of sporadic seizures and idiopathic chylothorax. The author suggests that the determination of whether an outbred strain or an inbred strain is most appropriate for a particular study is not easy, and the best way to decide is to perform outbred and inbred strains within a particular experiment. This is not routinely done, and the author asserts that more funding for such studies should be made available as it could greatly improve animal research if the benefits of inbred and outbred strains could be more widely demonstrated.

QUESTIONS

1. What are some key characteristics of inbred strains that differ from outbred strains?

a. They are homozygous

b. They are isogenic

c. They have less phenotypic variability

d. All of the above

2.  T/F. All outbred strains have the same level of heterozygosity

3.  T/F. Statistically significant differences between experimental groups may require significantly more animals when using outbred strains

ANSWERS

1. d

2. False

3. True

**O’Connor and Sargeant. Critical Appraisal of Studies Using Laboratory Animal Models, pp. 405-417**

Domain 3: Research; T3.9 Design and Conduct Research, Principals of experimental design and statistics including scientific method.

SUMMARY: This paper discussed using critical appraisal to assess laboratory animal studies. Critical appraisal is the systematic process used to identify the strengths and weaknesses of a research article in order to assess the usefulness and validity of its findings. The units of critical appraisal are the results from a study and the product of critical appraisal is a conclusion about the validity of the study.  In other words, it is a tool to assess the extent a study scientifically contributes to the profession.

There are two components of critical appraisal: external validity and internal validity:

1.   External validity is the extent of which the results can be inferred to a target population (i.e. human population).

2.  Intrinsic validity is how the results of a study population represent the source population.

This paper specifically focuses on critically appraising the internal validity of a study by assessing its systematic errors or biases. There are five important biases associated with animal studies: selection bias, performance bias, detection bias, attrition bias and reporting bias.

1.   Selection bias is systematic difference in baseline characteristics between the treatment groups being compared (example, gender). A tool used to reduce the risk of selection bias is randomization of the enrolled animals in a study.

2.   Performance bias is systematic differences between groups in the care that is provided (example, housing). A tool used to reduce the risk of performance bias is blinding the caregiver.

3.  Detection bias is systematic differences between treatment groups in which the outcome is assessed (i.e. knowing the treatment groups). A tool used to reduce the risk of detection bias is blinding the outcome assessor.

4.  Attrition bias is systematic differences in withdrawals from treatment groups. A tools used to reduce the risk of attrition bias are minimization of loss to follow-up.

5.  Reporting bias is systematic difference in variables reported (i.e. incomplete reporting). A tool used to reduce the risk of reporting bias is comprehensive reporting.

When a reviewer is critically appraising the internal validity of the study, he or she should determine whether the risk of each bias is high, low or unclear. Besides assessing the biases of the study, critically appraising the internal validity of study should assess random error.  After considering the biases and random errors of a study, the critical appraiser can then reach a conclusion of its internal validity.

QUESTIONS

1. What are the 2 components of critical appraisal?

2. True or False: External validity is how the results of a study population represent the source population.

3. What are five biases commonly associated with animal studies?

4. What is a tool used to reduce the risk of selection bias?

ANSWERS

1. External validity and internal validity

2. False. Internal validity is how the results of the study population represent the source population

3. Selection bias, performance bias, detection bias, attrition bias and reporting bias

4. Randomization

**Hooijmans et al. Meta-Analyses of Animal Studies: An Introduction of a Valuable Instrument to Further Improve Healthcare, pp. 418-426**

Domain 3: Research; T3.9 Design and Conduct Research, Principals of experimental design and statistics including scientific method.

SUMMARY: The authors of this paper assert that meta-analysis of animal studies is a valuable tool with myriad advantages such as: reduction in redundant animal studies, reduction in bias, more reliable conclusions, and greater potential for exploring possible sources of heterogeneity (vs. clinical trials). In addition, the authors highlight the main principles and practices of performing meta-analyses of animal studies. These are best summarized by the Figures provided in the paper.

1.  Systematic Review: all research evidence relevant to a specific question is identified, appraised, and synthesized in order to draw evidence-based conclusions.

a.  8 steps (Fig. 1)

2.  Meta-Analysis: valuable step in performing a SR

a.   10 Steps (Fig. 1)

b.  Results displayed in Forest Plot (Fig. 2)

3. Reasons to conduct MA

a.   Results of SR that contain MA are more robust vs. without MA

b. Knowledge of efficacy, adverse effects, etc. more comprehensive

c.  More reliable conclusions

d.  Increases power informs clinical trial design and improves patient safety

4. Differences between MA of human and animal studies

a.  Goal of MA in human trials: estimate overall effect size of an intervention in order to inform clinical decision making.

b.  Goal of MA in animal studies: more exploratory; results used to generate new hypotheses and guide design of clinical trials

c.  Animal studies more diverse (species, design, characteristics, etc.)

d.  Methods usually similar, however animal study may contain placebo and sham groups, generally much smaller and more heterologous than clinical trials

f.  Methodological quality usually poor; increases risk of bias

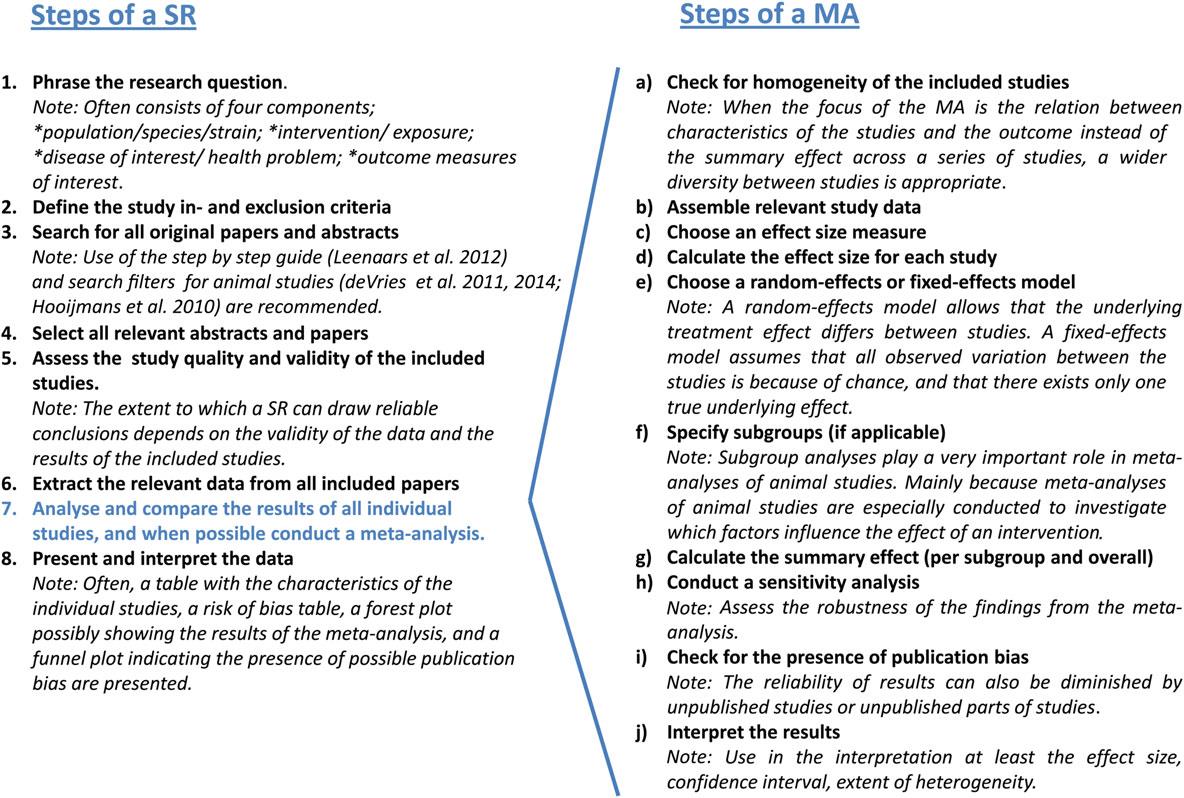


Figure 1: Steps of a Systematic Review & Steps of Meta-Analysis

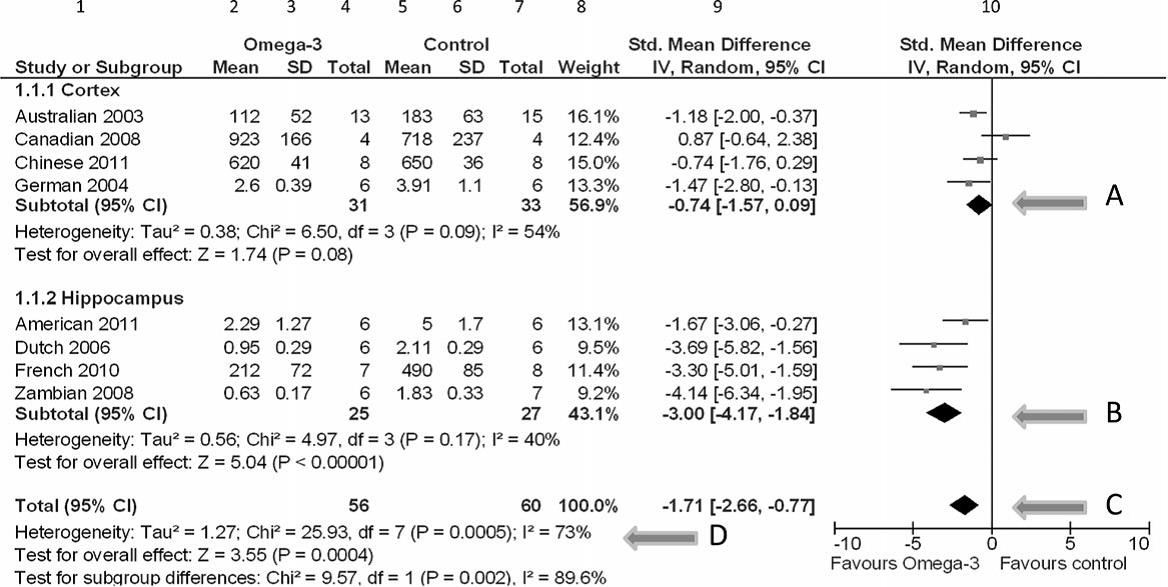


Figure 2: Example of a Forest Plot

QUESTIONS

1. What is one advantage of performing MA of animal studies in a SR?

2. True or False: The steps of performing a meta-analysis of animal studies are generally similar to the steps of performing meta-analysis of human clinical trials.

3. True or False: The quality of a meta-analysis is not dependent upon the quality of the primary studies.

4. True or False: Systematic Review of animal studies may increase the methodological quality of animal studies.

ANSWERS

1. Decreases bias, reduces animal numbers, more comprehensive knowledge, more robust, etc.
2. True

3. False

4. True

**De Vries et al. The Usefulness of Systematic Reviews of Animal Experiments for the Design of Preclinical and Clinical Studies, pp. 427-437**

Domain/Task:D3/T3

SUMMARY:The aim of this article is the systemic reviews (SR) of previously performed animal experiments and optimization of the design, conduct and analysis of the future experiments. If the animal experimentations are not properly designed, executed and analyzed, the results are likely unreliable and animals are wasted. SR are common practice in clinical research, especially for randomized controlled trials but are scarce in animal research.

The steps of SR includes:

1. Formulating a focused research question
2. Preparing a protocol
3. Defining inclusion and exclusion criteria
4. Systematically searching for original papers in at least two databases
5. Selecting the relevant papers
6. Assessing the quality/validity of the included studies
7. Extracting data
8. Data synthesis (if feasible, meta-analysis)
9. Interpreting the results

The SR characteristics are:

1. Specified and specific research question
2. Literature sources includes (more than one database)
3. Quality assessment includes critical appraisal on the basis of explicit quality criteria
4. Synthesis includes a summary quantitative (meta-analysis)

A proper experimental design may include, but are not limited to:

1. The choice of experimental and control groups
2. The determination of sample size
3. The choice of animal or disease model
4. The measures taken to reduce the introduction of bias

SRs of animal studies can contribute to:

1. Improving the methodological quality of experiments

2. More evidence-based choice of animal model

3. Evidence-based translation of animal data to the clinic

4. Implementing the 3Rs

I. Improving The Methodological Quality Of Experiments:

Methodological Quality: Refers to risk of bias, imprecision and lack of power.

Bias types are:

1. Selection Bias: the baseline characteristics of the experimental groups.
2. Performance Bias: the care for the animals or the administration of the intervention
3. Detection Bias: The way the outcomes are assessed
4. Attrition Bias: The way dropouts are handled

The threat to internal validity of an experiment can be reduced by randomization and blinding at three levels:

1. The allocation of animals to experimental groups
2. The administration of care and interventions during the experiment
3. The assessment of outcome

II. More Evidence-Based Choice Of Animal Model

Animal models include not only strain, species of laboratory animals but also to the way in which a disease or defect is induced. The selection of animal models sometimes are not evidence-based such as cost of buying, housing, ease of handling, availability of the tests, familiarity with a particular species. SRs support an evidence-based choice of animal models by providing a comprehensive overview of the models used so far, including their respective advantages and disadvantages.

III. Evidence-Based Translation Of Animal Data To The Clinic

It is becoming increasingly clear, however, that it is not straightforward to translate results found in laboratory animals to patients in clinical trials. This is partly due to fundamental biological/physiological differences between humans and other species. However, other avoidable factors related to:

1. Poor methodological quality
2. Differences in design between experimental animal studies and clinical trials
3. Publication bias

IV. Implementing the 3Rs

Replacement: SR may prevent unnecessary duplication, show that the study does not add substantially to current knowledge or indirectly lead to replacement by development or validation of alternatives. Evidence-based Toxicology Collaboration is pioneering this application of SRs in an assessment of the performance of the Zebrafish Embryo Test (ZET) in predicting the results of prenatal developmental toxicity.

Reduction: SRs contribute to reduction by making further use of the data already available and may improve the design of new experiments, leading to more reliable information from the same number of animals. In some instances SR may increase number of underpowered studies. However, such adequately powered experiments would produce more reliable data and results in fewer experiments and prevent waste of animals. SRs can reduce the number of animals by preventing the conduct of experiments not necessary to establish a certain effect of an intervention. Cumulative meta-analysis has been used for this purpose. In a cumulative meta-analysis, studies are sequentially included in the meta-analysis, and the point at which sufficient data exist to show stability of a treatment effect can be observed.

Refinement: SR can results in using tests with a lower burden of pain/distress, avoiding multiple tests, and shortening the duration of the test.

Progress in Implementing Systematic Reviews:

Two major research groups involved in the promotion of SRs of animal studies are:

1. The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES)
2. Systematic Review Centre for Laboratory animal Experimentation (SYRCLE)

QUESTIONS:

1. Which of the followings two statements are INCORRECT about SR? (2 answers)
   1. Specified and specific research question
   2. Selection of one independent reviewer
   3. Syntheses of a qualitative summary
   4. Literature search includes more than one database and explicit search strategy
2. Proper Experimental design choices may include which of the following statements?
   1. Experimental and control groups
   2. Sample size
   3. Measures taken to reduce bias
   4. All of the above
3. What is the benefit of SR?
   1. Improving the methodological quality of experiments
   2. More evidence-based choice of animal model
   3. Evidence-based translation of animal data to the clinic
   4. Implementing the 3Rs
   5. All of the above
4. Which of the following statement describes performance bias?
   1. The baseline characteristics of the experimental groups.
   2. The care for the animals or the administration of the intervention
   3. The way the outcomes are assessed
   4. The way dropouts are handled
5. All the following statements are example of non-evidence-based selection of animal models EXCEPT:
   1. Lower per diem cost
   2. Ease of handling of animals
   3. Comprehensive overview of all available models
   4. Available test for that species
   5. Technical experience with that animal model
6. Which of the following statements could be the reason for failure of translation research?
   1. Poor methodological quality
   2. Differences in design between experimental animal studies and clinical trials
   3. Publication bias
   4. Fundamental biological/physiological differences between humans and other species
   5. All of the above
7. Which study uses the Zebrafish Embryo Test (ZET)?
   1. Pregnancy test
   2. Pediatric leukemia
   3. Prenatal developmental toxicity
   4. None of the above
8. What kind of statistical test is commonly used in the SR data analysis?
   1. Cumulative meta-analysis
   2. Student t test
   3. Simple regression
   4. Chi-square
9. T/F. The SR data always leads directly in reduction of number of animal used in research.
10. T/F. The SR are common practice in clinical research, especially for randomized controlled trials but are scarce in animal research.

ANSWERS:

1. b and c, 2 reviewers, quantitative
2. d
3. e
4. b
5. c
6. e
7. c
8. a
9. F
10. T

**Garner. The Significance of Meaning” Why Do Over 90% of Behavioral Neuroscience Results Fail to Translate to Humans, and What Can We Do to Fix It?, pp. 438-456**

|  |  |  |  |
| --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | |

SUMMARY: The vast majority of drugs entering human trials fail. This problem (called “attrition”) is widely recognized as a public health crisis, and has been discussed openly for the last two decades.

Recent reviews of large-scale Food and Drug Administration, international industry, and public data indicate that clinical attrition rates range from approximately 80% to 97%, depending on the therapeutic area, with an average of approximately 90%.

Lack of efficacy identify as the single largest reason for why a drug fails in human trials.

Multiple recent reviews argue that animals may be just too different physiologically, anatomically, and psychologically from humans to be able to predict human outcomes, essentially questioning the justification of basic biomedical research in animals.

When models fail to predict human outcomes, perhaps we are not really modeling the human disease, but just modeling the model. The secret of using animal models is to understand their limitations.

This review argues that the philosophy and practice of experimental design and analysis is so different in basic animal work and human clinical trials that an animal experiment (as currently conducted) cannot reasonably predict the outcome of a human trial. This due to the underlying axiomatic differences in the philosophy and methodology of animal versus human research. In other words, if we want animal results to predict human outcomes, then we need to measure the same clinical phenomena in animals that we do in humans.

Attrition does reflect a lack of predictive validity of animal experiments, but it would be a tragic mistake to conclude that animal models cannot show predictive validity.

A variety of contributing factors to poor validity are reviewed: Face v Construct v Predictive, Internal v External, Convergent v Discriminant.

The need to adopt methods and models that are highly specific (i.e., which can identify true negative results) in order to complement the current preponderance of highly sensitive methods (which are prone to false positive results) is emphasized.

Concepts in biomarker-based medicine are offered as a potential solution, and changes in the use of animal models required to embrace a translational biomarker-based approach are outlined.

In essence, this review advocates a shift, where it treat every aspect of an animal experiment that it can as if it was a clinical trial in a human population.

 “Validation With Known Failures” Is Proposed:  New methods or models can be compared against existing ones using a drug that has translated (a known positive) and one that has failed (a known negative). Current methods should incorrectly identify both as effective, but a more specific method should identify the negative compound correctly.

By using a library of known failures the author suggest that he can empirically test the impact of suggested solutions such as enrichment, controlled heterogenization, biomarker-based models, or reverse-translated measures.

QUESTIONS

1. What is the attrition of Alzheimer drug development?

a.  87%

b.  55%

c.   76%

d.   96%

2. What field has the lowest attrition and what has the highest one?

3. What normal behavior phenomena considered as a marker to OCD?

ANSWERS

1. d. The attrition rate for Alzheimer’s drugs from 2002 to 2012 was 96.4%

2. Cardiovascular drugs consistently have the lowest attrition, while cancer drugs consistently have one of the worst success rates (e.g., 20% versus 5%)

3. At the same time, barbering is normal dominance behavior, but also widely used as an “OCD-like behavior”.

**Fry. Teaching Experimental Design, pp. 457-471**

Domain 6

SUMMARY: This article begins by reviewing issues that have been raised as early as 1957 regarding experimental design and the ability to draw informed conclusions based on the collected data. Various review articles are cited demonstrating that many published papers draw incorrect conclusions based on poor study design and/or inappropriate statistics. These citations are used to show the importance of teaching sound experimental design to both new and experienced researchers, and is the focus of this paper. The author’s discussion is divided into several sections including General Considerations, Structuring a Course or Module, and Case Studies.

The General Considerations section is subdivided into Basic Teaching Principles and Particular Considerations for Experimental Design. Each subsection begins with either a question or statement that includes a paragraph answering the question or commenting on the statement.

Basic Teaching Principles

* Who is it for?
* What Will Be the Scope?
* What Does it Hope to Achieve?
* How Will It Be Delivered and How Will It Engage the Audience?
* How Will It Be Assessed and for What Purpose?

Particular Considerations for Experimental Design

* Educating in Experimental Design Has a Different Starting Point from Teaching Statistics
* Particular Deficiencies Need to Be Addressed and a Range of Backgrounds Accommodated
* Technical Terms Used Pose Difficulties
* Attitude Change May Be an Important Consideration in Teaching about Studies Using Animals

The author then moves on to the next section, Structuring a Course or Module. This section is divided into Content, Delivery, and Assessment. Each section contains bolded points under each subsection that are discussed:

Content

* General Topics
* Scientific method
* Deciding upon an experimental question
* Different kinds of experiments
* Ethical considerations and the 3Rs: replacement, reduction, and refinement
* Why design?
* Definitions
* Basic Principles of Design
* Suitable comparisons and controls
* Replication
* Randomization and avoidance of bias
* The experimental unit
* Design and Analysis
* The importance of knowing how a design can be analyzed
* Severity considerations and planning for humane stopping-points
* Signal/noise ratio
* Sources of variability and reducing variability
* Blocking
* Expressing variability quantitatively
* Risk of false positives and false negatives: Type 1 and Type 2 errors
* Determining the numbers needed
* T-tests and analysis of variance
* Types of experimental design and choice of design
* Nested ANOVA
* Internal and external validity
* Types of data, nonparametric tests and contingency tables
* Assumptions behind parametric statistical tests
* Correlation and regression
* Additional Topics
* Planning a sequence of experiments
* Statistics programs/packages
* Presentation of data
* Comparison of Bayesian and frequentist approaches
* Further Reading
* Delivery
* Variety
* An Educational 3Rs: Repetition, Recapitulation, and Reinforcement
* Conveying Attitudes
* Online Delivery
* Assessment

The author concludes with several representative case studies, including three-day courses, half-day courses, and continuous professional development for animal technicians (i.e. CE). Topics discussed under three-day courses include Topics and Time Allocations, Delivery Methods, Assessments, Effectiveness, Participant Comments and Evaluation, and The Course as a Postgraduate Module. Half-day course topics include Topics, Delivery, Assessment, Evaluation, and Adaptation to a Two-Hour Time-Slot.

QUESTIONS

1.  What term refers to accepting experimental entities as independent when they are in fact not?

a. Bias

b. Parametric

c. Pseudoreplication

d. Nonparametric

2. The intention of which one of the following experiment types listed below is to provide preliminary information/data that can improve the conduct, quality, and efficiency of subsequent hypothesis-testing experiments?

a. Exploratory experiment

b. Analysis of variance (ANOVA)

c. Pilot experiment

d. Historical experiment

3.  Which one of the follow conditions is NOT considered unethical regarding experimental design?

a. Too many animals are utilized

b. Prolonged pain without analgesics

c. Animals are subjected to unnecessary severity

d. The design fails to acquire worthwhile data

4.  Which answer best describes the current state of experimental design regarding animal studies?

a. Poor experimental mental design is a new phenomenon, starting in the late 1980s

b. Poor experimental design may be widespread

c. Poor experimental design is rare in today’s animal studies

d.  >98% of published animal studies have acceptable experimental design

ANSWERS

1. c

2. c

3. b

4. b

**Festing. Randomized Block Experimental Designs Can Increase the Power and Reproducibility of Laboratory Animal Experiments, pp. 472-476**

SUMMARY: An excessive number of animal experiments are unreproducible even when they are well designed, executed, analyzed and documented. In many cases, is due to a lack of randomization, blinding, possible interactions treatment x environment, the usage of the wrong strain of animals or due to the 5% chance of getting a false positive result due to statistical sampling.

The randomized block design is when the experimental material is split up in a number of “mini-experiments”. Useful properties that these have include the possibility to spread the experiment over a period of time and/or space so it is more manageable or the increase of power of an experiment by matching the experimental units in each block. At the same time, it increases the external validity because each block samples a different environment or time period.

In order to determine the sample size, two methods can be used: power analysis using an estimate of the standard deviation based on an unblocked experiment or using the resource equation method being E= (total number of experimental units)-(number of experiments) and E should be about 10 and 20.

In the statistical analysis, each observation can be classified as “fixed” (treatment) and “random” (the block that needs to be removed in order to not cause noise in the statistical analysis). An individual observation will be made up of a grand mean, a deviation due to the treatment it receives, a deviation due to the block, and an individual deviation. Assuming a single treatment factor and a single block factor, the experiment is analyzed by a two-way analysis of variance without interaction. However, a factorial treatment structure can also be used. These designs there are not two observations that come from the same block and treatment so the estimate of the standard deviation is obtained from the square root of the error mean square in the analysis of variance.

An ANOVA makes three assumptions about the data: the experimental units are independent; the deviations from the mean have a normal distribution; and the variances are homogeneous. The ANOVA is quite robust to deviations from these assumptions. In a serious transformation from normality a log transformation of the raw data provides a slightly better fit.

If blocks separated in time have a good agreement this gives some assurance that the experiment is reproducible. Their usage would save money, animals, and resources.

QUESTIONS

1. A randomized block design is a family of experimental designs divided in blocks where in each block there is a single experimental unit to which each treatment is assigned.
2. The randomized block allows the removal of variables that are not of interest so if the relative values between blocks are unchanged, the treatment effect is more likely to be detected.
3. To obtain an increase of power of an experiment, the experimental unit has to be homogeneous.
4. The resource equation method is useful for more complex designs with many treatments that do not have an estimate of the standard deviation.
5. A well-designed experiment will always give the same results when repeated.
6. A log transformation of the raw data always provides a more robust ANOVA in normal distributed deviations.

ANSWERS

1. False
2. True
3. False
4. True
5. False
6. False

**Parker and Browne. The Place of Experimental Design and Statistics in the 3Rs, pp. 477-485**

Domain 3; T3 TT3.10

SUMMARY: The 3Rs –replacement, reduction, and refinement- are considered an ethical algorithm for animal experimentation.  Replacement can be either absolute where a total protocol goes from using animals to not using animals, or relative where one portion of animal use is substituted.  Alternative methods need validation to assess the relevance and reliability.  Reduction works to decrease the number of animals used without decreasing the quality of information gained from the study.  Calculation of appropriate sample sizes is the biggest contributor to reduction.  The sample size in calculated keeping in mind significance level or the probability of making a type I error (false positive), as well as the power or probability of rejecting the null hypothesis when it is false (the converse of this is a type II error).  To increase power, researchers can increase the sample size, decrease variation, or investigate alternative measures.  The sample size is also calculated keeping in mind the effect size or anticipated difference between group means and variability of two samples. The calculation is dependent on type of variables, structure of dependency, probability of missing data and inferences of key interest.  Minimizing bias can be done by randomization, blinding, and controlling confounding variables.  External validity is the extent to which results can be applied to a population.  Accurate reporting of protocols, methods, and results decreases unnecessary reproduction.  Reuse is also a consideration for reduction of overall animal numbers, but must be weighed against refinement. Refinement refers to minimizing or avoiding pain, distress or other adverse effects.  Reducing stress can be facilitated by training animals to comply with collecting physiological data.  Decreasing stress can decrease bias and variability in experiments.  Successful application of the 3Rs can be to the benefit of research.

QUESTIONS

1. What are the 3Rs of research?
   1. Replacement, reuse, and refinement
   2. Replacement, reduction, and redefining
   3. Reduce, reuse, and recycle
   4. Replacement, reduction, and refinement
2. What is a type I error?
   1. False negative
   2. False positive
   3. External validity
   4. Confounding variables
3. What is external validity?

ANSWERS

1. d. Replacement, reduction, and refinement
2. b. False positive
3. External validity is the extent to which results can be applied to a population

**Pearl. Making the Most of Clustered Data in Laboratory Animal Research Using Multi-Level Models, pp. 486-492**

Domain 3: Research

SUMMARY: When studying animals housed in group settings, one must account for data clustering or autocorrelation. Are outcomes of individual animals in a group setting truly independent? Multi-leveled models (also known as hierarchical or mixed models) are particularly useful in such situations because they allow analysis of variables on multiple levels and can account for any autocorrelation caused my group membership, repeated measures in time, or proximity in space. For example, a multi-level model would be appropriate for studying paternal diet on metabolic function and cardiovascular function in offspring. Using a 3-level hierarchical model consisting of the males themselves, different litters sired by the males, and pups within those litters, analysis of variance components could provide information on how much of the variation in a particular outcome is explained at the offspring, litter, and sire levels. Other options for analyzing clustered data include performing statistical analysis on a group-level statistic (e.g., mean group weights) and including an independent categorical variable in a regression model to account for group (e.g., fixed effects approach). However, these alternatives are limited by loss of statistical power, inability to account for confounding by group, limitation to 2-level modeling situations, and inability to examine distribution of variance in the outcome among hierarchical levels.

QUESTIONS

The paternal dietary habits example provides data for the impact of sire diet on the weight of offspring after controlling for effects of litter size and sex of offspring.

1. What is a variance partition coefficient (VPC)?

2. What is an intra-class correlation coefficient (ICC)?

3. What are some common statistical programs that support multi-level modeling?

ANSWERS

1. VPC can be calculated for each group by taking the variance (σ2) of the group divided by the variances of all of the groups added together. This explains how much of the variation in weight is explained at the level of each group.

a. VPCsire = σ2sire / σ2sire + σ2litter + σ2offspring

2. The ICC uses variance components to estimate correlation among individuals within a group. For example, to estimate the correlation among offspring within the same litter, you would perform the following calculation:

a. ICClitter = σ2sire + σ2litter / σ2sire + σ2litter + σ2offspring

3. STATA, SAS, R, SPSS, SYSTAT, OpenBUGS, MLwiN, HLM