The importance of good experimental design

Dr Derek Fry

(formerly Chief Inspector, Animals Scientific Procedures Inspectorate, UK)

EU Directive [First reading 2009 text]

- Project - "a programme of work having a defined scientific objective and involving one or more procedures"
- Evaluation to include
  - "evaluation of the objectives of the project, the predicted scientific benefits"
  - "assessment of the compliance of the project with the requirement of replacement, reduction and refinement"
  - "a harm-benefit analysis" of the project
- "Project authorisations shall be granted for a period not exceeding five years"

So evaluation

- Should cover not just individual experiments or protocols
- But also the whole experimental programme for up to 5 years.

..and design needs considering at two levels

- in the design of the programme
- in the design of the individual experiments
- and in both there is scope for Reduction and Refinement
- and need to consider possible replacements

Why consider the whole programme, not just individual experiments?

- As Russell and Burch (1959) pointed out "One general way in which great reduction may occur is by the right choice of strategies in the planning and performance of whole lines of research."
- Mapping out the whole programme can show where pilot experiments with small numbers or decision points in the programme can best be incorporated to save using animals, e.g. when worthwhile results are unlikely
- This can also highlight where non-animal experiments might be able to contribute to the aims
- The programme can also be planned to minimise severity

The need to evaluate projects may change the level of interest in planning

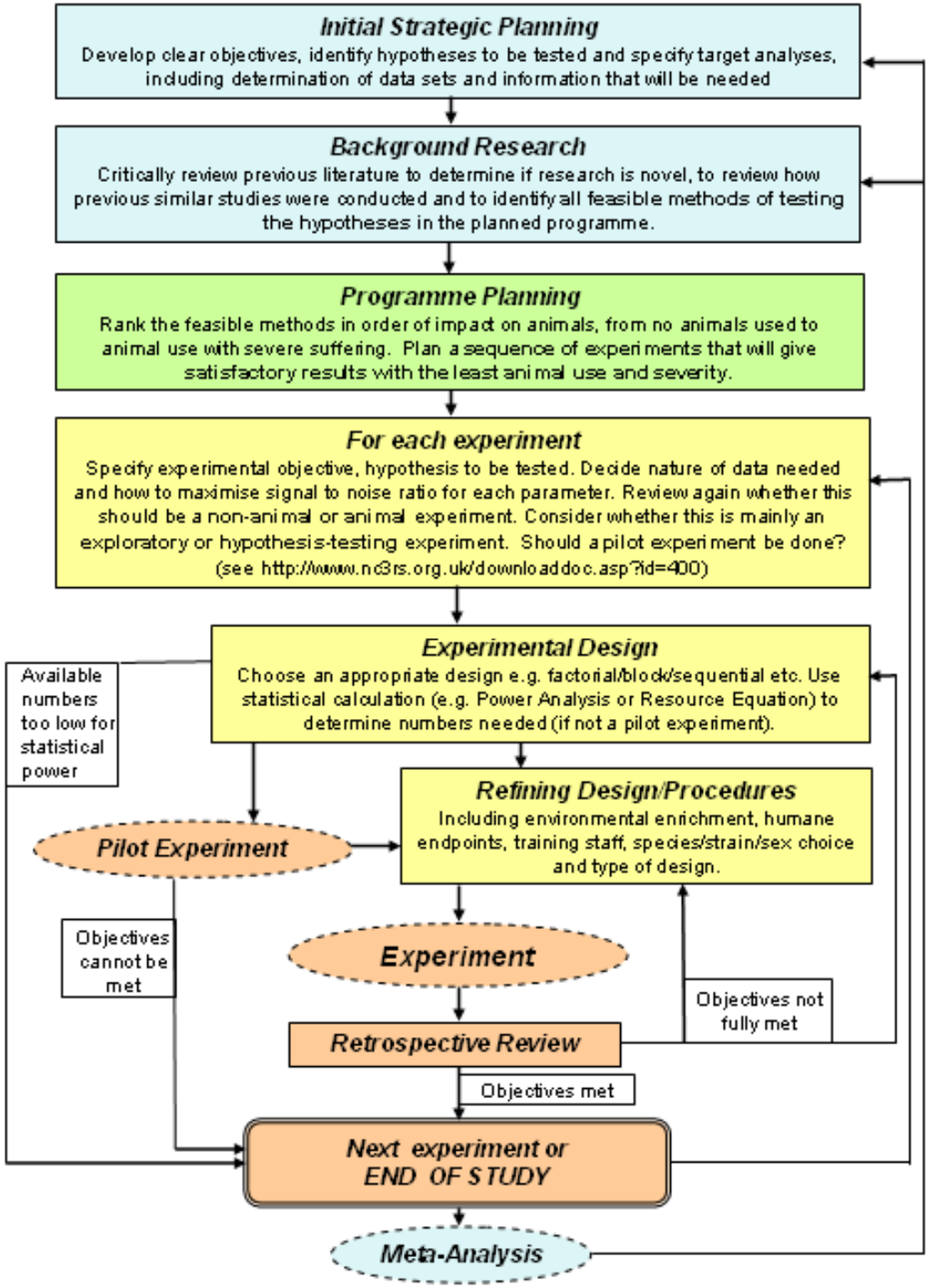
- There's not a lot of guidance but FRAME is developing some ...
- From discussions at UK meetings on design and from courses run for postgraduates from across Europe
- it has seen a need
- and produced a flow chart to help.

Flow chart

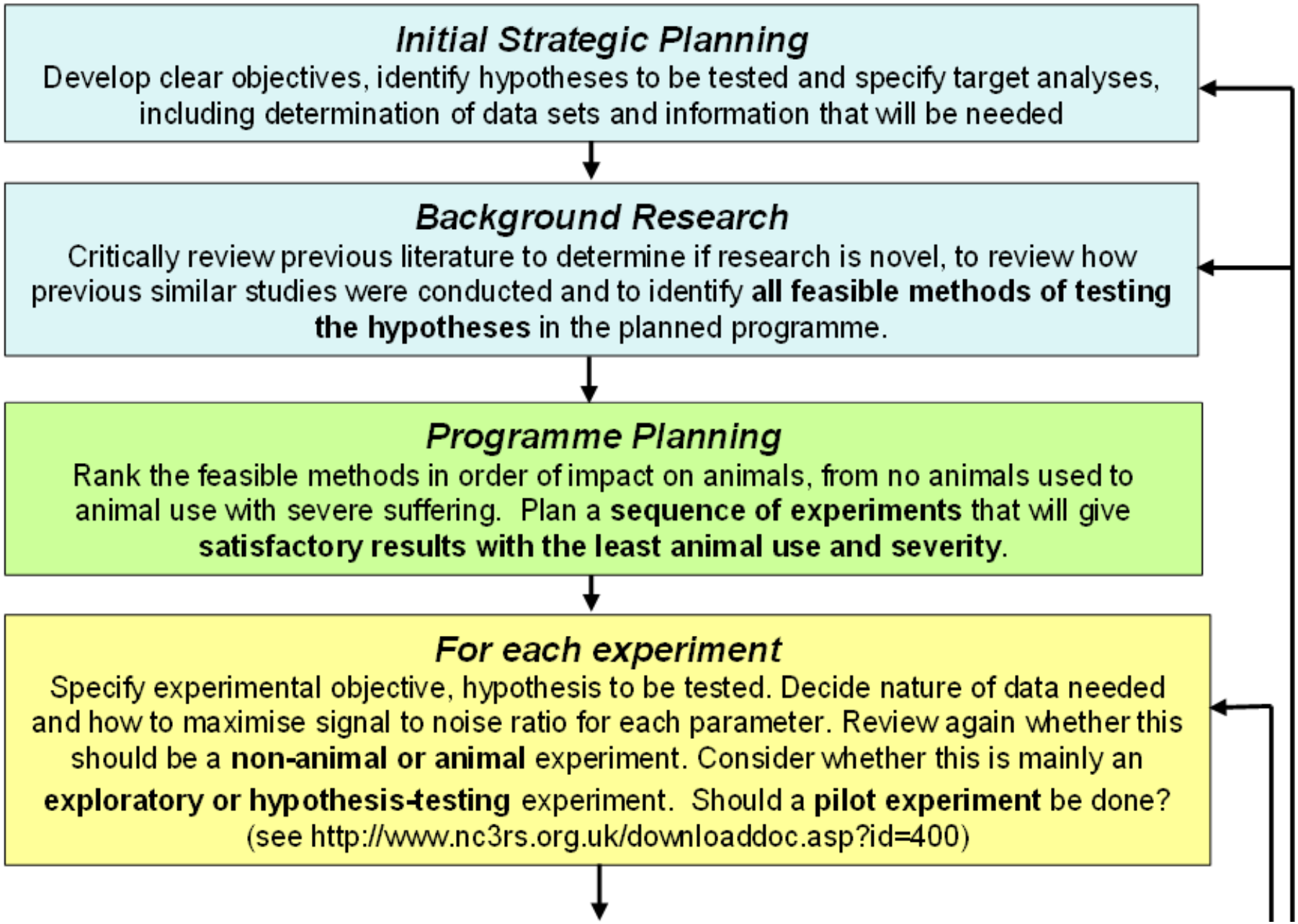
ATLA 37 (2009) has already published notes to this, concentrating on reduction

But it is also useful for refinement

And may be a helpful checklist for ethical evaluation



The planning stages



The planning stages

- ***Initial Strategic Planning***
  - Develop clear objectives,
  - identify hypotheses to be tested and
  - specify target analyses, including determination of data sets and information that will be needed

Failure to specify clear objectives is not uncommon

- In a detailed scrutiny of 271 published papers from a wide variety of journals Kilkenny et al. (2009) reported that 5% of the studies either did not describe the purpose of the study at all, or it was not clear to the assessors.

The planning stages

- ***Initial Strategic Planning***
  - Develop clear objectives,
  - identify hypotheses to be tested and
  - specify target analyses, including determination of data sets and information that will be needed
- ***Background Research***
  - Critically review previous literature regarding the need for the work
  - review how previous similar studies were conducted
  - Research the methods and techniques
  - identify all feasible methods of testing the hypotheses in the planned programme.
- ***Programme Planning***
  - Rank the feasible methods in order of impact on animals
  - Plan a sequence of experiments that will give satisfactory results with the least animal use and severity.

Sequence planning

- Can convert unknown unknowns to known knowns.
  - Plan to identify technical problems, unexpected adverse effects and suitable humane endpoints early
  - Include pilot experiments with few animals and good observation schedules
- Should include decisions/review points
  - E.g. with a new model
    - set criteria for acceptance and
    - a review point for whether criteria adequately met
    - decide whether, if not met, it is better to abandon the attempt to avoid further suffering, or to modify approach
- Should proceed from low to higher severity

Planning to minimise severity – a sequence of questions

- E.g.: Objective – to determine the effects of hypoxia (low oxygen) on sympathetic nerve activity
- What can be done without animals?
  - No relevant studies – needs vascular nerve network
- What can be done under terminal anaesthesia?
  - All the studies on effects of hypoxia up to 24h
- What can be done with only mild severity?
  - Studies on effects of prolonged mild hypoxia – chronic exposure then terminal anaesthesia
- What can only be done at more than mild severity?
  - Prolonged more severe hypoxia work
  - Confirmation studies with implanted electrodes

D J Fry May2010

Planning Example: Acute pancreatitis programme

- Acute pancreatitis
  - is very painful
  - carries substantial morbidity
  - can be fatal
- There are significant deficiencies in current treatment.
- Justified use of animals for improved treatment – whole animal needed for the interactions of multiple body systems
- Pancreatitis can be induced in mice and within 6h blood parameters show its severity and pancreas damage can be detected by microscopy

Possible severity-sensitive programme

- Initial experiment(s) using early blood changes and microscopy as measures to determine optimal induction.
- Timed-kill experiment for precise time course of the changes
- Confirmation in a few animals of progression to full blown acute pancreatitis
- Pharmacological studies using optimal induction and sampling arrangements as determined from the above.

Apparent sequence in 2008 paper in PNAS

- Initial experiments to find suitable arrangements for induction but with death as measure.
- Survival experiment with 10 mice
- Repeated in at higher dose till 30% die
- Experiment to show multiple organ failure
- Experiments on time course of the blood and microscopy changes
- Pharmacological studies

It does not take particular expertise

- To question such a sequencing of experiments
- And readily understand if severity was not being adequately considered by the researcher

Common Failings

- Programme aims unclear
- Background information insufficient or in error
- Programme planning does not take account of constraints (e.g animal house, people availability, time an experiment takes)
- Pilot experiments and review points not included
- Opportunities to use more refined or non-animal approaches missed

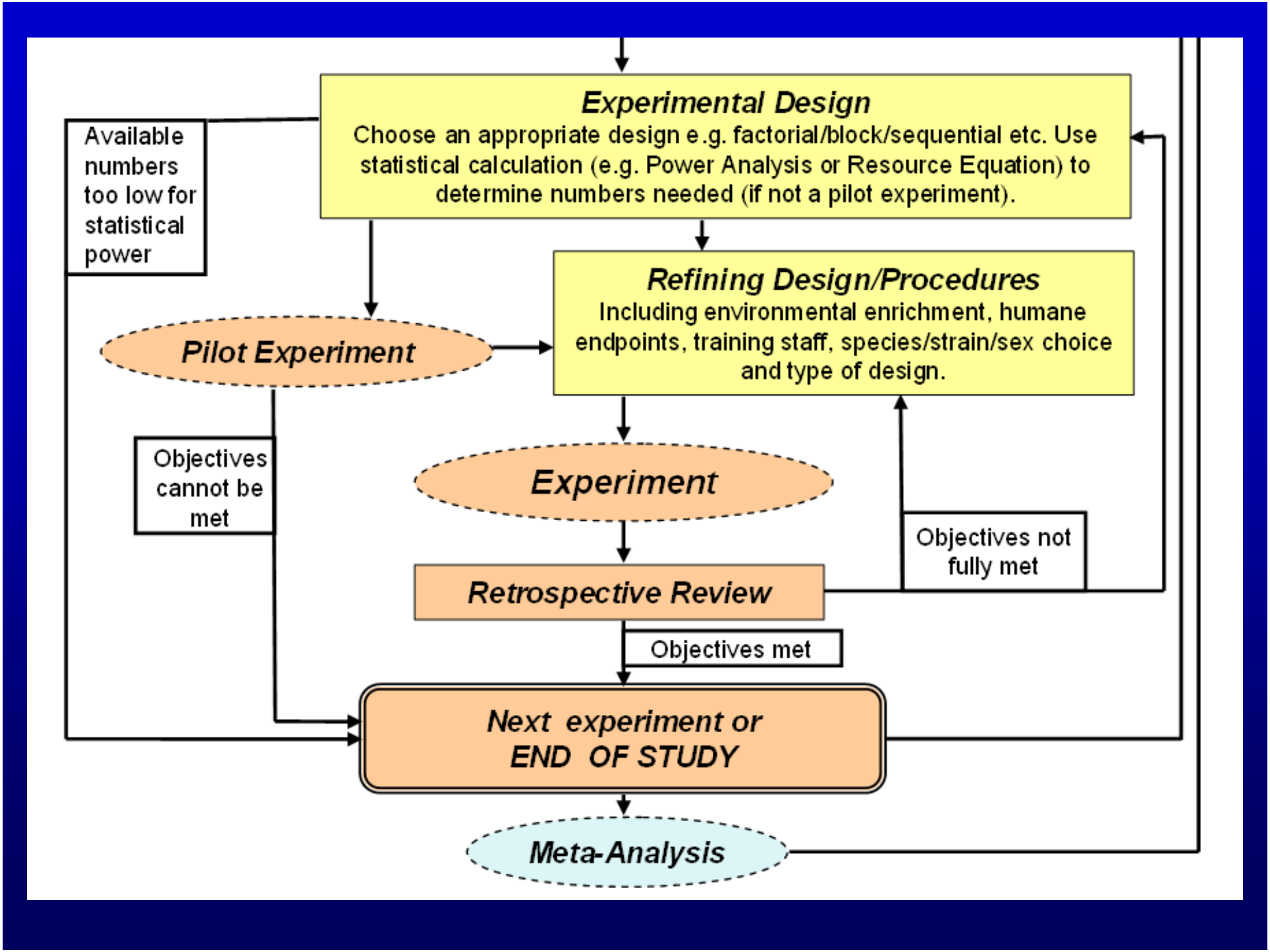
Pilot experiments

- Trials with small numbers
- Allow note of adverse effects, time to treatment effect, sources of variability, technical problems, etc.
- Guard against over optimism
- Adverse effects noted can help set severity endpoints
- Time to treatment effect can set objective-related endpoint

The Competent Authority can

- Include evaluation of the plan of work in the assessment of projects
- Expect evaluation to question whether non-animal possibilities have been explored and whether the severity of methods and techniques has been researched
- Include in the retrospective review, or assessment of an end-of-project report, consideration of the decision process during the work and its overall severity
- Encourage discussion of the topic, e.g. by meetings with case studies

And at the experiment level …


Experimental Design
Choose an appropriate design e.g. factorial/block/sequential etc. Use statistical calculation (e.g. Power Analysis or Resource Equation) to determine numbers needed (if not a pilot experiment).
Available numbers too low for statistical power
Refining Design/Procedures
Including environmental enrichment, humane endpoints, training staff, species/strain/sex choice and type of design.
Pilot Experiment
Objectives cannot be met
Experiment
Objectives not fully met
Retrospective Review
Objectives met
Next experiment or END OF STUDY
Meta-Analysis

Marshall Hall's Principles - as given in The Lancet (1847)

- "We should never have recourse to experiment in cases which observation can afford us the information required;
- No experiment should be performed without a distinct and definite object and without the persuasion that the object will be attained and produce a real and uncomplicated result;
- We should not needlessly repeat experiments and cause the least possible suffering, using the lowest order of animals and avoiding the infliction of pain;
- We should try to secure due observation so as to avoid the necessity for repetition."

Hall, M. (1847). On experiments in physiology as a question of medical ethics. *The Lancet*, **1**, 58-60.

Lack of clear objectives is not uncommon

- In FRAME's recent experimental design course each group of participants considered in detail the main experiment of a different published paper. The hypothesis being tested was clear in only 1 in 5 of the papers.
- The survey by Kilkenny et al. (2009) indicated unclear objectives in 1 in 20 papers.

But without clear objective(s)

- You can't assess
  - whether the experiment should be exploratory or testing an hypothesis
  - whether use of animals is necessary
  - what would be suitable controls
  - when it should be stopped because it has achieved or cannot achieve the objective
- … or design efficiently

Why design?

- To obtain valid results from which safe conclusions can be drawn
- To know how widely these may apply
- To use resources efficiently
- To minimise severity
- To ensure reproducibility

Experiments that do not keep to fundamental principles

- Include an unknown amount of uncertainty and bias so
- .. produce unreliable outputs
- .. which risk leading to erroneous conclusions
- .. which may take many other experiments to correct

And..

- It may not need an **expert** in experimental design to question proposed designs

Experimental Design is not Statistics

It calls for a combination of

- biological insight
  - to formulate a good experimental question
- logic
  - to devise a testable hypothesis
- common sense
  - to know what is feasible
- planning
  - to set out how best to perform the experiment
- and an appreciation of statistics!
  - To understand how the results can be properly analysed

A good experiment

- Is unbiased
  - Has independent repeats (replicates)
  - Which are randomly assigned to the different fixed experimental conditions
- Is precise
  - Has uniform material and/or control of variability
  - OR is a large experiment
- Has a wide range of applicability
  - Includes many controlled variables (e.g. sex, strain)
  - Allows interaction between the variables to be assessed
- Is simple to analyse
  - Keeps to a formal design
  - Has equal numbers in the sub-groups
- Allows uncertainty to be calculated
  - i.e. has independent repeats

[Based on Cox but with generally-used (and inexact) words]

Jargon – "design" terms

- Replication – repetition of measurement or observation in a way that each repeat can be independent of the others
- Precision – the extent of random scatter: the smaller the variability the greater the precision.
- Accuracy – the closeness of fit to the real situation
- Bias – a distortion likely to affect successive measurements
- Treatment – the experimental conditions fixed to test the hypothesis – e.g. for studying the effect of a drug each dose level of drug would be a "treatment" and so would the vehicle control

Fundamental principles

- Replication
  - Needs to be sufficient ('Power' of experiment)
- Appropriate controls or comparisons
  - Concurrent (usually), relevant
- Random assignment to treatments
  - Needs to be at every stage

Correct application of each fundamental principle is essential for a valid and successful experiment

- People without expertise in experimental design can detect when the fundamental principles are being ignored

Fundamental principles

- Apply to the 'experimental unit'
  - 'Experimental unit' can be assigned at random to any treatment; different units must be capable of receiving different treatments
  - Important to correctly define unit: common error to assume unit is individual animal when all animals in cage treated identically

Failures in design are common

- Kilkenny et al. 2009 found that "the experimental unit (e.g. a single animal or a group of animals) was not clearly identified in 13% of the 48 studies assessed in more detail".

- Only two out of 34 scientists attending FRAME's most recent experimental design course could correctly identify the experimental unit at the start of the course.
- Figures for previous courses were 18/33 10/52 3/37

Inefficient design is not uncommon, e.g.

- Factorial experiments, using treatment groups with mixed sex or age or strain for example, can gain two or more times the information from the same number of animals as those using single comparisons.
- Kilkenny et al. (2009) found that "only 62% (75/121) of all the experiments assessed that were amenable to a factorial design (and analysis) reported using one."
- They commented "it seems that a large number of the studies assessed did not make the most efficient use of the available resources (including the animals), by using the most appropriate experimental design"

There is a long list of failings seen repeatedly

- Experimental question poorly formulated
- Hypothesis unclear, or low discrimination between hypotheses
- Experiment planning fails to recognise constraints
- Experiment unit incorrect
- Inefficient design used
- Sample size too big or too small
- Units or measurements not independent
- Randomisation not done (or not done at all stages)
- No blinding
- Incorrect analysis
- Incorrect reporting and presentation

And that’s not a complete list!

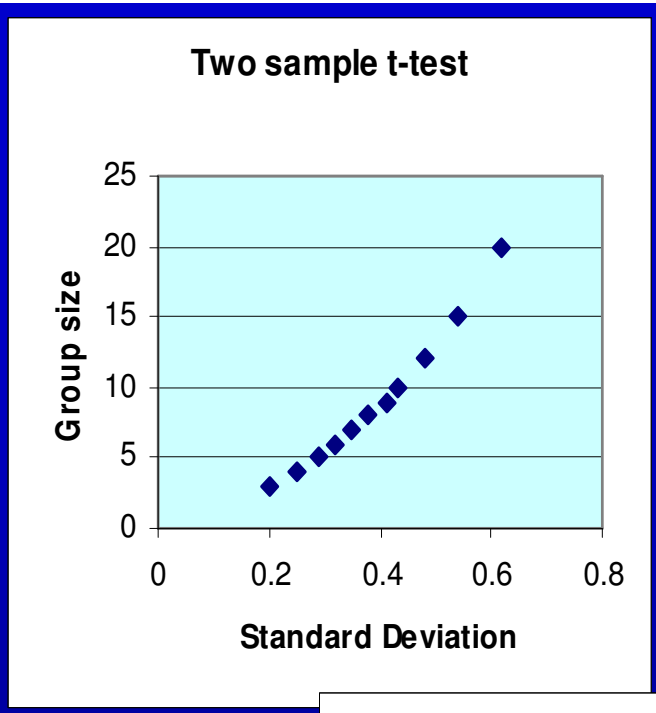
It doesn't include failure to refine

And it should …

- good science depends on good welfare
- physiological responses to suffering can reduce data quality and consistency and affect validity
- anxiety, pain, fear and distress can affect animal physiology
  - heart rate, blood pressure, body temperature, immune responses, blood biochemistry, brain complexity
- even minor welfare problems can influence the reliability, reproducibility and consistency of research

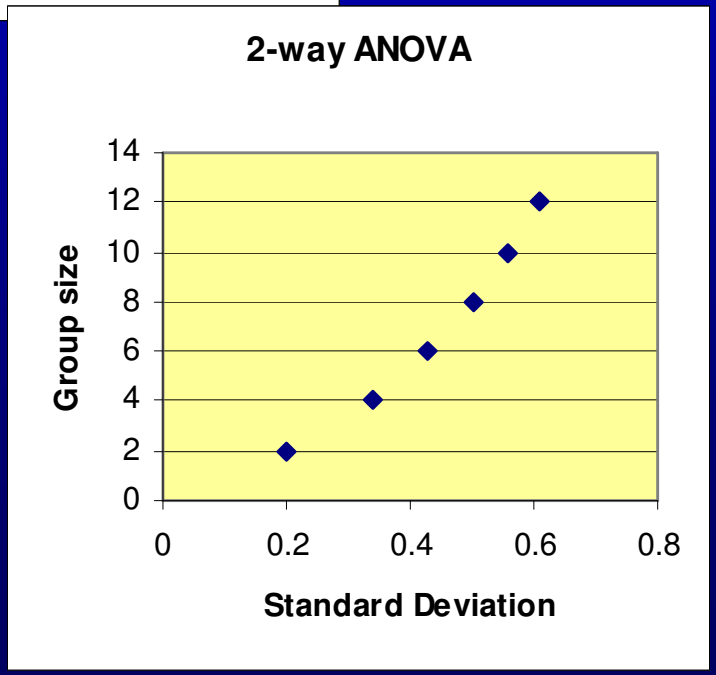
Variability has a large effect on numbers needed

Note – "standard deviation" is a measure of variability



- 2-sample t-test
  - Alpha .05 one-tail
  - Power 0.8
  - Control mean 1.0
  - Difference between sample means 0.5

- 2-way ANOVA
- As for a 2x2 factorial experiment



So failure in refinement can also be failure in reduction

D J Fry May2010

The Competent Authority can

- Ensure there is access to, in order of preference,
  - A biomedical scientist with good knowledge of statistics or
  - A statistician with good appreciation of biomedical experiments or clinical trials or
  - Any statistician!
- Expect applications to include the design of a typical experiment or the first main one proposed
- Have this design evaluated by a person or group able to judge experimental design
- Ensure retrospective review analyses a main experiment and has comment from a person or group able to judge experimental design

A Competent Authority can also

- Encourage wider appreciation of good design by promoting education in this area
- Fund scientists to attend specific experimental design courses
- Facilitate meetings in which scientists discuss their work with statisticians
- Set up a body which scientists can consult about design

Education

- Open Events
  - Raise awareness and improve knowledge,
  - May improve attitudes
- Seminars, lectures
  - Raise awareness and improve knowledge,
  - May improve attitudes (depending upon type of audience)
- Workshops
  - Raise awareness and improve knowledge,
  - May improve skills,
  - May improve attitudes (depending upon range of participants)
- Courses
  - Raise awareness and improve knowledge,
  - Improve skills,
  - Improve attitudes

Message: Poor design wastes resources – animals, time, money

- Persistent overestimating of animals needed by just 10% has been estimated at one UK university to cost the equivalent of one person's salary!
- Wasting money and time are bad, but wasting animals or increasing their suffering is worse
- Better not to do an experiment at all than to do it badly

And finally, to reduce suffering – and resource

- Every experiment should have points at which it will be stopped for humane reasons
- These "humane end-points" can be designed in.

Types of humane endpoints

- 1) severity cut-off - animal killed, or taken off experiment, to avoid further suffering
- 2) "objective-achieved" endpoint - sufficient data obtained for the purpose of the experiment to be met
- 3) an "objective-cannot-be-achieved" point - unlikely to be sufficient valid data if continued.

Useful References

- Festing, M.F.W., Baumans, V., Combes, R.D., Halder, M., Hendriksen, C.F.M., Howard, B.R., Lovell, D.P., Moore, G.J., Overend, P. & Wilson, M.S. (1998). Reducing the use of laboratory animals in biomedical research: problems and possible solutions. *ATLA* **26**, 283-301.
- Festing, M., Overend, P., Gaines Das, R., Cortina Borja, M. & Berdoy, M. (2002). *The Design of Animal Experiments: Reducing the Use of Animals in Research Through Better Experimental Design*, p39. London: Royal Society of Medicine Press Limited.
- Fry, D. J. (1998). Relating criteria for humane endpoints to objectives. In *Humane Endpoints in Animal Experiments for Biomedical Research* (ed. C. F. M. Hendriksen & D. Morton), pp. 54 - 57. London: Royal Society of Medicine Press Limited.
- Fry, D. (2004). Reduction by well-defined objectives. *ATLA* **32**(**Suppl. 1)**, 241–244.
- Fry, D. & Morton, D. (1999). Non-statistical experimental design as an aid to refinement. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (ed: M. Balls, A. M. van Zeller & M. Halder M), pp.1687 – 1690. Amsterdam: Elsevier.
- Gaines Das R, Fry D, Preziosi R and Hudson M (2009). Planning for reduction. ATLA 37, 27-32.
- Howard B, Hudson M and Preziosi R (2009). More is less: reducing animal use by raising awareness of the principles of efficient study design and analysis. ATLA 37, 33-42.
- Johnson PD & Besselsen DG (2002) Practical aspects of experimental design in animal research. *ILAR J.* 43(4), 223–232.
- Kilkenny C, Parsons N, Kadyszewski E, Festing MFW, Cuthill IC, Fry D, Hutton J, Altman DG. (2009) Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. PLoS ONE 4(11): e7824. doi:10.1371/journal.pone.0007824
- Morton DB (1998). The importance of non-statistical design in refining animal experimentation. ANZCCART Facts Sheet. ANZCCART News 1 l(June): 12. Osmond, Australia: ANZCCART.
- Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*, 238pp. London, UK: Methuen and Co.
- Shaw R, Festing MFW, Peers I, Furlong L. (2002) Use of factorial designs to optimize animal experiments and reduce animal use. *ILAR J.* 43(4), 223–232.
- US National Academy of Sciences (2008). *Recognition and Alleviation of Distress in Laboratory Animals,* pp63-86. The National Academies Press: Washington.