

Refinement, Reduction, and Replacement of Animal Use for Regulatory Testing: Future Improvements and Implementation Within the Regulatory Framework

Jon Richmond

Abstract

Many are critical of how regulatory testing practices have evolved and become established—critical both of the scientific rationale and the animal welfare costs. The test of whether we are more enlightened than our predecessors will be whether, armed with more powerful scientific insights and a better understanding of animal welfare, we can ensure that the best animal welfare and the best science drive and shape future developments in regulatory testing. Conducting the most humane animal-based regulatory testing requires establishing and maintaining a constructive dialogue between stakeholders and acknowledging the common ground that unites. Inclusive processes with stakeholders prepared to offer public, rational justifications for their policies and processes are essential if best practice is to be identified and implemented. There is general agreement that the best animal welfare results in the best science; that regulatory requirements based on an understanding of mechanisms and early relevant biomarkers result in elegant and valid science. Thus, “alternative” methods enabling replacement, reduction, or refinement (the 3Rs) are in reality often more scientifically “advanced” and scientifically valid methods. These principles provided the incentive and framework for recent initiatives in the United Kingdom to enhance the quality of the data prepared for regulatory submission while better protecting the welfare of the animals used. Some remaining 3R opportunities are explored in this paper, and some of the commonly encountered myths about regulatory testing and perceived barriers to change are challenged. Current “threats” may indeed offer opportunities for ensuring that sound science and the best animal welfare underpin developments in regulatory testing.

Key Words: animal testing alternatives; animal welfare; laboratory animal science; preclinical drug evaluation; regulatory requirements; risk assessment

Introduction

Planning and conducting the most humane and properly justified animal-based regulatory testing requires a constructive dialogue based on trust and respect between and among stakeholders. Inclusive processes, in

which all stakeholders are required to offer and sustain public and rational justifications for their interests, beliefs, policies, and positions, are essential if all of the relevant interests are to be balanced; best current practice identified, codified, and implemented; priorities determined; and progress made.

Best Welfare Is Best Science

The “best animal welfare” results in the “best science” (Russell and Burch 1959). As a better understanding of the key biological mechanisms and relevant, early, subclinical biomarkers develops, the scene is set for reliance on methods more refined than the current animal studies that rely on more empirical methods and findings based on significant morbidity or gross pathology. Strategies and methods based on an understanding of relevant biological mechanisms and early, appropriate biomarkers comprise elegant, sound science. “Alternative” methods taking account of advances in our understanding of biological systems facilitate replacement, reduction, and refinement of animal use. Such methods are often more valid and more reliable than those traditionally used in regulatory testing. They are also often quicker and less expensive than the methods that they supersede.

Acceptance of these principles has provided the incentive and framework for a number of recent initiatives in the United Kingdom aimed at both enhancing regulatory testing and better protecting the welfare of the animals used (Home Office 2001). More substantive progress requires a similar dialogue and change at the supranational level.

Those with (and those without) knowledge of science and/or laboratory animal welfare are often critical of how current practices have evolved and become established—critical both of the scientific rationale and of the costs in terms of animal welfare (Weisser and Hechler 1997). The challenge for us, armed with more powerful scientific insights and a better understanding of animal welfare issues than our predecessors, is to ensure that the best animal welfare and the best science drive and shape future developments in regulatory testing (Richmond 2000).

Some Common Ground?

I assume there is agreement that animal studies should be conducted only when (1) the objective is of sufficient im-

Jon Richmond, FRCSed., is Chief inspector at the Home Office, UK.

portance; (2) there is no nonsentient replacement alternative; (3) all relevant reduction and refinement strategies have been identified and implemented; (4) the design and conduct of the study minimizes the animal welfare cost as reflected in the total pain and distress that is caused, and not simply in the number of animals used; and (5) there is maximum scientific benefit. I also assume we agree that new regulatory test requirements should, from the outset, require the most relevant and humane test methods. I believe it is easier to factor in these aspects from the start than to change established requirements, guidelines, and practices.

Regulatory Requirements

Regulatory requirements aimed at protecting people and the environment are many and varied. Harmonization of test requirements and test methods, and availability and mutual recognition of test data and regulatory assessments, are required not only to minimize animal testing but also to ensure that responsible regulatory decisions are taken in an ethical, consistent, and timely manner. Even when test requirements are harmonized, perceived regulatory preferences and alleged regulatory prejudices may at times overwhelm the most relevant and refined test methods and protocols.

Political and public concerns about human safety and protection of the environment are increasing, raising expectations of progressively higher levels of safety and more transparency in the decision making process. The regulatory systems that underpin human and environmental safety must adapt and evolve to accommodate these expectations. In the case of impending chemical testing strategies, strategies that are evolving in different parts of the world should be coordinated. Every effort must also be made to ensure that thorough review of existing test data, mutual recognition of regulatory decisions, and active data sharing minimize the additional animal-based testing that will be performed. In the case of likely test requirements to evaluate endocrine disrupter potential, it is to be hoped that the most refined and relevant test methods will be identified and validated before formal test requirements are promulgated (Combes 1998). The current segregation of regulatory requirements by class of material has resulted in a plethora of regulatory requirements that must be maintained as technical progress is made.

Progress in biotechnology will have a significant impact on regulatory affairs. New healthcare products will be produced increasingly by biotechnology processes and may thus be classed and tested as “biologicals” rather than as classical chemical pharmaceuticals (EFPIA 2000). The prospect of xenotransplantation technologies and products being incorporated into clinical practice also highlights the need for revised or additional regulatory frameworks. Strictly speaking, they are not medical devices or classical organ or tissue allografts and may pose novel health risks

both to individual recipients and society. The possibility of cloning and stem cell technologies enabling tissue engineering and human and animal reproductive cloning also raises novel regulatory issues.

Although some people believe that regulatory requirements are rigid and inflexible, there is often, in fact, considerable flexibility in their interpretation and implementation, and a need to exercise expert judgment to determine the most appropriate test strategies and methods. This is rightly the case. However, when such judgment is exercised, for example in the selection of a nonrodent second species, it should be based on scientific considerations and the justification should be transparent. Concern remains that some decisions are based on custom and practice rather than on reflecting a reasoned, informed science-driven selection of the method most likely to give the most relevant result for the specific test material.

“Data-sharing powers” are not a feature of many current legislative frameworks—data sharing is generally encouraged rather than required. The forthcoming chemical testing strategies provide the strongest-ever incentive to enable and require data sharing. Without effective data-sharing protocols, the proposed strategies will not deliver the promised benefits within the deadlines being set. If progress is to be made, consideration must be given to overcome some of the obstacles to data sharing, to determine who should fund such initiatives, and to identify “honest brokers” who are best suited to manage such sensitive and valuable data.

There is also a need to consider how regulatory test requirements are maintained to accommodate future technical progress—both to enable the incorporation of suitably validated and more advanced test methods (only after their scope and limitations have been clearly defined) and to delete superseded tests (when the time is right). We need to reflect on whether there is the potential for shortening the time taken for technical progress to be reflected in test requirements and scientific practice.

Justify, Justify, Justify

The debate about the need to use animals for experimental and other scientific purposes has a long history (Barley 1999). Despite the strength of some of the views expressed, and the sincerity with which they are held, the debate is still taking place without the public and many opinion formers understanding the basic facts about the necessity for, and nature of, the animal use. Political and public confidence cannot be maintained unless a better informed debate is facilitated. It is important that all of us involved in requiring, commissioning, performing, and regulating animal testing, and assessing animal test data, publicly take responsibility for our roles and ensure that the facts are understood. Bodies that determine test requirements and agencies that take regulatory decisions should do more to raise public awareness of the purpose of testing, the insights that must be gained, and the importance of the decisions that

must be taken when the test data are considered. This increased activity requires educating the public, but not by trading propaganda with those who campaign emotively for nothing short of the immediate and unconditional abolition of the use of animals in science.

It is important to explain and justify WHAT testing has to be done, WHY it has to be done, and HOW it is to be done. If those involved in devising and implementing the test requirements do not do so, no one else will. If the test requirements are not demonstrably based on best science and best animal welfare, a sustainable case cannot be made. Providing this additional information will not reassure all of those who have concerns about the use of animals for this purpose, but it should raise the level of debate and allow the issues to be discussed on the basis of fact rather than ignorance and prejudice.

Alternative or Advanced?

A number of organizations are dedicated solely or primarily to the development and promotion of test methods that replace the use of living animals, reduce the number of animals used, or refine the ways in which the animals are produced, housed, and used. Nevertheless, the vast majority of the resource devoted to the 3Rs is provided, and the bulk of the resulting progress is made, by those who currently require animals for scientific purposes. The motive for incorporating the 3Rs is usually neither altruism nor public relations. Rather, methodological improvements are sought as a means of overcoming the technical limitations inherent in current animal models. To practicing scientists, these more elegant and relevant methods represent technical progress and are considered to be additional or advanced, rather than *alternative*, methods.

However, this progressive aspect of the 3Rs has not been communicated effectively to the public, and the belief persists that the scientific community is at best ignorant of, and at worst hostile to, such progress. Far from giving the scientific community the credit that is its due for leading the way, animal protection groups have persuaded some sectors of the public that the antivivisection movement, rather than the scientific community, is at the forefront of developing and promoting more refined methods.

Functional genomics and the availability of animals with specific genetic modifications will, in the context of regulatory testing, allow some existing things to be done better and enable some things that were not previously feasible. We should already be thinking about the impact of these technologies on regulatory testing, trying to identify priority areas where such advances can help solve existing problems, and considering how such models will be developed and incorporated into regulatory practice.

Many current test methods were developed with only a basic understanding of the gross endpoint or pathology of interest, and without knowledge of the mechanisms by which the biological effects are mediated. Methods based on

an understanding of the relevant mechanisms and endpoints according to early, relevant, subclinical biomarkers set the scene for more elegant, valid, and refined model systems. The murine local lymph node assay (NIEHS 1999) is a case in point: It relies on subclinical changes caused by the induction phase of the process that culminates in the development of acute contact dermatitis, whereas the traditional guinea pig maximization test relies on gross pathology and clinical changes. Another example is the use of telemetry systems to measure core temperature in laboratory rodents allowing the use of hypothermia, rather than serious morbidity and mortality, as the marker of overwhelming systemic infection (Soothill et al. 1992).

Much has been written of the validation of new test methods (Balls et al. 1995), the process by which the reliability and relevance of a procedure is established for a specific purpose. I believe it is essential for any authoritative statement of scientific validity to be linked clearly and inseparably to a candid and clear description of the scope and limitations of the test as validated. For example, the local lymph node assay (NIEHS 1999) has not yet been definitively validated against all of the classes of materials for which skin sensitization testing is required, and it is known that in some circumstances it can give false-positive and false-negative results. Nevertheless, many of those concerned about the welfare of animals used for regulatory testing mistakenly believe the current validation status of the test should require its use in all circumstances and for all test materials, and without the need to undertake supplementary animal tests to confirm negative results. Such misconceptions reinforce suspicion that only the conservatism and stubbornness of the scientific and regulatory communities prevent the immediate phasing out of the guinea pig maximization test.

All parties would benefit from a clearer understanding of when scientifically validated tests can be deemed to have become the “method most likely to produce a scientifically satisfactory result” and have become “reasonably and practically available” (EEC 1986).

Replacement

The greatest progress with replacement alternatives to date has been with respect to effects mediated by single-stage processes involving only a single cell or tissue type. Replacement strategies for processes that involve more dynamic and complex biological interactions are much more difficult. Key current replacement strategies should be considered to include harmonization of test requirements, facilitating data sharing, and ensuring that new regulatory test requirements do not invoke animal-based testing when in vitro methods will suffice.

Reduction or Minimization?

We need to think in terms of minimizing, rather than reducing, the numbers of animals used. Focusing purely on

decreasing numbers can lead to strategies that reduce the numbers used but produce a disproportionate increase in the pain and distress caused to the animals that are used (Richmond 1999).

Testing strategy is important. In the product development and regulatory contexts, one of the most effective means of minimizing the numbers of animals required is to develop tiered and hierarchical approaches to ensure that the right tests are done at the right time. The early identification and discarding of materials not destined for full-scale development comprise sound business sense, good science, and good animal welfare.

We should also reflect on the degree of precision required to undertake regulatory assessments. Although in some circumstances very precise information is required (e.g., when materials must be placed in rank-order by toxic potential), in many other situations information is required only to assign materials to broad categories. Although acute toxicity testing in accordance with Organization for Economic Cooperation and Development Guideline 401 (OECD 1987) can give a point estimate of the LD₅₀ and a good indication of the slope along which lethality can be expected (albeit the results are context specific rather than representing true biological constants), less precise information derived from OECD Guidelines 420, 423, and 425 (OECD 2001a-c) is sufficient for most regulatory needs requiring only that materials are assigned to various toxic classes or subsets.

Improved statistical methods are often cited as an important means of reducing the numbers of animals used. However, in my experience, it is more common to underestimate, rather than overestimate, the numbers of animals required. Nevertheless, sound statistical design plays a key role in ensuring that the right numbers of animals are used. Using too few animals is in many ways worse than using too many.

Experimental design is an art—and it is related to much more than statistics (Festing 2000). The need for positive and negative control groups (and the size of such groups relative to those receiving the true test material) might sometimes be questioned. In other circumstances, a proportionately large amount of additional information might be gained from the use of small additional satellite groups.

Refinement

The welfare cost to the animals used for regulatory testing has two components (Russell and Burch 1959): (1) the “direct costs” of the procedures applied and their consequences; and (2) the “contingent costs,” including the welfare-negative aspects of animal production and care. It is important to remember that welfare is compromised not only by the infliction of what is unpleasant, but also by the denial of what is pleasurable. The need to standardize biological variables and to have regular and easy access to study-animals largely determines how they are produced,

housed, cared for, and used. There is growing evidence that the standards of accommodation and care can have subtle yet profound effects both on the welfare of the animals used and on the data-streams collected during studies (Poole 1997). Because any unwanted stressor will have a negative effect both on animal welfare and science, it seems logical that they be identified and eliminated whenever possible.

One of the most significant changes I have witnessed over the last 10 yr relates to the housing of animals used for regulatory studies. The vast majority of dogs and nonhuman primates used for regulatory studies in the United Kingdom are now pair or group housed, although this arrangement is not yet the norm elsewhere. I congratulate those who tackled and overcame the previously “insurmountable problems” (e.g., “it has never been done that way”; “the data might be different”; “the studies might be compromised”; “the regulators won’t accept the data”; and “clients won’t want it”), and I urge others to do the same.

The best use of animals in science requires a multidisciplinary, team approach (Richmond 2000). No individual is truly expert in all of the disciplines involved. Although in the context of regulatory testing, those knowledgeable in regulatory affairs, toxicology, and xenobiotic metabolism are generally involved most actively, laboratory animal veterinarians and animal care staff should also play an active role in the planning and performance of tests.

A key strategy for refinement is to identify the mechanism of interest and to ensure that the earliest appropriate biomarkers are identified and used. This strategy results not only in better animal welfare but also in preventing contamination of the data stream with unwanted biological changes introduced by unwanted or unrelated secondary or tertiary effects, it is also better science.

Planning and implementing humane endpoints (Richmond 1999) are essential components of humane science. The Canadian Council on Animal Care has produced and published useful practical guidance on this topic (CCAC 1998). All reasonable steps should be taken to minimize the pain and distress caused. Endpoints should cover situations including animals experiencing more pain or distress than can be justified; when the objective has been realized; and when it is realized that the objective cannot be reached. It is important to appreciate that humane endpoints may thus be appropriate at times when the levels of pain and distress being experienced are not high.

Endpoints should be considered at the planning stage, when the rationale for the provision or withholding of specific, symptomatic, and supportive treatments should also be established. Such treatments are not necessarily appropriate in every case; however, justification is necessary if they are to be withheld. While studies are in progress, it is important to recognize evidence of possible welfare problems and to apply endpoint standards in a timely manner. These key concepts require staff training and competencies along with established observation schedules. When studies have been completed, the clinical and pathological findings should be reviewed to determine whether the clinical rec-

ords suggest that all of the expected clinical manifestations of the pathologies produced were detected and whether the clinical signs recorded were indicative of significant welfare problems. This information should be taken into account when future studies are planned. The aim should be continuous improvement.

Regulatory testing does not require death as an endpoint. Information from animals “found dead” on studies is less valuable than information from animals euthanized and promptly autopsied. It is my view that a key indicator of the technical and welfare standards of establishments that use animals for scientific purpose, and an important measure of their “culture of care,” is the number or proportion of animals that are found dead during studies. Generally such events indicate that opportunities for refinement have been missed and that useful scientific data might have been lost.

Cognitive neuroscientists will never be able to tell us precisely what animals actually experience. We are in the era of critical anthropomorphism, severity score sheets, and disturbance indices as means of assessing the relative merits of refinement strategies (Hendriksen and Morton 1999). Although such measures can be used to assign refinement options to general categories, we do not yet have the correct tools to discriminate objectively and reliably between subtly different degrees of refinement. That is not to decry the good work that is being done—it is a plea to do even more.

Barriers to Change

Test development, scientific validation, regulatory acceptance, and client uptake all take time. Opportunities to ease the existing bottlenecks should continue to be sought. That change “. . . might affect the results . . .” is undoubtedly true; however, it may be more of an indictment of the limitations and inadequacies of existing data sets than of the validity of the data obtained by more advanced methods.

Regulatory inertia is often stated as a barrier to change, and it is difficult to know whether this is a real or perceived problem. Whichever is the case, the solution, in part, is to establish more transparency and a better dialogue between stakeholders.

New technologies require new competencies and involve other costs. Within the European legislative framework, delay in introducing more refined methods that may produce scientifically satisfactory results cannot be permitted on the strength of arguments based on cost or convenience (EEC 1986). Lest others consider this approach to be a strange European eccentricity, I argue that it should be the case in any country whose science base is built on technical excellence and the ability to innovate and remain at the forefront of scientific endeavor.

Future Progress

A key part of making progress is to establish and maintain a constructive dialogue between all stakeholders. This has

already produced progress in the United Kingdom where a national guidance on humane regulatory testing (Home Office 2001) has been drawn up between the Home Office (which regulates animal research), the regulatory authorities responsible for regulatory testing, and the relevant professional bodies. Administrative concordats between the Home Office and the regulatory bodies also cover data sharing and the implementation of the 3Rs.

Opportunities

The forthcoming chemical testing strategies will test current commitments to (and provide once-in-a-lifetime opportunities for) harmonization of test requirements, data sharing, and promotion of the most humane, science-based test methods. The consideration being given to developing test requirements to predict endocrine disrupter activity allows the opportunity to recognize that the development of new test requirements is based on reason and science. The boom in the development of biotechnology-derived health care products sets the scene for informed discussion and review of how these products are best evaluated.

The Immediate Future

The principles outlined in the UK document on humane regulatory testing (Home Office 2001) and the OECD guidance on humane endpoints (OECD 2000) are applicable across all branches of regulatory testing involving living animals. Communication and dialogue based on transparency and reason are the keys to progress. Inclusive and reasoned process, and outputs that incorporate best science and best welfare, must be seen to underpin new regulatory testing requirements. Regulators and scientists must be seen to be enlightened and to take the credit for the progress that is being made with the 3Rs.

References

- Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, Reinhardt CA, Repetto G, Sladowski D, Spielmann H, Zucco F. 1995. Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM Workshop 5. *ATLA* 23:129-147.
- Barley JB. 1999. Animal experimentation, the scientist and ethics. *Anim Technol* 50:1-10.
- CCAC [Canadian Council on Animal Care]. 1998. CCAC guidelines on choosing an appropriate endpoint in experiments using animals for research, teaching and testing. (www.ccac.ca/english/gdlines/endpts/appopen.htm.)
- Combes R. 1998. Screening for endocrine disrupters: Time for more science and less politics. *ATLA* 26:735-739.
- EEC [European Economic Community]. 1986. Directive 86/609/EEC. The approximation of laws, regulations and administrative provisions of member states regarding the protection of animals used for experimental and other scientific purposes. *Off J Eur Comm* L358:1-28.

- EFPIA [European Federation of Pharmaceutical Industries Associations]. 2000. *The Pharmaceutical Industry in Figures*. 2000 ed. Brussels: EFPIA.
- Festing M. 2000. Doing better animal experiments: Together with notes on genetic nomenclature of laboratory animals. ANZCCART Fact Sheet, ANZCCART News 13(3).
- Hendriksen CFM, Morton DB, eds. 1999. Humane endpoints in animal experimentation for biomedical research. In: *Proceedings of the International Conference, 22-25 November 1998, Zeist, The Netherlands*. London: The Royal Society of Medicine Press.
- Home Office. 2001. *Animals (Scientific Procedures) Act 1986, Guidance on the Conduct of Regulatory Toxicology and Safety Evaluation Studies*, Home Office, February 2001. <<http://www.homeoffice.gov.uk/ccpd/regtox.htm>>.
- NIEHS [National Institute of Environmental Health Sciences]. 1999. *The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds: The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)*. Research Triangle Park. NIH Publication No. 90-4494. <<http://iccvam.niehs.nih.gov>>.
- OECD [Organisation for Economic Co-operation and Development]. 1987. *Guideline 401: Acute oral toxicity*. Paris: OECD.
- OECD [Organisation for Economic Co-operation and Development]. November 2000. *Guidance document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluation*. <http://www.oecd.org/ehs/>.
- OECD [Organisation for Economic Co-operation and Development]. 2001a. *Guideline 420: Draft revision: Acute oral toxicity—Fixed dose procedure*. Paris: OECD.
- OECD [Organisation for Economic Co-operation and Development]. 2001b. *Guideline 423: Draft revision: acute oral toxicity—Acute toxic class method*. Paris: OECD.
- OECD [Organisation for Economic Co-operation and Development]. 2001c. *Guideline 425: Draft revision: Acute oral toxicity—Modified up and down procedure*. Paris: OECD.
- Poole T. 1997. Happy animals make good science. *Lab Anim* 31:116-124.
- Richmond J. 1999. *Criteria for Humane Endpoints*. In: Hendriksen CFM, Morton DB, eds. *Humane Endpoints in Animal Experiments for Biomedical Research: Proceedings of the International Conference, 22-25 November 1998 Zeist, The Netherlands*. London: Royal Society of Medicine Press. p 26-32.
- Richmond J. 2000. The three Rs: A journey or a destination? *ATLA* 28: 761-773.
- Russell WMS, Burch RL. 1959. *The Principles of Humane Experimental Technique*. London: Methuen & Co. Ltd. [Reissued: 1992 Universities Federation for Animal Welfare, Herts, England]. <http://altweb.jhsph.edu/publications/humane_exp/het-toc.htm>.
- Soothill JS, Morton D, Ahmad A. 1992. The HID_{50} (hypothermia-inducing dose 50): An alternative to the LD_{50} for measurement of bacterial virulence. *Int J Exp Pathol* 73:95-98.
- Weisser K, Hechler U. 1997. *Animal Welfare Aspects in the Quality Control of Immunobiologicals—A Critical Evaluation of the Animal Tests in Pharmacopoeial Monographs*. Nottingham: FRAME.