Effects of the Introduction of *In Vitro* Assays on the Use of Experimental Animals in Pharmacological Research

Pieter M. Verbost,[1] Jan van der Valk[2] and Coenraad F.M. Hendriksen[2]

[1]Department of Pharmacology, Organon, Oss, The Netherlands; [2]Netherlands Centre for Alternatives to Animal Use (NCA), Department of Animals, Science and Society, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

Summary — The introduction of *in vitro* assays in pharmacological research has led to a reduction in the number of experimental animals used. But what has been the degree of this reduction, and when did it really start? This report describes the events in a medium-sized pharmaceutical company. Analysis of data collected over the last 12 years shows a five-fold reduction in the number of experimental animals used per compound synthesised. Compounds from compound libraries (large collections of randomly-synthesised molecules) that are being assessed for potential bioactivity in 'high-throughput screening' were not included in this analysis. Over the years, the (average) degree of discomfort for the animals in the experiments did not vary much; with variation generally observed from 1.5 to 2.0 (on a scale from 1–6). There was a peak in the discomfort score of experimental mice in 1997, which could be explained by the initiation of arthritis models that were subsequently refined, resulting in a lower degree of suffering. It might be concluded that the introduction of *in vitro* assays has indeed brought about a significant reduction in the number of experimental animals required to select a good compound (i.e. one that could progress to the preclinical toxicology phase). However, this development appears to have been neutralised by the low survival rate of new chemical entities in clinical studies, leading to a lower number of compounds per annum that actually reach the market place. Put in this 'productivity perspective', the number of experimental animals required to select a marketable drug has not much changed in the last decade.

Keywords: *alternatives,* in vitro *test, pharmacology, reduction.*

Address for correspondence: *P.M. Verbost, Department of Pharmacology, Organon, P.O. Box 20, 5340 BH Oss, The Netherlands.*
E-mail: *pieter.verbost@organon.com*

Introduction

The primary purpose of pre-clinical pharmacological research is to show the potency and efficacy of compounds to be further developed for the treatment of humans. Up to the 1990s, the standard procedure was to synthesise a compound for a certain purpose (e.g. treatment of rheumatoid arthritis or contraception) and administer it to an experimental animal, often a mouse, rat or rabbit, as a first test species. Receptor binding assays were available in the 1980s, but were still quite laborious and therefore not routinely used to (de-)select compounds. With the advent of medium-throughput *in vitro* assays (starting in the mid-1990s), and later the high-throughput screening (HTS) systems (in the late 1990s), a first selection was made based on *in vitro* data. At the same time, the rate of synthesis was increased significantly by the introduction of 'combichem-technology' (in the late 1990s), an automated synthesis applying 'clever' sets of chemical building blocks.

The research activities that were used for this analysis come from the pharmacology pipeline at the research site in The Netherlands of Organon, a medium-sized pharmaceutical company with research areas in anaesthesiology, cardiovascular pharmacology, gynaecology, immunology and psychiatry.

The objective of this overview is to evaluate the effects of the introduction of *in vitro* screening assays into the Organon pharmacological research programme on the numbers of animals (most commonly) used, and on the degree of discomfort conferred upon them (on a scale of 1–6). A correlation is made with the numbers of compounds synthesised by the chemists at Organon, and, secondly, with the number of new products that actually reached the market place.

The numbers presented reflect the major flow of events in pharmacological research, including some early toxicological research. It was not feasible to completely separate early toxicology research from pharmacological research. However, one could argue that 'early tox' makes up part of late pharmacology. Changes that are described here in the context of pharmacology have also occurred in toxicology, but this topic is addressed only briefly in the discussion, since it is beyond the scope of this overview.

Materials and Methods

Data sources

The data were retrieved from the corporate databases of the Pharmacology Department and the Medicinal Chemistry Department (Organon, Oss, The Netherlands), and relate to the five species — mouse, rat, guinea-pig, rabbit and dog — that represent 97–99% of the experimental animals used.

For comparison, the number of compounds analysed in HTS are also presented; these numbers have not been added to the 'compounds synthesised by the chemistry department'. Since they came from (often commercial) compound libraries, these compounds do not influence the outcome of the ultimate calculation of the use of experimental animals per synthesised compound.

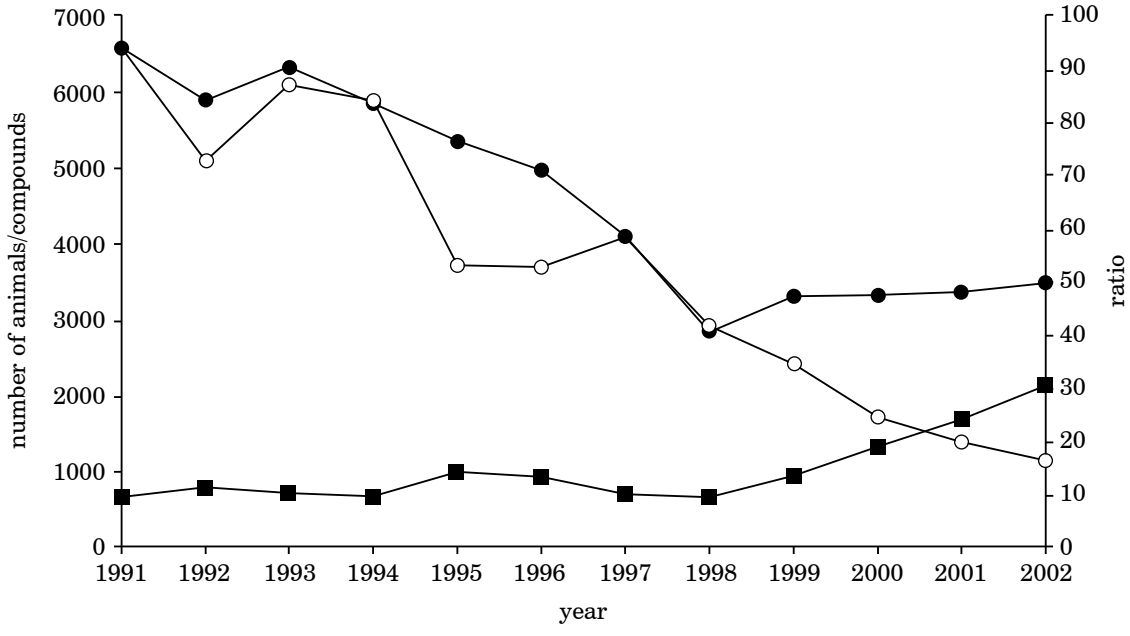
Strictly speaking, the numbers generated in this report refer to 'animal experiments', rather than the number of experimental animals used. Some of the experiments were performed with animals that were subsequently re-used. However, since the re-use of animals amounted to no more than 3–6% of the total, the two numbers are almost equal, so, for clarity in the presentation, we decided to state 'the number of experimental animals'. For the scoring of discomfort, the following categories were used (see also [1]): **1**: minor (e.g. a single injection); **2**: minor/mild (e.g. several injections, or several vaginal swabs); **3**: mild (e.g. surgery and recovery from anaesthesia); **4**: mild/serious (e.g. combination of treatments and surgery); **5**: serious (e.g. collagen-induced arthritis model in rat); and **6**: very serious (very exceptional, authorisation by a responsible authority required).

Results

The absolute number of animals used for pharmacological research has halved since the early 1990s to a level of about 33,000 — a level that appears to have stabilised during the last 5 years. This stabilisation may, at least partly, be explained by the approximate doubling of the rate of compound synthesis (from 800–900 before 1999, to a little over 2000 in 2002). However, measured over 12 years, the ratio of animals used per compound synthesised shows a 5-fold reduction (Figure 1).

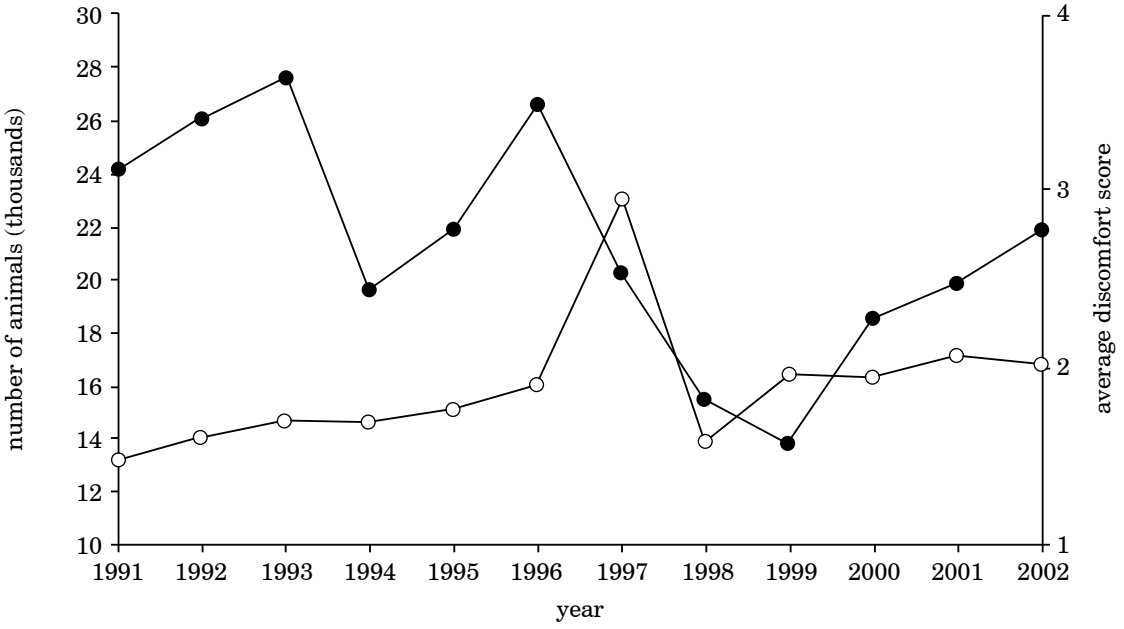
An average discomfort score was calculated by first multiplying the score by the number of animals rated with this score; these numbers for each score were then added together, then this value was divided by the total number of animals of the respective species. Overall, there was a tendency toward slightly higher average discomfort values, from 1.5 in the early 1990s to around 2 in the new century (see Figures 2–4).

Figure 1: Number of compounds synthesised, versus number of experimental animals used (summation of mice, rats, rabbits, guinea-pigs and dogs)



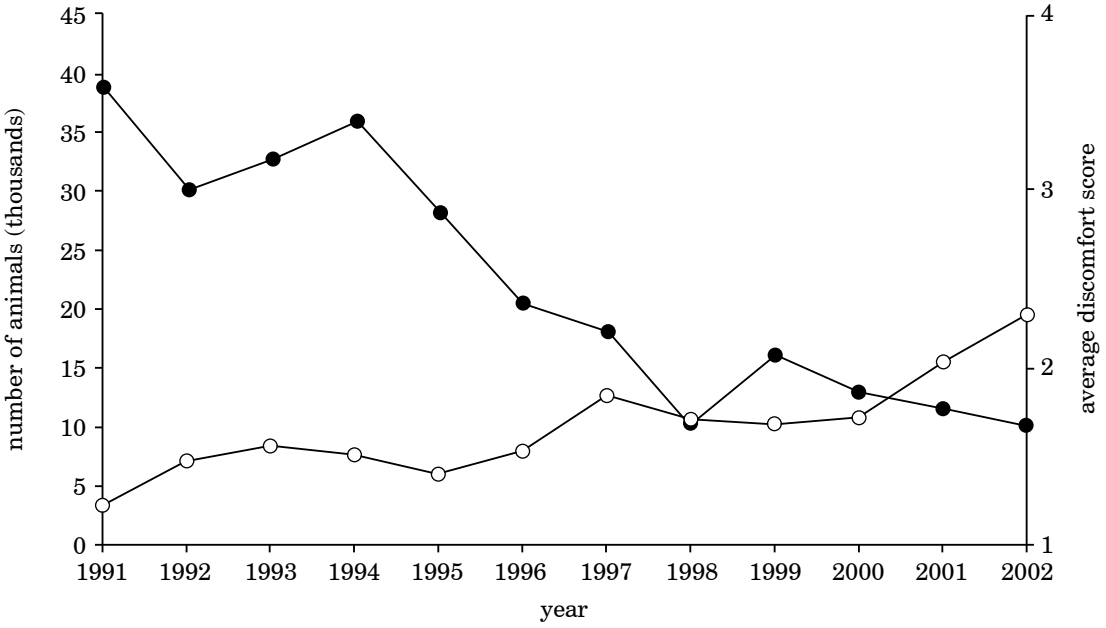
● = *number of animals/10;* ■ = *number of compounds;* ○ = *ratio of animals/compounds.*

Figure 2: Number of mice used, with average degree of discomfort, between 1991 and 2002



● = *animals;* ○ = *discomfort.*

Figure 3: Number of rats used, with average degree of discomfort, between 1991 and 2002



● = *animals;* ○ = *discomfort.*

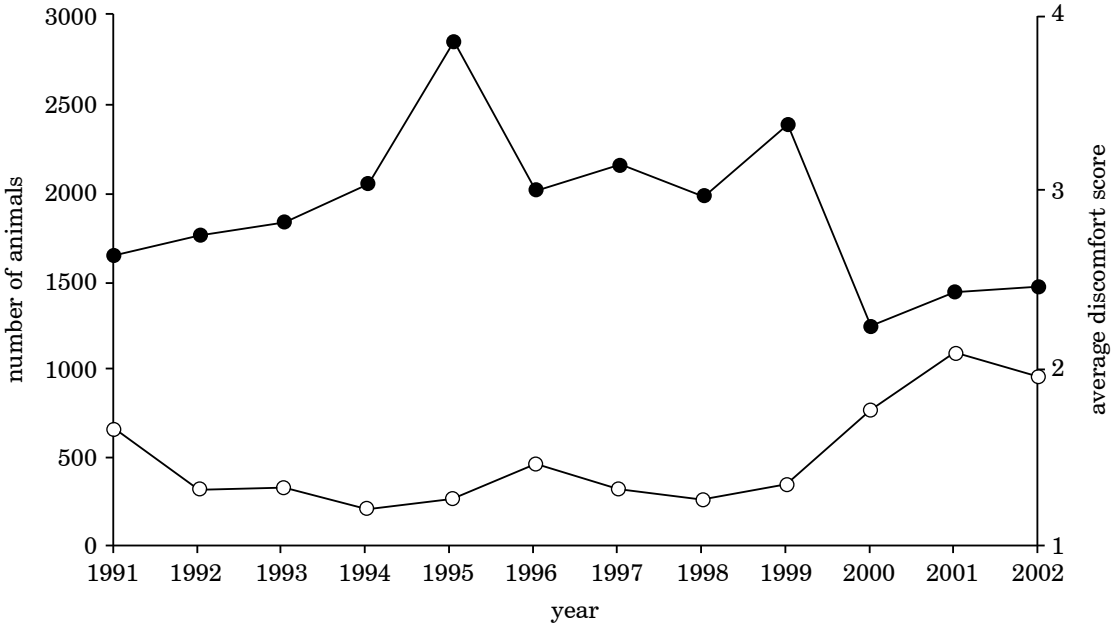
The peak in discomfort scores for experimental mice in 1997 could be attributed to an arthritis model that was refined shortly after its introduction. Perhaps the most drastic change in discomfort scoring occurred with rabbits: it increased from a stable 1.3 to around 2, in just two years. However, this was mainly caused by a change in scoring of a frequently-performed standard test, the McPhail assay (see Discussion), and the introduction of another model for studying angiogenesis in the endometrium, in which the animals are ovariectomised and subsequently treated with steroids daily for 10 days.

The utilisation of rats has dropped dramatically over the years, representing most of the reduction in animal use. In recent years, the number of mice used showed an unexpected rise, after an initial reduction, which can mainly be attributed to the introduction of knock-out and transgenic animals.
As stated above, the HTS of compounds from 'compound libraries' was not included in the current evaluation, because it does not require experimental animals. The number of compounds put through HTS has climbed tremendously since its introduction eight years ago. In 2002, the number increased to approximately 140 times the number of compounds actually synthesised by the Medicinal Chemistry Department (275,000 compared to 2,000; Figure 5). The positive hits from these screenings entered the normal lead optimisation process, involving the evaluation of medicinal chemistry activities (making changes to the molecule, plus the synthesis of sufficient quantities for testing), and subsequent *in vitro* and animal experiments.

Discussion

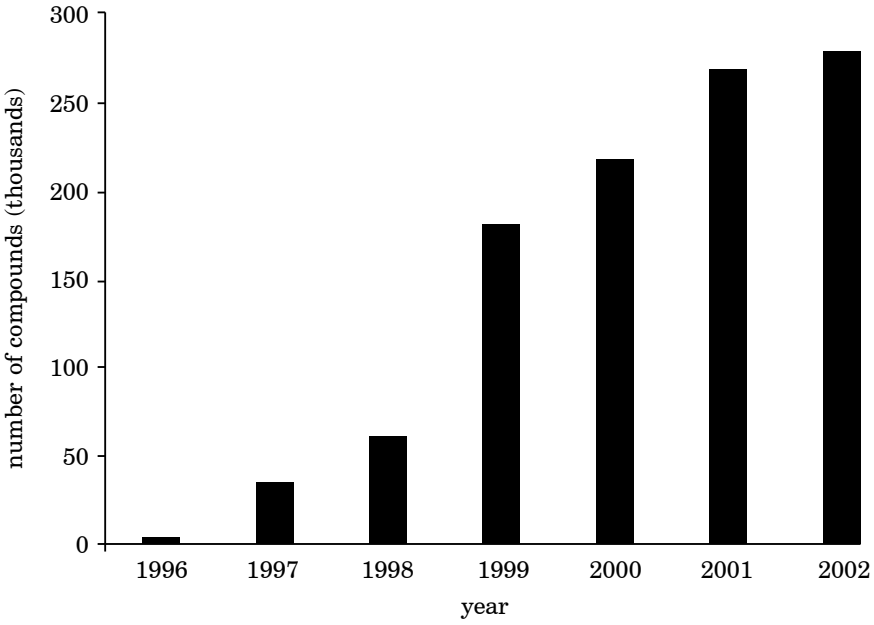
The data show a significant five-fold reduction over the last twelve years in the number of experimental animals used per compound tested in pharmacological research at Organon. This reduction coincided with the introduction of *in vitro* (primary) screening tests, indicating a causal relation. The introduction of HTS came a little later, but this was not included in the calculations presented here, because it precedes pharmacological research and does not require experimental animals. Its inclusion would strongly, but falsely, reduce the calculated number of animals per compound tested. The absolute number of animals used in Organon's pharmacological research has been reduced by 50% over a 12-year period, whereas the (overall) national reduction in animal use was about 20% in the same period (2). This reduction was mainly due to a decease in the use of rats. There was a similar initial downward trend in the number of mice used, but this has been reversed by the introduction of genetic modification approaches (knock-outs and transgenic mice).

Figure 4: Number of rabbits used, with average degree of discomfort, between 1991 and 2002



● = *animals;* ○ = *discomfort.*

Figure 5: Number of compounds tested in High-Throughput Screening (numbers NOT included in the calculations made for Figure 1)



The observed reduction in animal use is unmistakable. However, from a more holistic perspective, the data could be interpreted differently. What happened with respect to the 'delivery' of new molecular entities (NMEs) to the marketplace within the same time period? The trend in success rates is not encouraging. For a cohort of 19 companies (accounting for over 60% of global R&D expenditure), success rates between phase III and submission of a new drug application to a regulatory authority have fallen from 88% for NMEs entering phase III in 1994, to less than 50% for NMEs entering phase III in 1998 (3). This trend holds true for every company, irrespective of size and R&D spend. Analyses made world-wide confirm this picture (4): instead of around 40 NMEs in the early 1990s, the number of products first launched world-wide has dropped to around 30, a development which is inversely proportional to investment by the industry. The reason for the observed reduction in NMEs that are being introduced by pharmaceutical companies appears to be the increased difficulty in proving safety and efficacy in clinical studies. The investment escalation has almost entirely gone into the clinical phases (5). The preclinical toxicology phase has not changed significantly in the last decade. If anything, it has become more efficient, thanks to the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), which involves both regulators and research-based industry representatives from the EU, Japan and the US, and which took the initiative in scientific and technical discussions on the testing procedures required to assess and ensure (mainly) the safety and (also the quality and efficacy) of medicines.

According to the US Food and Drug Administration in 2004 (5), the success rate of drug discovery could (should) be improved by applying to the science of medical product development such new technologies as genomics, proteomics, bioinformatics systems, and new imaging technologies. Properly applied, these new technologies provide tools for detecting safety problems early, for identifying patients likely to respond to therapy, and can lead to new clinical endpoints. These new techniques are also attractive in terms of a further reduction in experimental animal use (e.g. toxicogenomics and biomarkers).

However, at this point in time, one has to conclude that, although the number of experimental animals invested per synthesised (research) compound has decreased, the net effect is that the number of experimental animals used per NME — the outcome relevant to the wider community — roughly stays the same. On the other hand, one could argue that, despite the fact that it is increasingly difficult to produce a NME, no more experimental animals are being used.

In the 12 years between 1991 and 2002, the average discomfort scores moderately climbed from 1.5 to around 2. This was most likely due to the fact that the general perception of the degree of suffering had been adjusted over the years. This way of presenting the average scores masks the ongoing

efforts to refine or replace animal experiments. It merely gives an interesting evaluation of the apparent evolution of the scoring process. An example of this evolution is the pattern for rabbits, which went up from 1.3 to around 2 in the average discomfort score. Rabbits were mainly used for a standard endocrinological assay, the McPhail test (6, 7), for determining progestagenic activity. In this test, immature rabbits are pretreated for 8 days with estrogen by daily subcutaneous (s.c.) injections, then progestin treatment starts (s.c. or orally [p.o.]) as 5 dosings over 3 days, with two blood samples being taken in this period. At some point, the Animal Ethical Committee decided that score 3 was more apt than score 2, which had the above-mentioned effect on the average score. In addition, albeit in smaller numbers, another test using rabbits was introduced, which involved ovariectomy. The aim of this test was to look at blood vessel growth in the endometrium following a 10-day treatment with progestin.

The only visible effect of refinement (in this way of presenting the discomfort scores) is in Figure 2, which shows the mouse data. As part of the rheumatoid arthritis programme, the 'delayed hypersensitivity test' (for the principle of method, see [8]) was introduced in 1997, with tetanus toxoid injection, rather than albumin injection, to create a model that was more predictive for the human situation. This method was subsequently refined by significantly reducing the tetanus toxoid dose that the animals received. This still made a good model, but resulted in much less severe swellings of the paws, in line with an observation of reduced ear swellings with lower tetanus dosages (9).

Developments in toxicology are not considered to be within the realm of this evaluation. However, similar efforts to diminish animal use are obviously ongoing in toxicology. The introduction of cell lines expressing different cytochrome P450 enzymes (10), and the banning of the LD50 test (11) are just two examples of the progress being made. It would be interesting to perform a similar study to evaluate the reduction in experimental animal use which has resulted from these developments.

Received 18.05.06; received in final form 30.10.06; accepted for publication 11.12.06.

References

1. De Greeve, P. (2000). Adverse effects: regulation and classification in Europe. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation 31B* (ed. M. Balls, A-M. van Zeller & M.E. Halder), pp. 1013–1021. Amsterdam, The Netherlands: Elsevier Science BV.
2. Anon. (2001 and 2002). *National Comment to the Statistical Report on the Use of Animals for Experimental and Other Scientific Purposes in The Netherlands* (ed. Dutch Food and Consumer Product Safety Authority/Inspectorate for Health Protection and Veterinary Public Health), 32pp. & 30pp. The Hague, The Netherlands: The Dutch Government.
3. Ashton, G.A. & Joshua, P.J. (2002). *Industry Success Rates 2002 — including Trends in Success Rates. Ref No. CMR02-175R*, 18pp. Epsom, Surrey, UK: Centre for Medicines Research International, Ltd.
4. EFPIA (2004). *The European Federation of Pharmaceutical Industries and Associations Report. The Pharmaceutical Industry in Figures, 2004 Edition; Main Trends*. Available at: http://www.efpia.org/6_publ/infigure2004b.pdf (Accessed 18.02.06).
5. Food and Drug Administration (2004). *Innovation Stagnation; Challenge and Opportunity on the Critical Path to New Medical Products*. Available at: http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (Accessed 18.02.06).
6. McPhail, M.K. (1934). The assay of progestin. *Journal of Physiology* **83**, 145–156.
7. Van der Vies, J. & De Visser, J. (1983). Endocrinological studies with desogestrel. *Drug Research* **33**, 231–236.
8. Jacobs, M.J., van den Hoek, A.E., van de Putte, L.B. & van den Berg, W.B. (1993). Aberrant tolerance induction with cationic antigens. *Scandinavian Journal of Immunology* **37**, 97–103.
9. Carrodeguas, L., Orosz, C.G., Waldman, W.J., Sedmak, D.D., Adama, P.W. & VanBuskirk, A.M. (1999). *Trans vivo* analysis of delayed-type hypersensitivity reactivity. *Human Immunology* **60**, 640–651.
10. Donato, M.T. & Castell, J.V. (2003). Strategies and molecular probes to investigate the role of cytochrome P450 in drug metabolism: focus on *in vitro* studies. *Clinical Pharmacokinetics* **42**, 153–178.
11. Botham, P.A. (2002). Acute systemic toxicity. *ILAR Journal* **43**, Suppl. S27–30.