



The ability of animal studies to detect serious post marketing adverse events is limited

Peter J.K. van Meer^{a,*}, Marlous Kooijman^b, Christine C. Gispen-de Wied^c, Ellen H.M. Moors^b, Huub Schellekens^{a,b}

^a Utrecht Institute of Pharmaceutical Sciences, Department of Pharmaceutics, Utrecht University, Universiteitsweg 99, 3584 CG, The Netherlands

^b Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Heidelberglaan 2, 3584 CS, Utrecht, The Netherlands

^c Medicines Evaluation Board, Graadt van Roggenweg 500, 3531 AH, Utrecht, The Netherlands

ARTICLE INFO

Article history:

Received 30 May 2012

Available online 12 September 2012

Keywords:

Predictive value

Animal studies

Non-clinical drug development

Serious adverse reactions

Pharmacovigilance

Regulatory science

ABSTRACT

The value of animal studies to assess drug safety is unclear because many such studies are biased and have methodological shortcomings. We studied whether post-marketing serious adverse reactions to small molecule drugs could have been detected on the basis of animal study data included in drug registration files. Of 93 serious adverse reactions related to 43 small molecule drugs, only 19% were identified in animal studies as a true positive outcome, which suggests that data from animal studies are of limited value to pharmacovigilance activities. Our study shows that drug registration files can be used to study the predictive value of animal studies and that the value of animal studies in all stages of the drug development should be investigated in a collaborative endeavour between regulatory authorities, industry, and academia.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

By law, new therapeutics have to be studied in animals before they can be tested in humans (US Congress, 1938; EU Parliament, 2004). An extensive set of guidelines, issued by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), offers practical advice on the planning and execution of these non-clinical studies, covering all aspects of drug safety. The vast experience with, and the volume of historical data from, animal studies have made laboratory animals the 'gold standard' for evaluating the safety of new drugs. However, the value of routine animal studies for drug development is increasingly debated (Hackam, 2007; Hartung, 2008; Lemon and Dunnett, 2005; Pound et al., 2004; Wall and Shani, 2008). Although the pharmaceutical industry and regulatory authorities rely on animal studies to predict the safety and efficacy of new therapeutics in humans, it is striking that few attempts have been made to demonstrate this predictive ability (Fletcher, 1978; Freireich et al., 1966; Owens, 1962; Rozenzweig et al., 1981; Schein et al., 1970).

Two large studies that assessed aspects of the predictive value of animal studies have been published in the last 20 years. In

1995, a Japanese industry consortium studied whether the pharmacological effects in animals of 104 therapeutics were associated with adverse reactions in humans. Of 43 established non-clinical pharmacological endpoints, 10 (23%) were statistically significant correlated with clinically relevant adverse reactions in humans (Igarashi et al., 1995). In 2000, Olson et al. evaluated the sensitivity of animal models to detect toxicity in humans, using 150 therapeutics. In total, the concordance between human and animal toxicity was 71% for rodent and non-rodent species, 63% for non-rodent species, and 43% for rodent species (Olson et al., 2000). Moreover, up to 43% of the toxicities identified in clinical trials were related to the pharmacological action of the drug under investigation and could have been anticipated on the basis of the drug's mechanism of action.

Most studies investigating the value of animal studies in drug development have methodological shortcomings. Datasets are often limited in size or scope. For instance, most studies have focused exclusively on anticancer therapeutics. Incorrect statistical definitions have also been used, leading to an overestimation of the utility of animal studies (Matthews, 2008). In the studies of Olson et al. and Igarashi et al., limited inclusion criteria were used which may have introduced selection bias. It is thus challenging to make an unbiased and comprehensive analysis of whether animal studies are of value in predicting short- and long-term clinical safety (Zbinden, 1991). This requires a method that minimizes bias and allows an evidence-based assessment. The study we performed is part of a larger project to assess the value of animal studies in drug

* Corresponding author. Fax: +31 30 251 7839.

E-mail addresses: p.j.k.vanmeer@uu.nl (P.J.K. van Meer), m.kooijman@uu.nl (M. Kooijman), cc.gispen@cbg-meb.nl (C.C. Gispen-de Wied), e.h.m.moors@uu.nl (E.H.M. Moors), h.schellekens@uu.nl (H. Schellekens).

development (Top Institute Pharma, 2010). We retrospectively studied whether animal studies that were part of the drug registration file of a new small molecule (SM) could have identified serious adverse reactions (SARs) which required a safety related regulatory action after market approval. Our method was set up to minimize bias in two ways in order to improve on previous designs; first, by using SARs requiring a safety related regulatory action as the starting point of our database which allowed random selection of therapeutics and secondly, by using the drug registration files to minimize publication bias.

2. Methods

For this study, chemically synthesized, drugs marketed in the European Union after 01-01-1985 and which prompted safety-related regulatory action between 01-01-1999 and 01-01-2010 were identified by searching the websites of the EMA and Medicines Evaluation Board for Product Safety Announcements or Direct Healthcare Practitioner Communications, including the Medicines Evaluation Board internal databases (European Medicines Agency; Medicines Evaluation Board; Mol et al., 2010). Safety-related regulatory actions were defined as either Product Safety Announcements or Direct Healthcare Practitioner Communications communicating a serious safety risk that necessitated changes being made to chapter 4 of the summary of product characteristics or box labelling as a result of new clinical or pharmacovigilance findings, and also market withdrawal due to safety reasons. Product Safety Announcements or Direct Healthcare Practitioner Communications for the same drug but with different warnings were pooled. SARs were identified and classified by organ class, as defined by the Medical Dictionary for Regulatory Activities. To preserve confidentiality, therapeutics were classified according to the Anatomical Therapeutic Chemical classification system (World Health Organization Centre for Drug Statistics Methodology).

Safety-related regulatory actions issued as class warnings or due to dosing interpretation error, production error, drug–drug interactions, viral resistance, contraindications due to lack of efficacy, and quality or safety-related regulatory actions derived from post-marketing animal studies were not included in this study. The non-clinical expert report in the drug registration file for the drugs investigated was obtained from the Medicines Evaluation Board. Primary and secondary pharmacodynamic data, safety pharmacology data, and single and repeat dose toxicology data were reviewed to identify *in vivo* events in any rodent or non-rodent species and at any dose or time point that could be considered to be associated with the SAR described in the Product Safety Announcement or Direct Healthcare Practitioner Communications. Other sections of the registration file, such as those dealing with carcinogenicity, reproductive and developmental toxicity, local tolerance, or special toxicology studies were also studied if these were relevant to the SAR. Non-clinical events, such as pathological, immunohistochemical, haematological, or biochemical changes, were considered true positive if they were causally identical to the corresponding clinical adverse reaction. An associated but not true positive event was one in which a pathological, immunohistochemical, haematological, or biochemical change occurred in the target organ, but which did not necessarily lead to an SAR in the animal species investigated. SARs without a corresponding non-clinical event were considered false negative. All non-clinical events identified were evaluated by the first author and were verified by three independent external experts (a non-clinical assessor, a toxicologist, and a medical doctor). The classification had to be unanimously agreed upon by the expert panel. Per drug investigated, the total number, associated, and true positive non-clinical findings were summed and grouped by the corresponding

highest level of system organ class and anatomical therapeutic class. Sensitivity was calculated as follows: $\text{sensitivity} = n_{\text{true positive}} / (n_{\text{true positive}} + n_{\text{false negative}})$.

3. Results

To identify SARs, we collected 244 Direct Healthcare Professional Communications and Product Safety Announcements issued between 01-01-1999 and 01-01-2010 from the websites of the European Medicines Association (EMA), and the Dutch Medicines Evaluation Board, including their internal databases (Mol et al., 2010). Duplicates, updates on existing issues, and press releases not relevant to this study were removed, leaving, 178 Direct Healthcare Professional Communications and Product Safety Announcements. Of these, 51 communications/announcements did not mention a regulatory action, and a further 37 did not meet our inclusion criteria. The remaining 90 communications/announcements informed healthcare practitioners of SARs associated with 49 drugs. The non-clinical expert report of 6 drugs could not be retrieved, and so the final database consisted of 43 drugs, for which 93 SARs were identified after market approval (Fig. 1). The drugs were distributed over eight anatomical therapeutic classes (Fig. 2). Of the 93 SARs, 59 (63%) did not have a non-clinical counterpart and were considered false negative, and 34 (37%) were accompanied by non-clinical events in the relevant target organ in the species, doses, and time points tested (Table 1). Most of these non-clinical events occurred at doses that were a multiple of the intended clinical dose and after prolonged exposure. In many cases, the incidence of these events was low and did not always occur in multiple species. In 18 of the 34 cases, the non-clinical events were identified as true positive events because they had the same mechanisms as the SAR. In the remaining 16 cases, the

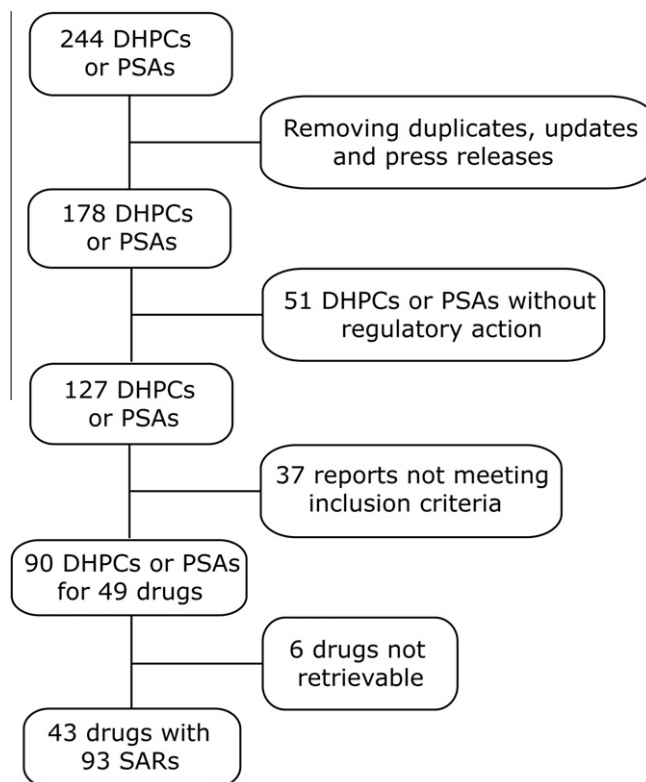


Fig. 1. Distribution of serious adverse reactions and small molecules over anatomical therapeutic chemical class. SAR, serious adverse reaction; SM, small molecule; $n_{\text{SAR}} = 93$; $n_{\text{SM}} = 43$.

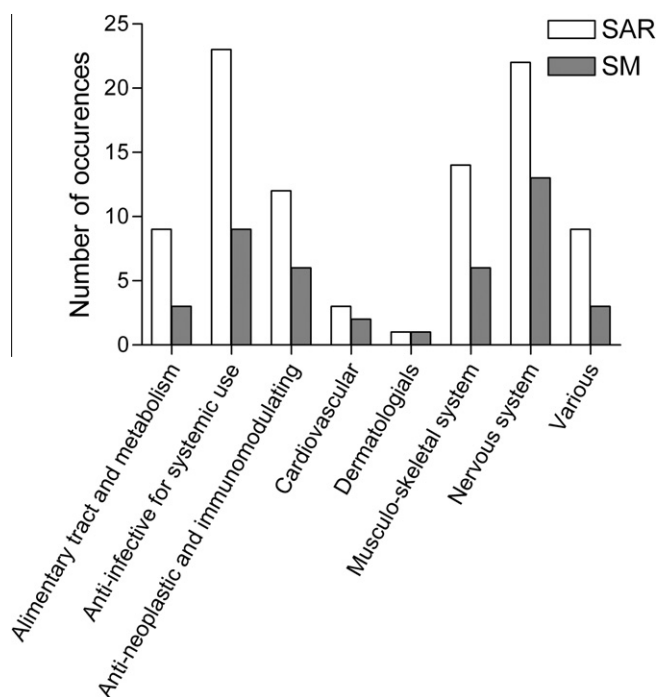


Fig. 2. Flowchart of the data collection leading to the selection of 43 drugs with 93 serious adverse events requiring a safety related regulatory action. DHPC, Direct Healthcare Practitioner Communication; PSA, Product Safety Announcement; SAR, serious adverse reaction.

Table 1

Detection rate and distribution of associated and predictive non-clinical events for serious adverse events by anatomical therapeutic class.

SM	SAR	Detected non-clinical event (%)	Associated non-clinical event (%)	True positive non-clinical event (%)
43	93	34 (37)	16 (17)	18 (19)

SAR, serious adverse reaction; SM, small molecule.

non-clinical events affected the relevant target organ but did not give rise to the reported SAR and were considered associated events (see box text for examples). Accordingly, the sensitivity of the animal studies for detecting SARs in humans was 19%.

4. Discussion

Our study was designed to minimize bias. We had a unique opportunity to access drug registration files, which contain all the experimental data generated by the company. Because animal studies described in the non-clinical section of the drug registration file are designed to give a complete overview of the safety and efficacy of a new therapeutic, this section contains both positive and negative data, thereby limiting publication bias. To minimize selection bias, we focused on adverse reactions serious enough to require regulatory action after marketing. Since Direct Healthcare Professional Communications were issued across almost all therapeutic classes (Mol et al., 2010), the drugs included in this study could be considered a random selection.

The definition of what represents a predictive outcome, and what does not, is an essential aspect of these kinds of retrospective studies. Olson et al. defined a true positive non-clinical event as one in which '...the same target organ was involved in humans and in animals in the judgment of the company clinicians and the toxicologists' (Olson et al., 2000). While identifying toxicity at the target

organ level in animals may be useful for evaluating the safety of a drug from a development perspective, it is inadequate when attempting to establish the predictive value, because toxicity in the target organ may give rise to several specific side effects in humans. Because we think that a stricter definition of true positive results is needed, we distinguished between target organ involvement and non-clinical events that were either identical to the SAR or causal to it. For this reason, non-clinical events which were related to the target organ but which did not give rise to a SAR by similar mechanisms were not considered true positive.

The study had some limitations. We restricted our analysis to true positive and false negative events identified in animal studies because we only studied those drugs that had received marketing approval. We did not have access to the safety data of drugs whose development was terminated in the non-clinical or clinical trial phase. As a result, we could only address the left portion of a 2×2 contingency table, meaning that the sensitivity of animal studies was the only assessable parameter. We used the non-clinical expert report included in the drug registration file. This report, written by an independent expert, summarizes the results of all animal studies that were conducted to support the drug application for a given indication. It is possible that the expert did not discuss data from individual study reports that might have identified a potential clinical adverse reaction. Moreover, several therapeutics did not meet the inclusion criteria, which limited the size of the dataset. We did not include SARs arising from healthcare practitioner errors, such as administration of the therapeutic through alternative, non-indicated, routes. We also did not include adverse events due to drug-drug interactions because interaction studies in animals are rarely conducted and if so, only for fixed combinations. Finally, our dataset was limited to small molecule drugs. It did not include biotech products and we do not yet know if there are differences in the frequency of detection of post-marketing SARs between these distinct classes of therapeutics.

One could argue that non-clinical studies are not designed to identify rare adverse reactions that appear after market approval. Although the number of animals used in non-clinical studies is relatively small, the studies are designed to find important side effects that are likely to occur in humans (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2009). However, as high doses are administered for a prolonged period to elicit a complete toxicological response in animals, this approach can also estimate potential toxicities that could occur in humans. In 18 cases, true positive events in animal studies correctly predicted post-marketing adverse reactions. Nevertheless, the low incidence rate, high doses, prolonged exposure, and species specificity were important reasons to assume that these events were unlikely to occur in the clinical trial population. In addition, the number of adverse reactions in animals that have no corollary in humans (false positives) increase with increasing dose, suggesting that over-exposure might not produce meaningful results (Igarashi et al., 1995). Overall, 63% of all SARs had no animal counterpart, not even at a target organ level, and less than 20% of SARs had a true positive corollary in animal studies. However, although animal data is not sensitive enough to detect post-marketing adverse events, they will remain useful to identify safe starting doses and to identify pharmacological effects that can be monitored during clinical trials. Data from non-clinical studies are nowadays added to risk management plans, which are developed by pharmaceutical companies to monitor the safety of the drug in the marketplace. Few drugs in this study required the submission of a risk management plan because it was not the policy to do so at the time of marketing approval (Committee for Medicinal Products for Human Use (CHMP), 2005). For most drugs, no new adverse effects will be identified after marketing authorization that requires a safety related regulatory action. Yet, our study suggests

that for the majority of drugs animal data in these plans will be redundant because animal data do not appear to be an optimal tool for prospectively assessing risk in humans.

5. Conclusion

We showed that the animal studies performed to evaluate the safety of new small molecule drugs are not sensitive enough to predict post-marketing SARs. Therefore, it is not relevant to include animal study data for prospective pharmacovigilance studies. While we only analysed a small set of data, the method used can be adapted to include all therapeutics on the market as well as those that are still in development. Such a study will enable a full assessment of the predictive value of animal studies in drug development.

The adoption of the precautionary principle by the regulatory authorities and the relative ease with which this burden of proof is accepted by the pharmaceutical industry – without attempts to improve the current paradigm – has created a stalemate in which animal studies, predictive or not, continue to exist with little room for innovation. Stakeholders in industry, academia, and regulatory agencies, need to critically assess animal studies and discuss their predictive value in all earnestness and with the scientific facts at hand. From this, possibilities based on scientific facts may develop which allow new technologies to be implemented that predict the safety and efficacy of therapeutics equal to or better than animal studies do. A way forward would be for the pharmaceutical industry to share clinical and laboratory data generated at all stages of product development with collaborating stakeholders, to enable a complete and transparent analysis of the predictive value of animal studies for drug development.

True positive non-clinical event: Peripheral neuropathy was a serious adverse reaction that led to a safety-related regulatory action. A 9-month repeat dose toxicity test in non-human primates showed minimal to mild degeneration of the sciatic nerve, consistent with axonopathy, in two female non-human primates receiving high doses of the study drug after 3 months of treatment and in one female non-human primate after 9 months of treatment. Axonopathy was more prevalent in the sciatic nerves of female non-human primates and in the spinal cord of male non-human primates than in controls, with nerve fibres showing signs of Wallerian degeneration.

Corneal perforation or ulceration was an adverse effect which occurred in less than 1:10,000 patients and was cause for a change in the SPC. In non-clinical safety studies, corneal atrophy and ulceration has been observed in high dose group beagle dogs.

Gastrointestinal perforation and bleeding was reason for a safety related regulatory action leading to a change of the SPC. Non-clinical studies showed that rats in the high dose group developed red lesions in the stomach and haemorrhage or inflammation of the stomach with gastric erosion. High dose treated rats also showed increased inflammation and haemorrhage of the intestine.

Rare cases of nephrogenic systemic fibrosis have been reported which resulted in a change in the SPC. In non-clinical studies, dose related renal cortical tubular cell vacuolation was observed in rats. Kidney weight of drug treated rats was increased compared to control treated animals although this was not considered to be drug related. Similar increased kidney weight increase was observed in monkey. Urothelial hyperplasia was observed in all drug treated animals. In atoxicity study in rats, one animal in the high dose treated group and two animals in the mid dose treated group showed

increased tubular mitotic rate. Degeneration with regeneration in the distal convoluted tubules and collecting ducts was noted in one rat from each dose group in a repeated dose toxicity study. Skin lesions were observed in a sub chronic repeated dose toxicity study in rats. The lesions were scabs or thickening of the skin with patchy hair loss. Similar skin reactions were observed in dog. Foveolar hyperplasia and focal increased interstitial connective tissue in glandular stomach were noted in two males. Mineralized areas in the superficial dermis were noted in 2 high dose treated rats.

False negative association: False negative associations have no non-clinical counterpart. For instance, a therapeutic received a safety-related regulatory action because of increased risk of myocardial infarction. In the corresponding safety pharmacology section of the non-clinical expert report, no discernable cardiovascular effects were observed in animal experiments or during repeated dose toxicity studies in multiple species.

Associated non-clinical event: A safety-related regulatory action was issued after reports of pure red cell aplasia. Animal studies showed that the haematopoietic and/or lymphoid systems were target organs in mice, rats, dogs, and monkeys dosed orally for up to 12 months. Haematopoietic toxicity in mice and rats was evidenced as decreased erythrocyte parameters. Anaemia occurred in both species. In mice increased granulocytic cells and megakaryocytes in bone marrow were also observed. Neonatal rats receiving the highest dose had reduced red blood cell parameters, reduced bone marrow cellularity, and increased splenic extramedullary haematopoiesis. In dogs and monkeys, haematological side effects were primarily decreased lymphocyte counts. However, while it was clear that the haematological system was the target organ and anaemia was a likely side effect, pure red cell anaemia could not be exclusively identified as an adverse effect in animals. Therefore, these non-clinical events were not considered true positive for this specific serious adverse reaction.

Increased risk for depression, including suicidal ideation and increased aggression was reported in addition to or in combination with increased incidence of insomnia which resulted in a safety related regulatory action. Interestingly, behavioural studies showed that drug treated mice showed anti-depressant like behaviour in the Porsolt forced swim test. In addition, drug treated mice showed anxiolytic behaviour in the elevated plus maze. Because the drug clearly had an effect on behaviour but was converse to that observed in humans, the non-clinical events were considered target organ related effects.

A safety related regulatory action was taken after reports of sudden onset of sleep which was associated with the use of a drug. In non-clinical studies, increased yawning was seen in drug treated rats. In both cats and rats REM sleep depression was observed in drug treated animals. Clearly, these observed effects indicate that the central nervous system, and in particular sleep, is affected but sudden onset of sleep was not observed. The effects were considered target organ related, but not predictive.

Conflict of interest

This work was supported by Top Institute Pharma [Project T6-301]. The views expressed in this article are the personal views of the authors and are not to be understood or quoted as being made on behalf of or reflecting the position of the Medicines Evaluation Board or any other regulatory agency, or one of its committees or working parties. PvM, MK, CGdW and EM declare no

conflict of interest. HS participated in meetings and publications sponsored by Amgen, Johnson & Johnson, Roche, Sandoz and Hospira. Part of his research is directly or indirectly sponsored by Roche and Amgen.

Acknowledgments

The authors wish to thank Gerard Mulder and Jan-Willem van der Laan for their expert contributions in the evaluation of non-clinical data. We are also grateful to Jane Sykes for her editorial support.

References

- Committee for Medicinal Products for Human Use (CHMP), 2005. Guideline on Risk Management Systems for Medicinal Products for Human Use, London.
- European Medicines Agency (online). Available from URL: <<http://www.ema.europa.eu>> (accessed 30.07.2012).
- EU Parliament, 2004. Directive 2004/27/EC of the European Parliament and of the Council of 31 March 2004 Amending Directive 2001/83/EC on the Community Code Relating to Medicinal Products for Human Use.
- Fletcher, A.P., 1978. Drug safety tests and subsequent clinical experience. *J. R. Soc. Med.* 71, 693–696.
- Freireich, E.J. et al., 1966. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother. Rep.* 50, 219–244.
- Hackam, D.G., 2007. Translating animal research into clinical benefit. *BMJ* 334, 163–164.
- Hartung, T., 2008. Thoughts on limitations of animal models. *Parkinsonism Relat. Disord.* 14 (Suppl. 2), S81–S83.
- Igarashi, T. et al., 1995. Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J. Toxicol. Sci.* 20, 77–92.
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2009. Guidance on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals M3(R2), EMA, London.
- Lemon, R., Dunnett, S.B., 2005. Surveying the literature from animal experiments. *BMJ* 330, 977–978.
- Matthews, R.A., 2008. Medical progress depends on animal models – doesn't it? *J. R. Soc. Med.* 101, 95–98.
- Medicines Evaluation Board (online). Available from URL: <<http://www.cbg-meb.nl>> (accessed 30.07.2012).
- Mol, P.G. et al., 2010. A decade of safety-related regulatory action in the Netherlands: a retrospective analysis of direct healthcare professional communications from 1999 to 2009. *Drug Saf.* 33, 463–474.
- Olson, H. et al., 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* 32, 56–67.
- Owens Jr., A.H., 1962. Predicting anticancer drug effects in man from laboratory animal studies. *J. Chronic Dis.* 15, 223–228.
- Pound, P. et al., 2004. Where is the evidence that animal research benefits humans? *BMJ* 328, 514–517.
- Rozenzweig, M. et al., 1981. Animal toxicology for early clinical trials with anticancer agents. *Cancer Clin. Trials* 4, 21–28.
- Schein, P.S. et al., 1970. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin. Pharmacol. Ther.* 11, 3–40.
- Top Institute Pharma, 2010. The Use of Animals in Pharmaceutical Development, vol. 2011. Top Institute Pharma.
- US Congress, 1938. Federal Food, Drug and Cosmetics Act.
- Wall, R.J., Shani, M., 2008. Are animal models as good as we think? *Theriogenology* 69, 2–9.
- World Health Organization Centre for Drug Statistics Methodology (online). Available from URL: <http://www.whocc.no/atc_ddd_index/> (accessed 30.07.2012).
- Zbinden, G., 1991. Predictive value of animal studies in toxicology. *Regul. Toxicol. Pharmacol.* 14, 167–177.