Considering the Three Rs in an Experimental Programme

Dr Derek Fry

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

Application of the 3Rs needs considering at two levels

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

So evaluation

- Should cover not just individual experiments or protocols
- But also the whole experimental programme.

Experimental strategy

- Fifty years ago Russell and Burch (1959) recognised that "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."
- They also recognised the ethical imperative to "reduce to an absolute minimum the amount of distress imposed ..".
- Reduction in animal usage reduces overall suffering by exposing fewer animals to adverse effects.
- But good programme planning can also minimise overall severity, and so help reduce distress to a minimum.

- 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 when worthwhile results are unlikely
- This can also highlight where non-animal experiments might be able to contribute to the aims

Guidance on programme planning?

- There is some e.g.
  - Strategy for reduction is covered to some extent in the handbook of ECVAM Workshop 29 - "Reducing the use of laboratory animals in biomedical research: problems and possible solutions" [Festing et al. (1998)]
  - A step-wise approach to refining an experimental programme was developed in a workshop at the 3rd Congress [Fry and Morton (1999)].
  - See also Johnson & Besselsen (2002), Morton (1998).
- But not a lot…
- Experimental design texts usually provide no guidance on how to design an individual experiment to minimise severity, and are silent on how to organise a sequence of experiments.

Is this an aspect missing from ethical assessments and researcher training?

- Ethical assessments may judge only protocols and miss a refinement strategy for the whole programme.
- Training and guidance on design may only cover the individual experiment e.g.

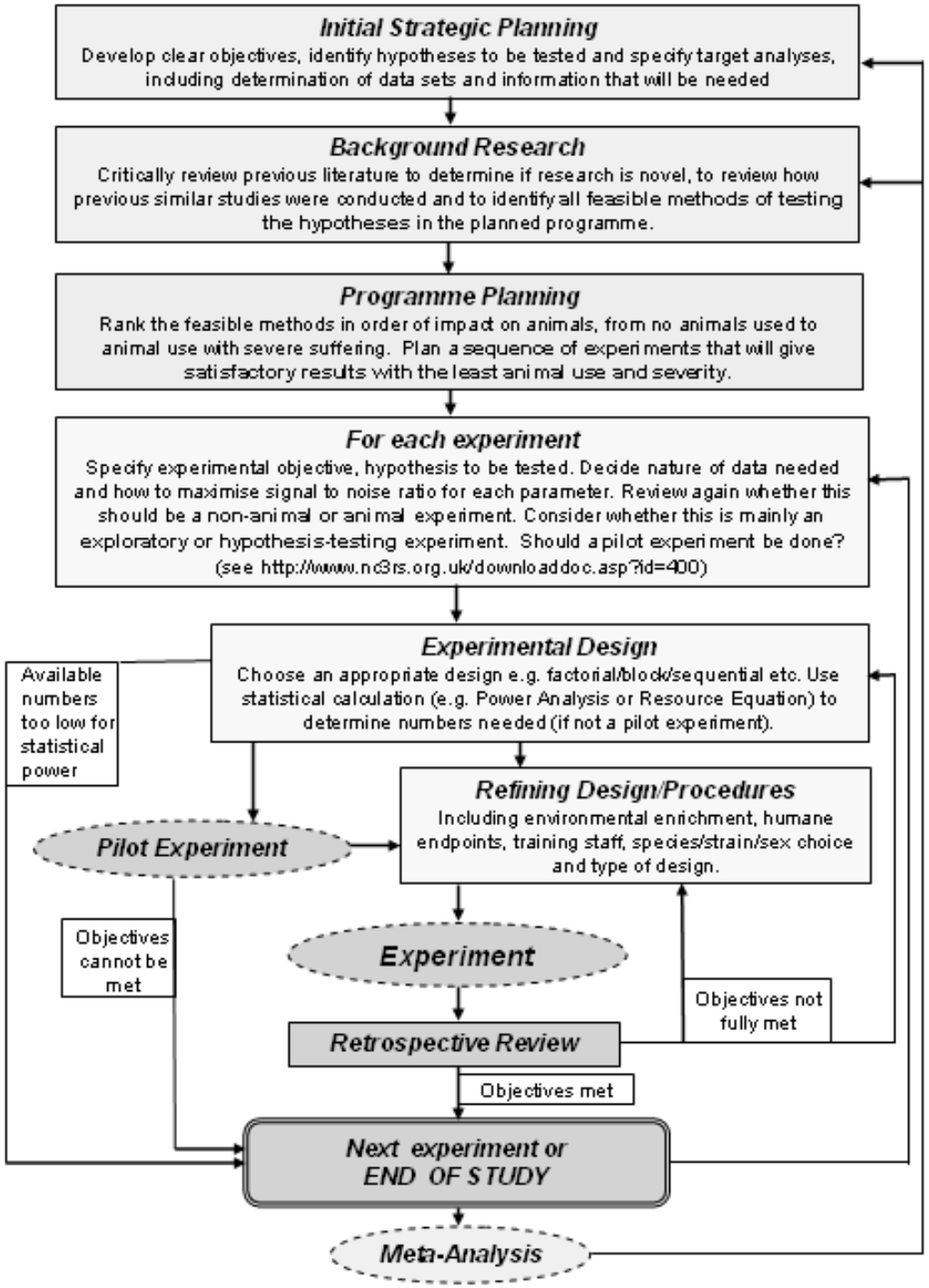
  FELASA Syllabus for scientists

- In UK since 1987 the assessment has been of experimental programmes of up to five years ("projects")
- And the project licence application form has included a special section on the plan of work
- In EU replacement for Directive 86/609 introduces "projects". Perhaps this could place more emphasis on ethical evaluation of overall programme severity.

Flow chart

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

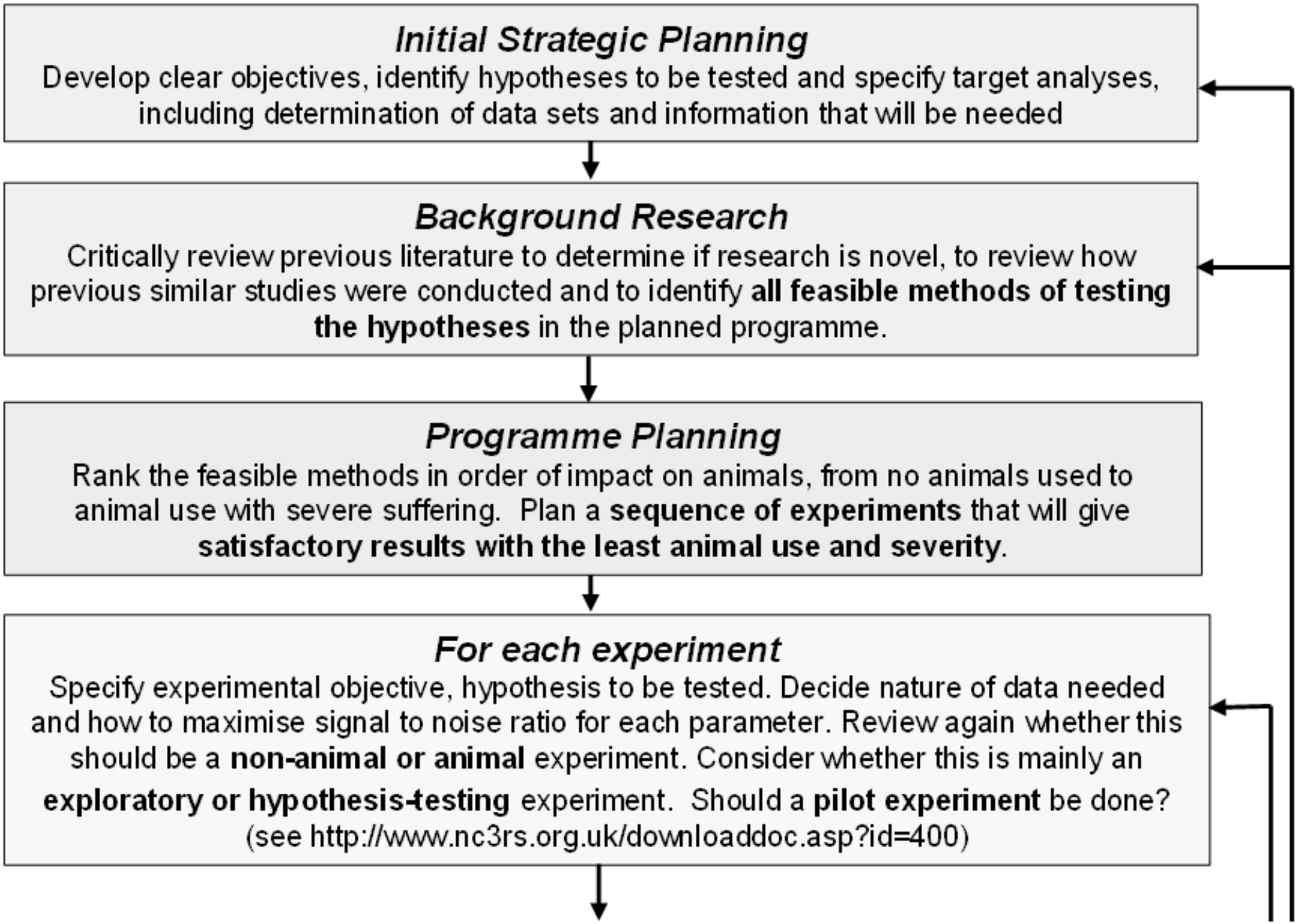
But it is also useful for refinement



This may be a useful checklist for ACC ethical evaluation

- ..and I will use it as a framework for indicating where ACCs could ask questions about the application of the 3Rs in proposals

The planning stages



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 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.

Example: Unclear objectives

- The next slide shows the only paragraphs dealing with the objectives of the programme in an abstract of proposed research submitted for evaluation
- Can you tell what the programme is attempting to achieve?

Example: Unclear objectives

- Certain strains of laboratory mice have conditions caused by mutations in the same genes implicated in the corresponding human diseases; and the conditions observed in these mice closely resemble the human disorder. In these diseases, examination of the brain cells (especially those termed neurones) shows that because of this defect, there is a damaging build-up of fatty substances within specific spaces in the cell.
- Our laboratory team has already developed an effective means to replace, by means of gene therapy, the function that is lacking in these diseases in genetically-altered mice that faithfully recapitulate the human condition.
- Other treatments (including those that work on the principle of the statin drugs, which reduce other types of harmful storage in cells) might cooperate with gene therapy to improve outcome: these would act by reducing the effects of storage due to the inherited defect or by enhancing the action of the residual gene product.

Example: Unclear objectives

- Probable aim of the programme - To determine effective additions to gene therapy treatment

Example: Unclear aims, objectives

- This was an abstract for the lay reader
- The scientist may be much better at providing aims and objectives in a scientific context
- But the abstract should alert evaluators to question whether what the programme is trying to achieve is clear.

The planning

- ***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 unknowns to 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

- What can be done without animals?
- What can be done under terminal anaesthesia?
- What can be done with only mild severity?
- What can only be done at more than mild severity?

Planning Example: Acute pancreatitis programme

- Acute pancreatitis
  - is very painful
  - carries substantial morbidity
  - can be fatal
  - is worse in obesity.
- There are significant deficiencies in current treatment.
- Mechanism by which increased adiposity worsens AP remains unknown.
- Justified use of animals for improved treatment – whole animal needed for the interactions of multiple body systems

Initial steps

- Overall aim – to investigate why acute pancreatitis is worse in obesity
  - To develop an obese animal model
  - To use pharmacological dissection to detect mechanisms and compare with non-obese
- Research indicates
  - intraperitoneal injection of a combination of two interleukins will induce AP
  - the ob/ob mouse could be a good obese model

Raised serum enzymes are early indicators

Indicators of pancreatitis severity also rise early

and acinar cell damage can be detected by histology

Possible severity-sensitive programme

- Pilot experiment(s) with
  - likely dose and combination of interleukins,
  - serial or post-mortem blood sampling for enzymes and post mortem histology
  - abdominal pain signs or set-time endpoint.
- Factorial experiment
  - different interleukin doses and combinations
  - early serum enzymes + histology as measures
  - determine optimal dose combination.
- Timed-kill experiment for precise time course of enzyme changes/histopathology
- 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

- Pilot to find suitable dose and interleukin combinations (? with death as measure)
- Survival experiment 10 obese, 10 non-obese, repeated. Obese die in 24-48h
- Repeated in non-obese at higher dose till 30% die
- Experiment to show multiple organ failure
- Experiments on time course of serum enzymes, markers and histology
- Pharmacological studies

An ACC could

- Question the sequencing of experiments
- And readily understand if severity was not being adequately considered by the researcher

An ACC could also

- Include evaluation of the plan of work in the assessment of proposals
- Request a retrospective review of a period of research, or an end-of-programme report, and use these to scrutinise the decision process during the work
- Encourage discussion of the topic, e.g. by local meetings with case studies

3Rs assessment of proposals

Prospective assessment

- Have the 3Rs questions been asked?
- What evidence is provided of effort to seek answers?
- What is the track record?

Ongoing assessment

- Is the work being reviewed against the 3Rs?
- What is being done to be up-to-date?

Retrospective assessment

- Could the experiments have been done without live animals?
- Could they have been done more efficiently?
- Could they have been done with less severity?

How does an ACC judge whether reasonable effort has but put into finding non-animal and refinement possibilities?

- Through what is given in a proposal, and through judgement of the track record and attitude of the proposer
- But with some difficulty
- Finding alternatives is complex and statements in applications of what was consulted and what search strategy was adopted are insufficient and may not be worth requiring
- Scrutiny by informed persons before during and after a programme of work is needed

So..

- Evaluators
- And the researchers themselves
- Should be able to pose relevant and probing questions on the application of the 3Rs

But also

- A culture of 3Rs questioning needs to be developed within the scientific community
- So that an ACC knows the peer group is expecting application of the 3Rs and thorough searching for alternatives
- And can trust the scientists to make efforts to find alternatives

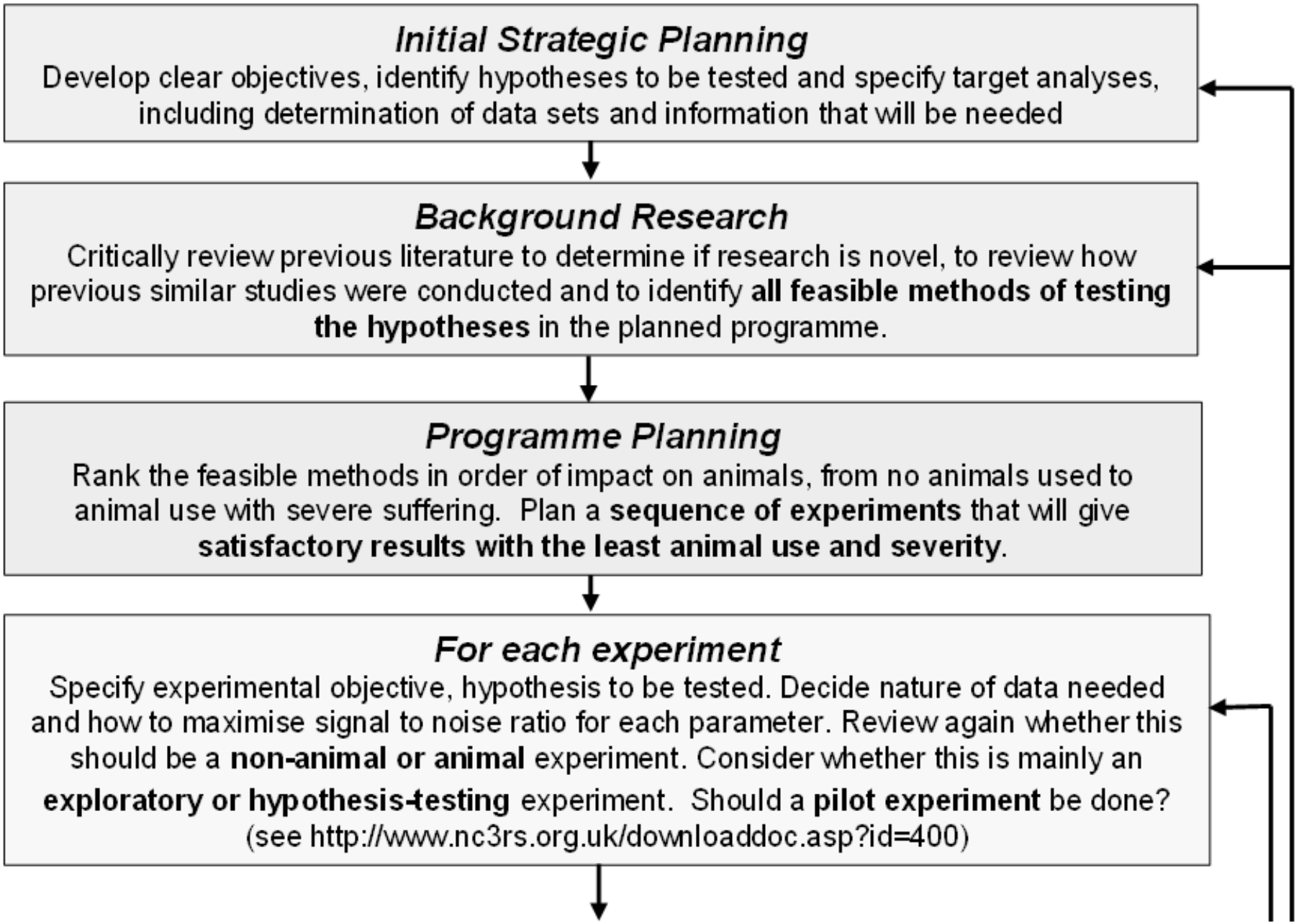
Changing attitudes

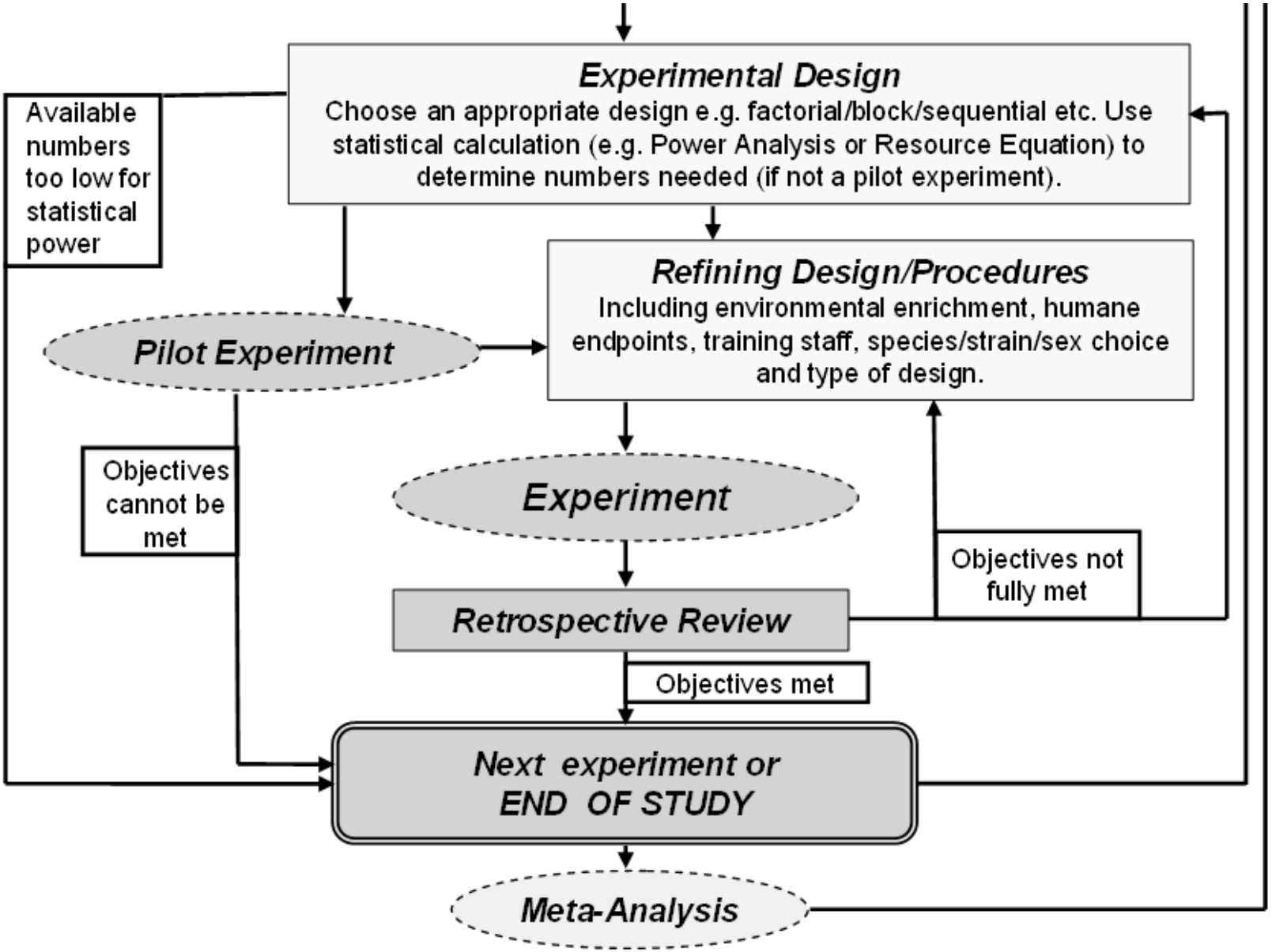
- E.g. for replacement put the question
- "If you could not use live animals what would you do?"
- In a workshop or event discuss scenarios for this question - realistic scientific problems in which replacement alternatives may not occur to the researcher
- E.g. for refinement have a (non-confrontational!) meeting between scientists, animal care staff and veterinarians exploring a general topic like pain relief, blood sampling

Finding alternatives involves

- Knowledge of information sources, skill to use them, and wish to do so
- Exploiting links and search engines
- Developing search strategies, or having tutored searching, or expert searches
- Using many information sources - colleagues, web, library, databases, learned societies
- Swapping anecdotes

The planning stages




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

An ACC can do much to question whether reduction and refinement are being applied at the experiment level …

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

Questioning about experimental objectives requires thought more than particular expertise

- ..and a probing committee may well help an experimenter to clarify his or her own thoughts
- .. and reduce and/or refine the proposed work

Clarity of objectives – example

- Stated objective - to determine the peak of growth factor response following injury
- Proposed design
  - kill groups daily from days 1 to 20 following injury
  - group size estimated 20 (statistician's advice; methods optimal and variance known)
  - 400 animals needed + 200 controls

Really two different objectives

- Objective 1 – to determine the timing of the peak
- Objective 2 – to determine the size of the peak

Really two different objectives

- Objective 1 – to determine the timing of the peak
  - kill 2 each day - 40 animals.
- Objective 2 – to determine the size of the peak
  - kill at 3 times around peak – 60 animals + day 0 and peak day controls – 100 animals in all
- So clarifying the objective could save 500 animals!

Also..

- It does not need expertise in experimental design to question proposed designs

Experimental Design is not Statistics

It calls for a combination of

- biological insight
- logic
- common sense
- planning
- and an appreciation of statistics!

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

Why design an experiment?

- 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

Fundamental principles

- Replication
  -
- Appropriate controls or comparisons
  -
- Random assignment to treatments
  -

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

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)
  - Allow 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

Members of an ACC may well be able to judge proposed experiments are poor

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

Inefficient design is not uncommon

- 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"

Inefficient design is not uncommon

Other examples –

- Experiments screening large numbers of potential drugs can use fewer animals by comparing many potential drug groups with small numbers against a single large control group, instead of the usual pattern of the same numbers of animals in all groups.
- Cross-over designs where different treatments are compared in the same animal at different time periods can halve or more the animals used. A food additive experiment which proposed using 96 animals was reduced to using 24 by this approach.

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

Many of these could be asked about by an ACC evaluating research work

- but that's not a complete list!

It doesn’t include failure to refine

And it should

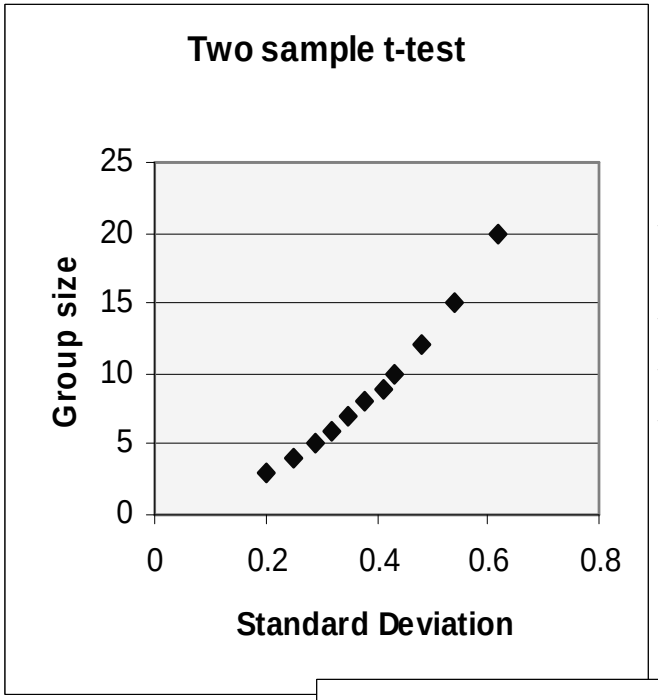
It doesn't include failure to refine

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

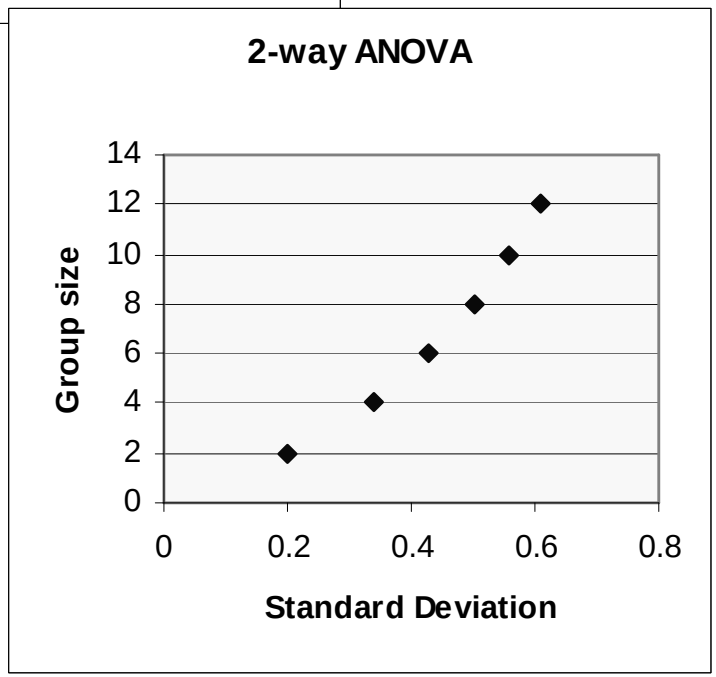
Variability has a large effect on numbers needed

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

Planning for minimal severity 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).
- ***Refining Design/Procedures***
  - Include environmental enrichment, humane endpoints, training of staff, species/strain/sex choice and type of design.

In conclusion

- An ACC can do a lot to judge how 3Rs principles are being applied, even without much expertise
- .. by seeing experiments in the context of the experimental programme
- .. and asking general questions on individual experiments

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.