THE SCIENTIFIC VALUE OF NON-CLINICAL ANIMAL STUDIES IN DRUG DEVELOPMENT

Peter Jan Karel van Meer

The studies presented in this thesis were performed under the framework of Top Institute Pharma, (project T6-301) which includes the Dutch Medicines Evaluation Board (CBG-MEB), The Organization of innovative Pharmaceutical Industries in Netherlands (Nefarma), the Life Science and health Initiative, and Utrecht University.

Printing of this thesis was kindly supported by:

Amgen

College ter Beoordeling van Geneesmiddelen

Het Nederlands Bijwerkingenfonds

Nefarma

Utrecht Institute for Pharmaceutical Sciences

Stichting Proefdiervrij

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 CBG or any other regulatory agency, or one of its committees or working parties.

van Meer P.J.K.

The scientific value of non-clinical animal studies in drug development

Thesis Utrecht University -with references- with a summary in Dutch

ISBN: 978-94-6182-359-5

Lay-out and printing: Off Page, www.offpage.nl

Cover concept and photography: P.M. van Meer and P.J.K. van Meer

Photograph: Magic 8-ball, the breakfast of champions

Cover: Off Page, www.offpage.nl

Chapter images are scans from the original Rider-Waite-Smith deck published in the US in 1909, and no longer under copyright. This is not the modified and recoloured US Games version published in 1971 (which is still under copyright).

THE SCIENTIFIC VALUE OF NON-CLINICAL ANIMAL STUDIES IN DRUG DEVELOPMENT

De wetenschappelijke waarde van niet klinische studies in geneesmiddelen ontwikkeling (met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan
de Universiteit Utrecht op gezag van de rector
magnificus, prof. dr. G.J. van der Zwaan ingevolge
het besluit van het college voor promoties in
het openbaar te verdedigen op woensdag
23 oktober des middags te 12.45 uur

door

Peter Jan Karel van Meer

geboren op 17 maart 1979 te Tilburg
te Tilburg

Promotoren:	Prof. Dr. H Schellekens
	Prof. Dr. B. Silva-Lima
	Prof. Dr. E.H.M. Moors
Co-promotoren:	Dr. C.C. Gispen-de Wied

TABLE OF CONTENTS

There was only one catch…Catch-22, which specified that a concern for one's own safety…was the process of a rational mind. Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions. Orr would be crazy to fly more missions and sane if he didn't, but if he was sane he had to fly them. If he flew them he was crazy and didn't have to; but if he didn't want to he was sane and had to.

-Joseph Heller, 1923-1999, *Catch-22*

1 GENERAL INTRODUCTION

INTRODUCTION

On the 25th of December in 1956 a child was born in Germany with flipper-like vestigial arms, a deformity known as phocomelia. By itself, it was considered to be a very rare incident. However, within a few years of this seemingly isolated case, thousands of children had been born with phocomelia. A request of one Australian doctor in the Lancet to report similar findings that he thought could be related to treatment with Thalidomide (marketed in the Netherlands as Softenon), quickly revealed this drug as the source of what is now known as the Thalidomide tragedy(*1*). No deformities were reported in animal studies common for that time with thalidomide(*2*). It was widely marketed and prescribed as a sedative (although, surprisingly, animals do not exhibit this effect) and to ameliorate nauseas, such as morning sickness in pregnant women, setting the stage for the disaster to come. The remarkable history of this drug –it was developed as an antidote to nerve gas during the second World War- would grow as it also became the spark that would lead to extensively regulated animal studies needed to demonstrate the safety and efficacy of a drug. This is because later efforts to determine the mechanism of phocomelia, showed that it only occurs in few species, but not rodent, and is not toxic in rodent at doses up to 4000 mg/kg(*3*). However, in the White New Zealand Rabbit, Thalidomide induced phocomelia, albeit at exposures far above the intended clinical dose(*4*). Nevertheless, it was considered that animal studies were needed as predictive models to prevent such adverse effects and future disasters, which was passed into United States of America (USA) legislation by the Kefauver Harris Amendment to the Federal Drug and Cosmetics (FDC) act in 1962(*5*).

DRUG DEVELOPMENT LEGISLATION AND GUIDELINES IN THE EUROPEAN UNION

By and large, we still rely on the same animal studies that were introduced by this legislation to evaluate safety and efficacy of pharmaceuticals today. All animal studies that are done to support the development of pharmaceuticals are governed by international and national legislation issued by countries or communities and international, pan-European and national guidelines issued by various organizations and regulatory authorities. In the European Union (EU) the rationale and requirements for animal testing in the development of pharmaceuticals is set out in Directive 2001/83/EC Annex I and protection of animals used for experimental and scientific purposes is governed by Directive 2010/63/EU(*6, 7*). Practically the entire process of safety testing of pharmaceuticals in animals has been extensively documented in international guidelines published by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), which is supported and supplemented in the EU by additional guidelines issued by the European Medicines Agency via the Committee for Medicinal Products for Human Use (CHMP). These guidelines have evolved over time to meet specific previously unaddressed challenges and reflect the state-of-the-art at that time (all ICH guidelines can be found at http://www.ich.org) and are used to guide the drug development and approval process(*8*). However, the scientific basis of the current paradigm on the predictability of animal studies for the effects

of drugs in man always has been and still is under discussion(*9-14*). Even for Thalidomide, the drug that initiated the extensive mandatory testing five decades ago, predictivity of both pharmacodynamic effects and toxicity is limited in numerous species(*3*).

LIMITATIONS OF ANIMAL STUDIES

Much of the uncertainty about the relevance for humans of animal studies stems from three factors: 1) sensitivity, 2) reproducibility and 3) predictivity. For example, researchers from Amgen attempted to repeat cancer treatment studies using mice. They succeeded in only six out of 53 cases, after repeated attempts(*15*). Similar results were obtained by researchers from Bayer(*16*). All these factors are inherently tied to the central drawbacks of animal studies, which is that it is a living model system that is being used to extrapolate results from. Animal model systems have high intrinsic variability, immutable differences within species and are also mechanistically not completely understood(*17*). Study design can influence all three factors(*18-20*). Finally, publication bias sustains the belief that animal studies are always relevant(*21-23*).

There is paucity in comprehensive studies on the predictive value of animal studies in the development of pharmaceuticals(*24-29*). The most recent study was published more than ten years ago and only included toxicology data(*30*). In addition, this pharmaceutical industry survey may have introduced bias by selecting only those compounds with toxic effects in animals. But even though the database of selected compounds this study was biased towards predictability, the results were disappointing. In retrospect, the concordance of human toxicity that was observed in any animal species was 71%. Rodent studies using rats or mice predicted only 43% of human toxicity, and non-rodents such as dogs only 63%. In fact, for studies using mice, a coin flip would predict toxicity with a similar success rate. The overall false negative rate for both rodents and non-rodents was about 30%.

In today's pharmaceutical landscape, predictive value of animal studies is not likely to be any higher and may even be lower. The advent of recombinant technologies in the early 1980s has introduced protein drugs (biotherapeutics) to the arsenal of available therapeutics(*31*). These highly complex proteins are also very species specific and are usually immunogenic in animals(*32-34*). Therefore, animal data for these types of products is also not likely to yield informative data. Animal studies with low predictive value might hamper the development of innovative pharmaceutical drugs because during development drugs may be wrongly identified as being too toxic and never reach the patient. A false sense of safety can be given when drugs were considered non-toxic in animal studies and serious adverse effects can occur after a drug has been granted a marketing authorization. For instance, in 2004 Vioxx (Rofecoxib), a COX-2 inhibitor, was withdrawn from the market after analysis of clinical trial data revealed that there was an increased risk of adverse cardiovascular effects(*35*). During non-clinical studies with Rofecoxib, such adverse effects were not observed(*36*). A role for COX-2 inhibition in decreasing the incidence of acute thromboembolic events and protection after myocardial infarction was suggested based on limited animal and human data(*37, 38*). However, animal studies with Rofecoxib have also shown damaging effects on the cardiovascular system(*39*).

The majority of animal studies that are conducted during drug development are the result of national and/or international legislation. Guidelines are, by definition, not binding and offer the possibility to adapt the program for non-clinical studies depending on the product and available model. However, in practice this hardly ever happens. Companies have little incentive to reduce the number of animal studies, because any deficit in these studies noted by one of the world's regulatory agencies leads to major delays in the registration process. In combination with the adoption of the precautionary principle by the regulatory authorities who also have no incentive to reduce the extent of animal data they request, this has created a stalemate in which animal studies, predictive or not, continue to exist with little room for innovative technologies to be adopted that are equally able to evaluate risk.

Estimates for the cost to develop a drug range from 161 to 1800 million dollars US (capitalized cost estimate, 2009 $)(*40*). The bulk of this cost lies in (large, multi-center) clinical trials(*41*). Nevertheless, a considerable part of the R&D budget is being spent on animal studies. These studies are expensive and time-consuming and if these do not have additional value because they poorly predict human outcomes, lead to destruction of capital and to unnecessary prolonged development times. The public and political discussion about the ethics and need of studies in animals is ongoing. The discussion is mainly based on perceptions and emotions and not facts but can have far reaching consequences(*42-44*). The evidence to support animal research is mostly supported by claims of necessity and is illustrated by cases where medical progress was advanced by animal studies(*45*). Similarly, opponents to animal experimentation will often present only anecdotal evidence or scarce data showing lack of predictive value of animal studies but do not offer acceptable alternatives(*46*). Thus, convincing data concerning the predictive value of animal studies in the development of pharmaceutics are lacking. The main reason for this is that existing databases with relevant non-clinical animal data and clinical data are confidential(*47-49*). These can be pharmaceutical industry databases or part of the marketing authorization submissions on file in regulatory agencies, neither of which are easily available for research.

AIM OF THE THESIS

Thus, there is a need for an independent and objective study of the predictive value of non-clinical animal studies based on scientific facts. To this end, Utrecht University has engaged in a public-private partnership with the pharmaceutical industry (represented by Nefarma) and the Medicines Evaluation Board (College ter Beoordeling van Geneesmiddelen, CBG-MEB) via Top Institute Pharma. This collaboration is unique, since it specifically allows the study of confidential databases, which include documents such as the marketing authorization application (MAA). This thesis is a result of this collaboration and offers a scientific analysis of the value of non-clinical animal studies in drug development, with a particular focus on biopharmaceuticals because these may offer immediate opportunities to reduce the need for animal studies based on their unique features. We present a structured method to study the relevance of animal studies based on the MAA. The scope of this thesis is limited to legislation and guidelines relevant to the European Union and to pharmaceuticals that are marketed in the EU. In this thesis, we evaluate the scientific basis of the current practices

and guidelines for the use of animal studies in pharmaceutical development and assess the consequences and implication for the regulatory guidelines when animal studies are not informative.

THESIS OUTLINE

Marketing authorization applications can be a valuable source to study the predictive value of animal studies because they contain all the experiments done to support registration of a drug. In Chapter 2 we demonstrate the value of these marketing authorization applications by using them to study whether new post marketing adverse effects of small molecule therapeutics could have been detected from non-clinical studies. The study is expanded to include biopharmaceuticals in Chapter 3. Biopharmaceutical drug development increasingly uses non-human primate as the primary species because they are often the only species available that are pharmacologically responsive. In Chapter 4 we evaluate the value of the non-human primate in the development of monoclonal antibodies. For biopharmaceuticals, immunogenicity has been identified as a factor that can influence interpretability. In Chapter 5 the immunogenicity of monoclonal antibodies in non-human primates is studied and the incidence and effects of anti-drug antibodies on non-clinical development are discussed.

Non-clinical studies are not only required for the development of innovative new drugs. Generic copies of biopharmaceuticals (biosimilars) are routinely assessed in animals to demonstrate that the copy is similar to its reference product. In Chapter 6 the value of non-clinical animal studies in the biosimilarity exercise are assessed in light of the new draft guideline on biosimilars. In addition to the first generation of biosimilars on the market today, monoclonal antibody biosimilars are set to enter the market. For these products, non-clinical animal studies may be even less relevant and in Chapter 7 we present alternatives to developing a monoclonal biosimilar using Bevacizumab (Avastin) as a case study. The findings of each chapter are put into perspective in a general discussion in Chapter 8 There, we identify redundancies and inefficiencies in the animal tests that are currently mandatory for the registration of new pharmaceuticals and make recommendations on improving the regulatory system for both the regulatory authorities and the pharmaceutical industry.

REFERENCES

1. W. G. McBride, Thalidomide and congenital abnormalities. *Lancet* 2, (1961).
2. Grunenthal Group, Questions and answers on Contergan and thalidomide. (2013).
3. R. Greek, N. Shanks, M. J. Rice, The history and implications of testing Thalidomide on animals. *The Journal of Philosophy, Science & Law* 11, 1 (2011).
4. S. Fabro, R. L. Smith, The teratogenic activity of thalidomide in the rabbit. *J Pathol Bacteriol* 91, 511 (Apr, 1966).
5. Anonymous, Federal food, drug and cosmetics act. (1938).
6. Anonymous, Directive 2001/83/EC of the European Parliament and the Council of 6 November 2001 on the Community code relating to medicinal products for human use. (2001).
7. Anonymous, Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. (2010).
8. G. Walsh, Drug approval in Europe. The EMEA gets good grades, but has room for improvement. *Nat Biotechnol* 17, 237 (Mar, 1999).
9. R. A. Matthews, Medical progress depends on animal models - doesn't it? *J R Soc Med* 101, 95 (Feb, 2008).
10. P. Pound, S. Ebrahim, P. Sandercock, M. B. Bracken, I. Roberts, Where is the evidence that animal research benefits humans? *Bmj* 328, 514 (Feb 28, 2004).
11. T. Hartung, Thoughts on limitations of animal models. *Parkinsonism Relat Disord* 14 Suppl 2, S81 (2008).
12. R. J. Wall, M. Shani, Are animal models as good as we think? *Theriogenology* 69, 2 (Jan 1, 2008).
13. D. G. Hackam, Translating animal research into clinical benefit. *Bmj* 334, 163 (Jan 27, 2007).
14. R. Lemon, S. B. Dunnett, Surveying the literature from animal experiments. *Bmj* 330, 977 (Apr 30, 2005).
15. C. G. Begley, L. M. Ellis, Drug development: Raise standards for preclinical cancer research. *Nature* 483, 531 (Mar 29, 2012).
16. F. Prinz, T. Schlange, K. Asadullah, Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10, 712 (Sep, 2011).
17. R. Greek, J. Greek, Evolution and animal models. *Comp Med* 52, 499 (Dec, 2002).
18. V. Bebarta, D. Luyten, K. Heard, Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad Emerg Med* 10, 684 (Jun, 2003).
19. P. Perel *et al.*, Comparison of treatment effects between animal experiments and clinical trials: systematic review. *Bmj* 334, 197 (Jan 27, 2007).
20. M. B. Bracken, Why are so many epidemiology associations inflated or wrong? Does poorly conducted animal research suggest implausible hypotheses? *Ann Epidemiol* 19, 220 (Mar, 2009).
21. E. S. Sena, H. B. van der Worp, P. M. Bath, D. W. Howells, M. R. Macleod, Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 8, e1000344 (Mar, 2010).
22. G. ter Riet *et al.*, Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. *PLoS One* 7, e43404 (2012).
23. M. Macleod, Why animal research needs to improve. *Nature* 477, 511 (Sep 29, 2011).
24. A. H. Owens, Jr., Predicting anticancer drug effects in man from laboratory animal studies. *J Chronic Dis* 15, 223 (Mar, 1962).
25. E. J. Freireich, E. A. Gehan, D. P. Rall, L. H. Schmidt, H. E. Skipper, Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 50, 219 (May, 1966).
26. P. S. Schein *et al.*, The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 11, 3 (Jan-Feb, 1970).
27. A. P. Fletcher, Drug safety tests and subsequent clinical experience. *J R Soc Med* 71, 693 (Sep, 1978).
28. M. Rozencweig *et al.*, Animal toxicology for early clinical trials with anticancer agents. *Cancer Clin Trials* 4, 21 (1981).
29. T. Igarashi, S. Nakane, T. Kitagawa, Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J Toxicol Sci* 20, 77 (May, 1995).

30. H. Olson *et al.*, Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56 (Aug, 2000).
31. G. Walsh, Pharmaceutical biotechnology products approved within the European Union. *Eur J Pharm Biopharm* 55, 3 (Jan, 2003).
32. P. J. Bugelski, G. Treacy, Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 6, 10 (Feb, 2004).
33. J. L. Bussiere, Species selection considerations for preclinical toxicology studies for biotherapeutics. *Expert Opin Drug Metab Toxicol* 4, 871 (Jul, 2008).
34. H. Schellekens, The immunogenicity of biopharmaceuticals. *Neurology* 61, S11 (Nov 11, 2003).
35. E. J. Topol, Failing the public health--rofecoxib, Merck, and the FDA. *N Engl J Med* 351, 1707 (Oct 21, 2004).
36. H. K. Heim, K. Broich, Selective COX-2 inhibitors and risk of thromboembolic events - regulatory aspects. *Thromb Haemost* 96, 423 (Oct, 2006).
37. M. R. Weir, R. S. Sperling, A. Reicin, B. J. Gertz, Selective COX-2 inhibition and cardiovascular effects: a review of the rofecoxib development program. *Am Heart J* 146, 591 (Oct, 2003).
38. M. C. LaPointe, M. Mendez, A. Leung, Z. Tao, X. P. Yang, Inhibition of cyclooxygenase-2 improves cardiac function after myocardial infarction in the mouse. *Am J Physiol Heart Circ Physiol* 286, H1416 (Apr, 2004).
39. Y. Cheng *et al.*, Role of prostacyclin in the cardiovascular response to thromboxane A2. *Science* 296, 539 (Apr 19, 2002).
40. S. Morgan, P. Grootendorst, J. Lexchin, C. Cunningham, D. Greyson, The cost of drug development: a systematic review. *Health Policy* 100, 4 (Apr, 2011).
41. J. A. DiMasi, H. G. Grabowski, The cost of biopharmaceutical R&D: is biotech different? *Managerial and Decision Economics* 28, 469 (2007).
42. PeTA. (04-15-2010).
43. M. Wadman, Animal rights: Chimpanzee research on trial. *Nature* 474, 268 (Jun 16, 2011).
44. P. M. Conn, J. V. Parker, The animal research war. *Faseb J* 22, 1294 (May, 2008).
45. E. F. Paul, J. Paul, *Why animal experimentation matters: The use of animals in medical research*. (Transaction Pub, 2001), vol. 2.
46. N. Shanks, R. Greek, J. Greek, Are animal models predictive for humans? *Philos Ethics Humanit Med* 4, 2 (2009).
47. P. C. Gotzsche, A. W. Jorgensen, Opening up data at the European Medicines Agency. *Bmj* 342, d2686 (2011).
48. G. Sinha, Trade secrets in balance as agencies issue new transparency rules. *Nat Biotech* 29, 98 (2011).
49. A. S. Kesselheim, M. M. Mello, Confidentiality laws and secrecy in medical research: improving public access to data on drug safety. *Health Affairs* 26, 483 (2007).

How fortunate we didn't have these animal tests in the 1940s, for penicillin would probably never been granted a license, and possibly the whole field of antibiotics might never have been realized.

-Alexander Fleming, 1881-1955

2

THE ABILITY OF ANIMAL STUDIES TO DETECT SERIOUS POST MARKETING ADVERSE EVENTS IS LIMITED

Peter J.K. van Meer, M. Kooijman,
Christine C. Gispen-de Wied,
Ellen H.M. Moors, Huub Schellekens

Published in *Regul Toxicol Pharmacol* 64, 345 (Dec, 2012)

ABSTRACT

The value of animal studies to the assessment of 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.

Keywords: predictive value, animal studies, non-clinical drug development, serious adverse reactions, pharmacovigilance, regulatory science

INTRODUCTION

By law, new therapeutics have to be studied in animals before they can be trialled in humans(*1, 2*). 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(*3-7*). 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(*8-12*).

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(*13*). 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(*14*). 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(*15*). 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(*16*). 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 development(*17*). 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.

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(*18-20*). 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(*21*).

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: sensitivity = $n_{\text{true positive}} /(n_{\text{true positive}} + n_{\text{false negative}})$.

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(*20*). 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 (Figure 1).

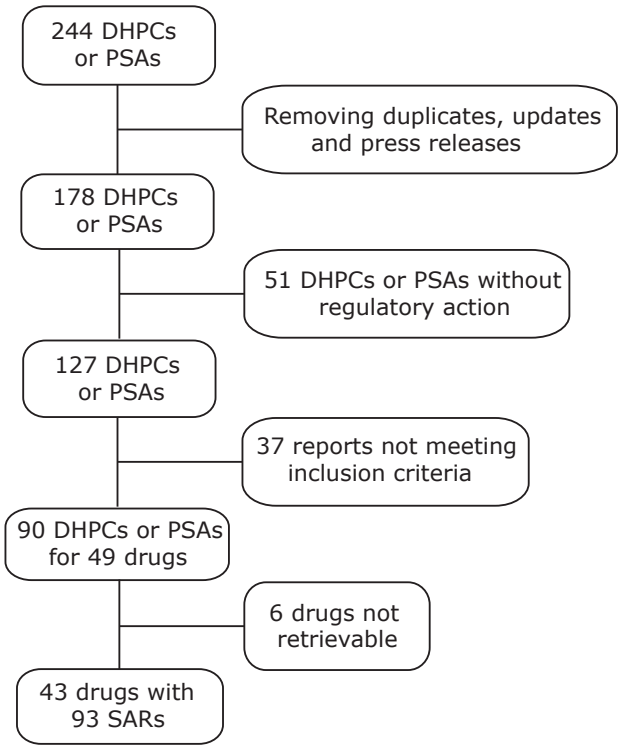


Figure 1. Flowchart of the data collection leading to the selection of 43 drugs with 93 serious adverse events requiring a safety related regulatory action. DHPC: Direct Healthcare Practitioner Communication; PSA: Product Safety Announcement; SAR: Serious Adverse Reaction.

The drugs were distributed over eight anatomical therapeutic classes (Figure 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

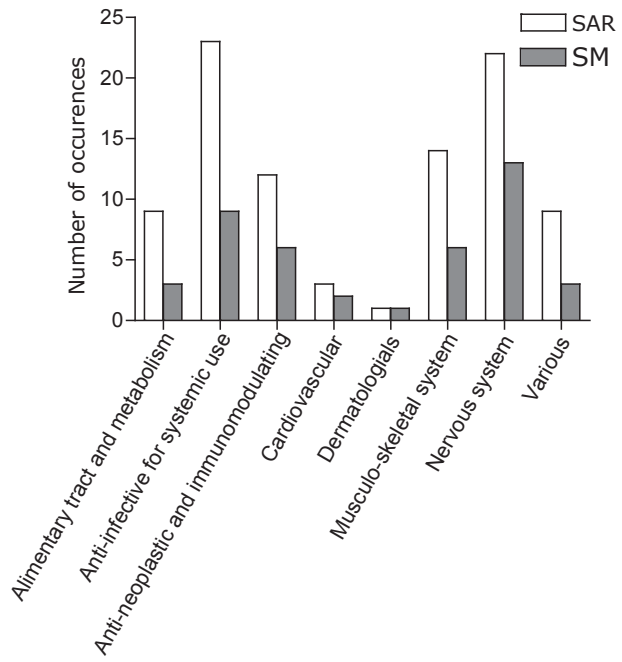


Figure 2. Distribution of serious adverse reactions and small molecules over anatomical therapeutic chemical class. SAR = Serious Adverse Reaction; SM = Small Molecule.

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

events because they had the same mechanisms as the SAR. In the remaining 16 cases, the 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%.

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 (*20*), 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'*(*14*). 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 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 2x2 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(*22*). 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(*13*). 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-

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 effet which occurred in less than 1:10.000 patients but 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 a toxicity 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.

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

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, behavioral 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.

policy to do so at the time of marketing approval(*23*). 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.

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.

ACKNOWLEDGEMENTS

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.

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.

REFERENCES

1. Anonymous, 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.(2004).
2. Anonymous, Federal food, drug and cosmetics act.(1938).
3. P. Pound, S. Ebrahim, P. Sandercock, M. B. Bracken, I. Roberts, Where is the evidence that animal research benefits humans? *Bmj* 328, 514(Feb 28, 2004).
4. T. Hartung, Thoughts on limitations of animal models. *Parkinsonism Relat Disord* 14 Suppl 2, S81(2008).
5. R. J. Wall, M. Shani, Are animal models as good as we think? *Theriogenology* 69, 2(Jan 1, 2008).
6. D. G. Hackam, Translating animal research into clinical benefit. *Bmj* 334, 163(Jan 27, 2007).
7. R. Lemon, S. B. Dunnett, Surveying the literature from animal experiments. *Bmj* 330, 977(Apr 30, 2005).
8. A. H. Owens, Jr., Predicting anticancer drug effects in man from laboratory animal studies. *J Chronic Dis* 15, 223(Mar, 1962).
9. E. J. Freireich, E. A. Gehan, D. P. Rall, L. H. Schmidt, H. E. Skipper, Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 50, 219(May, 1966).
10. P. S. Schein *et al.*, The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 11, 3(Jan-Feb, 1970).
11. A. P. Fletcher, Drug safety tests and subsequent clinical experience. *J R Soc Med* 71, 693(Sep, 1978).
12. M. Rozencweig *et al.*, Animal toxicology for early clinical trials with anticancer agents. *Cancer Clin Trials* 4, 21(1981).
13. T. Igarashi, S. Nakane, T. Kitagawa, Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J Toxicol Sci* 20, 77(May, 1995).
14. H. Olson *et al.*, Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56(Aug, 2000).
15. R. A. Matthews, Medical progress depends on animal models - doesn't it? *J R Soc Med* 101, 95(Feb, 2008).
16. G. Zbinden, Predictive value of animal studies in toxicology. *Regul Toxicol Pharmacol* 14, 167(Oct, 1991).
17. Top Institute Pharma.(2010), vol. 2011.
18. European Medicines Agency. vol. 2012.
19. Medicines Evaluation Board. vol. 2012.
20. P. G. Mol *et al.*, 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(Jun 1, 2010).
21. World Health Organization.
22. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.(EMA, London, 2009).
23. Committee for medicinal products for human use(CHMP).(London, 2005).

The sin which is unpardonable is knowingly and willfully to reject truth, to fear knowledge lest that knowledge pander not to thy prejudices.

-Aleister Crowley, 1875-1947, *Magick: Book 4, Liber ABA*

3 A RETROSPECTIVE EVALUATION OF NON-CLINICAL STUDIES PREDICTING NEW SERIOUS POST MARKETING ADVERSE EVENTS OF BIOPHARMACEUTICALS

Peter J.K. van Meer, Marlous Kooijman,
Christine C. Gispen-de Wied, Ellen H.M. Moors,
Huub Schellekens

To be submitted for publication

ABSTRACT

Animal studies are required to evaluate the safety of new medicines. However, the value of these studies is largely unknown. We used the marketing authorization application to evaluate the ability of animal studies to detect new adverse drug reactions of biopharmaceuticals post marketing. We identified 14 new adverse drug reactions for 13 biopharmaceuticals. Two adverse drug reactions had true positive counterparts in the animal data. Five other adverse drug reactions associated signals could be detected in the animal data. All adverse drug reactions with animal data counterparts were either related to the pharmacology of the biopharmaceutical or immune responses against it. Except for their timing and frequency, these adverse drug reactions can be expected and are better understood than idiosyncratic reactions of small molecule therapeutics. Pharmacovigilance activities for biopharmaceuticals should focus on the pharmacology of the target and unique characteristics of biopharmaceuticals. Animal data will not have added value in these activities.

Keywords: biopharmaceuticals, adverse drug reactions, non-clinical, animal studies, pharmacovigilance, predictive value

INTRODUCTION

Biopharmaceuticals are complex therapeutic proteins derived from recombinant biotechnologies. Unlike small molecule therapeutics (SMs), biopharmaceuticals are highly selective towards their intended target(*1*). Their half-life is also considerably longer but in contrast to SMs, biopharmaceuticals cannot be administered orally. The unique features of biopharmaceuticals lead to adverse effect profiles dissimilar to that of SMs(*2, 3*). Therefore, safety testing also requires an alternative approach to the standard and extensive package of animal studies required to evaluate the safety of SMs. This is reflected in guidelines for non-clinical safety testing of biopharmaceuticals. In particular, ICH S6, which was published in 2005 and has been recently updated, offers a flexible and case-by-case approach(*4*). Because certain studies may only be relevant for SMs, these can be omitted for biopharmaceuticals if adequately justified. For instance, metabolism studies are not required because biotherapeutics are catabolized and carcinogenicity is generally not required since proteins are not carcinogenic unless this is mediated by their pharmacology. Because of their complex nature, biopharmaceuticals are not active in most species and often the non-human primate is the only available responsive model(*5*). However, the value of primates in these studies is limited because the effects seen in primate safety studies are predominantly related to pharmacology and off-target effects are not probable or expected(*6, 7*). In addition the value of primate studies is reduced because the formation of antibodies to the biopharmaceutical, particularly in the case of monoclonal antibody products, can confound safety data and limit interpretability(*8-11*). So, the predictive value of animal studies in drug development is increasingly being questioned(*12-17*).

However, there are only few reviews available on animal-human correlations and those that exist may inadvertently have introduced publication and selection bias(*18-24*). We recently performed a study in which we attempted to minimize bias by using the non-clinical sections of the marketing authorization application to identify animal correlations to new adverse drug reactions of SMs requiring a safety related regulatory action post marketing(*25*). Only 19% of serious adverse reactions had an animal data counterpart that was considered true positive and an additional 17% had animal data that was considered associated but not predictive of the human adverse drug reaction. In this retrospective study, we report on the ability of safety studies in animals to detect adverse drug reactions requiring a safety related regulatory action of biopharmaceuticals after marketing.

METHOD OF REVIEW

Data was collected, analyzed and categorized in a similar method as reported previously(*25*). Briefly, Direct Healthcare Practitioner Communications (DHPCs) issued between 2000 and 2012 were identified on the website of the Medicines Evaluation Board (CMG-MEB, www.cbg-meb.nl) and adverse reactions requiring a safety related regulatory action were extracted. Safety-related regulatory actions were defined as DHPCs 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. The non-clinical summary of the marketing authorization application was used to find any animal data correlates related to the adverse reaction. All animal studies that were done to support marketing of the biopharmaceutical are identified non-clinical studies. Non-clinical events 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 a serious adverse reaction (SAR) in the animal species investigated. SARs without a corresponding non-clinical event were considered false negative. Finally, we noted whether the signals found in animal studies were related to the pharmacology of the biological or whether these were related to an immune response.

RESULTS

75 Direct Healthcare Professional Communications (DHPCs) were issued for 39 biopharmaceutical products between January 2001 and December 2012. 17 DHPCs concerning 13 products were issued to inform about 14 new adverse reactions leading to safety related regulatory actions such as market withdrawal of the product or addition of new warnings in the Summary of Product Characteristics (SPC). In the remaining 58 cases, DHPCs were issued to inform about a withdrawal from the market for commercial reasons, drug-drug interactions or new contra-indications, off-label use and dosing, delivery or production issues or to inform about known or new safety issues not requiring regulatory action and rewording of existing texts.

Non-clinical animal data from the marketing authorization application (MAA) could be accessed for 12 biopharmaceuticals (Table 1). In the case of Miacalcin (calcitonin) no MAA could be retrieved. Miacalcin was registered in The Netherlands in 1977 and a DHPC was issued on 22-08-2012 to inform about an increased risk of malignancy associated with frequent long term use of calcitonin, which led to a change in the SPC. 9 biopharmaceuticals were from the class of antineoplastic or immunomodulating products, 2 were blood and blood forming agents and 1 product was a musculo-skeletal agent.

No evidence in the non-clinical data was found for the occurrence of aplastic anemia and pancytopenia associated with the use of etanercept (Enbrel). No non-clinical data could be identified suggesting the occurrence of hepatosplenic T-cell lymphoma with the use of infliximab (Remicade) or adalimumab (Humira), both TNF-α inhibitors. Pure red cell aplasia resulting from binding of neutralizing erythropoetin (Eprex) antibodies to endogenous epoetin had no non-clinical counterpart. There was no non-clinical data suggesting acute respiratory distress, pleural effusion or acute pulmonary edema associated with the use of trastuzumab (Herceptin).

Anaphylactic reactions after infusion with tocilizumab (RoActemra) did not occur during non-clinical studies.

In the remaining seven cases non-clinical data could be found that was either associated with (five cases), or directly indicative (two cases) of the human serious adverse reaction (SAR). In these cases, the effects seen in animals were either related to the pharmacology

of the biopharmaceutical or related to an immune response (Table 2). Therefore, in total, a non-clinical signal or counterpart existed in 7 (50%) of the cases. The sensitivity of animal studies to detect new adverse drug reaction of biopharmaceuticals is 15% (2 cases).

Table 1. DHPC reported adverse drug reactions.

Active substance	Warning from DHPC
Bevacizumab	Hypersensitivity and infusion reactions
Botulinum toxin	Muscle weakness, dysphagia, aspiration as a result of peripheral spreading of botulin toxin
Etanercept	Risk for pancytopenia and aplastic anemia
Epoetin alpha	Pure red cell aplasia
Trastuzumab	Increased cardiac effects and rare pulmonary effects (Acute respiratory distress syndrome (ARDS), pulmonary infiltrates, pleural effusion and acute pulmonary edema)
Adalimumab	Hepatosplenic T-cell lymphoma
Rituximab	PML In RA patients and other auto-immune diseases
Efalizumab	PML, Guillain-Barré en Miller-Fisher, encephalitis, encephalopathy, meningitis, sepsis and opportunistic infections
Lepirudine	Severe/Fatal anaphylactic reactions
Infliximab	Heightened TBC and mortality
Infliximab	Risk for hepatosplenic T-cell lymphoma in young adults
Tocilizumab	Fatal anaphylaxis
Panitumumab	Severe keratitis and keratitis ulcerosa
Calcitonin	Increased risk of malignancies with long term use

Table 2. Adverse drug reactions of biopharmaceuticals are mediated by pharmacology or immune response and are expected.

Anatomic Therapeutic Class	Adverse reactions	True positive	Associated	Pharmacology	Immune response
Antineoplastic and immunomodulating	10	0	5	6	3
Blood and blood forming organs	2	1	0	0	2
Musculo-Skeletal System	1	1	0	1	0

DISCUSSION

We evaluated the ability of non-clinical animal studies to detect new adverse drug reactions after marketing authorization of biopharmaceuticals. This occurred in 50% of the cases that we studied. However, only in two cases (15%) the findings in animal studies correlated directly with the adverse drug reaction seen in humans. In the remaining five cases, animal data was found which corresponded to either the expected pharmacological action or immune responses that could lead to, but not predict, the adverse drug reaction detected

in humans. For example, increased tuberculosis infection has been reported with use of infliximab(*26*). This is a result of the pharmacological action of infliximab, which inhibits TNF-α and leads to a profound suppression of the immune system in pharmacologically responsive animals and humans(*27*). Nevertheless, animal data would not likely predict these findings because the low number of animals in a study lacks the statistical power to identify them with certainty or the studies are not lengthy enough to detect such adverse effects. The type of infection would also not be detected in a laboratory setting, since this is not something that can be studied.

Adverse effects with a non-clinical counterpart can fall within two classes, pharmacology mediated adverse effect or immune mediated adverse effect. Adverse effects not covered by these two classes might be caused by confounding factors related to the use of biopharmaceuticals. That is to say, biopharmaceuticals are rarely given as a monotherapy, are often administered in a hospital setting and patients receiving biotherapeutics can have more than one disease(*28*). For example, hepatosplenic T-cell lymphoma has been reported with the use of TNF-α inhibitors(*29*). An analysis of the REFURBISH study showed that in 91 cases of T-cell non-Hodgkin's lymphoma associated with the use of TNF-α inhibitors, 68% of the case involved exposure to a TNF-α inhibitor and another SM immunomodulator(*30*). The risk of T-cell non-Hodgkin's lymphoma was higher with TNF-α inhibitor used in combination with thiopurines but not with TNF-α inhibitor use alone. So while hepatosplenic T-cell lymphoma may still be driven by pharmacological effects of TNF-α inhibitor biotherapeutics, a confounding factor of thiopurines should not be ruled out. As a last example of confounding factors, concerns about calcitonin were raised after unlicensed use of calcitonin led to increased cases of prostate cancer(*31*). The indication of calcitonin was for postmenopausal osteoporosis and so the incidence of prostate cancer is orthogonal with the indication. However, a review of the data indicated an increased incidence of malignancy with the use of a nasal calcitonin formulation, the use of which was subsequently discontinued and warnings were added to the SPC. Calcitonin has been withdrawn from the market in the EU for commercial reasons in 2008(*32*).

Most of the new adverse drug reactions reported for biopharmaceuticals are very rare and predominantly arise from (exaggerated) pharmacology or immune reactions such as anaphylactic reactions. The latter type of reactions is only rarely observed in animal studies and the predictive value for humans is questionable. Previous studies have also reported that most adverse reactions for biologicals are related to infections, neoplasms and surgical and medical procedures(*33*). While we saw less new adverse drug reactions being added to the SPC of biotherapeutics compared to SMs, other studies note that there are roughly the same number of total adverse drug reactions identified for biopharmaceuticals and SMs(*34*). However, this may be because we only used those reported adverse effects which resulted in an amendment of the SPC to add the new adverse reaction or to withdraw the marketing authorization because of safety issues. Most DHPCs for biopharmaceuticals were issued to inform healthcare practitioners of known issues or drug-drug interactions, contra-indications or quality and manufacturing issues. In most cases, the benefit of using the biopharmaceutical continues to outweigh the risk and only one product has been withdrawn from the market as

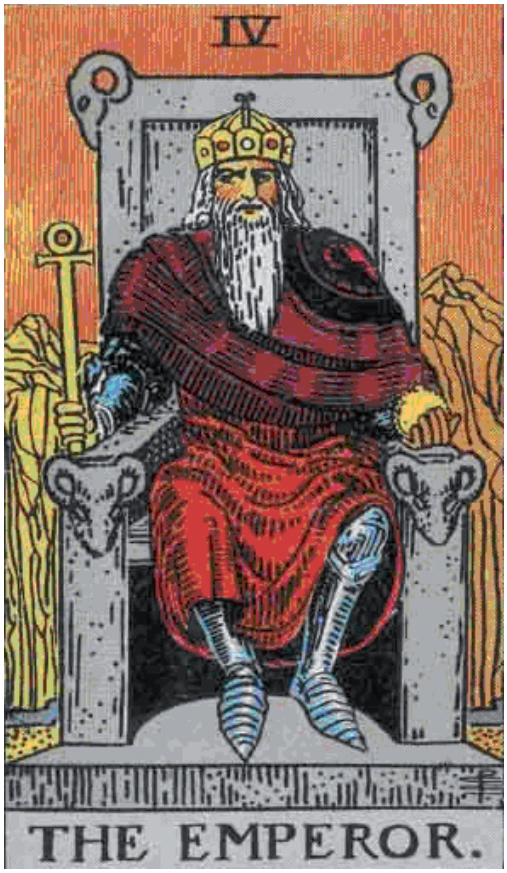
a result of safety concerns(*35*). At first glance, using animal data in risk management plans could be considered worthwhile because for biopharmaceuticals such data seems to be predictive. However, animal data will contribute minimally to pharmacovigilance activities because the adverse reactions can be understood and are expected. Therefore, companies rolling out pharmacovigilance strategies for biopharmaceuticals have been recommended to focus on understanding the target biology and intended mechanism of action to guide these efforts, and to take the unique features of biopharmaceuticals into account(*28, 36*). This is confirmed in our study as well. Pharmacovigilance activities would likely not benefit from having animal data available.

REFERENCES

1. D. J. Crommelin *et al.*, Shifting paradigms: biopharmaceuticals versus low molecular weight drugs. *Int J Pharm* 266, 3 (Nov 6, 2003).
2. P. Baldrick, Safety evaluation of biological drugs: what are toxicology studies in primates telling us? *Regul Toxicol Pharmacol* 59, 227 (Mar, 2011).
3. T. T. Hansel, H. Kropshofer, T. Singer, J. A. Mitchell, A. J. George, The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discov* 9, 325 (Apr, 2010).
4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Preclinical safety evaluation of biotechnology-derived pharmaceuticals ICH S6, (1997).
5. J. L. Bussiere, Species selection considerations for preclinical toxicology studies for biotherapeutics. *Expert Opin Drug Metab Toxicol* 4, 871 (Jul, 2008).
6. P. J. Bugelski, P. L. Martin, Concordance of Preclinical and Clinical Pharmacology and Toxicology of Therapeutic Monoclonal Antibodies and Fusion Proteins: 2-Cell Surface Targets. *Br J Pharmacol*, (Dec 13, 2011).
7. P. L. Martin, P. J. Bugelski, Concordance of Preclinical and Clinical Pharmacology and Toxicology of Monoclonal Antibodies and Fusion Proteins: 1-Soluble Targets. *Br J Pharmacol*, (Dec 13, 2011).
8. P. J. Bugelski, G. Treacy, Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 6, 10 (Feb, 2004).
9. J. Descotes, Immunotoxicity of monoclonal antibodies. *MAbs* 1, 104 (Mar-Apr, 2009).
10. R. Ponce *et al.*, Immunogenicity of biologically-derived therapeutics: assessment and interpretation of nonclinical safety studies. *Regul Toxicol Pharmacol* 54, 164 (Jul, 2009).
11. M. Kessler, D. Goldsmith, H. Schellekens, Immunogenicity of biopharmaceuticals. *Nephrol Dial Transplant* 21 Suppl 5, v9 (Oct, 2006).
12. A. Knight, Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. *Altern Lab Anim* 35, 641 (Dec, 2007).
13. J. Bailey, T. Capaldo, K. Conlee, M. Thew, J. Pippin, Experimental use of nonhuman primates is not a simple problem. *Nat Med* 14, 1011 (Oct, 2008).
14. R. A. Matthews, Medical progress depends on animal models - doesn't it? *J R Soc Med* 101, 95 (Feb, 2008).
15. N. Shanks, R. Greek, J. Greek, Are animal models predictive for humans? *Philos Ethics Humanit Med* 4, 2 (2009).
16. T. Hartung, Thoughts on limitations of animal models. *Parkinsonism Relat Disord* 14 Suppl 2, S81 (2008).
17. R. Lemon, S. B. Dunnett, Surveying the literature from animal experiments. *Bmj* 330, 977 (Apr 30, 2005).
18. A. H. Owens, Jr., Predicting anticancer drug effects in man from laboratory animal studies. *J Chronic Dis* 15, 223 (Mar, 1962).
19. E. J. Freireich, E. A. Gehan, D. P. Rall, L. H. Schmidt, H. E. Skipper, Quantitative comparison of toxicity of anticancer agents in

mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* 50, 219 (May, 1966).

20. P. S. Schein *et al.*, The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 11, 3 (Jan-Feb, 1970).
21. A. P. Fletcher, Drug safety tests and subsequent clinical experience. *J R Soc Med* 71, 693 (Sep, 1978).
22. M. Rozencweig *et al.*, Animal toxicology for early clinical trials with anticancer agents. *Cancer Clin Trials* 4, 21 (1981).
23. T. Igarashi, S. Nakane, T. Kitagawa, Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J Toxicol Sci* 20, 77 (May, 1995).
24. H. Olson *et al.*, Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56 (Aug, 2000).
25. P. J. van Meer, M. Kooijman, C. C. Gispen-de Wied, E. H. Moors, H. Schellekens, The ability of animal studies to detect serious post marketing adverse events is limited. *Regul Toxicol Pharmacol* 64, 345 (Dec, 2012).
26. J. Keane *et al.*, Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med* 345, 1098 (Oct 11, 2001).
27. European Medicines Agency, European Public Assessment Report of Infliximab (scientific discussion). (2005).
28. T. J. Giezen, A. K. Mantel-Teeuwisse, H. G. Leufkens, Pharmacovigilance of biopharmaceuticals: challenges remain. *Drug Saf* 32, 811 (2009).
29. C. A. Siegel, S. M. Marden, S. M. Persing, R. J. Larson, B. E. Sands, Risk of lymphoma associated with combination anti-tumor necrosis factor and immunomodulator therapy for the treatment of Crohn's disease: a meta-analysis. *Clin Gastroenterol Hepatol* 7, 874 (Aug, 2009).
30. P. Deepak *et al.*, T-cell non-Hodgkin's lymphomas reported to the FDA AERS with tumor necrosis factor-alpha (TNF-alpha) inhibitors: results of the REFURBISH study. *Am J Gastroenterol* 108, 99 (Jan, 2013).
31. Novartis, personal communication.
32. European Medicines Agency, Public statement on Forcaltonin - Withdrawal of the marketing authorisation in the European Union (2008).
33. T. J. Giezen *et al.*, Safety-related regulatory actions for biologicals approved in the United States and the European Union. *Jama* 300, 1887 (Oct 22, 2008).
34. H. C. Ebbers *et al.*, Differences between post-authorization adverse drug reactions of biopharmaceuticals and small molecules. *BioDrugs* 27, 167 (Apr, 2013).
35. European Medicines Agency, Public statement on Raptiva (efalizumab) - Withdrawal of the marketing authorisation in the European Union. , (2009).
36. H. C. Ebbers *et al.*, A cohort study exploring determinants of safety-related regulatory actions for biopharmaceuticals. *Drug Saf* 35, 417 (May 1, 2012).



Cuiusvis hominis est errare, nullius nisi insipientis in errore perseverare.

-Marcus Tullius Cicero, 106-43 B.C.

4 THE VALUE OF NON-HUMAN PRIMATES IN THE DEVELOPMENT OF MONOCLONAL ANTIBODIES

Peter J.K. van Meer, Marlous Kooijman,
Jan Willem van der Laan, Ellen H.M. Moors,
Huub Schellekens

Provisionally accepted for publication
(in an adapted format) in *Nature Biotechnology* (Sep 9, 2013)

ABSTRACT

Non-human primates (NHP) are often assumed to be the only relevant species to evaluate safety and efficacy of monoclonal antibodies (MAbs). However the scientific value of NHP in their development has never been established. Using drug registration files of all MAbs marketed in the European Union, we studied the value of using NHP to evaluate the safety and efficacy of these products. Inadequate justification of NHP use and ineffective study design led to a needless increase of NHP use. Immunogenicity further limited the value of NHP. As a predictive model NHP do not stand out because MAbs only exert their expected pharmacological action. Nevertheless, their use continues to increase. Therefore, a reevaluation of the need for routine studies with NHP to develop MAbs is urgently needed.

Keywords: non-human primate, non-clinical, biopharmaceuticals, monoclonal antibodies, drug development

INTRODUCTION

In the 1980s the first recombinant biotech products started coming onto the market. From discussions on the new challenges these products posed for drug safety testing, it became clear that non-clinical safety programs used for small molecule therapeutics would not be appropriate(*1, 2*). These discussions eventually led to the publication of a recently updated international guideline adopted by FDA, EMA and other regulatory bodies called 'ICH S6 - preclinical safety evaluation of biotechnology-derived pharmaceuticals'(*3*).

Non-clinical testing of small molecules is driven by an extensive set of guidelines which covers the non-clinical testing program. In contrast, ICH S6 offers a flexible, science based, and case-by-case approach to develop biotech products. For instance, animal studies to determine metabolism and genotoxicity are not required (therapeutic proteins are degraded to peptides and amino acids and it is not expected that therapeutic proteins interact with DNA). Carcinogenicity testing is only required if the mechanism of action of a therapeutic protein raises concerns. In the case of MAbs, most effects are highly species specific and often only non-human primates (NHP) possess the relevant target antigen. As a result, they are often considered the only relevant species for non-clinical studies. The use of NHP to assess the safety of monoclonal antibodies (MAbs) in drug development seems legitimate because they are often considered the only species sensitive to the adverse effects. However, the use of NHP poses ethical, practical and financial hurdles. But from a scientific point of view their use is also questionable(*4*).

Surprisingly few studies have been done to assess the value of NHP use in drug development and those have mainly focused on how to reduce their use(*5-7*). Some of these have suggested that, for small molecule therapeutics, NHP studies often under-predict serious toxicity and are less predictive for humans as assumed(*8*). This was also the case in the incident involving the phase I clinical trial with a CD28 superagonist monoclonal antibody, TGN-1412. There, six volunteers developed a severe immune reaction known as a 'cytokine storm' soon after receiving the first infusion. Because TGN-1412 does not act as a superagonist in NHP, this effect was not observed in safety studies and the product was considered safe in NHP(*9*).

Despite uncertainties over NHP use, their use in drug development has increased substantially has over the years because species specific MAbs with a broad range of indications are a fast growing class of products(*10*). Increasing regulatory demands are also believed to contribute to this rise(*5*). Scientific evidence is urgently needed to have an informed discussion on the use of NHP to develop biotech products. We performed a comprehensive study on the value of NHP in the development of MAbs. To perform this study we had unique access to confidential drug registration files, which contains the results of all animal studies done to support marketing authorization, of all MAbs approved in the European Union. To our knowledge, this source has not been used for analysis before. This is also the first comprehensive study to include all MAbs since muronomab (Orthoclone-OKT3) was registered in the late eighties.

MONOCLONAL ANTIBODY AUTHORIZATIONS IN THE EU

33 MAbs have been approved in the EU. IgG1 was the most common isotype (n=19) and seven products were Fab' fragments (Figure 1). Antineoplastic (n=8) or immunomodulatory (n=13) indications were dominant for older products, although often extension of indication

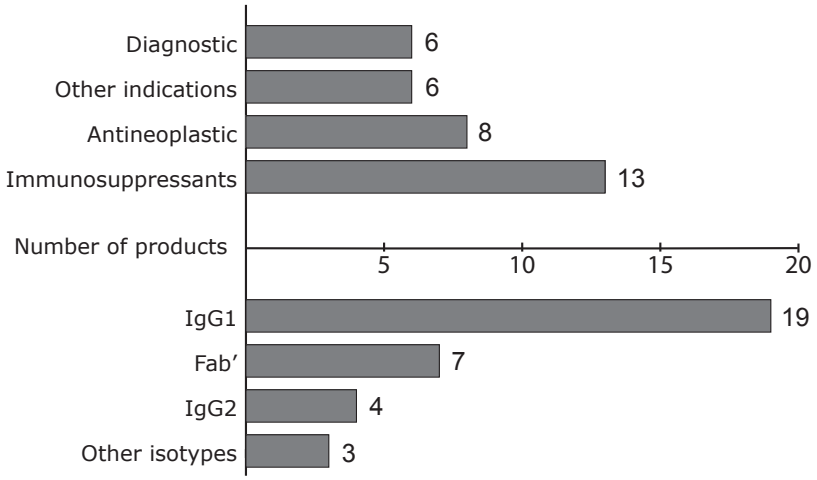


Figure 1. Number of products by indication and isotype (n=33). Above x-axis: antineoplastic and immunomodulatory agents form over 63% of all monoclonal antibodies on the market today. Aside from diagnostic agents, other monoclonal antibodies include a radiotherapeutic, an ophthalmological, an anti-thrombotic, an anti-osteoporotic and an anti-infective agent and a monoclonal antibody against obstructive airway disease. Below x-axis: monoclonal antibodies of the IgG1-isotype have been the primary source for development of therapeutic antibodies.

were sought for products later (for example, adalimumab, trastuzumab, infliximab). More recently, six indications other than antineoplastic, immunomodulatory or diagnostic agents have been registered (Figure 1). Early MAbs were predominantly murine and rodent-human chimera. Humanized MAbs, which are considered to have an improved immunogenicity profile, are currently the main form of therapeutic MAbs and fully human MAbs are a rapidly growing class of products. Six MAbs have been withdrawn from the market. For only one product (Efalizumab, Raptiva) this was related to safety issues which were identified after marketing authorization(*11*).

NHP USE IN NUMBERS

NHP were not used to assess the safety of six MAbs. Five of these were murine monoclonal antibodies, of which four have no target antigen in NHP because they are intended for diagnostic use in cancer. The fifth murine antibody is muronomab that targets to the T3 antigen of CD3 positive T-cells and is indicated for graft rejection therapy. This antigen is only present in humans, chimpanzees and gorillas. In agreement with the FDA, toxicity studies were not requested. Japanese regulatory authorities requested toxicity studies in rat and mouse with oral, *iv* and *sc* exposure which did not result in evidence of toxicity. Finally, for the sixth MAb, Eculizumab (Soliris), a humanized monoclonal antibody that targets C5 complement, non-clinical efficacy and safety testing was done using a surrogate antibody in mice (without signs of toxicity) because this MAb did not cross-react with its target in a

range of commonly used non-human primate and other species(*12*). 6,045 NHP were used in the non-clinical programs of the remaining 27 monoclonal antibodies (Table 1).

Table 1. The majority of NHP used in drug development is cynomolgus (86%). In only three cases (infliximab, efalizumab and omalizumab), chimpanzee is used during the non-clinical development program.

Species	Total number of NHP (products using this species for development)
Cynomolgus	5171 (24)
Rhesus	391 (6)
Marmoset	388 (2)
Chimpanzee	68 (3)
Baboon	18 (1)
African Green	9 (1)
Total NHP	6045

On average, a non-clinical program used 224±212 NHP with a median of 164 NHP. In one progam, one cynomolgus monkey was used to assess pharmacokinetics. The most NHP that were used in a non-clinical program was 755. Interestingly, the use of NHP increased as monoclonal antibodies became more human (Figure 2). There is a moderate correlation for increasing use of NHP in human MAb development over time (linear regression correlation coefficient r^2=0.698 p=0.0098, data not shown). For murine, chimera and humanized products

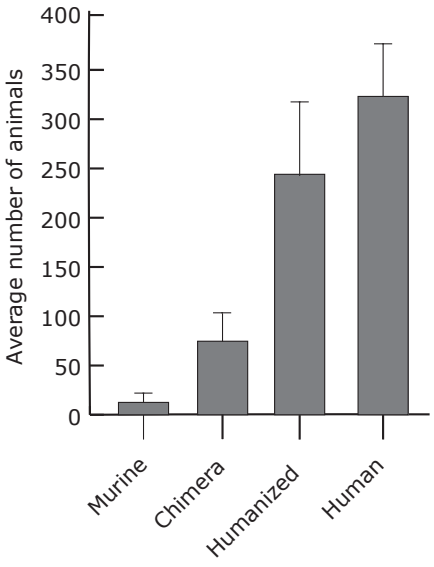


Figure 2. The average use of NHP per non-clinical programs increases as the MAb in development becomes more human ±SD. As humanization occurs more frequently in newer products, this increase may also reflect a temporal trend. The increase occurs irrespective of species specificity of the MAb. There is a significant difference between the group averages (non-parametric Kruskall Wallis test, p=0.001354. ANOVA with post-hoc Bonferroni comparison of means showed a significant difference between murine and human averages, p=0.01144).

no temporal correlation was found. Finally, except for the indication of diagnostic agents, where few NHP are used, there is no difference between NHP use for immunomodulatory, antineoplastic and other indications (data not shown).

SAFETY EVALUATION OF MABS IN NHP

15 MAbs (56%) were well tolerated in non-clinical studies using NHP, even when the animals were given high doses, and severe adverse effects were not noted. Other than the intended pharmacology, effects that were observed were generally mild and related to (secondary) pharmacology or injection site reactions. The remaining 12 MAbs induced more severe pharmacology or immune mediated reactions including death. Developmental and reproductive toxicity studies have been conducted in NHP for the safety assessment of 12 MAbs. This included all human MAbs except for votumumab, a diagnostic agent. In all but three cases, exposure to MAbs did not result in any effects on fertility or embryofetal development. In these three cases the effects were the result of the primary or secondary pharmacological action of the MAb (Box 1).

JUSTIFICATION OF NHP USE

For the non-clinical development of monoclonal antibodies choosing a relevant model species is paramount(13). Currently, ICH S6(R1) suggests the use of a species in which the product binds to its target and is pharmacologically active. A justification for the choice of species is also required if only one relevant species can be identified. The majority of study programs (78%, n=21) justified their choice for NHP by the ability of the MAb to bind to its target. But only in six cases the justification was also based on observed pharmacological activity following this binding.

In eight study programs (30%) the justification was either lacking, did not sufficiently take into account availability of other non-rodent species or NHP only provided a limited value. NHP were used in the case of Palivizumab (Synagis) because it was a commonly used species in monoclonal antibody development. In this case, NHP was used as a second, non-rodent, species next to rat and rabbit. The relevance of NHP use in this case was limited because Palivizumab does not have a target in NHP or human(14). in the case of votumumab (HumaSPECT), the non-clinical summary did not include a justification for the use of NHP. reduced affinity (Alemtuzumab, MabCampath)(15) for the target epitope or reduced activity (Certolizumab, Cimzia)(16) limited the relevance of NHP as a model species in two non-clinical programs. However, other non-rodent models were not available and so, NHP were considered as the only relevant option; despite identifying marmoset as the only relevant NHP species for non-clinical studies with Canakinumab (Ilaris), one pharmacokinetic study was conducted in rhesus monkeys(17). Catumaxomab (Removab) does not bind to NHP tissue. However, one cynomolgus monkey was used in a single dose immunogenicity/tolerance study(18). The affinity of Ranibizumab (Lucentis) for its target, VEGF-A, in NHP was not determined, although sequencing demonstrated an almost identical protein sequence to human and predicted a high homology with the

human target(19). Natalizumab (Tysabri), a α-4-integrin inhibitor, binds to its target in several species; in addition to NHP, the dog, pig, ferret and guinea pig were all found to be a suitable model species with similar binding affinities for one target as humans. Dog (safety pharmacology) and guinea pig (multiple sclerosis model, reproductive toxicity) were accordingly used in non-clinical studies. Nevertheless, the non-clinical program included extensive testing in NHP in addition to studies in rodents. The latter were considered to have limited value because the MAb did not bind to rodent α-4-integrin(20). Although species specificity of the MAb or its target often justified the use of one species, only one MAb (Basiliximab, Simulect) has been developed using only NHP. In the majority of cases rodents were extensively used as the second species for safety assessment even though these were not always a relevant species. Similarly, non-rodent species other than NHP were also used although the MAb was not necessarily pharmacologically active in those species.

Box 1 *Reproductive and developmental effects of MAbs are also pharmacologically mediated*

Panitumumab (Vectibix), an EGFR inhibiting antibody, increased the duration of the menstrual cycle(*24*). This effect may be pharmacological since EGFR is involved in endometrial growth and differentiation and may play a role in fetal implantation(*25, 26*) but could also have been a secondary effect to weight loss in these animals since there was no effect on female reproductive organs. Similarly, treatment with Bevacizumab (Avastin), a VEGF inhibitor, resulted in reduced ovarian function in NHP(*27*). In the case of Natalizumab (Tysabri), a significantly increased abortion and stillbirth rate was observed(*20*). This may be related to involvement of integrins (the target of natalizumab) in placental and embryofetal development(*28, 29*). However, in a recent study, cynomolgus monkeys exposed to Natalizumab until gestation day 100 did not exhibit these adverse developmental effects(*30, 31*).

DIFFERENT SPECIES OF NHP IN MAB DEVELOPMENT

If NHP were considered as the primary model species, the non-clinical program generally followed or exceeded those studies outlined in the ICH S6 guideline. The NHP species in most non-clinical programs was the cynomolgus macaque. Other species such as the rhesus macaque and the marmoset were also used as the primary model species in some programs. Eight non-clinical programs used two or more NHP species. This included the use of NHP in routine non-clinical studies, biocomparability studies and as animal models of disease such as thrombosis, organ transplantation and choroidal neovascularization. In three non-clinical programs chimpanzees were used. In these cases poor study design or the low number of animals in the different study groups limited the value of the studies (Box 2).

IMMUNOGENICITY IN NHP STUDIES

Anti-drug-antibodies (ADAs) may be formed in NHP as an auto-immune response to exposure to therapeutic MAbs which can influence the activity of these products in several ways; ADAs may neutralize the effect of the MAb by binding with high affinity to their active site; by affecting pharmacokinetic behavior of the MAb by clearing it from the body; or by enhancing the pharmacological effect of MAbs. MAbs were immunogenic in most non-clinical programs which used NHP (93% n=25). ADAs influenced pharmacokinetic or pharmacodynamic parameters of MAbs in the majority of (78%, n=21) of the programs. For one of the products, immune responses led to anaphylactic shock. Pharmacokinetic or pharmacodynamic profiles changed to due to formation of ADA in animals to some extent but could affect a considerable number of monkeys, including leading to a limitation of the duration of long term toxicity studies in five programs.

OTHER FACTORS INFLUENCING THE VALUE OF NHP

Immunogenicity and species specificity were not the only factors which reduced the value of NHP studies. Wild caught monkeys were used as a model species in two non-clinical programs. But wild caught NHP may be inappropriate because their age, genetic background, exposure to pathogens, environmental and social conditions are not known. In both non-clinical programs, high inter-animal variability interfered with interpretation of study results. Study design also limited the scientific quality of the data. Some studies were initiated that could be considered irrelevant a priori, for example because the route of administration in the study was oral, even though MAbs are not taken up by the gut; the MAb was known not to be active in NHP; or developmental and reproductive toxicity studies may not have been necessary because the MAb was developed as an antineoplastic intended for co-treatment with cytostatics. In a few cases, studies were requested by regulatory authorities that were not relevant.

All studies with NHP used necropsy to evaluate the effects of the MAb at a target organ level with the exception of the programs that used chimpanzees. Necropsy was performed in control and treatment groups before and after a recovery period to study the mechanism of action, effects at the tissue level and reversibility of adverse effects. In most cases, necropsy confirmed findings from non-invasive methods such as clinical observations and blood chemistry but did not reveal any new adverse effects.

DISCUSSION

One important limitation of our analysis is that we have only used data from MAbs which received a marketing authorization; we did not have access to data from MAbs that were never marketed. The number of NHP per MAb and other metrics concerning NHP use are a conservative estimate since we could not always recover all the data for all studies.

When most companies started their non-clinical development phase, little was known about sustained effects of therapeutic MAbs in a complex organism such as NHPs. How MAbs mediate safety issues either through on-target and off-target pharmacology, toxicity,

Box 2 *Use of chimpanzees in MAb development*

In cases where great apes are used the utmost care should be taken to ensure that the experiments are of the highest quality. This was not the case for three products that used chimpanzees in non-clinical study programs initiated between 1991 and 2000. In two cases, chimpanzees were used as the primary test species due to the species specificity of the target antigen whereas in one case chimpanzees were used in one study alongside other NHP species to elucidate the mechanism of an adverse effect. A common feature of programs where chimpanzees were used as a primary species was an insufficient *a priori* justification of their use and a scientific study rationale. The subsequent problematic study design and conduct led to a limited value of their use to evaluate the safety of these products. Likely because of ethical considerations, cost and/or difficulties in handling chimpanzees, few animals were used in the non-clinical programs. Low number of chimpanzees study or per group within a study negatively influenced the scientific value due to lack of statistical power. Studies as small as one animal per study or one animal per group were recorded. However, such studies may still be useful in hazard identification. Even so, regulatory authorities considered one of the programs to be of marginal value for the human safety assessment due to uncertainties in the conduct of experiments and their study design(*32*).

The value of chimpanzees for safety assessment of another product was not considered optimal for the intended patient population and in some cases study design may have masked treatment related effects. As a result, a homolog was tested in rodents to generate a more complete safety profile. Finally a study used chimpanzees to study the cause of an adverse effect although it was not clear whether any of the NHP species in this study represented the human situation. Therefore, little justification existed for using a contentious species such as the chimpanzee.

complement activation or immune responses was not as well understood as it is today. The use of a relevant NHP model to investigate the non-clinical fate of MAbs has led to the current body of work, spanning almost 30 years and a wide variety of antibodies and targets. This current state of the art continues to evolve and would not be as complete without NHP studies and for this reason alone their past use has not been altogether valueless. Still, there remain unaddressed challenges ahead of us if a science driven development of innovative MAbs is to be realized.

The use of NHP in research is rising because of an increasing development of MAbs but a scientific basis for their increase is lacking. The use of NHP is primarily justified by the expression of the relevant target epitope with a comparable tissue distribution to human. NHP have been used even if they lack this attribute. The widespread use of rodent species in non-clinical programs is also surprising. ICH S6 explicitly allows for exclusion of multiple species testing in favor of testing in one scientifically relevant species if this can be justified. The choice for two species testing may be driven by risk averse behavior and

uncertainty about regulatory demands. The strategy to simply follow the guidances rather than develop a limited but science based non-clinical program is further supported by the fact that most non-clinical programs generally followed or exceeded the studies outlined by the guidelines.

An increase in the use of NHP was observed as MAbs became more human, but increased humanization does not necessarily restrict the number of species which possess the relevant target antigen. Likely, the increase in their use occurred because NHP are perceived to better predict pharmacokinetic behavior and to show less immunogenicity due to the close relationship with human or because of increased affinity of the MAb in NHP. Nevertheless, most products were immunogenic in NHP, in some cases leading to limitations in study duration or loss of dose groups that could not be evaluated because of clearing ADAs, which further compromises the predictive value of these models. In addition, sub-optimal study design and the use of NHP when it was not a relevant species limited the value of NHP studies.

The main reason NHP are of very limited value is because the adverse events induced by MAbs are highly predictable. Those effects are either mediated by the pharmacology of the MAb which may be exaggerated by dose or exposure or they are mediated by immune responses mounted against the therapeutic MAb(21-23). This was confirmed in our analysis where we have also not found any likely off-target effects. This included findings from post-mortem and developmental and reproductive toxicity studies. However, there is still some remaining value for NHP use in MAb development. In science driven proof of concept studies they may be informative or they can be used in small confirmatory studies once the target pharmacology is characterized *in vitro*.

In conclusion, the value of extensive NHP use in routine safety and efficacy studies for the non-clinical assessment of MAbs is scientifically debatable and may, in that way, also miss a moral basis. Incentives for pharmaceutical companies to develop biotech products without the use of animals may help in facilitating the much needed innovation in non-clinical drug development. Dialogue between pharmaceutical industry and regulatory authorities on non-clinical requirements should be increased. More importantly however, an in depth reevaluation of the value of NHP in all stages of drug development, including products which did not reach the market, is long overdue.

METHODS

Monoclonal antibodies (MAbs), and immunoglobulin fragments, which received a marketing authorization in the EU or one of its member states up to 01-06-2011 were identified from websites of the EMA, Medicines Evaluation Board (CBG-MEB) and literature(*33, 34*). For each MAb the drug registration file (common technical document, CTD) was located at the archives of the CBG-MEB in The Hague. For one MAb (Xolair, Omalizumab), the non-clinical overview was obtained from the EMA. The non-clinical overview and tabulated non-clinical summaries contained in the CTD were used as the primary source for this study. The total number of non-human primates (NHP) per product is the sum NHP per study described in the tabulated summary and includes F_1-offspring generated in developmental

and reproductive toxicity testing, if applicable. For studies providing ranges of animals used, the range average was used. If no accurate count could be obtained and individual study reports in module 4 of the CTD could not be accessed, counts were left blank. Qualitative data was extracted, summarized and categorized by the first author.

STATISTICS

Because some products did not use NHP, our samples contain 0 as an outcome. Therefore, we added 1 to all our NHP counts for statistical analysis. SPSS (version 16.0 for Windows, IBM, Armonk, NY, USA) was used to conduct the non-parametric Kruskall-Wallis test with a Bonferroni post-hoc test to identify variance between groups. Prism (Prism 4 for Windows, v.4.03, GraphPad Software Inc., La Jolla, CA, USA) was used for linear regression.

CONFLICT OF INTEREST

P.J.K. van Meer, M. Kooijman, JW van der Laan and E.H.M. Moors declare no conflict of interest. H. Schellekens 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.

ACKNOWLEDGEMENTS

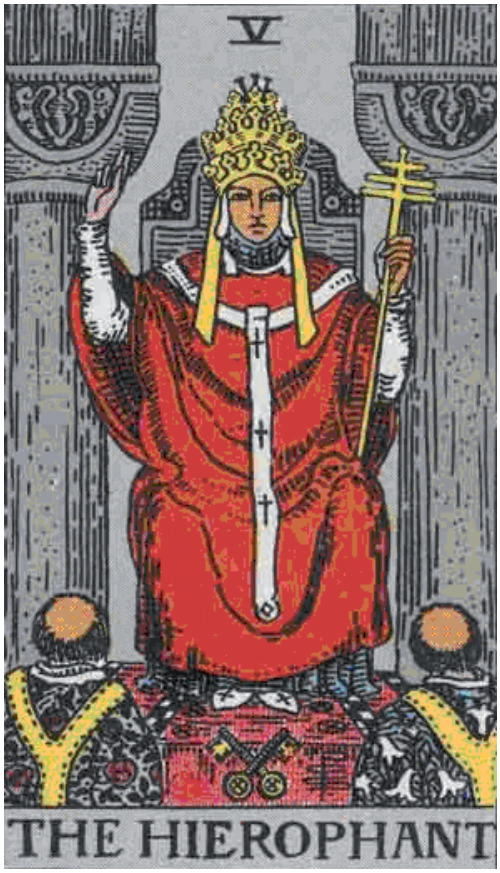
This research was performed under the framework of Top Institute Pharma, (project T6-301) which includes the Medicines Evaluation Board, The Organization of innovative Pharmaceutical Industries in The Netherlands (Nefarma), the Life Science and health Initiative, and Utrecht University. PvM and MK performed this work with the TI Pharma grant. 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.

REFERENCES

1. K. Teelmann, C. Hohbach, H. Lehmann, Preclinical safety testing of species-specific proteins produced with recombinant DNA-techniques. An attempt to transfer current experience into future testing strategies. *Arch Toxicol* 59, 195 (Dec, 1986).
2. M. Kooijman, P. J. van Meer, E. H. Moors, H. Schellekens, Thirty years of preclinical safety evaluation of biopharmaceuticals: Did scientific progress lead to appropriate regulatory guidance? *Expert Opin Drug Saf* 11, 797 (Sep, 2012).
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Preclinical safety evaluation of biotechnology-derived pharmaceuticals ICH S6, (1997).
4. J. Bailey, T. Capaldo, K. Conlee, M. Thew, J. Pippin, Experimental use of nonhuman primates is not a simple problem. *Nat Med* 14, 1011 (Oct, 2008).
5. K. L. Chapman, N. Pullen, L. Andrews, I. Ragan, The future of non-human primate use in mAb development. *Drug Discov Today* 15, 235 (Mar, 2010).
6. J. Clarke *et al.*, Duration of chronic toxicity studies for biotechnology-

derived pharmaceuticals: is 6 months still appropriate? *Regul Toxicol Pharmacol* 50, 2 (Feb, 2008).

7. R. H. Bettauer, Systematic review of chimpanzee use in monoclonal antibody research and drug development: 1981-2010. *Altex* 28, 103 (2011).
8. J. Bailey, Non-human primates in medical research and drug development: a critical review. *Biogenic Amines* 19, 235 (2005).
9. Expert Group on Phase One Clinical Trials (Chairman: Professor Gordon W. Duff), *Expert Group on Phase One Clinical Trials: Final report*. (TSO, Norwich, 2006).
10. S. Aggarwal, What's fueling the biotech engine-2009-2010. *Nat Biotechnol* 28, 1165 (Nov, 2010).
11. European Medicines Agency, Public statement on Raptiva (efalizumab) - Withdrawal of the marketing authorisation in the European Union. , (2009).
12. European Medicines Agency, Initial marketing authorization documents - scientific discussion on eculizumab. (2007).
13. J. L. Bussiere, Species selection considerations for preclinical toxicology studies for biotherapeutics. *Expert Opin Drug Metab Toxicol* 4, 871 (Jul, 2008).
14. European Medicines Agency, European Public Assessment Report of Palivizumab (scientific discussion). (2005).
15. European Medicines Agency, European Public Assessment Report of Alemtuzumab (scientific discussion). (2005).
16. European Medicines Agency, European Public Assessment Report of Certolizumab (scientific discussion). (2009).
17. European Medicines Agency, European Public Assessment Report of Canakinumab (scientific discussion). (2009).
18. European Medicines Agency, European Public Assessment Report of Catumaxomab (scientific discussion). (2009).
19. European Medicines Agency, European Public Assessment Report of Ranibizumab (scientific discussion). (2007).
20. European Medicines Agency, European Public Assessment Report of Natalizumab (scientific discussion). (2007).
21. P. Baldrick, Safety evaluation of biological drugs: what are toxicology studies in primates telling us? *Regul Toxicol Pharmacol* 59, 227 (Mar, 2011).
22. T. J. Giezen *et al.*, Safety-related regulatory actions for biologicals approved in the United States and the European Union. *Jama* 300, 1887 (Oct 22, 2008).
23. T. T. Hansel, H. Kropshofer, T. Singer, J. A. Mitchell, A. J. George, The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discov* 9, 325 (Apr, 2010).
24. European Medicines Agency, European Public Assessment Report of Panitumumab (scientific discussion). (2007).
25. Z. P. Yue, Z. M. Yang, S. J. Li, H. B. Wang, M. J. Harper, Epidermal growth factor family in rhesus monkey uterus during the menstrual cycle and early pregnancy. *Mol Reprod Dev* 55, 164 (Feb, 2000).
26. J. L. Cameron, Regulation of reproductive hormone secretion in primates by short-term changes in nutrition. *Rev Reprod* 1, 117 (May, 1996).
27. European Medicines Agency, European Public Assessment Report of Bevacizumab (scientific discussion). (2006).
28. A. Beauvais-Jouneau, J. P. Thiery, Multiple roles for integrins during development. *Biol Cell* 89, 5 (Mar, 1997).
29. J. T. Yang, H. Rayburn, R. O. Hynes, Cell adhesion events mediated by alpha 4 integrins are essential in placental and cardiac development. *Development* 121, 549 (Feb, 1995).
30. N. G. Wehner, G. Shopp, S. Oneda, J. Clarke, Embryo/fetal development in cynomolgus monkeys exposed to natalizumab, an alpha4 integrin inhibitor. *Birth Defects Res B Dev Reprod Toxicol* 86, 117 (Apr, 2009).
31. N. G. Wehner *et al.*, Postnatal development in cynomolgus monkeys following prenatal exposure to natalizumab, an alpha4 integrin inhibitor. *Birth Defects Res B Dev Reprod Toxicol* 86, 144 (Apr, 2009).
32. European Medicines Agency, European Public Assessment Report of Infliximab (scientific discussion). (2005).
33. A. L. Nelson, E. Dhimolea, J. M. Reichert, Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* 9, 767 (Oct, 2010).
34. G. Walsh, Pharmaceutical biotechnology products approved within the European Union. *Eur J Pharm Biopharm* 55, 3 (Jan, 2003).



The truth is rarely pure and never simple.

-Oscar Wilde, 1854-1900, *The Importance of Being Earnest*

5 IMMUNOGENICITY OF MABS IN NON-HUMAN PRIMATES DURING NON-CLINICAL SAFETY ASSESSMENT

Peter J.K. van Meer, Marlous Kooijman, Vera Brinks,
Christine C. Gispen-de Wied, Beatriz Silva-Lima,
Ellen H.M. Moors, Huub Schellekens

Published (in an adapted format) in *MAbs* 5, 810 (Sep 1, 2013)

ABSTRACT

The immunogenicity of biopharmaceuticals used in clinical practice remains an unsolved challenge in drug development. Non-human primates (NHPs) are often the only relevant animal model for the development of monoclonal antibodies (mAbs), but the immune response of NHPs to therapeutic mAbs is not considered to be predictive of the response in humans because of species differences. In this study, we accessed the drug registration files of all mAbs registered in the European Union to establish the relative immunogenicity of mAbs in NHPs and humans. The incidence of formation of antidrug-antibodies in NHPs and patients was comparable in only 59% of the cases. In addition, the type of antidrug–antibody response was different in NHP and humans in 59% of the cases. Humanization did not necessarily reduce immunogenicity in humans. Immunogenicity interfered with the safety assessment during non-clinical drug development when clearing or neutralizing antibodies were formed. While important to interpret the study results, immunogenicity reduced the quality of NHP data in safety assessment. These findings confirm that the ability to compare relative immunogenicity of mAbs in NHPs and humans is low. Furthermore, immunogenicity limits the value of informative NHP studies.

Keywords: immunogenicity, non-human primates, monoclonal antibodies, antidrug-antibodies, non-clinical, safety assessment, drug development, regulatory science

INTRODUCTION

The advent of recombinant technology some 30 years ago led to revolutionary novel methods of drug development that allowed the production of any protein-based drug in cell culture. These protein-based biopharmaceuticals offer important advantages over classical small-molecule drugs, such as longer half-lives and very high specificity. In contrast to small molecules, which are metabolized, therapeutic proteins are degraded into their constituent amino acids. As a result, most adverse effects are a result of exaggerated pharmacodynamics(*1*). An unsolved challenge in biopharmaceutical development is that these proteins ultimately become immunogenic in some patients, provoking an immune response. The immunogenicity of the agents is dependent on specific properties, such as protein folding, aggregation, post-translational modifications, and the presence of B and T cell epitopes. In addition, the presence of impurities in the formulation, the route of administration, mode of action, patient population, and treatment regimen may all affect immunogenicity(*2*). In clinical practice, a drug-evoked immune response can lead to a host of side effects, such as serum sickness, hypersensitivity, and injection site reactions or, in some rare cases, hazardous cross-reactivity with endogenous proteins.(*3*) More commonly, the immune response leads to a loss of drug efficacy because of the development of neutralizing or clearing antidrug antibodies (ADAs)(*4*).

The engineering of proteins may yield potentially marked reductions in immunogenicity of protein-based drugs(*5*). It is, however, difficult to evaluate the immunogenicity of mAbs because there are few robust and predictive bioinformatics approaches or *in vitro* screens to measure and characterize the immune response. Bioinformatics approaches have been developed that can identify immunogenic T cell epitopes(*6*). and removal of these T cell epitopes is suggested to reduce immunogenicity(*7*). Harding et al. have shown that removal of $CD4^+$ T-helper cell epitopes from V-region peptides of the chimeric antibody cetuximab by humanizing these peptides results in a reduction in immunogenic potential(*8*). T cell activation assays could also be used to measure the potential of protein drugs to evoke an immune response(*9, 10*), but, in non-clinical safety assessment, these studies are not required and laboratory animals are routinely used to evaluate immunogenicity. The predictive value of immunogenicity measured in common animal models such as rodents and dogs, however, is low because these models generally overpredict immunogenicity in humans(*11*). Non-clinical immunotoxicity studies in animals are also considered inadequate to evaluate safety issues related to immunotoxicity such as hypersensitivity and auto-immunity(*12*). The shortcomings of animal studies are reflected in international and European immunogenicity guidelines(*13, 14*). Although the assessment of immunogenicity in non-clinical studies is not recommended as a way to estimate the response in humans, animals may be useful to study some aspects of immunogenicity, such as determining the relative immunogenicity of a biosimilar compared to its reference product(*15*) and to interpret the findings from animal studies(*16-18*). Besides the low predictive value of immunogenicity in animals, a major handicap is that assays used to assess the immunogenicity of therapeutic proteins are not standardized. A recent industry survey showed that several assays are being used

that, although complying with general guidelines, often yield variable results that cannot be compared because of different assay formats. Moreover, the lack of a reference standard, amongst others, makes these assays semi-quantitative(*19*). This makes direct comparisons of immunogenicity between products and species particularly challenging, if not impossible. The relative immunogenicity of mAbs in humans and animals has been assessed in the past(*11, 17*). Here, we provide an overview and comparison of the immunogenicity in NHPs and humans of all mAbs approved for use in the European Union (EU) through 2010. We also studied the influence of immunogenicity on the ability to interpret non-clinical study findings. For this study, we had access to the marketing authorization applications, which contain all animal studies done to support marketing authorization of mAbs approved in the EU.

IMMUNOGENICITY IN NHPS

Of 33 mAbs in our analysis (Table 1), the safety of 27 mAbs was evaluated in NHPs. MAb29 and MAb31 were not immunogenic in NHP, whereas the remaining 25 MAbs (93%) were. Moreover, the presence of ADAs led to changes in the pharmacokinetic or pharmacodynamic profiles of all but four mAbs: MAb8, MAb13, MAb14 and MAb17. For the remaining 21 products, immunogenicity influenced pharmacokinetics and pharmacodynamics with varying magnitude and severity (Table 2). Repeated dose studies for five MAbs (MAb7, MAb9, MAb14, MAb15 and MAb21) were limited in duration by immunogenicity.

MURINE ANTIBODIES

Three murine mAbs were assessed in NHPs. MAb5 was highly immunogenic because all NHPs developed clearing antibodies by the end of a 1-month repeated dose study. MAb7 was moderately immunogenic in NHPs, with repeated dose studies being restricted to 2 weeks because ADA development was expected to interfere with the safety assessment. MAb8 had immunogenic potential only because ADAs were detected after the last dose in an escalating repeated dose study and it was assessed in only one NHP.

CHIMERIC ANTIBODIES

The safety of five chimeric mAbs was assessed in NHPs. MAb13 had low immunogenic potential and a low-titer ADA response was measured in one control group chimpanzee that was accidentally dosed with MAb13. MAb10 and MAb11 had moderate immunogenic potential. In a 2-week repeated dose study with MAb11, the death of one animal with high titers of ADAs was attributed to thrombocytopenia. Two animals with high ADA titers that received a second MAb11 dose rapidly developed thrombocytopenia. MAb9, MAb12 were highly immunogenic. Repeated dose studies with MAb9 were limited to 8 weeks because of the development of ADAs, which resulted in the rapid clearance and decreased pharmacodynamics. The considerable ADA response made it difficult to generate conclusive data on the effects of long-term treatment in NHPs.

Table 1. Monoclonal antibodies approved in the European Union from 1988 to 2010

INN	Brand name	Approval date
Abciximab	Reopro, CentoRx	1995**
Adalimumab	Humira, Trudexa	2003
Alemtuzumab	MabCampath	2001#
Anti-melanoma antibody fragments*	Tecnemab-K1	1996#
Arcitumomab*	CEA-scan	1996#
Basiliximab	Simulect	1998
Besilesomab*	Scintimun	2010
Bevacizumab	Avastin	2005
Canakinumab	Ilaris	2009
Catumaxomab	Removab	2009
Certolizumab pegol	Cimzia	2007
Cetuximab	Erbitux	2004
Daclizumab	Zenapax	1999#
Denosumab	Prolia	2010
Eculizumab	Soliris	2007
Efalizumab	Raptiva	2004#
Golimumab	Simponi	2009
Ibritumomab tiuxetan	Zevalin	2004
Igovomab*	Indimacis 125	1996#
Infliximab	Remicade	1999
Muromonab	Orthoclone-OKT3	1988**
Natalizumab	Tysabri	2006
Ofatumumab	Arzerra	2010
Omalizumab	Xolair	2005
Palivizumab	Synagis	1999
Panitumumab	Vectibix	2007
Ranibizumab	Lucentis	2007
Rituximab	Mabthera	1998
Sulesomab*	Leukoscan	1997
Tocilizumab	RoActemra	2009
Trastuzumab	Herceptin	2000
Ustekinumab	Stelara	2009
Votumumab*	Humaspect/ Oncospect CR	1998

INN: International nonproprietary name. The infixes that immediately precede -mab indicate the sequence source: u, human; zu, humanized; xi, chimeric; o, mouse; axo, rat/mouse. Chimeric antibodies are antibodies which are part rodent, often the variable region, and part human, often the constant region.
*Diagnostic/imaging agent.
**Country specific approval
#Withdrawn from use in the European Union.

Table 2. Incidence and response level, either clearing or neutralizing, of anti-drug antibodies in non-human primates

	Incidence of clearing or neutralizing antibodies in NHP		
ADA incidence in NHP	Low	Intermediate	Majority
Low (0-6%)	MAb12	MAb21	MAb30
	MAb14		
	MAb16		
	MAb17		
	MAb25		
	MAb29		
	MAb31		
Intermediate (6-45%)	MAb9	MAb7	MAb20
	MAb26	MAb10	MAb22
	MAb23	MAb24	MAb32
	MAb27		MAb33
High (>45%)			MAb5
			MAb8
			MAb11
			MAb13
			MAb15
			MAb18
			MAb28

ADA: Anti-drug antibodies
NHP: Non-human primate

HUMANIZED ANTIBODIES

Eleven of the 12 humanized mAbs included in this study were evaluated in NHP. MAb14, MAb16, MAb17, MAb21, and MAb25 had low immunogenicity in NHPs, whereas MAb20, MAb22, MAb23, and MAb24 had moderate immunogenicity. In the case of MAb21, the duration of meaningful repeated dose studies was limited to one month. Anti-MAb24 antibodies were only measured in repeated dose studies. Interestingly, serum concentrations of MAb24 were increased in NHPs positive for anti-Mab24 antibodies. In addition, perivascular sheathing in some NHPs was associated with high anti-Mab24 antibody titers. MAb15 and MAb18 were highly immunogenic. Antibodies to MAb15 developed in most NHPs within two weeks of single or multiple doses and increased clearance. Immunogenicity was reduced when the dose was increased. In repeated dose studies with MAb18, antibodies to MAb18 were always associated with rapid clearance of the drug. Reliable estimates of pharmacokinetic parameters could only be obtained after the first dose because the development of ADAs interfered with the distribution and pharmacokinetics.

HUMAN ANTIBODIES

NHPs did not develop an immune response to MAb29 or MAb31. MAb30 was poorly immunogenic, and MAb26, MAb37, MAb32 and MAb33 were moderately immunogenic in NHPs. In the case of MAb30, clearing antibodies developed only in single dose, but not repeated dose, toxicity studies. Antibodies to MAb26 were detected after repeated dosing. In some cases, the presence of ADAs was associated with increased plasma clearance. Two NHPs developed clinical hemolytic anemia that may have been secondary to high antibody titers. It is likely that ADAs developed more often in MAb26-treated animals because a direct Coombs' test, which is used to determine autoimmune hemolytic anemia, suggested that most positive animals were slowly developing anemia. The pharmacokinetic profile of MAb33 was affected by the development of ADAs after repeat dosing, leading to an inverse dose–response relationship. Similarly, clearing antibodies to MAb32 were detected more frequently in NHPs receiving low doses. There, the presence of ADAs was associated with an increased clearance and reduced half-life and anti-MAb32 antibodies developed in up to 97% of the animals after a single low dose. One human MAb, MAb28, was highly immunogenic in NHPs and led to the formation of binding and neutralizing antibodies, and higher doses were needed to maintain exposure. Antibodies to MAb28 were formed in more than 50% of animals tested.

COMPARISON OF IMMUNOGENICITY IN HUMANS AND NHPS

Data on the clinical immunogenicity of all mAbs was included in the SPC (29 products) or EPAR (4 products). The induction of ADAs to 20 mAbs affected clinical efficacy, altered pharmacokinetic profiles, or caused adverse effects. The presence of ADA to nine mAbs did not have consequences, and 4 mAbs did not give rise to ADA development (Table 3). For seven mAbs the incidence of ADAs in NHPs overpredicted the induction of ADAs in humans, and for four mAbs the reverse, underprediction was the case. Sixteen mAbs had comparable ADA incidences in both NHPs and humans. The ADA response was similar for 9 mAbs, including 4 mAbs that did not cause an ADA response in either NHPs or humans. Two mAbs (MAb15 and MAb22) induced clearing and neutralizing antibodies in NHPs, but neutralizing antibodies only in humans. In NHPs, ADAs were more often directed against the Fc-region, resulting in clearing antibodies (17 out of 27 cases) whereas in humans, ADAs were most often formed against the complementarity-determining region (CDR), resulting in neutralizing antibodies (10 out of 33 cases).

DISCUSSION

Minimizing immunogenicity remains a considerable challenge in the development of mAbs. While the humanization of mAbs has been successful in reducing the immunogenicity of some products, clinically relevant immunogenicity can still occur despite such modifications(*20-22*). Most mAbs in the clinic can be categorized as negligible or tolerably immunogenic. The onset of ADA formation in the clinic usually occurs after multiple injections which can cover months of treatment. For physicians, treatment management should include frequent monitoring for neutralizing ADA and when these occur, treatment should be stopped or the patient should

Table 3. Incidence and effect of anti-drug antibodies in clinical trials in comparison with non-clinical data

Product	Clinical immunogenicity	ADA response clinical	Non-clinical immunogenicity	ADA response non-clinical
Murine antibodies				
MAb1	Marked	Reduced efficacy due to interference	Not available	Not available
MAb2	Negligible	Diminished efficacy and allergic or hypersensitivity reactions	Not available	Not available
MAb3	Negligible	None	Not available	Not available
MAb4	Negligible	Diminished efficacy	Not available	Not available
MAb5	Tolerable	Diminished efficacy possible	High	Clearing
MAb6	Marked	Neutralizing and hypersensitivity	Not available	Not available
MAb7	Tolerable	Unknown	Intermediate	Clearing
MAb8	Marked	Neutralizing	High	None
Chimeric antibodies				
MAb9	Tolerable	Allergic or infusion site reactions in few patients	High	Clearing
MAb10	Tolerable	Clearing (in few patients)	Intermediate	Clearing
MAb11	Tolerable	Thrombocytopenia	Intermediate	Thrombocytopenia
MAb12	Tolerable	Unknown	High	Clearing
MAb13	Marked	Neutralizing and hypersensitivity	Low	None
Humanized antibodies				
MAb14	Negligible	None	Low	None
MAb15	Negligible	Positive Coombs' test, Neutralizing antibodies	High	Clearing and neutralizing
MAb16	Negligible	Allergic reaction in 1 patient	Low	Clearing
MAb17	Negligible	None	Low	None
MAb18	Tolerable	Neutralizing	High	Clearing
MAb19	Tolerable	None	Not available	Not available
MAb20	Tolerable	Clearing	Intermediate	Clearing
MAb21	Tolerable	None	Low	Neutralizing and anaphylaxis
MAb22	Negligible	Neutralizing and hypersensitivity	Intermediate	Clearing and neutralizing
MAb23	Tolerable	Clearing	Intermediate	Clearing
MAb24	Tolerable	Possible role in inflammation	Intermediate	Perivascular sheathing, increased exposure
MAb25	Negligible	None	Low	Clearing
Human antibodies				
MAb26	Negligible	Unknown	Intermediate	Anemia
MAb27	Tolerable	Neutralizing and binding	Intermediate	Clearing
MAb28	Negligible	None	High	Neutralizing
MAb29	Negligible	None	Low	None
MAb30	Tolerable	Neutralizing	Low	Clearing
MAb31	Negligible	None	Low	None
MAb32	Tolerable	Neutralizing and infusion reactions	Intermediate	Clearing
MAb33	Tolerable	Neutralizing	Intermediate	Clearing

switch to a new treatment(*23*). Immunogenicity of single use products such as diagnostics is generally not an issue, but it should be considered that when ADA to the diagnostic agent develops, they can negatively influence the imaging. Immunogenicity can also result in profound adverse effects after only few administrations. For example, thrombocytopenia is seen in 1% of patients treated with abciximab which is caused by antibodies specific to the murine derived CDR regions of abciximab. The incidence of this effect could be increased fourfold after administration of abciximab to patients a second time(*24*).

Interestingly, the level of humanization did not appear to influence the ADA incidence in humans (Table 4). This is surprising because the aim of humanization is to reduce immunogenicity. The merits of humanization have been questioned before(*25*). An interesting hypothesis put forward by Clark suggests that the basic idea of humanization to create self-like-proteins is flawed because every B cell clone with a unique specificity also has a unique V-region sequence, and it not likely that tolerance to each clone exists for every new sequence. A complete converse immunological concept is that B cell clones provoke anti-idiotype responses forming an antibody network that regulates immune responses. An equilibrium of these opposites more reflects the real situation. Therefore, immunogenicity and humanization of the variable region are not necessarily correlated and decreasing immunogenicity is not a simple matter of increasing the sequence homology(*25*, *26*). There also appeared to be little difference between the relative immunogenicity of chimeric, humanized, and human mAbs in NHPs. This is not unexpected because mAbs are probably readily recognized as non-self in NHPs because of species differences in major histocompatibility complex classes and T cell subsets. Most ADA responses in NHPs were directed against the Fc-region (anti-isotype) of the mAbs, resulting in enhanced clearance. In some cases, loss of efficacy and adverse effects were reported after the induction of ADAs. Conversely, in humans, antibodies were more often directed against the CDR (anti-

Table 4. Immunogenicity in non-human primates vs. human

			Immunogenicity in NHP		
		Not evaluated in NHP	Low	Intermediate	High
Immunogenicity in clinical trials	Low	Murine Murine Murine	Humanized Humanized Humanized Humanized Human Human	Humanized Human	Humanized Human
	Intermediate	Humanized	Humanized Human	Murine Chimeric Chimeric Humanized Humanized Humanized Human Human Human	Murine Chimeric Chimeric Humanized
	High	Murine Murine	Chimeric		Murine

idiotype), resulting in neutralization of the function of the antibody and loss of efficacy. This may occur because the CDR, which is a unique sequence, is the most foreign region of a mAb in humans, whereas both the CDR and Fc regions are foreign in NHPs.

Because it is difficult to compare directly the immunogenicity of different products in different species, their relative immunogenicity is used to make between-species and -product comparisons. Even with this type of normalization, immunogenicity in laboratory animals is not considered predictive of immunogenicity in humans(*18, 27, 28*). This was confirmed by our analysis, with only 59% of the tested antibodies having comparable incidence of immunogenicity in NHPs and humans. While the incidence of ADAs was comparable in some cases, immunogenicity in NHPs over-predicted the immunogenicity of 30% of the mAbs and underpredicted the immunogenicity of 11% of the mAbs in humans. Bugelski and Treacy established immunogenic classes of recombinant therapeutic proteins based on their source(*11*). Prokaryotic and mammalian protein would have very low homology to human protein and these are generally highly immunogenic. Other classes were novel constructs and chimeric, humanized or human antibodies. The homology of these proteins is high, although immunogenicity could be variable. And even NHP studies had limited ability to predict immunogenicity with a trend to over-predict, despite the perceived extensive identity of V-regions (93% at amino acid level for the VH-framework regions and 88-99% for V_K of cynomolgus monkeys to human antibody sequences). Similarly, comparison of macaque V-regions with that of humans revealed identity between 84% and 97%; however, here differences were considered to possibly result in increased immunogenic response(*29*).

The presence of ADAs often interfered with the assay used to detect mAb concentrations in serum. Therefore, pharmacokinetic or safety data should be interpreted with caution. Improvements in assay design could partially overcome some of these difficulties(*30, 31*); however, proper validation of assays is impeded by the lack of relevant reference standards, and other animal species are often used as controls, which makes the assay less specific. Animals are also the source of antisera needed to develop and validate assays for antibodies to be used during clinical development.

Safety and dose-finding studies make use of laboratory animals, but the potential for immunogenicity complicates the interpretation of kinetic and toxicity data, especially because the development of ADAs can alter or abolish exposure and in some cases result in loss of efficacy. Immunogenicity is not a problem if study groups are of sufficient size and a sufficient number of animals do not develop ADAs, but ethical constraints typically limit NHP studies to small sample sizes. This means that safety studies are particularly difficult to interpret if most or all of the NHPs develop a significant ADA response. For example, in the case of MAb18 the majority of NHPs developed clearing ADAs after the first dose, which affected the interpretability of study results. Therefore, the value of using NHPs during non-clinical drug development will be limited if there is a significant immune response to the test substance. Continuing the study beyond this point will not yield relevant data, and subsequent long-term studies should be reconsidered(*32*).

Our study had some limitations. Grouping the NHP data into three operative categories is a necessary over-simplification of immunogenicity. In addition, the various studies differed in their reporting of the rate and effect of ADA development. We used the scale established

by Hwang and Foote to classify the immune response, and as the immunogenicity of mAbs is probably higher in animals (the mAbs are foreign) than in humans, we chose to increase the ranges threefold(33). This choice could be considered arbitrary; however, higher or lower ranges would lead to either an under- or over-estimation of immunogenicity in NHPs, respectively. Even though our dataset included all mAbs approved in the EU through 2010, there were not enough samples to perform statistical analyses. Therefore, we could only observe and describe trends. Lastly, we only investigated mAbs that received marketing authorization. Inclusion of mAbs that failed during drug development or regulatory review would have provided a larger study cohort, but sufficient immunogenicity data for these mAbs are not publically available. In conclusion, the results of this study suggest that the immunogenic response in NHPs is poorly predictive of the response in humans, even when using broad categories of immunogenicity. The development of clearing or neutralizing antibodies against the test mAb in NHPs might limit exposure or the duration of repeated dose studies, which in turn can influence the reliability and interpretability of pharmacokinetic, pharmacodynamic and safety data. Lastly, it is difficult to compare the immunogenicity across products and species because of species differences and limits in assay technology. Therefore, NHPs may not be a suitable species for testing mAbs that are immunogenic in NHP, even if these are the only species available.

STUDY METHODS

The drug registration files of mAbs and immunoglobulin fragments approved in the EU through 2010 (Table 1) were accessed at the Dutch Medicines Evaluation Board. The non-clinical summary and overview, including the tabulated study reports, were used to evaluate the immunogenicity of the mAbs in NHPs. This evaluation was done by assessing the presence of antidrug antibodies in serum from mAb-treated NHPs. Reporting of immunogenicity in animal studies is not standardized in marketing authorization applications, which could be quantitative or qualitative. To normalize the data and to enable comparison between products and species, the incidence of ADAs was classified into three categories, namely, low, intermediate, or high. The incidence of ADAs was scored regardless of titer, time of occurrence, and persistence of the response. The incidence was scored as 'Low' when no or fewer than 6% of animals were positive for ADAs, 'Moderate' when between 6% and 45% of animals were positive for ADAs, and 'High' when more than 45% of the animals were positive for ADAs. These categories were based on those established by Hwang and Foote and were three times higher than similar categories in humans based on higher baseline immunogenicity of NHPs(*33*). Three authors individually categorized immunogenicity of the mAbs. When opinions deviated, the mAbs were discussed to come to a final distribution. If changes in pharmacokinetic or pharmacodynamic profiles occurred, the type and incidence of these changes were recorded. Lastly, ADA incidence and the rate of pharmacokinetic or pharmacodynamic changes were aggregated in a 3x3 matrix, with ADA incidence in the rows and rate of pharmacokinetic or pharmacodynamic changes in the columns (Table 2).

The Summary of Product Characteristics (SPC) or the European Public Assessment Report (EPAR) was used to determine the immunogenicity of therapeutic mAbs in humans,

defined as the proportion of patients in Phase 3 clinical trials that were positive for ADAs. When these could not be retrieved other online sources, such as company websites or web databases were used. The categories established by Hwang and Foote were used: negligible if immunogenicity was seen in 2% of patients or less, tolerable when the incidence was between 2% and 15%, and marked if immunogenicity occurred in more than 15% of the patients. Changes in efficacy mentioned in the SPC or EPAR as a result of the development of ADAs were noted. Immunogenicity in NHPs was considered predictive if human and NHP immunogenicity fell into the same operative category. MAbs evaluated in NHP were anonymized.

CONFLICT OF INTEREST

P.J.K. van Meer, M. Kooijman, V. Brinks, C. Gispen-de Wied, B. Silva-Lima and E.H.M. Moors declare no conflict of interest. H. Schellekens 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

This research was performed under the framework of Top Institute Pharma, (project T6-301), which includes the Medicines Evaluation Board, Nefarma (the association for innovative medicines in The Netherlands), the Life Science and Health Initiative, and Utrecht University. PvM and MK performed this work with the TI Pharma grant. 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.

REFERENCES

1. P. Baldrick, Safety evaluation of biological drugs: what are toxicology studies in primates telling us? *Regul Toxicol Pharmacol* 59, 227 (Mar, 2011).
2. H. Schellekens, Factors influencing the immunogenicity of therapeutic proteins. *Nephrol Dial Transplant* 20 Suppl 6, vi3 (Jun, 2005).
3. H. Schellekens, Immunogenicity of therapeutic proteins: clinical implications and future prospects. *Clin Ther* 24, 1720 (Nov, 2002).
4. M. Kessler, D. Goldsmith, H. Schellekens, Immunogenicity of biopharmaceuticals. *Nephrol Dial Transplant* 21 Suppl 5, v9 (Oct, 2006).
5. G. J. Wolbink, L. A. Aarden, B. A. Dijkmans, Dealing with immunogenicity of biologicals: assessment and clinical relevance. *Curr Opin Rheumatol* 21, 211 (May, 2009).
6. E. Koren *et al.*, Clinical validation of the "in silico" prediction of immunogenicity of a human recombinant therapeutic protein. *Clin Immunol* 124, 26 (Jul, 2007).
7. A. S. De Groot, P. M. Knopp, W. Martin, De-immunization of therapeutic proteins by T-cell epitope modification. *Dev Biol (Basel)* 122, 171 (2005).
8. F. A. Harding, M. M. Stickler, J. Razo, R. B. DuBridge, The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. *MAbs* 2, 256 (May-Jun, 2010).

9. P. Gaitonde, S. V. Balu-Iyer, *In vitro* immunogenicity risk assessment of therapeutic proteins in preclinical setting. *Methods Mol Biol* 716, 267 (2011).
10. F. R. Brennan *et al.*, Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies. *MAbs* 2, 233 (May-Jun, 2010).
11. P. J. Bugelski, G. Treacy, Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Curr Opin Mol Ther* 6, 10 (Feb, 2004).
12. J. Descotes, Immunotoxicity of monoclonal antibodies. *MAbs* 1, 104 (Mar-Apr, 2009).
13. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Preclinical safety evaluation of biotechnology derived pharmaceuticals S6(R1). *(online)*, (2011).
14. Committee for medicinal products for human use, Guideline on immunogenicity assessment of biotechnology-derived therapeutic proteins. (2008).
15. V. Brinks, W. Jiskoot, H. Schellekens, Immunogenicity of therapeutic proteins: the use of animal models. *Pharm Res* 28, 2379 (Oct, 2011).
16. D. Wierda, H. W. Smith, C. M. Zwickl, Immunogenicity of biopharmaceuticals in laboratory animals. *Toxicology* 158, 71 (Feb 2, 2001).
17. R. Ponce *et al.*, Immunogenicity of biologically-derived therapeutics: assessment and interpretation of nonclinical safety studies. *Regul Toxicol Pharmacol* 54, 164 (Jul, 2009).
18. S. J. Swanson, J. Bussiere, Immunogenicity assessment in non-clinical studies. *Curr Opin Microbiol* 15, 337 (Jun, 2012).
19. B. Gorovits, Antidrug antibody assay validation: industry survey results. *AAPS J* 11, 133 (Mar, 2009).
20. G. M. Bartelds *et al.*, Development of antidrug antibodies against adalimumab and association with disease activity and treatment failure during long-term follow-up. *JAMA* 305, 1460 (Apr 13, 2011).
21. M. K. de Vries *et al.*, Decreased clinical response to infliximab in ankylosing spondylitis is correlated with anti-infliximab formation. *Ann Rheum Dis* 66, 1252 (Sep, 2007).
22. J. D. Isaacs *et al.*, CAMPATH-1H in rheumatoid arthritis--an intravenous dose-ranging study. *Br J Rheumatol* 35, 231 (Mar, 1996).
23. H. Schellekens, The immunogenicity of therapeutic proteins. *Discov Med* 9, 560 (Jun, 2010).
24. B. R. Curtis, J. Swyers, A. Divgi, J. G. McFarland, R. H. Aster, Thrombocytopenia after second exposure to abciximab is caused by antibodies that recognize abciximab-coated platelets. *Blood* 99, 2054 (Mar 15, 2002).
25. D. R. Getts, M. T. Getts, D. P. McCarthy, E. M. Chastain, S. D. Miller, Have we overestimated the benefit of human(ized) antibodies? *MAbs* 2, 682 (Nov-Dec, 2010).
26. M. Clark, Antibody humanization: a case of the 'Emperor's new clothes'? *Immunol Today* 21, 397 (Aug, 2000).
27. I. C. Buttel *et al.*, Taking immunogenicity assessment of therapeutic proteins to the next level. *Biologicals* 39, 100 (Mar, 2011).
28. E. Koren, L. A. Zuckerman, A. R. Mire-Sluis, Immune responses to therapeutic proteins in humans--clinical significance, assessment and prediction. *Curr Pharm Biotechnol* 3, 349 (Dec, 2002).
29. P. Thullier, S. Chahboun, T. Pelat, A comparison of human and macaque (Macaca mulatta) immunoglobulin germline V regions and its implications for antibody engineering. *MAbs* 2, 528 (Sep-Oct, 2010).
30. G. Shankar, C. Pendley, K. E. Stein, A risk-based bioanalytical strategy for the assessment of antibody immune responses against biological drugs. *Nat Biotechnol* 25, 555 (May, 2007).
31. M. Wadhwa, R. Thorpe, Strategies and assays for the assessment of unwanted immunogenicity. *J Immunotoxicol* 3, 115 (Sep 1, 2006).
32. K. Chapman *et al.*, Preclinical development of monoclonal antibodies: considerations for the use of non-human primates. *MAbs* 1, 505 (Sep-Oct, 2009).
33. W. Y. Hwang, J. Foote, Immunogenicity of engineered antibodies. *Methods* 36, 3 (May, 2005).

Wisdom comes from experience. Experience is often a result of lack of wisdom.

-Terry Pratchett

6 CONTRIBUTION OF ANIMAL STUDIES TO EVALUATE THE SIMILARITY OF BIOSIMILARS TO REFERENCE PRODUCTS

Peter J.K. van Meer, Hans C. Ebbers, Marlous Kooijman,
Christine C. Gispen-de Wied, Beatriz Silva-Lima,
Ellen H.M. Moors, Huub Schellekens

To be submitted for publication

ABSTRACT

The European Union was the first region to establish a regulatory framework for biosimilars. Animal studies, although limited, are needed according the EU guidelines to confirm similarity to a reference product in terms of pharmacodynamics, potency and safety. Animal studies however did not identify biological differences between biosimilars and the reference product, despite differences in quality. In addition, the animal studies included in refused and withdrawn biosimilar applications did not reflect the differences identified in clinical trials. Our analysis shows that animal studies lack the sensitivity to confirm biosimilarity.

Key words: biosimilars, biosimilarity exercise, non-clinical, animal studies, pharmacodynamics, repeat dose toxicity, regulatory science, guidelines

INTRODUCTION

When small molecule drugs go off-patent, generic copies are allowed on the market after these copies are shown to be pharmaceutically equivalent to the originator product and their bioavailabilities lie within acceptable predefined limits. For small molecules, pharmaceutical equivalence is established through physico-chemical characterization, while bioavailability is usually demonstrated in randomized, crossover studies in a limited number of subjects(*1*). Animal studies are not required. Medicinal products of recombinant biotechnology are relatively large and complex proteins that are difficult to fully characterize(*2*). In addition, biopharmaceuticals are complex mixtures of closely related molecules, the exact composition of which is dependent on the product's manufacturing process resulting in differences in, for example, protein aggregation and glycosylation patterns(*3, 4*). In addition, it these differences may influence the pharmacodynamics, kinetics or toxicity parameters(*5-8*). Therefore, tailored regulatory requirements have been developed for the authorization of competing version of biopharmaceuticals, so-called biosimilars.

Compared to small molecule therapeutics, more studies are required to establish the similarity of the biosimilar in terms of quality, safety and efficacy compared to a reference product to allow marketing authorization(*9*). The European Union (EU) was the first region to adopt legislation that allows the registration of biosimilars based on an abbreviated marketing authorization application(*10*). Overarching guidelines have been released that lay down, quality, non-clinical and clinical issues for biosimilars(*11-13*). In addition product specific guidelines have been released(*14-22*). In the current guideline on non-clinical and clinical issues of biosimilars, emphasis is placed on performing comparative non-clinical studies that are sensitive enough to detect differences between the biosimilar and the reference product. This case-by-case approach effectively limits the number of animal studies needed (Table 1)(*12*). Recently, a concept paper on the revision of the biosimilar guideline was released in which a risk-based approach is suggested, opting for fewer or perhaps even no animal studies(*23*). However, it remains unclear what the contribution is of animal studies in establishing biosimilarity. Therefore, we have assessed non-clinical animal study programs of the marketing authorization applications of all biosimilar products registered in the EU or submitted for authorization until January 1st 2013.

Table 1. Animal studies recommended to demonstrate biosimilarity in the current guideline(*12*)

Type of study	Type of investigation	Necessity of study
Pharmacodynamics	Pharmacodynamic effect or activity relevant to clinical application	Recommended
Repeat dose toxicity	Toxicity profile including toxicokinetics and determination of antibody titres, cross reactivity and neutralizing capacity	Recommended
Specific concerns	If relevant (e.g. local tolerance), addressed in the same repeat dose toxicity	Recommended
Routine toxicological studies	Safety pharmacology, reproduction toxicology, mutagenicity, carcinogenicity	Not recommended unless indicated of results of repeat dose studies

METHODS

European Public Assessment Reports (EPARs) were obtained from the website of the European Medicines Agency (EMA, www.ema.europa.eu). Data from the EPAR were supplemented with data from Marketing Authorization Applications (MAAs) and scientific advice for registered biosimilar products. MAAs and scientific advice were obtained from the database of the Medicines Evaluation Board (CBG-MEB, Utrecht, The Netherlands). Non-clinical studies were tabulated and categorized by type of study and inclusion of a reference group. Categories for study type were ordered by studies recommended in the biosimilar guideline, including pharmacodynamics, repeat dose toxicity and local tolerance or studies not recommended in the biosimilar guideline such as single dose toxicity, safety pharmacology, repeat dose pharmacokinetics or toxicokinetics, developmental and reproductive toxicity studies including embryo-fetal development and peri- and postnatal development, immunogenicity studies and special toxicity studies. Only scientific advice pertaining to the non-clinical development or animal studies was included. The manufacturer's questions and the responses of the Committee for Medicinal Products for Human Use (CHMP) answer were analyzed.

OVERVIEW OF BIOSIMILAR APPLICATIONS

Fourteen biosimilars, which were all copies of endogenous proteins, were registered in the EU up to 2012. Several biosimilars are registered under different trade names and effectively, seven biosimilar applications have been developed (Table 2). Alpheon (interferon α-2a) was

Table 2. Biosimilars currently on the EU market

Medicine Name	Active Substance	Development code	Status	Authorisation date
Omnitrope(*29*)	Somatropin	EP2000	Authorised	12-4-2006
Valtropin*(*30*)	Somatropin	LBD-009	Withdrawn	24-4-2006
Binocrit(*31*)	Epoetin alfa	HX-575	Authorised	28-8-2007
Abseamed(*32*)	Epoetin alfa	HX-575	Authorised	28-8-2007
Epoetin alfa Hexal(*33*)	Epoetin alfa	HX-575	Authorised	28-8-2007
Retacrit(*34*)	Epoetin zeta	SB309	Authorised	18-12-2007
Silapo(*35*)	epoetin zeta	SB309	Authorised	18-12-2007
Ratiograstim(*36*)*	Filgrastim	XM-02	Authorised	15-9-2008
Biograstim(*37*)	Filgrastim	XM-02	Authorised	15-9-2008
Tevagrastim(*38*)	Filgrastim	XM-02	Authorised	15-9-2008
Filgrastim ratiopharm(*39*)	Filgrastim	XM-02	Withdrawn	15-9-2008
FilgrastimHexal(*40*)	Filgrastim	EP2006	Authorised	6-2-2009
Zarzio(*41*)	Filgrastim	EP2006	Authorised	6-2-2009
Nivestim(*42*)	Filgrastim	PLD-108	Authorised	8-6-2010

*Marketed in the USA as Valtropin (filgrastim) via a new drug application and Ratiograstim as Tbo-Filgrastim (filgrastim) via a biological licence application

refused market entry and the marketing authorization application for Insulin Human Rapid Marvel (insulin) was withdrawn by the manufacturer (*24, 25*). The manufacturer of Epostim (epoetin) withdrew its biosimilar application following a request from CHMP for additional data(*26*). Filgrastim Ratiopharm (filgrastim) and Valtropin (somatropin) received marketing authorization, but were withdrawn from the market for commercial reasons in April 2011 and May 2012 respectively. Neither of these products was marketed in any EU country(*27, 28*).

The submission of the marketing authorization application of Ratiograstim was intended as a full application, but was registered in the EU as a biosimilar. Similarly, for the marketing authorization application of Valtropin the manufacturer referred to the biosimilar guideline for the development of their product(*30*). However, the submitted non-clinical package was a full application containing a safety pharmacology study in non-human primates, developmental and reproductive toxicity studies, single and repeat dose pharmacokinetic and toxicity studies (in a rodent and non-rodent animal model), and genotoxicity and antigenicity studies. Extensive direct comparisons between Valtropin and its reference product were not done on the basis of comparable pharmacological activity in parallel pharmacodynamics studies.

IN VITRO COMPARABILITY

In addition to the thorough physicochemical characterization of the biosimilar protein required to evaluate the quality of the product, *in vitro* receptor binding assays and biological activity assays were performed for Binocrit, Retacrit, Filgrastim Hexal, Nivestim and Ratiograstim (Table 3). Similarity of the biosimilar to the reference product of both receptor binding affinity and *in vitro* potency measured by various bioassays was established for all these products.

Table 3. *In vitro* similarity exercise

Product	Similar	Results
Omnitrope	Not known	No *in vitro* potency assays were available, only bioassays.
Valtropin	Not known	No *in vitro* potency assays were available, only bioassays.
Binocrit	Yes	Characterization of receptor binding and signal transmission. Dose response curves of reference and biosimilar were similar.
Retacrit	Yes	Receptor binding, proliferation and second messenger activation was evaluated. Similarity at the level of receptor binding as well as in functional assays was demonstrated.
Filgrastim Hexal	Yes	*In vitro* potency of the product samples was comparable to that of the reference product. Comparative receptor binding studies showed similar affinity to the receptor.
Nivestim	Yes	A bioassay and receptor binding assay was performed to evaluate PD. Receptor binding affinity and effects of the biosimilar and the reference product were similar.
Ratiograstim	Yes	Relative receptor binding and biological activity showed similar binding affinities of the biosimilar and reference product and are equally effective in inducing cellular proliferation.

USE OF ANIMALS

We identified that 5593 animals were used in 59 studies for seven distinct non-clinical biosimilarity exercises, including studies required according to the European or United States Pharmacopoeia monographs. Most animals were rodent (rat and mouse) and to a lesser extent guinea pigs and rabbits. Dog was predominantly used as a non-rodent species. For two products, 6 studies included non-human primates; 104 cynomolgus and 12 rhesus monkeys. A median of 8 (interquartile range, IQR 4-12) animal studies were done to support demonstration of comparability with a minimum of 3 and a maximum of 24 (Table 4). For all but one product, the EPARs list the same non-clinical studies that are described in the dossier.

Table 4. The number of animal studies in a non-clinical program according to the marketing authorization application and alignment of the non-clinical program through regulatory dialog via scientific advice.

Product	Number of studies (in EPAR)	Scientific advice	Non-clinical program sufficient
Omnitrope	8	Yes	Yes
Valtropin	24	No	No scientific advice was given
Binocrit	4	Yes	Yes, but shorter study in lower species would also be sufficient
Retacrit	4	Yes	No, animal studies are recommended in the guideline and can only be omitted if equivalence can be demonstrated
Ratiograsatim	12	Yes	No, local tolerance and immunogenicity assessment should be included, a full complement of non-clinical animal studies not needed for biosimilar application
Filgrastim Hexal	4	Yes	Yes, but *in vitro* data should be included
Nivestim	3	Yes	Yes, but *in vitro* data should be included

PHARMACOLOGICAL SIMILARITY *IN VIVO*

Pharmacological biosimilarity was almost always demonstrated (Table 5). Only in the case of Ratiograstim, one PD study in a rodent disease model did not show similar potency between the biosimilar and reference compound. However, there was a tendency towards comparable potencies in terms of the *in vivo* biological activity(*36*). In animals receiving low and high dose biosimilar filgrastim, male animals had a statistically significant lower CD4+ counts at the end of the first period than males treated with the reference product. In repeat dose studies however, results were similar.

TOXICOLOGICAL SAFETY PROFILE AND LOCAL TOLERANCE *IN VIVO*

Biosimilar products were well tolerated. Local tolerance studies generally did not include a reference group, but for those studies that did, no differences were observed.

Table 5. Similarities and differences in biosimilarity in pharmacodynamics and repeat dose toxicity studies

Product	Study	Similarity is demonstrated	Note
Binocrit	PD	Yes	No differences in PD.
Binocrit	RDT	Yes	Comparable PD effects, no unexpected findings or differencs in safety profile. Marked increase in LDH in male dogs given biosimilar at week13.
Retacrit	PD	Yes	Increase in PD marker without consistent differences in PD effects and within 80-125% of stated reference potency. Marked within-group variability was observed in all drug treated groups.
Retacrit	RDT	Yes	Substantial treatment related mortality in the high dose groups of both biosimilar and reference compound and was comparable. Dosing had to be stopped prior to the end of the study. Toxicology findings were comparable, but biosimilar caused slight reduction in uriniary specific gravity and significant increase in urine volume in males after 6 weeks compared to the reference.
Ratiograstim	PD	Yes	In a PD animal model of disease, possible differences in potency were measured. Statistical meta-analysis revealed inconsistencies. Both biosimilar and reference product induced expected pharmacological effects to a similar extent in the animal disease model with a tendency towards comparable potencies. In healthy animals, low and high dose biosimilar treated male animals had a statistical significant lower mean outcome for a PD marker at the end of the first period than males treated with the reference product.
Ratiograstim	RDT	No comparator	A higher ADA response to the biosimilar compared to the reference product in the lowest dose at the end of the study, but a lower ADA response in the highest dose at both time points.
Omnitrope	PD	Yes	No significant difference.
Omnitrope	RDT	Yes	Comparable to other products on the market.
Filgrastim Hexal	PD	Yes	No differences in PD.
Filgrastim Hexal	RDT	Yes	Similar findings were obtained for the biosimilar and reference product.
Nivestim	PD	Yes	No differences in PD.
Nivestim	RDT	Yes	No differences in toxicology and observed effects are similar between biosimilar and reference. No unexpected toxicology.
Valtropin	PD	Yes	Potency of the biosimilar was within the established limits although the overall potency was lower for the biosimilar which was considered to be a result of the variability of the PD assay.
Valtropin	RDT	No comparator	Primate studies had high inter-group variability although no ADA was detected.

Safety profiles of biosimilars obtained from repeat dose toxicity studies showed dose dependent expected pharmacology mediated adverse effects (Table 5). The repeat dose studies never resulted in unexpected or new toxicities for the biosimilar or the reference product. One manufacturer also examined safety via the oral route.

The safety profiles of biosimilars were generally considered similar in incidence and response to that of the reference compound. Only in a few cases, minor differences were

observed. In male, but not female, dogs treated with Binocrit, there was a marked increase in lactate dehydrogenase (LDH), a marker for hemolysis. Male, but not female, rats treated with Retacrit had slight reductions in urinary specific gravity and significant increase in urine volume and a significant difference in values for reticulocytes and red blood cells between animals in the high dose groups of the biosimilar and reference product after 6 weeks. These results were not seen in dog and it was concluded that it was likely the result of higher potency of the biosimilar product. Variability inherent to animal studies was also considered as a source of these differences, which were not considered to be toxicologically meaningful(*34*).

The assessment of antibody formation to the biosimilar during repeated administration was assessed and in general the incidence of (neutralizing) anti-drug antibody formation was similar for the biosimilar and its reference product. In ratiograstim treated animals, there was a higher ADA response in the lowest dose at the end of a repeat dose study study compared to the reference product, but a lower ADA response in the highest dose at both time points compared to the reference products. It was considered that this was likely a reflection of the normal inter-animal variability, rather than being a real difference in immunogenicity between the two products.

Every non-clinical program included at least one pharmacodynamics study with a reference compound, a repeat dose study with a reference compound and local tolerance study (Table 6). For three biosimilars, the local tolerance study did not include a reference compound group. 59 animal studies have been done, using 5593 animals. None of these studies showed differences that precluded clinical trials and overall, all animal studies supported the claim of biosimilarity.

Table 6. Types of animal studies conducted to support the biocomparability exercise

Type of study	Programs that included a study	Number of studies
Recommended		
Pharmacodynamics*	7	15
Repeat dose toxicity	7	11
Local tolerance	7	8
Not recommended		
Single dose toxicity	2	8
Safety pharmacology	2	3
Repeat dose pharmacokinetics or toxicokinetics	3	5
Developmental and reproductive toxicity studies⁺	1	4
Special toxicity	1	1
Immunogenicity/antigenicity	2	4

SCIENTIFIC ADVICE

All but one out of seven products specified that scientific advice was sought on non-clinical issues in their MAA or the EPAR. Interestingly, the manufacturer of Retacrit asked about the necessity of any pre-clinical studies as most endpoints could be studied in humans as well(*34*). The CHMP's opinion was that, in principle, they agreed. However, omission of safety testing in animals would depend on demonstration of a number of characteristics and being equivalent to the reference product. Additional studies, *in vitro* and *in vivo*, were requested, which were subsequently performed by the manufacturer. The sufficiency of the non-clinical testing strategy was discussed for all other programs that requested scientific advice (Table 4). Other questions in scientific advice dealt with more specific issues related to study design, regulatory requirements or suggestions to reduce regulatory requirements. Most non-clinical programs were considered sufficient by the CHMP although comments were made or additional *in vivo* studies were requested for 4 products (including Retacrit). Manufacturers were always compliant with these requests. Conversely, regulatory authorities have only rarely been non-compliant with the given scientific advice. For example, regulatory authorities did not consider a pharmacodynamics assay to be sensitive enough and requested another, additional pharmacodynamics study, even though the program was initially considered adequate during scientific advice. Manufacturer positions were generally endorsed and when this was not the case, recommendations were generally made to adapt the suggested development program. For instance, when one manufacturer suggested using only one dose group of the reference compound, the CHMP considered it necessary to include more than one dose because the purpose of the repeated dose study was also to compare PK, PD and unexpected toxicity. Establishing a dose response was considered important for a proper comparison, although for PK a single dose group would suffice. Nevertheless, the recommendation was to study the PD of the reference compound at three dose levels similar to the test compound. Another example concerns a manufacturer of a similar product who suggested using only two dose groups for their repeat dose study, which was accepted by the CHMP. Manufacturers have also overestimated regulatory expectations. This is illustrated by a case where a 13-week repeat dose toxicity study in dogs had been conducted and was considered adequate to fulfill the regulatory requirements. But regulatory authorities did suggest that a 4-week rat repeat dose toxicity study would have also produced data of equal relevance.

REFUSED OR WITHDRAWN BIOSIMILARS

Three products were reviewed by the EMA and were either withdrawn by the applicant or refused by the CHMP. Alpheon (interferon alfa 2a) was refused on the European market. The major objections to the claim of biosimilarity were based on the quality and clinical dossier. Non-clinical studies were also evaluated, but not considered to be a major objection. Animal data was not considered to allow a clear conclusion regarding similarity of the biosimilar and the reference product because of a number of deficiencies in the submitted 4-week repeat dose toxicity study in monkeys. Biosimilar insulin, produced by Marvel, was withdrawn by

the manufacturer because there were major objections based on quality and clinical data. The non-clinical data package included extensive *in vitro* PD studies in isolated cell lines to evaluate binding activity, functional binding and activation of downstream proteins. Animal studies consisted of single and repeat dose toxicity studies in rodent and a local tolerance study in rabbit. Marvel's insulin was well tolerated without unexpected or new toxicity. However, because of limitations in the *in vitro* assays and in the design of the animal studies, the claim of biosimilarity was considered invalid.

The marketing authorization application of Epostim was withdrawn by the manufacturer because major objections were raised by the EMA. The epoetin copy was not considered to be comparable to the reference product in terms of quality and potency. Furthermore, consistency of manufacturing was not demonstrated. It was recommended not to request additional non-clinical clarification until the major quality issues were resolved. Repeat dose studies in rat and dog showed different PK parameters and potential formation of neutralizing anti-drug antibodies in one species, but this was not characterized. Overall, the studies did not reveal significant differences in pharmacokinetics. Safety profiles of the biosimilar and the reference products were similar. However, mortality of animals of the biosimilar was higher.

DISCUSSION

The marketing authorization applications of all registered, refused and withdrawn biosimilars were used for this study. Our results show that none of the animal studies submitted within biosimilar applications showed relevant differences to the reference product. The animal data contribute little to the overall biosimilarity exercise. The clinical relevance of the differences that are detected is questionable and did not preclude clinical trials and invariably, animal data supported the claim of similarity for marketed biosimilars. Our data suggest that animal studies are likely not sensitive enough to detect these differences or translate them into measurable endpoints because of the inherent variability of animal models and the variability of assays.

Even in products with quality differences between the reference and biosimilar product in terms of level of impurities, host cell type, formulation or post translational modification, all of the animal studies supported biosimilarity, while differences were observed in clinical PK profiles(*43*). In addition, considerable differences in potency, content and isoform profile have been reported for several epoetins, including biosimilar epoetins(*44*).

The unique properties of biopharmaceutical copies called for new regulatory guidance to establish similarity to the products already on the market, the first of which came into effect in the EU in 2004. In the early regulatory guidelines, relatively extensive animal studies have been considered essential to determine similarity in potency, pharmacological effects and safety of new biosimilars. However, most, if not all, companies started the development of their biosimilar before the first biosimilar guideline appeared. Non-clinical strategies focused on existing guidelines for biopharmaceuticals, such as ICH S6, common scientific sense and scientific advice of regulatory authorities(*45*). For both Valtropin and Ratiograstim, a full package of non-clinical studies was done, which may have been because both products were also submitted for marketing in the US and Valtropin in Korea as well.

Scientific advice was sought for almost all biosimilar products and the non-clinical program of the manufacturer was the starting point of these discussions. The type of questions predominantly focused on whether the non-clinical study package was sufficient. The responses of the regulatory authority were generally affirmative, although in some cases additional studies were requested to meet the requirements set out in the biosimilars guideline. Manufacturers have proposed omitting certain animal studies, using a single species, route of administration or using fewer dose groups. In some cases this resulted in reduced, scientifically justified non-clinical development plans. Therefore, scientific advice allowed companies to align their development strategies with regulatory expectation. In addition, the regulatory dialogue was driven by experience of biosimilar development of the pharmaceutical industry and has increased the knowledge of regulatory authorities. This may have facilitated the revision of the current biosimilar guidelines. However, earlier contact with regulatory authorities through scientific advice could have increased regulatory intelligence even further and should be considered for future development strategies.

Alternative strategies to evaluating similarity of new biotherapeutics copies have been proposed(*43*). For example, the regulatory demand for comparative studies should be foregone because comparative clinical trials to demonstrate similarity preclude informative data due to practical consequences such as undefined acceptance ranges for pharmacokinetic parameters(*43*). As our data concerning the refused and withdrawn applications show, inconsistencies in quality, manufacture and control of the process between the biosimilar and the reference product are also reflected in animal studies and clinical trials. This further stresses the use of a step-wise, risk-based approach to demonstrate biosimilarity and full characterization of the product in terms of quality, manufacture and control is the starting point for any subsequent studies and can be facilitated by continuing improvements of analytical and *in vitro* studies(*46*).

The first generation biosimilars, which are copies of endogenous proteins, were developed in the absence of detailed regulatory guidance. Manufacturers faced unknown regulatory requirements and detailed guidance on the development of biosimilars. Where at first non-clinical animal studies were always requested to demonstrate the similarity of a biosimilar to its reference product, in new draft documents of guidelines for biosimilar development, the limited relevance of animal studies is acknowledged(*23*). The next generation of biosimilars includes exogenous proteins, such as monoclonal antibodies. These are highly species specific products and the non-human primate is often the only available relevant species(*47*). But the relevance of studies with biosimilar applications with these animals is doubtful(*48, 49*). The concept paper may reflect the current European regulatory thinking on the limited contribution of animal studies with regard to the next generation of biosimilars(*23*).Therefore, the critical re-evaluation of the current biosimilar guideline is timely. Indeed, the draft guidances on biosimilars which were released for consultation in May 2013, emphasize a step-wise development of biosimilars where non-clinical animal studies are even further limited, if any are to be done at all depending on preceding steps(*50, 51*). Such reassessments should be initiated for all non-clinical guidelines.

CONFLICT OF INTEREST

P.J.K. van Meer, H. Ebbers, M. Kooijman, C. Gispen-de Wied, B. Silva-Lima and E.H.M. Moors declare no conflict of interest. H. Schellekens 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

This research was performed under the framework of Top Institute Pharma, (project T6-301), which includes the Medicines Evaluation Board, Nefarma (the association for innovative medicines in The Netherlands), the Life Science and Health Initiative, and Utrecht University. PvM and MK performed this work with the TI Pharma grant. 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.

REFERENCES

1. European Medicines Agency, Guideline on the investigation of bioequivalence. (2010).
2. D. J. Crommelin *et al.*, Shifting paradigms: biopharmaceuticals versus low molecular weight drugs. *Int J Pharm* 266, 3 (Nov 6, 2003).
3. H. C. Mahler, W. Friess, U. Grauschopf, S. Kiese, Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci* 98, 2909 (Sep, 2009).
4. M. C. Manning, D. K. Chou, B. M. Murphy, R. W. Payne, D. S. Katayama, Stability of protein pharmaceuticals: an update. *Pharm Res* 27, 544 (Apr, 2010).
5. N. Jenkins, L. Murphy, R. Tyther, Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. *Mol Biotechnol* 39, 113 (Jun, 2008).
6. G. Walsh, R. Jefferis, Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology* 24, 1241 (2006).
7. R. Jefferis, Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* 8, 226 (Mar, 2009).
8. A. S. Rathore, Follow-on protein products: scientific issues, developments and challenges. *Trends Biotechnol* 27, 698 (Dec, 2009).
9. L. Zuniga, B. Calvo, Biosimilars approval process. *Regul Toxicol Pharmacol* 56, 374 (Apr, 2010).
10. Anonymous, Directive 2001/83/EC of the European Parliament and the Council of 6 November 2001 on the Community code relating to medicinal products for human use. (2001).
11. European Medicines Agency, Guideline on similar biological medicinal products. (2005).
12. European Medicines Agency, Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2006).
13. European Medicines Agency, Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues. (2006).
14. European Medicines Agency, Guideline on similar biological medicinal products containing interferon beta. (2013).
15. European Medicines Agency, Guideline on non-clinical and clinical development of similar biological medicinal products containing recombinant human follicle stimulating hormone (r-hFSH). (2013).
16. European Medicines Agency, Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues. (2012).
17. European Medicines Agency, Guideline on non-clinical and clinical development of similar biological medicinal products

containing recombinant erythropoietins (Revision). (2010).

18. European Medicines Agency, Reflection paper: Non-clinical and clinical development of similar medicinal products containing recombinant interferon alfa. (2009).
19. European Medicines Agency, Guideline on non-clinical and clinical development of similar biological medicinal products containint low-molecular-weight-heparins. (2009).
20. European Medicines Agency, Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues - Guidance on similar medicinal products containing recombinant human soluble insulin. (2006).
21. European Medicines Agency, Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues - Guidance on similar medicinal products containing somatropin. (2006).
22. European Medicines Agency, Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues - Guidance on similar medicinal products containing recombinant granulocyte colony stimulating factor. (2006).
23. European Medicines Agency, Concept paper on the revision of the guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2011).
24. European Medicines Agency, Questions and answers on the recommendation for refusal of marketing application for Alpheon. (2006).
25. European Medicines Agency, Questions and answers on the withdrawal of the marketing authorization application for Inuslin human rapid Marvel, Insulin human long Marvel, Insulin human 30/70 mix Marvel, International non-proprietary name (INN): human insulin. (2006).
26. European Medicines Agency, Questions and answers on the withdrawal of the marketing authorisation application for
27. Epostim (epoetin alfa). (2011).
28. European Medicines Agency, Public statement on Filgrastim ratiopharm (filgrastim): Withdrawal of the marketing authorisation in the European Union. (2011).
29. European Medicines Agency, Public statement on Valtropin (somatropin): Withdrawal of the marketing authorisation in the European Union. (2012).
30. European Medicines Agency, European Public Assessment Report for Omnitrope. (2006).
31. European Medicines Agency, European Public Assessment Report for Valtropin. (2007).
32. European Medicines Agency, European Public Assessment Report for Binocrit. (2007).
33. European Medicines Agency, European Public Assessment Report for Abseamed. (2007).
34. European Medicines Agency, European Public Assessment Report for Epoetin alfa Hexal. (2007).
35. European Medicines Agency, European Public Assessment Report for Retacrit. (2008).
36. European Medicines Agency, European Public Assessment Report for Silapo. (2008).
37. European Medicines Agency, European Public Assessment Report for Ratiograstim. (2008).
38. European Medicines Agency, European Public Assessment Report for Biograstim. (2008).
39. European Medicines Agency, European Public Assessment Report for Tevagrastim. (2008).
40. European Medicines Agency, European Public Assessment Report for Filgrastim Ratiopharm. (2008).
41. European Medicines Agency, European Public Assessment Report for Filgrastim Hexal. (2009).
42. European Medicines Agency, European Public Assessment Report for Zarzio. (2009).
43. European Medicines Agency, European Public Assessment Report for Nivestim. (2010).
44. H. Schellekens, E. Moors, Clinical comparability and European biosimilar regulations. *Nat Biotechnol* 28, 28 (Jan, 2010).
45. V. Brinks *et al.*, Quality of original and biosimilar epoetin products. *Pharm Res* 28, 386 (Feb, 2011).

46. M. J. Parnham, J. Schindler-Horvat, M. Kozlovic, Non-clinical safety studies on biosimilar recombinant human erythropoietin. *Basic Clin Pharmacol Toxicol* 100, 73 (Feb, 2007).
47. A. O'Connor, M. Rogge, Nonclinical development of a biosimilar: the current landscape. *Bioanalysis* 5, 537 (Mar, 2013).
48. J. L. Bussiere, Species selection considerations for preclinical toxicology studies for biotherapeutics. *Expert Opin Drug Metab Toxicol* 4, 871 (Jul, 2008).
49. P. J. Declerck, Biosimilar monoclonal antibodies: a science-based regulatory challenge. *Expert Opin Biol Ther* 13, 153 (Feb, 2013).
50. H. C. Ebbers *et al.*, Measures of biosimilarity in monoclonal antibodies in oncology: the case of bevacizumab. *Drug Discov Today*, (May 18, 2013).
51. European Medicines Agency, Draft guideline on similar biological medicinal products. (2013).
52. European Medicines Agency, Draft guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2013).

Whenever you find yourself on the side of the majority, it is time to pause and reflect

-Mark Twain, 1835-1910

7

MEASURES OF BIOSIMILARITY IN MONOCLONAL ANTIBODIES IN ONCOLOGY: THE CASE OF BEVACIZUMAB

Hans C. Ebbers, Peter J.K. van Meer, Ellen H.M. Moors,
Aukje K. Mantel-Teeuwisse, Hubert G. M. Leufkens,
Huub Schellekens

Published (in an adapted format)
in *Drug Discov Today*, (May 18, 2013)

ABSTRACT

Biosimilars have been on the European market since 2006 and experience with their use is increasing. The next wave of biopharmaceuticals that are about to lose patent protection consists of more complicated products, including many monoclonal antibodies. Guidance has been released on the particulars of a biosimilarity exercise involving these products. Considerable challenges are posed by anticancer products and there is ongoing controversy on which basis to establish biosimilarity for such products. An especially challenging product is bevacizumab (Avastin®). Based on data available for the innovator product (bevacizumab) we will discuss strengths and weaknesses of preclinical and clinical models and explore the application of novel endpoints to the biosimilar comparability exercise.

BACKGROUND

Legislation allowing the approval of competitor versions of biopharmaceuticals, so-called biosimilars, has been in place in the European Union (EU) since 2004. The foundation of the European biosimilar framework is to provide a full quality dossier, demonstrating that the product is manufactured according to regulatory standards. In addition, comparative analytical, pre-clinical and clinical studies should demonstrate comparable quality, safety and efficacy to a product authorized in the EU(1). Thus far 14 products, including epoetins, filgrastims and recombinant human growth hormones, have received marketing authorization as a biosimilar. The patents of some of the top selling monoclonal antibodies (MABs) are about to expire and there is a lot of interest in developing biosimilar versions of these products(2, 3). Recently, a guideline outlining the approval criteria for biosimilar MABs was released in the EU, which states that a case-by-case approach will be adopted to determine the requirements of the comparability exercise(4).

Box 1 About bevacizumab

Bevacizumab is a 149 kDa humanized IgG1 monoclonal antibody that selectively binds and inhibits vascular endothelial growth factor (VEGF), a 45 kDa homodimeric glycoprotein. VEGF binds and activates a receptor tyrosine kinase stimulating the growth of blood vessels (angiogenesis), which has a central role in the growth, invasion and metastasis of tumors. Overexpression of VEGF has been found in several tumor types and has been linked to a worsened prognosis in patients(61). Bevacizumab was first authorized in the US in 2004, for the treatment of metastatic colon and rectal cancer, in combination with 5-fluorouracil-based chemotherapy and Europe followed in 2005. In Europe, bevacizumab is currently authorized as an add-on therapy to various chemotherapy agents in the treatment of non-small-cell lung carcinoma, colorectal neoplasms, renal cell carcinoma, ovarian neoplasms and breast neoplasms. In the US bevacizumab is no longer indicated for the treatment of metastatic breast cancer, but is indicated as second line treatment as monotherapy for the treatment of glioblastoma. The patent of Avastin will expire in the US in 2017 and in the EU in 2019(3).

The primary sequence of the biosimilar antibodies should be identical to the reference product and the product attributes should be comparable for biosimilars(5). However, differences in post translational modifications such as glycosylation or oxidation may affect a product's immunogenic potential. In addition, subtle differences in effector functions and pharmacokinetics and pharmacodynamics (PK/PD) of two products may occur as a result of differences in the production cell lines and/or other production processes. For example, levels of fucosylation may influence the antibody's ability to induce antibody dependent cellular cytoxicity and high mannose concentrations have been linked to altered serum half-life(6). Furthermore, it may be particularly challenging to establish comparable

efficacy and safety for anticancer products based on preferred endpoints like overall survival or progression free survival. Therefore, novel endpoints, other than patient benefit per se, may be acceptable to establish biosimilarity(*4*). The questions that arise are: which strategy is required to establish biosimilarity in monoclonal antibodies used in an oncology setting? Which instruments are optimal to detect relevant differences between a biosimilar MAB and its reference product? Currently, there are six MABs authorized as anticancer agent. We have chosen one product, bevacizumab (Avastin®), to assess challenges in establishing biosimilarity of MABs licensed as anticancer agents (for more information about bevacizumab, see Box 1). Based on the available preclinical and clinical experience of bevacizumab we will review possibilities and challenges of a comparability exercise for MABs in oncology. We assume that the product is sufficiently similar in terms of quality and focus our review on *in vitro*, *in vivo* and clinical studies.

PRECLINICAL *IN VITRO*

Nonclinical similarity studies start with *in vitro* studies to demonstrate comparable functional binding. For the development of bevacizumab, specific binding of bevacizumab (and its parent murine antibody, A4.6.1) to vascular endothelial growth factor (VEGF) was demonstrated by various *in vitro* affinity binding assays such as radioligand binding, immunoprecipitation assays and surface plasmon resonance assays(*7-9*). These assays allow a quantitative assessment of the affinity of the antibody to its target. Because VEGF is a soluble target, evaluation of antibody dependent cellular cytotoxicity (ADCC) and complement dependent cytotoxicity (CDC) is not required. Potent inactivation of VEGF by bevacizumab and A4.6.1 was demonstrated in several assays which model hallmarks of angiogenesis(*10*). Cell sorting and counting assays demonstrate that endothelial cells (ECs) could be dose dependently inhibited after incubation with bevacizumab. In addition, an ex ovo model of angiogenesis, the chick embryo chorioallantoic membrane (CAM) assay, offered visible semi quantitative confirmation of inhibition of vessel formation after treatment with bevacizumab(*7*). Bevacizumab demonstrated similar efficacy and potency to A4.6.1 *in vitro*. For example, both A4.6.1 and bevacizumab showed 90% inhibition of proliferation of capillary endothelial cells at 500 ng/ml. The ED50 of A4.6.1 and bevacizumab were comparable at 48 ± 8 and 50 ± 5 ng/ml, respectively(*11*). Likewise, these assays could be used to demonstrate similarity. Additional assays could include inhibition of migration of EC in a scratch wound assay, where closure of an inflicted scratch in an EC culture after stimulation with VEGF is measured. A sensitive *in vitro* assay is the 3D spheroid sprouting assay which simulates the complex behavior of EC in angiogenesis(*12*). Functional *in vitro* studies are highly accurate, sensitive and reproducible and allow for a detailed comparison of the affinity, potency and downstream signaling modulation of anti VEGF antibodies.

PRECLINICAL *IN VIVO*

The need for *in vivo* studies is determined based on the outcome of the *in vitro* program and may include pharmacodynamic (PD), pharmacokinetic (PK) and safety studies. If novel

excipients are used in formulating the biosimilar or information of their use in the clinic is limited, local tolerance studies may also be required. For bevacizumab, PD studies were performed in severe combined immune deficient (SCID), athymic and nude mouse or athymic rat xenograft models of various tumors. In these models, bevacizumab reduced tumor growth and increased survival(*13-19*). However, for a biosimilarity exercise, xenograft models may have limited value as many variables can influence the outcome, including the choice of mouse strain, tumor type and size at implantation and the implantation site itself, all of which add to the complexity of interpreting the study results and may be difficult to reproduce(*20, 21*). Directed *in vivo* angiogenesis assays, which can quantitatively compare inhibition of angiogenesis in tumors, could be used as an alternative PD model, which would reduce the number of test animals required(*22*). *In vivo* pharmacokinetics studies should allow for a quantified comparison between the reference and biosimilar. Single dose studies, including internal comparability studies between different bevacizumab formulations and batches, were conducted in mouse and rat and met bioequivalence criteria. Because bevacizumab only binds to cynomolgus monkeys and rabbit VEGF, these were the two model species of choice for most repeated dose PK and toxicity studies. In rabbits, bevacizumab induced the formation of antibodies after eight days. Therefore, non-human primates may be the only valid alternative for longer term PK studies. However, ethical (and economical) considerations lead to the use of fewer non-human primates in a study, which may affect the sensitivity of the *in vivo* study. Although many monoclonal antibodies are immunogenic in nonhuman primates, bevacizumab generally did not induce formation of antidrug antibodies(*11*).

Three repeated dose toxicity studies were conducted in nonhuman primates to evaluate the safety of bevacizumab(*11*). In these studies, bevacizumab was generally well tolerated and main side effects were related to the pharmacology of the drug including physial displasia of the distal femur, degeneration of the cartilage matrix and disruption of the menstrual cycle in female monkeys. Two common side effects of bevacizumab in clinical trials were hypertension and proteinuria. Interestingly, no effects of bevacizumab on blood pressure or urinalysis were observed in preclinical safety studies with non-human primates(*11*). Repeated dose safety studies in nonhuman primates are not likely to add substantially to demonstrate biosimilarity and are also not recommended(*23*).

CLINICAL PHARMACOKINETICS AND PHARMACODYNAMICS

The clearance of monoclonal antibodies is determined by a non-specific clearance pathway mediated by its Fc-region and a specific clearance pathway, which is dependent on target binding through their Fab regions. The non-specific pathway generally shows linear clearance, whereas the specific pathway can be saturated, dependent on dose of the MAB and the expression of the target(*24*). To date, there are no PD biomarkers for clinical efficacy of bevacizumab. Given the long elimination half-life and the possibility of immunogenicity of MABs, the EMA accepts parallel group PK studies, rather than cross-over studies(*4*). Bevacizumab is not cytotoxic and it should be feasible to study PK in volunteers. However, all PK studies for bevacizumab presented in the European Public Assessment Report (EPAR)

involved patients rather than healthy volunteers, which may indicate that regulatory bodies consider the risks associated with bevacizumab unacceptable for healthy volunteers. Single dose studies are preferred, although ethical considerations may prevent performing a single dose study in patients, because current standard of care requires multiple doses of bevacizumab to be administered.

Expression levels of VEGF are low and PK studies investigating bevacizumab doses ranging from 0.1 mg/kg- 10 mg/kg, found linear PK for all doses studied, with a terminal half-life of +/-20 days (range 11-50) regardless of concomitant chemotherapy(*25-27*). Based on limited available data, there is no interaction between various chemotherapy agents (including carboplatin and paclitaxel) and bevacizumab. PK studies for currently authorized biosimilars studied the authorized dose of the comparator product and guidelines recommended to study PK using the lowest possible dose(*25*). Therefore, the recommended starting dose for colorectal cancer of 5 mg/kg will be a likely candidate for PK studies. Eight PK studies were included in the authorization dossier of bevacizumab, two of which examined the authorized dose for colorectal cancer of 5mg/kg every two weeks. The clearance rate was calculated at 2.79 ± 0.849mL/kg/day (mean ± SD), the mean area under the curve was calculated at 2009 ± 653 µg• day/mL (mean ± SD), while the median elimination half-life was estimated to be 19.9 days(*28*). Although there is considerable inter-individual variability in availability of bevacizumab, even at 2.5 mg/kg dosing, it is expected that >98% of VEGF is bound to bevacizumab(*28*). Therefore, even significant variability in clearance is not likely to influence efficacy. None of the studies identified antibodies to bevacizumab.

CLINICAL EFFICACY AND SAFETY

Given the absence of validated PD markers, confirmatory clinical efficacy studies will be needed. Commonly applied endpoints are preferred in the biosimilar monoclonal antibodies guideline, including overall survival (OS) or progression free survival (PFS), but it is recognized that these may not be sensitive enough for establishing comparability. To illustrate the challenges in sensitivity we calculated sample sizes required for trials in indications that are licensed in the EU to demonstrate non-inferiority, here defined as retention of 50% of the efficacy of the innovator over placebo(*29*). More stringent equivalence trials are preferred by European regulators, but we chose to determine sample sizes for time-to-event analyses assuming an optimistic scenario (Table 1). Source data were obtained from pivotal trials described in the European Public Assessment Reports and publications of these studies that have appeared in peer reviewed scientific medical journals. Bevacizumab was approved in combination with irinotecan and 5-fluorouracil (IFL), for the treatment of metastatic colorectal cancer. However, the standard of care has evolved and so has the relative improvement in progression free survival (PFS) and overall survival (OS). European guidance on the evaluation of anticancer medicinal products in man stipulates that the choice of reference regimen should be selected from the best available, evidence based therapeutic options(*30*). This means that a widely used, but not necessarily licensed, regimen with a favorable benefit-risk profile should be chosen. At this moment, European practice guidelines recommend using oxaliplatin + 5-FU combination or capecitabine + 5-FU

Table 1. Sample size calculations of for non-inferiority trials in various authorized indications and possible novel endpoints.

		1 year survival rate	Placebo efficacy (median)	Innovator efficacy (median)	Hazard Ratio (95% CI)	NI margin	Sample size required	reference
Time to event Endpoint								
Colorectal Cancer	PFS	30.4%	8.0 months	9.4 months	0.83 (0.72-0.95)‡	δ= 1.10	**3454**	(32)
	OS**	78.9%	19.9 months	21.3 months	0.84 (076-1.03)‡	δ= 1.10*	**8546***	
Renal Cell Carcinoma	PFS	42.9%	5.4 months	10.2 months	0.63 (0.52-0.75)	δ= 1.29	**544**	(62)
	OS	74.2%	21.3 months	23.3 months	0.91 (0.76-1.10)	δ= 1.13*	**3838 ***	
Non-small Cell Lung Cancer	PFS	17.3%	4.5 months	6.2 months	0.66 (0.57-0.77)	δ= 1.25	**622**	(63)
	OS	51.6%	10.3 months	12.3 months	0.79 (0.67-0.92)	δ=1.13	**2502**	
Epithelial Ovarian cancer	PFS	55.7%	10.3 months	14.1 months	0.72 (0.63-0.82)	δ=1.20	**1258**	(64)
	OS	91.0%	39.4 months	39.8 months	0.90 (0.72-1.13)	δ=1.06*	**52378 ***	
Breast Cancer‡	PFS	34.2%	5.7 months	8.6 months	0.69 (0.56-0.84)	δ=1.22	**828**	(65)
	OS	81.6%	21.2 months	29 months	0.85 (0.63-1.14)	δ= 1.09*	**11104 ***	
Other endpoints†								
Hypertension			6.4%	18.9%	-	δ=1.97	**300**	(32)
AMD			62.2%	94.5%	-	δ=1.26	**66**	(52,53)

AMD= Age related macular degeneration
Sample sizes were calculated using R statistics (Version 2.15, The R-foundation). Time-to-event analyses were calculated using the *plansurvct.func* function created by Filleron et al.(*66*) See supplementary information for an example of a sample size calculation. All calculations were based on a non-inferiority trial assuming 1 control subject per experimental subject, an accrual interval of 12 months, and additional follow-up after the accrual interval of 24 months, α=0.025 and β= 0.80, no dropout is assumed and overall 1-years survival was measured from the primary publications. NI margins (δ) were computed based on the ratio innovator/placebo using the 50% retention method of the point estimate of innovator efficacy vs. placebo as described by Tanaka et al(*29*).
*Overall survival was not significantly different in various clinical studies, so care must be taken to interpret a "Non-inferiority" margin.
** 97.5% confidence interval was provided
‡ Breast cancer based only on arm receiving bevacizumab in combination with capecitabine (EU authorized indication).
†Sample sizes for hyperertension and AMD were calculated as binary endpoints using the 'R' gsDesign package for a non-inferiority trial with equal efficacy, 1-sided alpha level of 2.5% and power of 80%.

combination plus bevacizumab as first line treatment for colorectal cancer(*31*). Therefore, for the calculation of the required sample size in colorectal cancer patients we included study NO16966(*32*). For all other authorized indications we included those trials that included treatment regimens that were included in respective treatment guidelines, which were mostly the pivotal trials (Table 2).

Trials designed to establish non-inferiority for PFS require sample sizes that range from 544 patients in the metastatic renal cell carcinoma indication to 3454 in advanced

Table 2. Incidence of commonly reported adverse events in bevacizumab clinical Phase III trials.

Event (%)	Colorectal cancer	(32)	Renal cell carcinoma	(62)	Non-small cell lung cancer	(63)	Ovarian Cancer	(64)	Breast cancer‡	(65)
	Placebo	Bevacizumab	Placebo	Bevacizumab	Placebo	Bevacizumab	Placebo	Bevacizumab	Placebo	Bevacizumab
Hypertension										
(any)	6.4	18.9	9	26	NA	NA	10	26	NA	NA
(Grade 3)	1.2	3.7	<1	3	0.5	6.8	<1	6	1.0	10.1
Proteinurea										
(any)	NA	NA	3	18	NA	NA	2	4	NA	NA
(Grade 3)	0	<1	0	7	0	2.6	<1	1	0	2.2
Bleeding										
(Any)	25.9	30.5	9	33	NA	NA	11	38	NA	NA
(grade 3 or 4)	1	2	<1	3	0.7	4.4	<1	1	0.5	0.2
Neutropenia										
(Any)	NA	NA	7	7	NA	NA	29	28	NA	NA
(grade 3 or 4)	NA	NA	2	4	16.8	25.5	15	17	1.0	1.2
Thrombotic event										
Venous thrombotic event	4.9	7.8	<1	3	NA	NA	4	7	3.5	5.0
Arterial thrombotic event	1.0	1.7	<1	1	NA	NA	1	4	1.5	1.5
Gastrointestinal perforation	<1	<1	0	1	NA	NA	<1	1	0	0
Wound healing complications	<1	<1	1	1	NA	NA	2	5	0	0.7

Grade 2/3 Hypertension= requiring medication. NA= Not available. ‡Breast cancer based only on arm receiving bevacizumab in combination with capecitabine (EU authorized indication).

metastatic colorectal cancer. Sample sizes required to demonstrate non-inferiority using OS as endpoint, are 2502 patients for non-small cell lung cancer and many more patients for the other authorized indications. Bevacizumab did not show significant improvement in OS versus the respective control groups, and it is not possible to establish a valid non-inferiority margin for these endpoints. In principle, in order to be able to establish non-inferiority it is recommendable to perform comparative studies in patients with poor prognosis that benefit most from bevacizumab, e.g. patients with renal cell carcinoma. The non-inferiority margin used in our example is illustrative and does not necessarily reflect clinical relevance, which is required by regulatory authorities(*33*). In various trials, the improvement in PFS was not supported by similar improvements in OS (at least not for the duration if the trial). Other widely used efficacy endpoints in oncology, such as Response Evaluation Criteria in Solid Tumors (RECIST), may not be good predictors of efficacy, because treatment with bevacizumab does not result in tumor shrinkage(*34, 35*). Accepted clinical endpoints may thus be unsuitable to establish biosimilarity.

SAFETY EVENTS AS POSSIBLE ENDPOINT

The monoclonal antibody guideline states that safety outcomes may also be considered a measure of biosimilarity(*23*). In order to be suitable for a biosimilarity exercise, safety issues need to be relatively frequent and occur relatively early. In Table 2 commonly reported adverse events for bevacizumab are presented. Hypertension is one of the most common side effects of bevacizumab occurring in up to 22.4% of the patients in phase III clinical studies. There are several hypotheses about the biological mechanisms underlying this effect, including a reduction of VEGF induced NO production and systemic changes to the vascular network(*36*). Several studies have investigated hypertension as a predictor of clinical response. A single arm observational study reported that the median PFS was 14.5 months in patients with grade 2/3 hypertension vs. 3.1 months in patients without grade 2/3 hypertension(*37*). Two other studies found a PFS survival of 15.1 months in patients with hypertension vs. 8.3 in patients without and 10.5 months vs. 5.3 months(*38, 39*). Interestingly, the median time to onset of hypertension was 1 month, and 95% of the cases developed with 6 months. Another study failed to show an improved radiological and clinical response rate in patients with poor-risk colorectal liver-only metastases that developed hypertension(*40*). A prospective analysis of 218 patients, including 184 colorectal cancer patients found that patients who developed grade 2/3 hypertension had improved overall survival versus patients without hypertension (median PFS 29.9 months vs. 17.2 months) (*41*). Other smaller studies also reported an improved outcome in patients that developed hypertension(*42*). A retrospective analysis of the pivotal studies of bevacizumab found that hypertension was predictive in study AV2107g, but not in study NO16966(*43*). Most studies investigating hypertension were retrospective and did not use standardized methods. Although hypertension may be a suitable biomarker, its etiology and relation to clinical outcome remains uncertain(*44*). Nevertheless, hypertension seems to be a pharmacological, dose dependent ADR that may have a sufficiently high incidence to be a suitable surrogate endpoint of a clinical comparability study. A trial choosing hypertension as a primary

endpoint would require significantly less patients; 150 patients per arm may have enough power to demonstrate non-inferiority (Table 2).

OTHER SAFETY ENDPOINTS

A meta-analysis reported that proteinurea occurred in as many as 41-63% of patient receiving bevacizumab, is dose dependent and associated with the occurrence of hypertension(*45*). Proteinurea has been associated with improved OS and PFS in a limited number of case series(*46*). A meta-analysis including bevacizumab for various solid cancers did not identify an association between high grade proteinurea and OS or PFS(*47*). Although proteinurea may have a high incidence, its relation to the pharmacology of bevacizumab is less well-established. Also, bleeding events have been reported frequently in patients receiving bevacizumab, mainly mucocutaneous hemorrhage (20%-40%) and epistaxis (22-34%)(*11*). Bleeding events have not been linked to bevacizumab efficacy and increases in bleeding events were mainly observed in non-small cell lung cancer patients and to a lesser extent in colorectal cancer patients(*48*). Other adverse events occur too infrequent to be suitable markers for a clinical comparability study.

OTHER BIOMARKERS

Various candidates, including baseline VEGF levels, circulating VEGF levels, capillary density and ICAM-1 levels have been investigated as biomarkers predictive for VEGF activity, but thus far none has been identified that reliably predicts clinical efficacy(*34, 49, 50*). Other studies have focused on identifying prognostic biomarkers that could help to identify patients that benefit more from bevacizumab, including genotypes, baseline values of various proteins, baseline proteinurea, cell counts, miRNAs, but so far this search has not yielded viable candidates(*34*). While such a biomarker may have clinical importance, its use in a comparability exercise is of questionable value, as a trial including patients that respond well to treatment will increase rather than decrease the number of patients required to establish non-inferiority.

UNAUTHORIZED ENDPOINTS

It has been demonstrated that bevacizumab is efficacious in the treatment of age related macular degeneration (AMD). The efficacy of bevacizumab was compared with ranibizumab (Lucentis®), a PEGylated anti-VEGF Fab fragment authorized for the treatment of AMD in two large trials involving >1800 patients(*51, 52*). Both trials concluded that monthly administered bevacizumab has equivalent efficacy to ranibizumab. The effectiveness of bevacizumab in AMD is considerable and beneficial effects are seen within 12-24 weeks of treatment. A non-inferiority trial to establish retention of 50% of the effect of the originator vs. placebo would require 66 patients, assuming efficacy measured as percentage of patients losing visual acuity according to Rosenfeld et al. (Table 2)(*53*). However, in AMD bevacizumab is administered intravitreally at doses that are only a fraction of those used in oncology. This results in a very different distribution profile, which makes it challenging to extrapolate the safety profile to the cancer indication.

DISCUSSION

Here we have reviewed various challenges and opportunities for a biosimilarity exercise of a biosimilar anti-VEGF antibody. A comparability exercise requires a stepwise approach and each step has strengths and weaknesses. *In vitro* assays are most sensitive to find differences in pharmacological activity. They are easily quantifiable and may be more specific and sensitive than studies in animals(*54*). Several assays exist that have a strong discriminative power. However, it is not always clear how much difference can be accepted. For example, ranibizumab shows considerable differences to bevacizumab in various *in vitro* assays, but comparable clinical efficacy in the treatment of AMD(*52, 55-57*). However, *in vitro* assays cannot evaluate changes that occur by post translational modification *in vivo*. Internal PK comparability for bevacizumab studies have been successfully conducted in rat and rabbit, but the sensitivity and reproducibility of *in vivo* studies is limited and results should be interpreted with caution. Most common *in vivo* PD models for cancer are rodent xenografts, which do not optimally represent the situation of advanced metastatic cancer. *In vivo* safety studies in non-human primates did not reveal limiting toxicities and are not likely to generate additional information. Therefore, clinical studies will be required regardless of the outcome of *in vivo* studies.

PK studies are likely to be performed in patients at a dose of 5 mg/kg. However, validated PD markers that are indicative of efficacy currently do not exist for anti-VEGF therapy. PK shows linear kinetics and is an important measure to demonstrate comparability. Demonstrating similar PK behavior is even more important for MAbs in oncology given the challenges in establishing similarity based on clinical efficacy outcomes. Nevertheless, clinical efficacy and safety studies will likely be required. As a principle, a clinical endpoint should be selected that fulfills the following requirements: (1) sensitive enough to detect small differences, (2) measurable with sufficient precision, (3) clinically relevant for the target population. Given the long time it requires to measure efficacy and the relatively modest benefit of bevacizumab as addition to existing chemotherapy regimens it requires large trials that follow patients for long times to establish clinical comparability based on non-inferiority. Performing clinical studies in patients at high risk of disease progression, such as non-small cell lung cancer or renal cell carcinoma, which are likely to benefit from bevacizumab may reduce the sample size required, but will still require substantial numbers of patients that need to be followed up for a considerable amount of time. Furthermore, it may be challenging to recruit a sufficient number of patients for trials in some less prevalent cancers. When designing a non-inferiority trial, identifying likely responders will not reduce the required sample size. Hypertension could be suitable surrogate measure because it is widely recognized as a pharmacological effect that has also been linked to improved clinical outcome and becomes apparent within a matter of months. Other adverse events like proteinurea and/or bleeding events may also be considered, but their discriminative power is lower and they lack clinical rationale. Although choosing hypertension as a surrogate endpoint will significantly reduce the amount of patients required in a comparability study, there are also considerable challenges. There is limited prospective data studying the association hypertension and clinical outcome and the assessment of hypertension in the various studies has been far from standardized(*44*).

Another route may be to assess efficacy of bevacizumab in AMD. Establishing comparability in AMD has obvious appeal, as it is highly efficacious and it can be established relatively quickly with fewer patients, when compared to accepted outcomes in cancer trials. Although AMD is not an authorized indication of bevacizumab, the EMA guidelines leave room for establishing comparability in non-authorized indications. However, "the applicant should justify that the model is relevant as regards efficacy and safety, and sensitive to demonstrate comparability in the indication(s) applied for"(*4*). If it is considered that the mechanism of action is the same in AMD, it is conceivable that a three armed-trial comparing a biosimilar to both bevacizumab and a third arm including ranibizumab to ensure comparability to an authorized indication could be an acceptable surrogate for anti-VEGF activity in humans. Alternatively a third arm including placebo may be chosen to validate the chosen non-inferiority margin as recommended by current guidelines(*30*). While this may demonstrate comparable efficacy, questions may remain on the comparability of the safety of the product when administered systematically in an oncology setting. Any residual concerns about the safety and efficacy of the product at the level of the prescribers and patients may prevent the uptake of biosimilars in clinical practice(*58*).

None of the studies involving bevacizumab identified antibodies against the product. However, it is possible that a biosimilar version of bevacizumab induces antibodies. In this case it is likely that the product's pharmacokinetics and thus its efficacy and possibly safety profile is affected. In this case it will be hard to conclude that a product is biosimilar. Therefore it is of the utmost importance that immunogenicity is studied using validated assays and taken into account in the overall analysis of efficacy and safety in a pre-planned manner.

Our case study demonstrates that anticancer products such as bevacizumab may also force regulatory authorities to reconsider the objective of the comparability exercise. Currently, European legislation requires a product to 'demonstrate biosimilarity', but this could be altered to 'exclude excessive dissimilarity'(*59*). A similar position has been adopted for erythropoietins. The key concern of erythropoietins is that a biosimilar product is more immunogenic than the innovator product. Pre-licensing data is expected only to identify "excessive immunogenicity" and it is recognized that the real frequency of pure red cell aplasia, a severe consequence of immunogenicity for erythropoietins, can only be established post authorization(*60*). Likewise, for anticancer monoclonal antibodies it may be conceivable that a biosimilar is granted authorization on an extensive *in vitro* and *in vivo* package, comparable PK, comparable immunogenicity and clinical efficacy using a non-validated PD marker, provided that sufficient long term efficacy data will be obtained post authorization, for example through long term follow up of patients included in the clinical efficacy trials, or through post authorization observational studies.

CONCLUDING

It may be conceptually challenging to demonstrate that two products are similar. Clearly there is no "one size fits all" answer to the challenges posed by complicated products such as bevacizumab. Comparability of such products needs to be investigated at all levels. In order to allow the authorization of biosimilars in oncology at realistic costs requires novel

approaches to the clinical development program that may conflict with the existing biosimilar framework. Care must be taken to ensure that an alternative approach generates relevant data to establish comparable efficacy and safety. While the biosimilarity exercise needs to convince regulatory bodies to grant marketing authorization, ultimately, prescribers and patients need to feel confident that a biosimilar product is as safe and efficacious as the original product.

ACKNOWLEDGEMENTS

We would like to thank J. Tanniou, MSc. and prof. dr. K.C. Roes for providing statistical advice and prof. dr. J.H.M. Schellens for providing valuable insights into the clinical aspects of bevacizumab.

This study was performed in the context of the Escher project (T6-202), a project of the Dutch Top Institute Pharma. All declare no conflict of interest relevant to the subject matter or materials discussed in the manuscript. The department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, employing authors Hans C. Ebbers, Aukje K. Mantel-Teeuwisse and Hubert G.M. Leufkens has received unrestricted research funding from the Netherlands Organisation for Health Research and Development (ZonMW), the Dutch Health Care Insurance Board (CVZ), the Royal Dutch Pharmacists Association (KNMP), the private-public funded Top Institute Pharma (www.tipharma.nl, includes co-funding from universities, government, and industry), the EU Innovative Medicines Initiative (IMI), EU 7th Framework Program (FP7), the Dutch Medicines Evaluation Board, the Dutch Ministry of Health and industry (including GlaxoSmithKline, Pfizer, and others). Huub Schellekens participated in meetings and publications sponsored by Amgen, J&J, Roche, Sandoz and Hospira. Part of his research is directly or indirectly sponsored by Roche, Amgen and Sandoz.

REFERENCES

1. M. McCamish, G. Woollett, The state of the art in the development of biosimilars. *Clin Pharmacol Ther* 91, 405 (Mar, 2012).
2. A. Mullard, Can next-generation antibodies offset biosimilar competition? *Nat Rev Drug Discov* 11, 426 (Jun, 2012).
3. H. Ledford, 'Biosimilar' drugs poised to penetrate market. *Nature* 468, 18 (Nov 4, 2010).
4. European Medicines Agency, Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues. EMA/CHMP/BMWP/403543/2010. (2010).
5. A. Beck, S. Sanglier-Cianferani, A. Van Dorsselaer, Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal Chem* 84, 4637 (Jun 5, 2012).
6. R. Jefferis, Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* 8, 226 (Mar, 2009).
7. K. J. Kim, B. Li, K. Houck, J. Winer, N. Ferrara, The vascular endothelial growth factor proteins: identification of biologically relevant regions by neutralizing monoclonal antibodies. *Growth Factors* 7, 53 (1992).
8. L. G. Presta *et al.*, Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res* 57, 4593 (Oct 15, 1997).

9. L. G. Presta *et al.*, Humanization of an antibody directed against IgE. *J Immunol* 151, 2623 (Sep 1, 1993).
10. S. M. Weis, D. A. Cheresh, Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat Med* 17, 1359 (2011).
11. European Medicines Agency, European Public Assessment report for Avastin. (2006).
12. P. Correa de Sampaio *et al.*, A heterogeneous *in vitro* three dimensional model of tumour-stroma interactions regulating sprouting angiogenesis. *PLoS One* 7, e30753 (2012).
13. P. Borgstrom, M. A. Bourdon, K. J. Hillan, P. Sriramarao, N. Ferrara, Neutralizing anti-vascular endothelial growth factor antibody completely inhibits angiogenesis and growth of human prostate carcinoma micro tumors *in vivo*. *Prostate* 35, 1 (Apr 1, 1998).
14. A. Gossmann *et al.*, Dynamic contrast-enhanced magnetic resonance imaging as a surrogate marker of tumor response to anti-angiogenic therapy in a xenograft model of glioblastoma multiforme. *J Magn Reson Imaging* 15, 233 (Mar, 2002).
15. L. Hu *et al.*, Vascular endothelial growth factor immunoneutralization plus Paclitaxel markedly reduces tumor burden and ascites in athymic mouse model of ovarian cancer. *Am J Pathol* 161, 1917 (Nov, 2002).
16. O. Melnyk, M. Zimmerman, K. J. Kim, M. Shuman, Neutralizing anti-vascular endothelial growth factor antibody inhibits further growth of established prostate cancer and metastases in a pre-clinical model. *J Urol* 161, 960 (Mar, 1999).
17. S. Mesiano, N. Ferrara, R. B. Jaffe, Role of vascular endothelial growth factor in ovarian cancer: inhibition of ascites formation by immunoneutralization. *Am J Pathol* 153, 1249 (Oct, 1998).
18. S. Z. Soffer *et al.*, Combination antiangiogenic therapy: increased efficacy in a murine model of Wilms tumor. *J Pediatr Surg* 36, 1177 (Aug, 2001).
19. R. S. Warren, H. Yuan, M. R. Matli, N. A. Gillett, N. Ferrara, Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. *J Clin Invest* 95, 1789 (Apr, 1995).
20. C. G. Begley, L. M. Ellis, Drug development: Raise standards for preclinical cancer research. *Nature* 483, 531 (Mar 29, 2012).
21. L. R. Kelland, Of mice and men: values and liabilities of the athymic nude mouse model in anticancer drug development. *Eur J Cancer* 40, 827 (Apr, 2004).
22. C. J. Sweeney *et al.*, The antiangiogenic property of docetaxel is synergistic with a recombinant humanized monoclonal antibody against vascular endothelial growth factor or 2-methoxyestradiol but antagonized by endothelial growth factors. *Cancer Res* 61, 3369 (Apr 15, 2001).
23. European Medicines Agency, Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues. (2012).
24. R. Deng, F. Jin, S. Prabhu, S. Iyer, Monoclonal antibodies: what are the pharmacokinetic and pharmacodynamic considerations for drug development? *Expert Opin Drug Metab Toxicol* 8, 141 (Feb, 2012).
25. European Medicines Agency. (2006).
26. M. S. Gordon *et al.*, Phase I safety and pharmacokinetic study of recombinant human anti-vascular endothelial growth factor in patients with advanced cancer. *J Clin Oncol* 19, 843 (Feb 1, 2001).
27. K. Margolin *et al.*, Phase Ib trial of intravenous recombinant humanized monoclonal antibody to vascular endothelial growth factor in combination with chemotherapy in patients with advanced cancer: pharmacologic and long-term safety data. *J Clin Oncol* 19, 851 (Feb 1, 2001).
28. J. F. Lu *et al.*, Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol* 62, 779 (Oct, 2008).
29. S. Tanaka, Y. Kinjo, Y. Kataoka, K. Yoshimura, S. Teramukai, Statistical issues and recommendations for noninferiority trials in oncology: a systematic review. *Clin Cancer Res* 18, 1837 (Apr 1, 2012).
30. European Medicines Agency, Guideline on the evaluation of anticancer medicinal products in man (CPMP/EWP/205/95/Rev.3/Corr.2). (2005).
31. H. J. Schmoll *et al.*, ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann Oncol* 23, 2479 (Oct, 2012).
32. L. B. Saltz *et al.*, Bevacizumab in combination with oxaliplatin-based chemotherapy as first-

line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol* 26, 2013 (Apr 20, 2008).

33. European Medicines Agency, Guideline on the choice of the non-inferiority margin (EMEA/CPMP/EWP/2158/99). (2005).
34. A. M. Jubb, A. L. Harris, Biomarkers to predict the clinical efficacy of bevacizumab in cancer. *Lancet Oncol* 11, 1172 (Dec, 2010).
35. H. Hurwitz *et al.*, Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 350, 2335 (Jun 3, 2004).
36. J. J. Mourad, G. des Guetz, H. Debbabi, B. I. Levy, Blood pressure rise following angiogenesis inhibition by bevacizumab. A crucial role for microcirculation. *Ann Oncol* 19, 927 (May, 2008).
37. M. Scartozzi *et al.*, Arterial hypertension correlates with clinical outcome in colorectal cancer patients treated with first-line bevacizumab. *Ann Oncol* 20, 227 (Feb, 2009).
38. P. Osterlund *et al.*, Hypertension and overall survival in metastatic colorectal cancer patients treated with bevacizumab-containing chemotherapy. *Br J Cancer* 104, 599 (Feb 15, 2011).
39. A. De Stefano, C. Carlomagno, S. Pepe, R. Bianco, S. De Placido, Bevacizumab-related arterial hypertension as a predictive marker in metastatic colorectal cancer patients. *Cancer Chemother Pharmacol* 68, 1207 (Nov, 2011).
40. A. Dewdney, D. Cunningham, Y. Barbachano, I. Chau, Correlation of bevacizumab-induced hypertension and outcome in the BOXER study, a phase II study of capecitabine, oxaliplatin (CAPOX) plus bevacizumab as peri-operative treatment in 45 patients with poor-risk colorectal liver-only metastases unsuitable for upfront resection. *Br J Cancer* 106, 1718 (May 22, 2012).
41. E. Tahover *et al.*, Hypertension as a predictive biomarker in bevacizumab treatment for colorectal cancer patients. *Med Oncol* 30, 327 (Mar, 2013).
42. B. Burnette, A. Grothey, in *ASCO Meeting Abstracts*. (2010), vol. 28, pp. e14066.
43. H. Hurwitz *et al.*, Analysis of early hypertension (HTN) and clinical outcome with bevacizumab (BV). *J Clin Oncol* 28, 3039 (2010).
44. L. Horsley, K. Marti, G. C. Jayson, Is the toxicity of anti-angiogenic drugs predictive of outcome? A review of hypertension and proteinuria as biomarkers of response to anti-angiogenic therapy. *Expert Opin Drug Metab Toxicol* 8, 283 (Mar, 2012).
45. X. Zhu, S. Wu, W. L. Dahut, C. R. Parikh, Risks of proteinuria and hypertension with bevacizumab, an antibody against vascular endothelial growth factor: systematic review and meta-analysis. *Am J Kidney Dis* 49, 186 (Feb, 2007).
46. E. Saloustros, N. Androulakis, L. Vamvakas, D. Mavroudis, V. Georgoulias, Favorable clinical course of patients experiencing bevacizumab-induced proteinuria. *Case Rep Oncol* 3, 368 (2010).
47. S. Wu, C. Kim, L. Baer, X. Zhu, Bevacizumab increases risk for severe proteinuria in cancer patients. *J Am Soc Nephrol* 21, 1381 (Aug, 2010).
48. S. Hapani, A. Sher, D. Chu, S. Wu, Increased risk of serious hemorrhage with bevacizumab in cancer patients: a meta-analysis. *Oncology* 79, 27 (2010).
49. A. Gerger, M. LaBonte, H. J. Lenz, Molecular predictors of response to antiangiogenesis therapies. *Cancer J* 17, 134 (Mar-Apr, 2011).
50. N. Steeghs *et al.*, Reversibility of capillary density after discontinuation of bevacizumab treatment. *Ann Oncol* 21, 1100 (May, 2010).
51. U. Chakravarthy *et al.*, Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration: one-year findings from the IVAN randomized trial. *Ophthalmology* 119, 1399 (Jul, 2012).
52. D. F. Martin *et al.*, Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med* 364, 1897 (May 19, 2011).
53. P. J. Rosenfeld *et al.*, Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med* 355, 1419 (Oct 5, 2006).
54. A. M. Goodwin, *In vitro* assays of angiogenesis for assessment of angiogenic and anti-angiogenic agents. *Microvasc Res* 74, 172 (Sep-Nov, 2007).
55. L. Yu, X. H. Liang, N. Ferrara, Comparing protein VEGF inhibitors: *In vitro* biological studies. *Biochem Biophys Res Commun* 408, 276 (May 6, 2011).
56. A. Barzelay, A. Lowenstein, J. George, A. Barak, Influence of non-toxic doses of bevacizumab and ranibizumab on endothelial

functions and inhibition of angiogenesis. *Curr Eye Res* 35, 835 (Sep, 2010).

57. A. Klettner, J. Roider, Comparison of bevacizumab, ranibizumab, and pegaptanib *in vitro*: efficiency and possible additional pathways. *Invest Ophthalmol Vis Sci* 49, 4523 (Oct, 2008).
58. H. C. Ebbers, T. Pieters, H. G. Leufkens, H. Schellekens, Effective pharmaceutical regulation needs alignment with doctors. *Drug Discov Today* 17, 100 (Feb, 2012).
59. Anonymous, Directive 2003/63/EC amending Directive 2001/83/EC of the European Parliament and of the Council on the Community code relating to medicinal products for human use. (2003).
60. European Medicines Agency, Overview of comments received on the Guideline on non-clinical and clinical development of similar biological medicinal products containing recombinant erythropoietins (EMEA/CHMP/BMWP/301636/2008). (2011).
61. N. Ferrara, H. P. Gerber, J. LeCouter, The biology of VEGF and its receptors. *Nat Med* 9, 669 (Jun, 2003).
62. B. Escudier *et al.*, Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* 370, 2103 (Dec 22, 2007).
63. A. Sandler *et al.*, Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* 355, 2542 (Dec 14, 2006).
64. T. J. Perren *et al.*, A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med* 365, 2484 (Dec 29, 2011).
65. N. J. Robert *et al.*, RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2-negative, locally recurrent or metastatic breast cancer. *J Clin Oncol* 29, 1252 (Apr 1, 2011).
66. T. Filleron, J. Gal, A. Kramar, Designing group sequential randomized clinical trials with time to event end points using a R function. *Comput Methods Programs Biomed* 108, 113 (Oct, 2012).

Some men see things as they are and say why. I dream things that never were and say why not.

-Robert F. Kennedy, 1925-1968

8 DISCUSSION AND PERSPECTIVES

INTRODUCTION

The predictive value of animal studies in drug development is largely unknown and has only rarely been the subject of appropriate study. In 1962 Owens published a study on the toxicity of 23 chemotherapeutic agents of which data were available in rodent, dog or monkey which he compared to the observed toxicity in human. Specific classes of toxicity relating to bone marrow, G.I tract, liver and kidney toxicity in animals were generally predictive for human toxicity, whereas nervous system toxicity was poorly identified in animal studies and skin toxicity was not predicted in any species(*1*). A few years later, anticancer therapeutics were also the focus of a study where the toxicity of 25 anticancer therapeutics in dogs and monkeys were compared to adverse effects in humans. Both species were useful in identifying toxicities in various target organs but renal and hepatic toxicities were overpredicted in these species(*2*). In 1978, Fletcher used clinical trial certificates and product licenses of 45 drugs, which were under consideration by the British Committee on Safety of Medicines to study the relation between animal and clinical data. Few adverse effects were reported in both animal and humans but many side effects were seen either in animals or in humans, but not both. More recently, industry initiatives have attempted to evaluate the value of animal studies to predict adverse reactions. One of these studies showed that of 43 established non-clinical pharmacological endpoints, 10 were statistically significant correlated with clinically relevant adverse reactions in humans(*3*). Another study evaluated the concordance of animal models to detect toxicity in humans. Retrospectively, the concordance between human and animal toxicity was 71% when looking for matching toxicity in any species. This concordance decreased to 43% when looking only at rodent data(*4*). Fletcher was the first to use what is now known as the marketing authorization application for his study in 1978 and pointed out that a larger study and a more sophisticated analysis would provide a better evaluation of the relationships between animal toxicity and human events(*5*). To our knowledge, our study is the first work to use his suggestion to date.

Animal studies are required by law to demonstrate the safety and efficacy of new drugs(*6*). However, animals have known limitations such as high intrinsic variability, immutable differences within species and mechanistically, are not completely understood(*7*). Therefore, animal studies may not, and in fact do not, always predict pharmacological effects or safety risks for humans. Unpredictive animal studies create a loss of resources because time and money is spent to perform and evaluate uninformative studies and impedes the efficiency of the development of innovative drugs. Pharmaceuticals that are falsely labeled as toxic never reach clinical trials and, conversely, pharmaceuticals that are considered safe can lead to serious risk for volunteers and patients. Novel and complex diseases and the entry of largely human specific biopharmaceuticals on the market further reduce the value of animal studies. In this thesis the predictive value of regulatory non-clinical animal studies in drug development was studied. As a source for the study marketing authorization applications were used, which include all the animal studies that have been performed by the manufacturer to support registration of the drug. Whereas the current discussion on animal studies mostly centers on political and ethical arguments, we aim to offer a scientific basis of the predictive value of animal studies in drug development. In this chapter, we

identify weaknesses of regulatory animal testing paradigms in predicting safety aspects that translate to the clinical dossier and discuss the opportunities this creates to re-evaluate current guidelines on non-clinical animal studies based scientific facts.

BIAS, UNTAPPED DATA AND ANIMAL STUDIES IN PHARMACOVIGILANCE

Scientific publications on animal studies with drugs under development are available as one source for study. However, to rely exclusively on these sources would introduce considerable bias because non-clinical safety studies in animals are not published often. Negative data is also more likely not to be published(*8, 9*). Publication bias of animal studies in stroke research was recently revealed by a meta-analysis of several systematic reviews involving 525 unique publications where it contributed to a considerable overestimation of the effect size(*10*). Publication bias is also reported by Begley and Ellis in their investigation of cancer studies using mouse models(*11*). European Public Assessment Reports are an abbreviated summary of the marketing authorization application for the public, contain safety studies in animals and have been used as a source to study clinical comparability of biosimilars(*12*). While this source contains both positive and negative data, it is not complete because it is a summary of the most important findings. The marketing authorization application (MAA) on the other hand, contain all the studies that are done to demonstrate the safety and efficacy of a drug under investigation, but these are largely confidential and data included in these documents is difficult to obtain(*13*).

Because both the proof of concept and risk profiles are established in animal studies, the non-clinical sections of the MAA contain all positive and negative data and effectively limits a part of the publication bias. The collaboration between Utrecht University and the Dutch Medicines Evaluation Board created an unprecedented opportunity to use the MAA as a source for our studies. However, only using these data means that products that were terminated because of adverse findings in non-clinical studies or clinical trials and that did not receive marketing authorization are not studied. Combined with the high attrition rates in non-clinical and clinical research, this also effectively limits the available development programs that can be studied.

To demonstrate the value of the MAA as a source, we studied whether new adverse drug reactions that were identified after marketing authorization could be detected in the non-clinical sections of the MAA for small molecule therapeutics (SMTs) in Chapter 2. Overall, 63% of all serious adverse reactions (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. Signals for these adverse events were seen in only few animals or were considered species-specific. In addition, non-clinical studies are not powered to detect rare adverse events, but include exposures that are the multiple of the intended clinical dose to elicit the most complete toxicological response possible. The incidence rate of adverse effects in animals was low and adverse reactions developed after administration of high doses and/or prolonged exposure. Therefore, it is plausible to assume that these findings have little clinical relevance and that these events are unlikely to occur in the clinical trial population.

In addition, the number of adverse reactions in animals that does not translate to a human adverse drug reaction (false positives) increases with increasing dose, so that over-exposure might not produce meaningful results(*3*). Similarly, in Chapter 3 we showed that new adverse drug reactions for biopharmaceuticals that were detected after marketing authorization had an animal counterpart in 50% of the cases. For biopharmaceuticals all adverse drug reactions were related to the pharmacology of the target or were immune responses to the biopharmaceutical. Therefore, while the exact adverse drug reactions that would occur could not be predicted, these reactions could have been expected based on the pharmacological effect of the biopharmaceutical. For example, progressive multifocal leukoencephalopathy (PML) has been associated with the use of the immunosuppressor Raptiva (efalizumab)(*14, 15*). While no animal ever developed PML, profound and prolonged suppression of the immune system was observed in animal studies(*16*). This immune suppression does allow latent viruses like John Cunningham virus, the probable cause of PML, to reactivate(*17*). Animal data would probably not predict these findings because animal studies are not appropriate, adequately powered or lengthy enough to detect these adverse effects.

Today, data from non-clinical studies are added to risk management plans when these have identified a clear potential risk that should be monitored after marketing(*18*). In addition, there is an increase in non-clinical post marketing commitments that are being requested. In part, this reflects differences in perceived risk by pharmaceutical companies and regulatory authorities or how the relevance of existing guidelines is perceived(*19*). This may reflect the increased regulatory demand to closely monitor the safety of drugs and the establishment of the Pharmacovigilance Risk Assessment Committee (PRAC) at the European Medicines Agency (EMA)(*20*). When clinical data is available, this should be considered more informative than animal studies and non-clinical postmarketing commitments can be foregone, save for, for example, juvenile toxicity studies to support pediatric trials. However, the value of juvenile toxicity studies is still being studied and these studies are not always informative or needed(*21-23*).

Animal data will likely not lead to better management of risk for pharmaceuticals. The signals from animal studies must already be considered relevant for the human situation to be included in the risk management plan. High safety margins, low incidence and species-specific responses all lead to reduced clinical relevance and rare adverse effects are therefore unlikely to be noticed or monitored. At first glance, using animal data in risk management plans could be considered worthwhile because for biopharmaceuticals such data is often predictive or can be explained by findings from animal studies. But as was the case with SMTs, in all likelihood animal data will contribute minimally to pharmacovigilance activities with biopharmaceuticals because those adverse events are dependent on pharmacology or immune reactions.

Companies rolling out pharmacovigilance strategies for biopharmaceuticals have been recommended to focus on understanding the target biology and intended mechanism of action to guide these efforts, and to take the unique features of biopharmaceuticals into account(*24, 25*). This is confirmed in our studies. Adverse reactions for biopharmaceuticals are relatively easy to explain because of the inherent attributes of these products. But the

timing of adverse drug reactions of biopharmaceuticals is not predictable. Having animal data available is not likely to improve this. Therefore, animal data are not an optimal tool for prospective risk management in patients using biopharmaceuticals or small molecules but only serves a retrospective purpose.

THE FUTURE OF NHP STUDIES IN THE DEVELOPMENT OF MONOCLONAL ANTIBODIES IS LIMITED BECAUSE OF EXPECTED PHARMACOLOGY AND IMMUNOGENICITY

The MAAs were also used to evaluate the use of non-human primates (NHP) in the development of monoclonal antibodies (mAbs). Non-clinical development of biopharmaceuticals is guided by the international guideline ICH S6(*26*). This guideline offers a flexible, science based, and case-by-case approach to develop biotech products and does not require all studies that are needed for the development of SMTs.

For biopharmaceutical products, such as mAbs, NHPs are often considered to be the only relevant species for safety testing because NHP possess the relevant target antigen and other species do not(*27, 28*). But their value has been questioned since the start of the development of biopharmaceuticals(*29, 30*). The vast body of knowledge on mAbs and their development, which is available to us today, is a product of a continuous and iterative learning process. In Chapter 4 we describe that even though this knowledge is also based on NHP studies, it suggests that their future use has considerable limits and even may not be informative at all. We found that the use of NHP is most often justified by being the only pharmacologically responsive species. However, NHP have been used in the development of mAbs even if they lack this attribute, for example because they are considered the common species for the development of mAbs. ICH S6 allows the use of a single species if this can be justified(*26*). Despite the availability of this opportunity the use of two species is standard practice, similar to the classical small molecule approach, rather than a rare occurrence, and one which arose from scientific necessity. This may be due to risk aversion to be non-compliant with regulations or uncertainty about regulatory demands. This facilitates a "check-box approach" despite the fact that the additional data is not always scientifically informative or needed. Earlier dialogue with regulatory authorities through scientific advice may increase regulatory intelligence in these cases. In addition to two species testing, there is also additional evidence for lack of a real case-by-case approach because most non-clinical programs generally followed or exceeded the study requirements outlined by the guidelines.

mAbs have a very high affinity for their target and off-target effects are not probable or expected(*31, 32*). Our study confirms that all adverse events induced by mAbs are highly predictable because they are either mediated by the pharmacology of the MAb and can be exaggerated by dose or exposure. In addition, adverse effects can also be mediated by immune responses against the mAb(*33, 34*). Our findings in Chapter 4 also relate to *post-mortem* results as well as developmental and reproductive toxicity studies. Investigations into placental transfer of NHP in reproductive toxicity studies with biopharmaceuticals have shown similar findings(*35*). A recent analysis of non-clinical animal studies to evaluate

biopharmaceuticals showed that rodent studies using surrogate mAbs were no more or no less predictive than studies with NHP being administered the clinical mAb and no evidence was found that NHP have better predictive value(*31, 32*). Human adverse effects were less detectable because of low sample size in NHP studies, the laboratory setting, complications of the disease, which are not present in animals, or simply not being able to measure descriptive endpoints such as headache in animals(*31, 32*).

In Chapter 5, we found that the interpretation, and therefore the relevance, of studies in NHP could be limited by immunogenicity. Immunogenicity can occur via a number of variables, including the physicochemical characteristics of the mAb, the manufacture and formulation, duration of treatment and can even be based on the patient population(*36*). Currently, the main strategy to reduce immunogenicity is humanization, where parts of, or the whole mAb is engineered to contain human sequences. Nevertheless, immunogenicity that can influence the effectiveness of a humanized mAb in the clinic can still occur(*37-40*).

In primates, humanization had little effect in reducing immunogenicity. This is not unexpected because non-NHP mAbs are recognized by the immune system as non-self.

In 98% of the development programs, immunogenicity occurred at one point in the program. This is not necessarily a problem. The anti-drug antibody response does not always have a pharmacodynamic or pharmacokinetic effect unless the ADA response is directed against the Fc-region, or the complement determining region, respectively. In addition, if the incidence of ADA formation occurs in a small subset of animals this also does not have to lead to difficulties in the interpretation of study results. However, only 33% of mAbs in development had a low incidence of immunogenicity. mAbs were moderately immunogenic in 41% of the cases and highly immunogenic in 26%. Most ADA responses in NHPs were directed against the Fc-region (anti-isotype) of the mAbs, resulting in enhanced clearance. In some cases, loss of efficacy and adverse effects were reported after the induction of ADAs. Increased clearance, neutralizing ADA, or both were detected in 78% of the programs. ADA could affect all dose treated groups or specific dose groups. In those cases, the low dose groups were predominantly affected. Increased clearance of the mAb by ADAs led to shortened repeat dose toxicity studies in 19% of the cases. More importantly, immunogenicity led to warnings to cautiously interpret the pharmacokinetic and pharmacodynamic data because the number of animals with maintained exposure was low, dose groups had been abolished at one stage, or animals required increased doses to maintain exposure after clearing ADA were detected.

Methodological issues have also limited the value of experiments in NHP. Wild caught or old monkeys have been used in safety studies. For wild caught monkeys, the age, genetic background, exposure to pathogens, environmental and social conditions are unknown and may introduce high inter-animal variability. Old monkeys can introduce adverse effects related to age and this makes it difficult to attribute the observed effects as a drug related or as an effect of age with confidence. Low sample sizes are common in studies with NHP due to the ethical considerations and cost involved with running these experiments(*28*). However, this can limit the interpretability of data generated in this species, particularly when functional ADAs are induced.

The knowledge base for the development of mAbs, to which NHP data has contributed, continues to evolve. But our analysis suggests that the role of NHPs should be limited in the future. The effects that are observed are either related to the pharmacology of the mAb or an immune response until ADAs interfere with the interpretation of study results. Therefore, extensive animal studies as they are done now are not recommended for mAbs. A single repeat dose study would equally be able to show expected pharmacology as long as ADA are not prevalent.

THE CHANGE OF REGULATION ON THE USE OF ANIMALS IN THE DEVELOPMENT OF BIOPHARMACEUTICALS: THE CASE OF BIOSIMILARS

Not only innovative drugs are required to have animal data to support the marketing authorization. Because biopharmaceuticals are relatively large and complex proteins, these are difficult to fully characterize(*41*). The final biopharmaceutical product is often a complex mixture of closely related molecules. The exact composition of this mixture can be influenced by protein aggregation and glycosylation patterns(*42, 43*). These differences may ultimately influence the pharmacodynamics, kinetics or toxicity parameters(*44-46*). To guide the marketing authorization of these biopharmaceutical copies, so-called biosimilars, guidelines have been developed by the EMA(*47-49*). Animal studies are considered necessary in the biosimilarity exercise and should be included in an assessment of the pharmacodynamics and the safety profile after repeated administration of the biopharmaceutical(*48*). In the current guideline on non-clinical and clinical issues of biosimilars, comparative non-clinical studies that are sensitive enough to detect differences between the biosimilar and the reference product are requested(*48*). This case-by-case approach effectively limits the number of animal studies needed. However, it remains unclear what the contribution has been of animal studies in establishing biosimilarity. In chapter 6 we assess the non-clinical programs of the marketing authorization applications of all biosimilar products registered in the EU or submitted for authorization until January 1st 2013. We found that animal studies submitted for biosimilar applications do not show relevant differences compared to the reference product. Animal studies were all considered adequate to demonstrate pharmacodynamic similarity of the biosimilar to the reference product but only rarely identified differences in pharmacodynamic or toxicity studies with questionable clinical relevance. Several biosimilar applications have known quality differences compared to the reference product in terms of level of impurities, host cell type, formulation or post-translational modification(*12, 50*). Yet all of the non-clinical animal studies supported biosimilarity, while differences were observed in clinical PK profiles(*12*). Our data suggest that animal studies are likely not sensitive enough to detect these differences or to translate them into measurable endpoints because of low number of animals in a study and the inherent variability of animal models (also see Chapter 1). In addition, safety studies in animals never revealed new or unexpected toxicity. However, when considerable differences are observed between the quality of the biosimilar and reference product, these are far more indicative of dissimilarity and are later reflected in animal and human data(*51, 52*).

These differences usually lead to major objections by regulatory authorities(*51, 52*). This further stresses the use of a step-wise approach to demonstrating biosimilarity, where full characterization and similarity to the reference product should be demonstrated in terms of quality manufacture and control and should be the starting point for any subsequent studies. All companies started the development of the first generation biosimilars before the first biosimilar guideline appeared. Non-clinical strategies were developed by focusing on existing guidelines for biopharmaceuticals, such as ICH S6, common scientific sense and scientific advice of regulatory authorities(*53*). The new guidance on biosimilars came into effect in the EU in 2004(*54*). In the early regulatory guidelines, animal studies have been considered essential to determine similarity in potency, pharmacological effects and safety of new biosimilars(*48*). However, the next generation of biosimilars, which includes monoclonal antibodies, is on the doorstep of regulatory approval. These products are highly species specific and animal studies may not be feasible or informative. We performed a hypothetical biosimilarity exercise in chapter 7 to illustrate opportunities and challenges of biosimilar mAb development in oncology using Bevacizumab (Avastin) as an example. For mAbs, the use of *in vitro* assays may be more informative than animal studies because these assays are quantifiable and may be more specific and sensitive to identify differences than animal studies(*55*). Finding a relevant species for Bevacizumab can be challenging when attempting to demonstrate pharmacodynamic similarity. Most common *in vivo* PD models for cancer are rodent xenografts, which do not optimally represent the situation of advanced metastatic cancer(*11, 56*). On the other hand, safety studies in non-human primates did not reveal limiting toxicities and are not likely to generate additional information(*57*). Therefore, clinical studies will be required regardless of the outcome of *in vivo* studies. This study demonstrates that anticancer products such as bevacizumab may require regulatory authorities to reconsider the objectives of the comparability exercise.

Recently, the draft guidance for biosimilars has been released for consultation(*58, 59*). The draft guidance adopts a step-wise, risk-based approach, where the manufacturer should consider the feasibility and relevance of animal studies and favors *in vitro* approaches where available. When combined with getting scientific advice from regulatory authorities and performing 'scientific common sense' studies, our study underwrites this position, although when the quality of a biopharmaceutical application can be characterized and is similar, it should also mean that animal studies can be foregone completely as is the case for small molecule generics.

IMPLICATIONS FOR THE FUTURE USE OF NON-CLINICAL ANIMAL STUDIES IN DRUG DEVELOPMENT

This thesis would not have been possible without the unrestricted access to the MAA that was made available by the Dutch Medicines Evaluation Board, with support of Nefarma. To further support the limited value of animal studies in drug development, investigational brochures should be included as a source for future studies. These documents form the basis on which first-in-man trials are approved and it is at this junction that the decision has to be made whether the data from animal studies translates to progression into the clinic.

While pharmaceuticals that have not reached this stage are still not included, it is another important part of the data source that has to be analysed, for example, to study how often animal data led to safety concerns that precluded a clinical trial.

These data and those from the MAA are, to a certain extent, sensitive with regards to commercial interests and privacy of patients and must be used with care(*60*). But making data available for unrestricted and serious study, including data from projects that have been terminated before marketing submission, is needed. There is some progress in this regard. The EMA is in the process of developing a policy on the proactive publication of clinical data, which should come into force on January 1st 2014(*61*). If clinical trial data can be made public, it may be that there are also no real objections that stand in the way of publishing non-clinical data as well. It takes effort and trust of both regulatory authorities and pharmaceutical companies to offer non-clinical and clinical data to the academic community for study. This mountain of data must be organized and analysed which in itself is a challenge. Promising approaches such as 'big data' analysis are being developed that may be used to this end(*62-64*). Systematic reviews and meta-analysis can be used to evaluate and improve the quality of study designs, reveal reporting transparency and can provide evidence based decision making in species selection for future experiments(*65*). Together with academia, the pharmaceutical industry and the regulatory authorities, the complete record of animal studies can be studied and held to the light. As a result of these thorough studies, the discussion of the predictive value of animal studies is facilitated and with the scientific facts at hand.

Nevertheless, even if MAAs are made freely available for study, the vast majority of animal data remains in the private libraries of pharmaceutical companies. Several academic groups have successfully collaborated with pharmaceutical industry companies to study data from these sources(*22, 66-69*). This has equally proven to be valuable and several of these studies have made recommendations to limit the use of animals in drug development. In a few cases, the studies using public and private data have prompted the revision of the guideline(*69, 70*). For example, the use of acute toxicity studies, where lethality is an endpoint, has been challenged by the European Partnership for Animal Alternatives (EPAA) since the data from these animal studies is not considered informative(*69*). This position has been adopted by the European regulatory authorities and the ICH and acute single dose toxicity studies are no longer a requirement(*71*).

However, most of the analyses of non-clinical animal studies have not fully challenged the scientific validity of animal studies but suggest ways to reduce or refine their need. There is a growing body of evidence that animal studies have their limitations, which in turn limits the efficiency of drug development(*72-78*). Or, to quote the FDA's critical path initiative to improve drug development, it is because '*developers are forced to rely on the tools of the last century to evaluate this century's advances*'(*79*, p.12). These limitations can even lead to the irrelevance of extensive animal studies, for example when informative studies are impossible due to immunogenicity, adverse effects are expected or animal studies are not sensitive enough to detect small differences between products. This creates difficulties for both the pharmaceutical industry which is obliged to meet regulatory requirements

and for the regulatory authorities that must ensure that the submitted data is adequate in assessing the safety and efficacy of a drug. At the same time, the use of animals for research is guided by very strict laws in the EU(*6, 80*). For example, the use of animal procedures should be restricted to areas which may ultimately benefit human or animal health, or the environment(*80*). In addition, the use of NHP is only allowed if they are essential for the benefit of human beings and no other alternative replacement methods are available(*80*). In addition, Directive 89/609/EC Article 7 states that '*When an experiment has to be performed, the choice of species shall be carefully considered and, where necessary, explained to the authority'(80).* These laws out rank guidelines, which are not mandatory, but suggestions to proceed and deviations are possible when scientifically justified. Clearly, uninformative studies in NHP are not of benefit for humans. Therefore, the use of NHP may not have an ethical and lawful basis. The EU has implemented a ban on the use of animals for the testing of cosmetics(*81*). Similarly, such a ban could be imposed for the development of new drugs. This is a rather drastic approach for complex products such as pharmaceuticals, but may ultimately provide a common starting point for serious discussions with pharmaceutical industry and regulatory authorities. Better communication with regulatory authorities and pharmaceutical industry can initially help to develop a testing strategy that is based on scientifically relevant data. For example, to justify the use of a single relevant species that can confirm the expected pharmacological effects in one repeat dose study and negating the need for rodent studies. In this way, regulatory flexibility makes it possible to let scientific common sense lead development after reaching consensus with regulatory authorities when the state-of-the-art is ahead of content in the guideline. This would allow a phased transition towards animal free development.

Another suggestion to improve non-clinical animal study use is the introduction of reporting guidelines for non-clinical animal studies to improve the quality of study design(*82*). In addition, non-clinical study registries can be set up to improve reporting of animal studies. In combination with scientific advice from regulatory authorities, this could aid in the design of a science based non-clinical program. At the same time, it discourages the use of "check-box" approaches and requires a careful consideration of the need for animal studies. An additional benefit is that pharmaceutical companies are less likely to run into regulatory hurdles or conduct irrelevant studies if the need and design of pivotal non-clinical studies are discussed with regulatory authorities prior to the start of the non-clinical animal testing program.

Where no relevant model exists, a slow and cautious entry into fist-in-man studies is the only scientifically relevant option. These can be preceded by alternative technologies such as -omics, bio-informatics approaches, tissue culture assays, the use of stem-cells, and chip-based assays. However, these technologies still need to be further developed and validated to have real value in evaluating the safety profile of new innovative medicines.

Micro-dosing studies in humans can already be done to evaluate pharmacokinetic data for a new pharmaceutical product(*83-85*). Advances in experimental medicine create opportunities to study the pharmacodynamic effects and disposition of an investigational drug in humans the 'ultimate model' system or to identify new biomarkers, thereby

increasing the confidence in the relevance of drug targets, which are poorly predicted by animal efficacy studies(*86*). Another instance where direct entry to clinical trials is possible is when mAbs are being developed that are copies of marketed products, since quality issues are more likely to be indicative of similarity. The European legislation currently requires a product to 'demonstrate biosimilarity', but this could be altered to 'exclude excessive dissimilarity', which can also be shown *in vitro*.

Regulatory authorities are increasingly becoming aware of the limitations of animal research. In September 2010, the Dutch Medicines Evaluation Board issued a statement on their website that '*the Medicines Evaluation Board strives for drug development without the use of animals*'(*87*). The EMA has created an expert group to advise on 3Rs (refinement, reduction and replacement) in the development of medicinal products in order to '*Eliminate[e] repetitious and unnecessary animal testing*'(*88*). A thorough revision of the existing guidelines and pharmacopoeia monographs would be an excellent starting point to achieve a first reduction of animal studies. The withdrawal of the note for guidance on single dose toxicity and the draft guidance on non-clinical and clinical issues in biosimilar development, which states that there may not be any relevance for NHP studies, demonstrate that this is possible(*71*). Governments can aid in the acceleration of using innovative non-clinical testing programs which use limited but high quality animal data or even no animal data by offering incentives such as additional data protection for approved products. This approach has been successful in generating much needed data on drug disposition in children via pediatric clinical trials and orphan drug registration(*89, 90*).

CONCLUDING REMARKS

As we have shown in this thesis, animal studies to assess the safety of new drugs are not always needed and can be, in fact, unscientific, uninformative or irrelevant. A thorough revision of regulatory guidelines on non-clinical issues will be helpful to identify unnecessary animal studies and improve the efficiency of non-clinical drug development.

Increased and earlier dialogue between pharmaceutical industries and regulatory authorities through scientific advice could help to improve regulatory intelligence and reduce requirements for animal studies in non-clinical programs. This dialogue has been shown to lead to withdrawal of unnecessary guidelines and reduce unnecessary use of animals in drug development. But further dialogue is needed, since the path towards animal free drug development is not the responsibility of a single party but a joint scientific effort. This may require regulatory authorities to rethink their reliance on the precautionary principle, since it does not require scientific consensus and does not promote innovation. On the other hand, the pharmaceutical industries will need to invest in developing new and improving on existing technologies to provide informative information on the safety of new drugs. Governments have an important role in incentivizing these activities, which includes the use of political leverage as well as economical incentives, and promoting further research.

Privately held data in the form of marketing authorization application or animal study reports of drugs hold a wealth of information that can contribute to the discussions on the need of animal studies, revision of guidelines and overall decision making process in

drug development policy. This data needs to be made available for research. With perfect hindsight and the scientific facts at hand, the pharmaceutical industry, regulatory authorities and academia have an opportunity to reconsider when and how (and for how much longer) we use animals in drug development. This does require considerable effort and trust by pharmaceutical companies and regulatory authorities alike but the trade-off is an efficient non-clinical drug development process which is based on science.

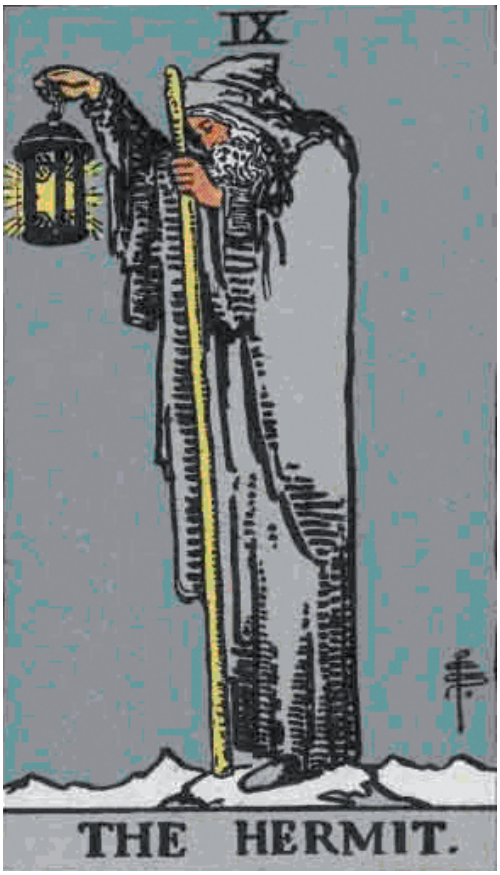
REFERENCES

1. A. H. Owens, Jr., Predicting anticancer drug effects in man from laboratory animal studies. *J Chronic Dis* 15, 223 (Mar, 1962).
2. J. D. Schein, A. Gentile, K. W. Haase, Development and evaluation of an expanded hearing loss scale questionnaire. *Vital Health Stat 2*, 1 (Apr, 1970).
3. T. Igarashi, S. Nakane, T. Kitagawa, Predictability of clinical adverse reactions of drugs by general pharmacology studies. *J Toxicol Sci* 20, 77 (May, 1995).
4. H. Olson *et al.*, Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56 (Aug, 2000).
5. A. P. Fletcher, Drug safety tests and subsequent clinical experience. *J R Soc Med* 71, 693 (Sep, 1978).
6. Anonymous, Directive 2001/83/EC of the European Parliament and the Council of 6 November 2001 on the Community code relating to medicinal products for human use. (2001).
7. R. Greek, J. Greek, Evolution and animal models. *Comp Med* 52, 499 (Dec, 2002).
8. J. P. Ioannidis, Why most discovered true associations are inflated. *Epidemiology* 19, 640 (Sep, 2008).
9. M. B. Bracken, Why are so many epidemiology associations inflated or wrong? Does poorly conducted animal research suggest implausible hypotheses? *Ann Epidemiol* 19, 220 (Mar, 2009).
10. E. S. Sena, H. B. van der Worp, P. M. Bath, D. W. Howells, M. R. Macleod, Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 8, e1000344 (Mar, 2010).
11. C. G. Begley, L. M. Ellis, Drug development: Raise standards for preclinical cancer research. *Nature* 483, 531 (Mar 29, 2012).
12. H. Schellekens, E. Moors, Clinical comparability and European biosimilar regulations. *Nat Biotechnol* 28, 28 (Jan, 2010).
13. P. C. Gotzsche, A. W. Jorgensen, Opening up data at the European Medicines Agency. *Bmj* 342, d2686 (2011).
14. European Medicines Agency, Public statement on Raptiva (efalizumab) - Withdrawal of the marketing authorisation in the European Union. , (2009).
15. N. Kothary *et al.*, Progressive multifocal leukoencephalopathy associated with efalizumab use in psoriasis patients. *J Am Acad Dermatol* 65, 546 (Sep, 2011).
16. European Medicines Agency. (2009).
17. A. Bellizzi *et al.*, New Insights on Human Polyomavirus JC and Pathogenesis of Progressive Multifocal Leukoencephalopathy. *Clin Dev Immunol* 2013, 839719 (2013).
18. S. Frau, M. Font Pous, M. R. Luppino, A. Conforti, Risk Management Plans: are they a tool for improving drug safety? *Eur J Clin Pharmacol* 66, 785 (Aug, 2010).
19. L. M. Reeve, Non-clinical Post-Marketing Commitments for newly licenced pharmaceuticals. *Regul Toxicol Pharmacol*, (Jul 7, 2009).
20. J. J. Borg, G. Aislaitner, M. Pirozynski, S. Mifsud, Strengthening and rationalizing pharmacovigilance in the EU: where is Europe heading to? A review of the new EU legislation on pharmacovigilance. *Drug Saf* 34, 187 (Mar 1, 2011).
21. I. Leconte *et al.*, Value of juvenile animal studies. *Birth Defects Res B Dev Reprod Toxicol* 92, 292 (Aug, 2011).
22. G. P. Bailey, D. Marien, The value of juvenile animal studies "What have we learned from preclinical juvenile toxicity studies? II". *Birth Defects Res B Dev Reprod Toxicol* 92, 273 (Aug, 2011).

23. L. De Schaepdrijver *et al.*, Real life juvenile toxicity case studies: the good, the bad and the ugly. *Reprod Toxicol* 26, 54 (Sep, 2008).
24. H. C. Ebbers *et al.*, The role of Periodic Safety Update Reports in the safety management of biopharmaceuticals. *Eur J Clin Pharmacol* 69, 217 (Feb, 2013).
25. T. J. Giezen, A. K. Mantel-Teeuwisse, H. G. Leufkens, Pharmacovigilance of biopharmaceuticals: challenges remain. *Drug Saf* 32, 811 (2009).
26. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Preclinical safety evaluation of biotechnology derived pharmaceuticals S6(R1). *(online)*, (2011).
27. J. L. Bussiere, Species selection considerations for preclinical toxicology studies for biotherapeutics. *Expert Opin Drug Metab Toxicol* 4, 871 (Jul, 2008).
28. K. Chapman *et al.*, Preclinical development of monoclonal antibodies: considerations for the use of non-human primates. *MAbs* 1, 505 (Sep-Oct, 2009).
29. K. Teelmann, C. Hohbach, H. Lehmann, Preclinical safety testing of species-specific proteins produced with recombinant DNA-techniques. An attempt to transfer current experience into future testing strategies. *Arch Toxicol* 59, 195 (Dec, 1986).
30. M. Kooijman, P. J. van Meer, E. H. Moors, H. Schellekens, Thirty years of preclinical safety evaluation of biopharmaceuticals: Did scientific progress lead to appropriate regulatory guidance? *Expert Opin Drug Saf* 11, 797 (Sep, 2012).
31. P. J. Bugelski, P. L. Martin, Concordance of Preclinical and Clinical Pharmacology and Toxicology of Therapeutic Monoclonal Antibodies and Fusion Proteins: 2-Cell Surface Targets. *Br J Pharmacol*, (Dec 13, 2011).
32. P. L. Martin, P. J. Bugelski, Concordance of Preclinical and Clinical Pharmacology and Toxicology of Monoclonal Antibodies and Fusion Proteins: 1-Soluble Targets. *Br J Pharmacol*, (Dec 13, 2011).
33. P. Baldrick, Safety evaluation of biological drugs: what are toxicology studies in primates telling us? *Regul Toxicol Pharmacol* 59, 227 (Mar, 2011).
34. T. T. Hansel, H. Kropshofer, T. Singer, J. A. Mitchell, A. J. George, The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discov* 9, 325 (Apr, 2010).
35. N. Pentsuk, J. W. van der Laan, An interspecies comparison of placental antibody transfer: new insights into developmental toxicity testing of monoclonal antibodies. *Birth Defects Res B Dev Reprod Toxicol* 86, 328 (Aug, 2009).
36. H. Schellekens, Factors influencing the immunogenicity of therapeutic proteins. *Nephrol Dial Transplant* 20 Suppl 6, vi3 (Jun, 2005).
37. G. M. Bartelds *et al.*, Development of antidrug antibodies against adalimumab and association with disease activity and treatment failure during long-term follow-up. *JAMA* 305, 1460 (Apr 13, 2011).
38. M. K. de Vries *et al.*, Decreased clinical response to infliximab in ankylosing spondylitis is correlated with anti-infliximab formation. *Ann Rheum Dis* 66, 1252 (Sep, 2007).
39. J. D. Isaacs *et al.*, CAMPATH-1H in rheumatoid arthritis--an intravenous dose-ranging study. *Br J Rheumatol* 35, 231 (Mar, 1996).
40. C. L. Krieckaert, M. T. Nurmohamed, G. J. Wolbink, Methotrexate reduces immunogenicity in adalimumab treated rheumatoid arthritis patients in a dose dependent manner. *Ann Rheum Dis*, (May 14, 2012).
41. D. J. Crommelin *et al.*, Shifting paradigms: biopharmaceuticals versus low molecular weight drugs. *Int J Pharm* 266, 3 (Nov 6, 2003).
42. H. C. Mahler, W. Friess, U. Grauschopf, S. Kiese, Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci* 98, 2909 (Sep, 2009).
43. M. C. Manning, D. K. Chou, B. M. Murphy, R. W. Payne, D. S. Katayama, Stability of protein pharmaceuticals: an update. *Pharm Res* 27, 544 (Apr, 2010).
44. R. Jefferis, Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* 8, 226 (Mar, 2009).
45. N. Jenkins, L. Murphy, R. Tyther, Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. *Mol Biotechnol* 39, 113 (Jun, 2008).
46. G. Walsh, R. Jefferis, Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology* 24, 1241 (2006).

47. European Medicines Agency, Guideline on similar biological medicinal products. (2005).
48. European Medicines Agency, Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2006).
49. European Medicines Agency, Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues. (2006).
50. V. Brinks *et al.*, Quality of original and biosimilar epoetin products. *Pharm Res* 28, 386 (Feb, 2011).
51. European Medicines Agency, Questions and answers on the recommendation for refusal of marketing application for Alpheon. (2006).
52. European Medicines Agency, Questions and answers on the withdrawal of the marketing authorization application for Inuslin human rapid Marvel, Insulin human long Marvel, Insulin human 30/70 mix Marvel, International non-proprietary name (INN): human insulin. (2006).
53. M. J. Parnham, J. Schindler-Horvat, M. Kozlovic, Non-clinical safety studies on biosimilar recombinant human erythropoietin. *Basic Clin Pharmacol Toxicol* 100, 73 (Feb, 2007).
54. European Medicines Agency, Guideline on comparability of medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2004).
55. A. M. Goodwin, *In vitro* assays of angiogenesis for assessment of angiogenic and anti-angiogenic agents. *Microvasc Res* 74, 172 (Sep-Nov, 2007).
56. L. R. Kelland, Of mice and men: values and liabilities of the athymic nude mouse model in anticancer drug development. *Eur J Cancer* 40, 827 (Apr, 2004).
57. European Medicines Agency, European Public Assessment report for Avastin. (2006).
58. European Medicines Agency, Draft guideline on similar biological medicinal products. (2013).
59. European Medicines Agency, Draft guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. (2013).
60. H. G. Eichler, E. Abadie, A. Breckenridge, H. Leufkens, G. Rasi, Open clinical trial data for all? A view from regulators. *PLoS Med* 9, e1001202 (2012).
61. R. Steinbrook, The European Medicines Agency and the Brave New World of Access to Clinical Trial DataBrave New World of Access to Clinical Trial Data. *JAMA Internal Medicine* 173, 373 (2013).
62. S. Davis, M. Ratcliffe, Big Data, Big Challenges, Big Opportunities for Big Pharmaceutical and Biotechnology. (2012).
63. V. Marx, Biology: The big challenges of big data. *Nature* 498, 255 (2013).
64. R. Abbott, Big Data and Pharmacovigilance: Using Health Information Exchanges to Revolutionize Drug Safety. *Iowa Law Review, Forthcoming*, (2013).
65. R. B. de Vries, P. Buma, M. Leenaars, M. Ritskes-Hoitinga, B. Gordijn, Reducing the number of laboratory animals used in tissue engineering research by restricting the variety of animal models. Articular cartilage tissue engineering as a case study. *Tissue Eng Part B Rev* 18, 427 (Dec, 2012).
66. K. L. Chapman *et al.*, The design of chronic toxicology studies of monoclonal antibodies: implications for the reduction in use of non-human primates. *Regul Toxicol Pharmacol* 62, 347 (Mar, 2012).
67. E. S. Trepakova, J. Koerner, S. D. Pettit, J. P. Valentin, A HESI consortium approach to assess the human predictive value of non-clinical repolarization assays. *J Pharmacol Toxicol Methods* 60, 45 (Jul-Aug, 2009).
68. K. Chapman *et al.*, A global pharmaceutical company initiative: An evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. *Regul Toxicol Pharmacol*, (Apr 19, 2013).
69. S. Robinson *et al.*, A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. *Regul Toxicol Pharmacol* 50, 345 (Apr, 2008).
70. J. Clarke *et al.*, Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? *Regul Toxicol Pharmacol* 50, 2 (Feb, 2008).
71. European Medicines Agency, Questions and answers on the withdrawal of the 'Note for guidance on single dose toxicity'. (2010).

72. R. A. Matthews, Medical progress depends on animal models - doesn't it? *J R Soc Med* 101, 95 (Feb, 2008).
73. P. Pound, S. Ebrahim, P. Sandercock, M. B. Bracken, I. Roberts, Where is the evidence that animal research benefits humans? *Bmj* 328, 514 (Feb 28, 2004).
74. T. Hartung, Thoughts on limitations of animal models. *Parkinsonism Relat Disord* 14 Suppl 2, S81 (2008).
75. R. J. Wall, M. Shani, Are animal models as good as we think? *Theriogenology* 69, 2 (Jan 1, 2008).
76. D. G. Hackam, Translating animal research into clinical benefit. *Bmj* 334, 163 (Jan 27, 2007).
77. R. Lemon, S. B. Dunnett, Surveying the literature from animal experiments. *Bmj* 330, 977 (Apr 30, 2005).
78. J. Kimmelman, A. J. London, Predicting harms and benefits in translational trials: ethics, evidence, and uncertainty. *PLoS Med* 8, e1001010 (Mar, 2011).
79. Food and Drug Administration, Innovation or stagnation: challenge and opportunity to new medicinal products. (2004).
80. Anonymous, Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. (2010).
81. European Commission, Communication from the commission to the European parliament and council on the animal testing and marketing ban and on the state of play in relation to alternative methods in the field of cosmetics. (2013).
82. J. Kimmelman, Ethics, ambiguity aversion, and the review of complex translational clinical trials. *Bioethics* 26, 242 (Jun, 2012).
83. A. I. Graul, Promoting, improving and accelerating the drug development and approval processes. *Drug News Perspect* 20, 45 (Jan-Feb, 2007).
84. I. R. Wilding, J. A. Bell, Improved early clinical development through human microdosing studies. *Drug Discov Today* 10, 890 (Jul 1, 2005).
85. G. Lappin, R. Noveck, T. Burt, Microdosing and drug development: past, present and future. *Expert Opin Drug Metab Toxicol* 9, 817 (Jul, 2013).
86. B. H. Littman, S. A. Williams, The ultimate model organism: progress in experimental medicine. *Nat Rev Drug Discov* 4, 631 (Aug, 2005).
87. College ter Beoordeling van Geneesmiddelen. (2010).
88. European medicines Agency, Mandate, objectives and rules of procedure for the joint CVMP/CHMP ad-hoc expert group on the application of the 3Rs (replacement, reduction and refinement) in regulatory testing of medicinal products (JEG 3Rs). (2011).
89. H. E. Heemstra, R. L. de Vrueh, S. van Weely, H. A. Büller, H. G. Leufkens, Orphan drug development across Europe: bottlenecks and opportunities. *Drug discovery today* 13, 670 (2008).
90. K. Hoppu, Paediatric clinical pharmacology—at the beginning of a new era. *European journal of clinical pharmacology* 64, 201 (2008).



If you can dream - and not make dreams your master;
If you can think - and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build 'em up with worn-out tools;
…
Yours is the Earth and everything that's in it,
And - which is more - you'll be a Man, my son!

-Rudyard Kipling, 1865-1936, *If*

9 SUMMARY

SUMMARY

Animal studies are considered needed as predictive models to evaluate safety and efficacy of new pharmaceuticals and are required by law. However, the scientific basis of the current paradigm on the predictability of animal studies for the effects of drugs in man is under discussion. Much of the uncertainty about the relevance for humans of animal studies stems from three factors: 1) sensitivity, 2) reproducibility and 3) predictivity. The adoption of the precautionary principle by the regulatory authorities and the need to meet regulatory requirements to avoid delays by the pharmaceutical industry, without attempts to improve on the current paradigm, has created a stalemate in which animal studies, predictive or not, continue to exist with little room for innovation. Thus, there is a need for an independent and objective study of the predictive value of non-clinical animal studies based on scientific facts.

In this thesis, we evaluated the scientific basis of the current practices and guidelines for the use of animal studies in pharmaceutical development and assessed the consequences and implication for the regulatory guidelines when animal studies are not informative. An important benefit of this study is that data from marketing authorization applications (MAAs) could be used. MAAs can be a valuable source to study the predictive value of animal studies because they contain all the experiments done to support registration of a drug. In Chapter 2 we demonstrate the value of these marketing authorization applications by using them to study whether new post marketing adverse effects of small molecule therapeutics could have been detected from non-clinical studies. 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. Signals for SARs were seen in only few animals or were considered species-specific. The incidence rate of adverse effects in animals was low and adverse reactions developed after administration of high doses and/or prolonged exposure. The study was expanded to include biopharmaceuticals in Chapter 3. There, we showed that new adverse drug reactions for biopharmaceuticals that were detected after marketing authorization had an animal counterpart in 50% of the cases. For biopharmaceuticals all adverse drug reactions were related to the pharmacology of the target or were immune responses to the biopharmaceutical. Therefore, these reactions could have been expected based on the pharmacological effect of the biopharmaceutical. Animal data are included in risk management plans to help monitor safety issues in the patient population. But because of the high safety margins at which serious adverse reactions occur in animals, low incidence and species-specific responses, these all lead to reduced clinical relevance. Therefore, these signals are unlikely to be noticed and animal data will likely not lead to better management of risk for pharmaceuticals but only has value for retrospective studies.

Biopharmaceutical drug development increasingly uses non-human primate as the primary species because they are often the only species available that are pharmacologically responsive. In Chapter 4 we evaluate the value of the non-human primate in the development of monoclonal antibodies. The use of NHP was most often justified by being the only pharmacologically responsive species. However, NHP have been used in the development of mAbs even if they lack this attribute. The use of two species is standard practice, similar to the classical small molecule approach, rather than a rare occurrence, and one which arose

from scientific necessity. Earlier dialogue with regulatory authorities through scientific advice may increase regulatory intelligence in these cases. In addition to two species testing, there is also additional evidence for lack of a real case-by-case approach because most non-clinical programs generally followed or exceeded the study requirements outlined by the guidelines. Methodological issues have also limited the value of experiments in NHP. For example, the use of wild caught or old monkeys may result in uninformative studies because age, genetic background, exposure to pathogens, environmental and social conditions are unknown and may introduce high variability or confounders. Low sample sizes are common in studies with NHP due to the ethical considerations and cost involved with running these experiments. However, this can limit the interpretability of data generated in this species. Our study confirms that all adverse events induced by mAbs are highly predictable because they are either mediated by the pharmacology of the mAb and can be exaggerated by dose or exposure. Off-target effects are not probable or expected. In addition, adverse effects can also be mediated by immune responses against the mAb, which could influence the interpretability of data from NHP studies. In **Chapter 5** the immunogenicity of monoclonal antibodies in non-human primates was studied. In primates, humanization of the antibody, a technique thought to reduce immunogenicity, had little effect in reducing immunogenicity. In 98% of the development programs, immunogenicity occurred at one point in the program. mAbs were moderately immunogenic in 41% of the cases and highly immunogenic in 26%. Most ADA responses in NHPs were directed against the Fc-region (anti-isotype) of the mAbs, resulting in enhanced clearance. Increased clearance, neutralizing ADA, or both were detected in 78% of the programs. Immunogenicity led to warnings to cautiously interpret the pharmacokinetic and pharmacodynamic data because the number of animals with maintained exposure was low, dose groups had been abolished, or animals required increased doses to maintain exposure after clearing ADA were detected. Therefore, the results from Chapter 4 and 5 suggest that the role of NHPs should be limited in the future. The effects that are observed are either related to the pharmacology of the mAb or an immune response until ADAs interfere with the interpretation of study results. Extensive animal studies as they are done now are not recommended for mAbs. A single repeat dose study would equally be able to show expected pharmacology as long as ADA are not prevalent.

Non-clinical studies are not only required for the development of innovative new drugs. Generic copies of biopharmaceuticals (biosimilars) are routinely assessed in animals to demonstrate that the copy is similar to its reference product. In **Chapter 6** the value of non-clinical animal studies in the biosimilarity exercise are assessed in light of the new draft guideline on biosimilars. Animal studies submitted for biosimilar applications did not show relevant differences compared to the reference product. Animal studies were all considered adequate to demonstrate pharmacodynamic similarity of the biosimilar to the reference product and never revealed unknown or unexpected toxicity. Only rarely were differences in pharmacodynamic or toxicity observed but the clinical relevance of these differences was questionable. Despite demonstrating similarity in animal studies, several biosimilar applications have known quality differences compared to the reference product. Our data

suggest that animal studies are likely not sensitive enough to detect these differences or to translate them into measurable endpoints because of low number of animals in a study and the inherent variability of animal. However, when considerable differences are observed between the quality of the biosimilar and reference product, these are far more indicative of dissimilarity and are later reflected in animal and human data. This further stresses the use of a step-wise approach to demonstrating biosimilarity, where full characterization and similarity to the reference product should be demonstrated in terms of quality manufacture and control and should be the starting point for any subsequent studies.

The next generation of biosimilars, which includes monoclonal antibodies, is on the doorstep of regulatory approval. These products are highly species specific and animal studies may not be feasible or informative. In Chapter 7 we performed a hypothetical biosimilarity exercise to illustrate opportunities and challenges of biosimilar mAb development in oncology using Bevacizumab (Avastin) as an example. We suggest that the use of *in vitro* assays to develop a Bevacizumab biosimilar may be more informative than animal studies because these assays are quantifiable and may be more specific and sensitive to identify differences than animal studies. Informative animal studies with Bevacizumab can be challenging when attempting to demonstrate similarity because common *in vivo* PD models for cancer do not optimally represent the situation of advanced metastatic cancer. Animal studies in a relevant species, the non-human primate, are not likely to generate new information. Therefore, clinical studies will be required regardless of the outcome of *in vivo* studies. This study demonstrates that anticancer products such as bevacizumab may require regulatory authorities to reconsider the objectives of the comparability exercise.

Recently, the draft guidance for biosimilars has been released for consultation. The draft guidance adopts a step-wise, risk-based approach, where the manufacturer should consider the feasibility and relevance of animal studies and favors *in vitro* approaches where available. When combined with getting scientific advice from regulatory authorities and performing 'scientific common sense' studies, our study underwrites this position, although when the quality of a biopharmaceutical application can be characterized and is similar, it should also mean that animal studies can be foregone completely as is the case for small molecule generics.

In Chapter 8 we discuss the findings of this thesis and offer perspectives on improving non-clinical drug development based on scientific facts. To this end, the use of the MAA has been invaluable. Data from the MAA generally remain undisclosed. But making these data available for unrestricted and serious study is needed. However, the vast majority of animal data remains in the private libraries of pharmaceutical companies. The revision of guidelines or withdrawal of unnecessary guidelines has been achieved using proprietary pharmaceutical data demonstrates the value of these sources and should be made available for further study.

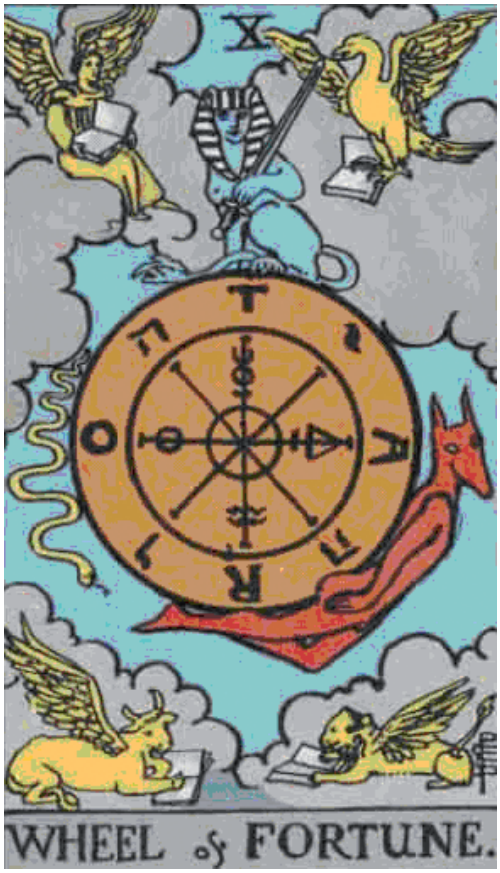
To reduce non-clinical animal studies in the short-term and improve the quality of the studies that are done, the introduction of reporting guidelines for non-clinical animal studies and non-clinical study registries has been suggested. These can be set up to improve the quality of the study design and the reporting of animal studies. In combination with early

scientific advice from regulatory authorities, this could aid in the design of a science based non-clinical program. At the same time, it discourages the use of "check-box" approaches. An additional benefit is that pharmaceutical companies are less likely to run into regulatory hurdles or conduct irrelevant studies. Where no relevant model exists, a slow and cautious entry into fist-in-man studies, preceded by micro-dosing studies, is the only scientifically relevant option. These can be preceded by alternative technologies such as -omics, bio-informatics approaches, tissue culture assays, the use of stem-cells, and chip-based assays.

There is a growing body of evidence that animal studies have their limitations and can be, in fact, unscientific, uninformative or irrelevant. This, in turn, increases the cost and limits the efficiency of drug development. The EU has implemented a ban on the use of animals for the testing of cosmetics. Similarly, such a ban could be imposed for the development of new drugs. This is a rather drastic approach for complex products such as pharmaceuticals, but may ultimately provide a common starting point for serious discussions with pharmaceutical industry and regulatory authorities. This would allow a phased transition towards animal free development.

A thorough revision of regulatory guidelines on non-clinical issues will be helpful to identify unnecessary animal studies and improve the efficiency of non-clinical drug development. Increased and earlier dialogue of pharmaceutical industries with regulatory authorities through scientific advice could help to improve regulatory intelligence and reduce requirements for animal studies in non-clinical programs. However, regulatory authorities will have to rethink their reliance on the precautionary principle, since it does not require scientific consensus and does not promote innovation. On the other hand, the pharmaceutical industries will need to invest in alternative technologies to provide informative information on the safety of new drugs. Governments have an important role in incentivizing these activities, which includes the use of political leverage as well as economical incentives, and promoting further research.

All of this does require considerable effort and trust by pharmaceutical companies and regulatory authorities alike but the trade-off is an efficient non-clinical drug development process which is based on science.



Courage is what it takes to stand up and speak; courage is also what it takes to sit down and listen."

-Winston Churchill, 1874-1965

10 NEDERLANDSE SAMENVATTING

Dierproeven worden nodig geacht om de veiligheid en werkzaamheid van nieuwe medicijnen te voorspellen en worden voorgeschreven door de wet. De wetenschappelijke basis van de voorspelbaarheid van dierproeven voor de mens wordt echter in twijfel getrokken. Een groot deel van de onzekerheid rond de relevantie van dierproeven voor de mens komt voort uit drie factoren: 1) de gevoeligheid, 2) de reproduceerbaarheid en 3) de voorspelbaarheid. De aanvaarding van het voorzorgsbeginsel door de regelgevers enerzijds en de noodzaak om aan de eisen te voldoen en vertraging te vermijden door de farmaceutische industrie anderzijds, zonder pogingen om de huidige methodes te verbeteren, zorgt voor een patstelling waarin dierproeven, voorspelbaar of niet, blijven bestaan zonder ruimte te creëren voor innovatie. Er is dus onafhankelijk en objectief onderzoek nodig om de voorspelbaarheid van dierproeven te bestuderen.

In dit proefschrift hebben we de wetenschappelijke basis voor de huidige aanpak en richtlijnen onderzocht en evalueren we de consequenties voor richtlijnen als dierproeven niet informatief blijken te zijn. Een belangrijk element in dit project was dat we gebruik hebben gemaakt van de registratie dossiers. Dit is een belangrijke bron omdat registratie dossiers alle studies bevatten die zijn gedaan om de registratie van een nieuw medicijn te onderbouwen. In **Hoofdstuk 2** hebben we registratie dossiers gebruikt om te onderzoeken of nieuwe bijwerkingen die na markt introductie zijn vastgesteld ook in de dierproeven gedetecteerd hadden kunnen worden. In 63% van de serieuze bijwerkingen kon geen signaal in de dierproeven gevonden worden. In minder dan 20% is er een positief voorspelbaar signaal gevonden in de dierproeven. Deze signalen werden echter alleen in een paar dieren gezien of werden als soort specifiek beschouwd. De incidentie van bijwerkingen die werd gezien in dieren was laag en ontwikkelden zich pas na toediening van hoge doses en/of na lange blootstelling. De studie werd uitgebreid met biofarmaceutica in **Hoofdstuk 3**. Daar hebben we aangetoond dat nieuwe bijwerkingen voor biofarmaceutica een correlerend signaal hadden in 50% van de gevallen. Alle bijwerkingen waren wel gerelateerd met de farmacologie van het biofarmaceutisch product of een immuun respons daarop. Op deze bijwerkingen had dus ook op basis van de te verwachten farmacologische of immuun effecten geanticipeerd kunnen worden. Data uit dierproeven worden geincludeerd in risico management plannen om de veiligheid van medicijnen te monitoren. Maar omdat serieuze bijwerkingen alleen gezien zijn bij hoge dosis, de lage incidentie en soort specifieke respons, hebben dierdata verminderde klinische relevantie Daarom leiden signalen uit dierdata niet noodzakelijkerwijs op verbeterde farmacovigilantie, maar hebben ze alleen waarde in retrospectief onderzoek.

Voor het ontwikkelen van biofarmaceutica worden steeds meer niet-humane primaten gebruikt omdat dit vaak de enige diersoort is die een relevante farmacologische respons hebben op het medicijn. In **Hoofdstuk 4** evalueren we de waarde van het gebruik van niet-humaan primaten (NHP) bij het ontwikkelen van monoklonale antilichamen. Het gebruik van NHP werd voornamelijk verantwoord doordat het de enige diersoort was die een farmacologische respons kon opwekken. Maar NHP zijn ook gebruikt als ze deze kwaliteit niet bezaten. Het gebruik van meerdere diersoorten was een standaard gebruik en was vergelijkbaar met de klassieke kleine molecuul ontwikkeling. Eerdere dialoog met

regulatoire autoriteiten via wetenschappelijk advies kan de regulatoire kennis verbeteren in deze gevallen. Daarnaast volgden de meeste ontwikkeling strategieën de richtlijnen of voerden meer studies uit, waardoor er van een echte case-by-case aanpak geen sprake was. De methodologie van de dierproeven kon ook de waarde van NHP gebruik verminderen. Het gebruik van wild gevangen of oude NHP kon bijvoorbeeld tot oninformatieve studies leiden omdat de leeftijd, genetische achtergrond, blootstelling aan pathogenen en omgeving of sociale factoren onbekend zijn en tot hoge variabiliteit en uitbijters. Lage aantallen NHP in een studie zijn gebruikelijk, en komt door ethische en economische overwegingen. Dit kan de interpreteerbaarheid van de data beperken. Onze studie stelt vast dat vrijwel alle bijwerkingen zeer voorspelbaar zijn omdat ze gestuurd worden door farmacologische effecten van het antilichaam dat kan worden versterkt door de dosis of duur van blootstelling. Niet doel eiwit gestuurde effecten zijn niet waarschijnlijk of verwacht. Daarnaast konden bijwerkingen ook gestuurd worden door immuun respons tegen het antilichaam wat ook de interpreteerarbeid van data uit NHP studies kon beïnvloeden. In **Hoofdstuk 5** hebben we de immunogeniteit van monoklonale antilichamen in NHP bestudeerd. Het humaniseren van een antilichaam, een techniek die de immunogeniteit moet verminderen, leidde in NHP niet tot het verminderen van de immunogeniteit. In 98% van de ontwikkelingsprogramma's is immunogeniteit vastgesteld op een bepaald punt in het programma. Antilichamen waren mild immunogeen in 41% van de gevallen en hoog immunogeen in 26%. De meeste anti-drug antilichamen (ADA) in NHP werden gevormd tegen het Fc-gebied van de antilichamen wat resulteert in verhoogde klaring. Verhoogde klaring, neutralisatie van de farmacologische werking van het antilichaam of beide werd gedetecteerd in 78% van de programma's. Immunogeniteit leidde tot waarschuwingen om de pharmacokinetische en pharmacodynamische data voorzichtig te interpreteren, omdat het aantal dieren waarin blootstelling gewaarborgd kon blijven laag was, behandelde groepen zijn uitgevallen of de dieren verhoogde doses nodig hadden om blootgesteld te blijven nadat anti-drug antilichamen werden gedetecteerd. De resultaten uit **Hoofdstuk 4 en 5** suggereren dus dat de rol van NHP voor toekomstige ontwikkeling van antilichamen beperkt kan worden. Uitgebreide dierproeven zoals ze nu gedaan worden zijn niet informatief voor monoklonale antilichamen. Een enkele lange termijn studie zou even informatief zijn in het demonstreren van de verwachte farmacologie, zolang immuun respons en anti-drug antilichamen niet prevalent zijn.

Ook voor niet innovatieve medicijnen zijn niet-klinische studies vereist. Generieke kopieën van biofarmaceutica (biosimilars) worden routine matig geëvalueerd in dieren om te demonstreren dat de kopie vergelijkbar is met het referentie product. In **Hoofdstuk 6** hebben we de waarde van dierstudies om de vergelijkbaarheid tussen een biosimilar en het referentie product aan te tonen bestudeerd. Dierproeven die ingediend zijn om biosimilar registratie te ondersteunen lieten geen relevante verschillen zien vergeleken met het referentie product. . Dierproeven waren allemaal in staat om pharmacodynamische vergelijkbaarheid van de biosimilar aan te tonen en vertoonden geen onbekende of onverwachte toxiciteit. Alleen in enkele gevallen werden verschillen gezien vergeleken met het referentie product, maar de klinische relevantie van deze verschillen is twijfelachtig.

Ondanks dat vergelijkbaarheid is aangetoond in dierproeven, is het bekend dat verschillende biosimilars kwaliteitverschillen hebben vergeleken met hun referentie product. Onze data suggereren dat dierproeven niet gevoelig genoeg zijn om deze verschillen te detecteren of om ze te vertalen in meetbare eindpinten door de lage aantallen dieren in de dierproef en de inherente variabiliteit van dieren. Maar als significante kwaliteitsverschillen worden waargenomen tussen de biosimilar en het referentie product, is dit meer indicatief van onvergelijkbaarheid die later ook weerspiegeld worden in dier en humane data. Dit onderstreept het gebruik van een stapsgewijze aanpak om vergelijkbaarheid aan te tonen. Daarbij zou volledige karakterisering en vergelijkbaarheid van de biosimilar aangetoond moeten worden in termen van kwaliteit en productie. Dit zou een startpunt moeten zijn van mogelijke vervolg studies.

De volgende generatie biosimilars, waaronder ook monoklonale antilichamen, staan op het punt te worden goedgekeurd door regulatoire autoriteiten. Deze producten zijn zeer soort afhankelijk en dierproeven zijn daarvoor mogelijk moeilijk uit te voeren of niet informatief. In **Hoofdstuk 7** hebben wij aan de hand van een hypothetische ontwikkeling van een biosimilar voor Bevacizumab (Avastin) de uitdagingen en mogelijkheden beschreven voor het ontwikkelen van een biosimilar monoklonaal antilichaam voor gebruik in oncologie. We stellen daarin voor dat het gebruik van *in vitro* proeven niet alleen kwantificeerbaar zijn, maar ook specifieker en gevoeliger kunnen zijn om eventuele verschillen te detecteren, in tegenstelling tot een dierproef. Informatieve dierproeven met Bevacizumab om vergelijkbaarheid aan te tonen kunnen een uitdaging zijn, omdat gebruikelijke *in vivo* pharmacodynamische modellen voor kanker niet goed in staat zijn gevorderde metastatische kanker te modelleren. Dierproeven in een relevante diersoort, de NHP, voegen waarschijnlijk geen of beperkte informatie toe. Daarom zullen klinische studies nodig zijn, ongeacht de uitkomst van dierproeven. Deze studie suggereert dat antikanker medicijnen zoals Bevacizumab regulatoire autoriteiten in staat stellen om de vereisten van een vergelijkbaarheids studie pakket te herevalueren.

Recentelijk is er een concept richtlijn voor het ontwikkelen van biosimilars vrijgegeven voor consultatie. Het concept start nu vanuit een stapsgewijze, op risico evaluatie gebaseerde aanpak, waar de producent de haalbaarheid en relevantie van dierstudies moet evalueren en waar de aandacht ligt op het doen van *in vitro* studies om vergelijkbaarheid aan te tonen, waar dat enigszins mogelijk is. Indien gecombineerd met het verkrijgen van wetenschappelijk advies van regulatoire autoriteiten en het uitvoeren van wetenschappelijk zinvolle studies kan deze nieuwe positie onderschreven worden. Maar als de vergelijkbaarheid van de kwaliteit tussen de biosimilar en van het referentie product vormt daarbij het uitgangspunt vormt, zouden dierproeven volledig kunnen worden overgeslagen, zoals het geval is voor de generieke kleine synthetische moleculen.

In **Hoofdstuk 8** bespreken we de belangrijkste resultaten van dit proefschrift en doen we voorstellen en aanbevelingen om niet-klinisch onderzoek voor het ontwikkelen van medicijnen te verbeteren. Data uit registratie dossiers is een zeer belangrijke bron geweest voor dit onderzoek. Deze bron is over het algemeen vertrouwelijk, maar zou beschikbaar gemaakt moeten worden voor verder onderzoek. De meeste data over dierproeven is echter

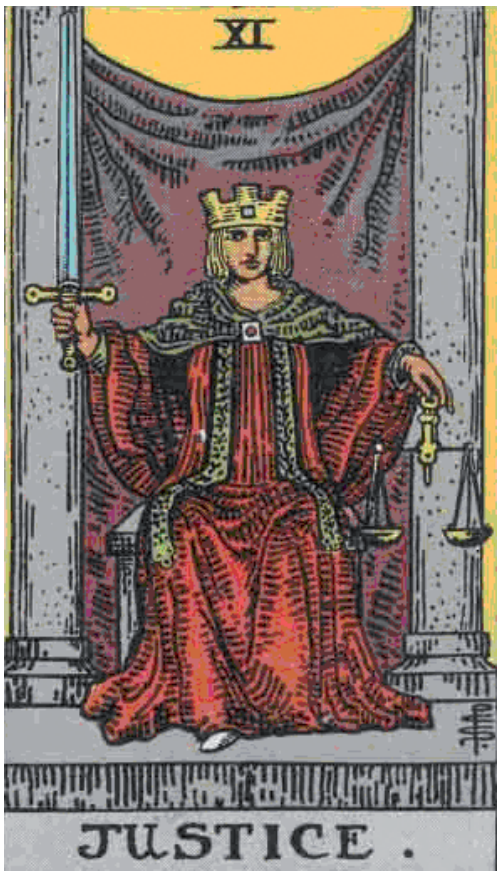
te vinden in private databases van farmaceutische bedrijven. De revisie of het schrappen van richtlijnen die is bereikt door deze bronnen te bestuderen toont hun waarde aan en daarom zouden deze bronnen ook beschikbaar gemaakt moeten worden voor verder onderzoek.

Om op korte termijn niet-klinische dierproeven te verminderen en de kwaliteit ervan te verbeteren zijn een aantal suggesties gedaan. Richtlijnen voor het rapporteren van data uit dierproeven en een registratie plicht van dierproeven is in het verleden voorgesteld. In combinatie met het eerder aanvragen van wetenschappelijk advies kan dit helpen om een wetenschappelijke aanpak van niet-klinische studies te ontwikkelen. Tegelijkertijd, ontmoedigt dit de zogenaamde 'checkbox' aanpak. Een bijkomend voordeel is dat farmaceutische bedrijven minder regulatoire obstakels tegenkomen of irrelevante studies zullen uitvoeren. Daar waar geen relevante dier modellen bestaan, is een langzame en voorzichtige start in de mens de enige wetenschappelijk relevante optie. Deze kunnen uiteraard worden voorgegaan door micro-dosing studies en andere studies zoals –omics, bio-informatica, cel en weefsel kweek experimenten, het gebruik van stam cellen en chip-assays.

Er is toenemend bewijs dat dierproeven hun beperkingen hebben en niet wetenschappelijk, niet informatief of irrelevant kunnen zijn. Dit kan de kosten van het ontwikkelen van medicijnen verhogen en is niet efficiënt. De EU heeft een ban ingesteld op het doen van dierproeven om cosmetica te testen. Zo een ban zou ook kunnen worden ingesteld voor het doen van dierproeven om geneesmiddelen te ontwikkelen. Dat is een drastische aanpak voor complexe producten zoals medicijnen, maar kan uiteindelijk leiden tot een gezamenlijk startpunt voor serieuze discussies met de farmaceutische industrie en regulatoire autoriteiten. Dit zou een gefaseerde transitie naar proefdier vrije geneesmiddelen ontwikkeling kunnen bevorderen.

Een grondige revisie van bestaande richtlijnen over niet-klinische studies is nodig om onnodige dierproeven te identificeren en de efficiëntie van niet-klinische studies in geneesmiddelen ontwikkeling te verbeteren. Eerder en meer intensieve dialoog tussen de farmaceutische industrie en regulatoire autoriteiten door wetenschappelijk advies kan helpen om kennis ven regelgeving te verbeteren en de eisen voor niet-klinische dierproeven mogelijk verminderen. Dat betekent wel dat regulatoire autoriteiten hun afhankelijkheid van het voorzorgsprincipe zullen moeten herzien omdat daar een wetenschappelijke consensus geen vereiste is en het innovatie in de weg staat. Aan de andere kant zal de farmaceutische industrie moeten investeren in nieuwe en alternatieve technologie om informatieve data te genereren over de veiligheid en werkzaamheid van nieuwe geneesmiddelen. Overheden hebben een belangrijke rol om deze activiteiten te stimuleren door bijvoorbeeld politieke druk uit te oefenen, economische impulsen te geven en verder onderzoek te ondersteunen.

Al deze activiteiten vergen enorme inspanningen waarvoor veel vertrouwen nodig is bij zowel farmaceutische industrie en regulatoire autoriteiten. Maar daar staat een efficiënt en goedkoper niet-klinische ontwikkeling fase tegenover dat is gebaseerd op wetenschap.



The fact that we live at the bottom of a deep gravity well, on the surface of a gas covered planet going around a nuclear fireball 90 million miles away and think this to be normal is obviously some indication of how skewed our perspective tends to be.

-Douglas Adams, 1952-2001, *The Salmon of Doubt*

& APPENDIX

GLOSSARY

3R	Refinement, reduction and replacement (3Rs plur.)
ADA	Anti-drug antibody (ADAs plur.)
ADCC	Antibody-dependent cell-mediated cytotoxicity
AMD	Age-related macular degeneration
ARDS	Acute respiratory distress syndrome
ATC	Anatomical Therapeutic Chemical classification system
CAM	Chick embryo chorioallantoic membrane
CBG-MEB	College ter Beoordeling van Geneesmiddelen-Medicines Evaluation Board
CD	Cluster of differentiation
CDC	Complement-dependent cytotoxicity
CDR	Complement determining region
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
COX-2	Cyclooxigenase 2
CTD	Common technical document (CTDs plur.)
DHPC	Direct Healthcare Practitioner Communication (DHPCs plur.)
EC	Endothelial cells (ECs plur.)
ED50	Effective dose at 50%
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
EPAA	European Partnership for Animal Alternatives
EPAR	European Public Assessment Report (EPARs plur.)
EU	European Union
Fab'-fragment	Fragment antigen binding
Fc-region	Fragment crystallizable region
FDA	Food and Drug Agency
FDC act	Federal Drug and Cosmetics Act
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
IFL	irinotecan and 5-fluorouracil
IgG	Immunoglobulin G
IQR	inter-quartile range
MAA	Marketing authorization application (MAAs plur.)
mAb	monoclonal antibody (mAbs plur.)
miRNA	micro RNA (miRNAs plur.)
NHP	Non-human primate (NHPs plur.)
NO	Nitric oxide
OS	Overall survival
PD	Pharmacodynamics
PEG	Polyethyleneglycol

PFS	Progression free survival
PK	Pharmacokinetics
PML	progressive multifocal leukoencephalopathy
PRAC	Pharmacovigilance Risk Assessment Committee
PSA	Product Safety Announcement (PSAs plur.)
RA	Rheumatoid arthritis
RDT	Repeat dose toxicity
RECIST	Response Evaluation Criteria in Solid Tumors
SAR	Serious adverse reaction (SARs plur.)
SD	Standard deviation
SM	Small molecule therapeutic (SMs plur.)
SPC	Summary of Product Characteristics
TBC	Tuberculosis
TNF-a	Tumor necrosis factor alpha
USA	United States of America
VEGF-A	Vascular endothelial growth factor A
VH-framework	Heavy chain variable framework
V-region	Variable region

LIST OF CO-AUTHORS

Vera Brinks
Utrecht Institute of Pharmaceutical Sciences, Department of Pharmaceutics, Utrecht University, The Netherlands

Hans C Ebbers
Utrecht Institute of Pharmaceutical Sciences, Department of Pharmaceutics, Utrecht University, The Netherlands

Christine C. Gispen-de Wied
Cellege ter Beoordeling van Geneesmiddelen-Medicines Evaluation Board, Utrecht, The Netherlands

M. Kooijman
Copernicus Institute of Sustainable Development, Innovation studies, Utrecht University, The Netherlands

Jan Willem van der Laan
College ter Beoordeling van Geneesmiddelen-Medicines Evaluation Board, Utrecht, The Netherlands

Hubert G. M. Leufkens
Utrecht Institute of Pharmaceutical Sciences, Department of pharmaceutics, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands
College ter Beoordeling van Geneesmiddelen-Medicines Evaluation Board, Utrecht, The Netherlands

Aukje K. Mantel-Teeuwisse
Utrecht Institute of Pharmaceutical Sciences, Department of pharmaceutics, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands

Ellen H.M. Moors
Copernicus Institute of Sustainable Development, Innovation studies, Utrecht University, The Netherlands

Huub Schellekens
Utrecht Institute of Pharmaceutical Sciences, Department of pharmaceutics, Utrecht University, The Netherlands
Copernicus Institute of Sustainable Development, Innovation studies, Utrecht University, The Netherlands

Beatriz Silva Lima
iMED-UL, Department of pharmacological sciences, University of Lisbon, Lisbon, Portugal

LIST OF PUBLICATIONS

Publications relevant to this thesis

M. Kooijman, P. J. K. van Meer, C. C. Gispen-de Wied, E. H. M. Moors, M. P. Hekkert, H. Schellekens, The risk-based approach to ATMP development - Generally accepted by regulators but infrequently used by companies. **Regul Toxicol Pharmacol**, (Aug 7, 2013).

P. J. K. van Meer, M. Kooijman, V. Brinks, C. C. Gispen-de Wied, B. Silva-Lima4, E. H. M. Moors, H. Schellekens, Immunogenicity of mAbs in non-human primates during nonclinical safety assessment. **MAbs** 5, (Jun 6, 2013).

H. C. Ebbers, P. J. K. van Meer, E. H. M. Moors, A. K. Mantel-Teeuwisse, H. G. M. Leufkens, H. Schellekens, Measures of biosimilarity in monoclonal antibodies in oncology: the case of bevacizumab. **Drug Discov Today**, (May 18, 2013).

P. J. K. van Meer, M. Kooijman, C. C. Gispen-de Wied, E. H. Moors, H. Schellekens, The ability of animal studies to detect serious post marketing adverse events is limited. **Regul Toxicol Pharmacol** 64, 345 (Dec, 2012).

M. Kooijman, P. J. K. van Meer, E. H. M. Moors, H. Schellekens, Thirty years of preclinical safety evaluation of biopharmaceuticals: Did scientific progress lead to appropriate regulatory guidance? **Expert Opin Drug Saf** 11, 797 (Sep, 2012).

P. J. K. van Meer, M. Kooijman, JW. van der Laan, E. H. M. Moors, H. Schellekens, The value of non-human primates in the development of monoclonal antibodies. Provisionally accepted for publication in **Nature Biotechnology**.

P. J. K. van Meer, H. C. Ebbers, M. Kooijman, C. C. Gispen-de Wied, B. Silva-Lima, E. H. M. Moors, H. Schellekens, Contribution of non-clinical animal studies to evaluate the similarity of biosimilars to reference products. To be submitted

P. J. K. van Meer, M. Kooijman, C. C. Gispen-de Wied, E. H. M. Moors, H. Schellekens, A retrospective evaluation of non-clinical studies predicting new serious post marketing adverse events of biopharmaceuticals. To be submitted

M. Kooijman, P. J. K. van Meer, E. H. M. Moors, H. Schellekens, M. P. Hekkert, How institutional logics hamper innovation: the case of animal testing. To be submitted

M. Kooijman, D. van der Zwam, P. J. K. van Meer, E. H. M. Moors, H. Schellekens, M.P. Hekkert, How institutional logics enable innovation: the case of animal testing. To be submitted

Other publications

J. A. Siegel, T. S. Benice, P. Van Meer, B. S. Park, J. Raber, Acetylcholine receptor and behavioral deficits in mice lacking apolipoprotein E. **Neurobiol Aging** 32, 75 (Jan, 2011).

L. Villasana, C. Poage, P. van Meer, J. Raber, Passive avoidance learning and memory of 56Fe sham-irradiated and irradiated human apoE transgenic mice. **Radiats Biol Radioecol** 48, 167 (Mar-Apr, 2008).

S. F. Acevedo, T. Pfankuch, P. van Meer, J. Raber, Role of histamine in short- and long term effects of methamphetamine on the developing mouse brain. **J Neurochem** 107, 976 (Nov, 2008).

P. van Meer, T. Pfankuch, J. Raber, Reduced histamine levels and H3 receptor antagonist induced histamine release in the amygdala of Apoe-/- mice. **J Neurochem** 103, 124 (Oct, 2007).

P. van Meer, S. Acevedo, J. Raber, Impairments in spatial memory retention of GFAP apoE4 female mice. **Behav Brain Res** 176, 372 (Jan 25, 2007).

G. Bongers, K. M. Krueger, T. R. Miller, J. L. Baranowski, B. R. Estvander, D. G. Witte, M. I. Strakhova, P. van Meer, R. A. Bakker, M. D. Cowart, A. A. Hancock, T. A. Esbenshade, R. Leurs, An 80-amino acid deletion in the third intracellular loop of a naturally occurring human histamine H3 isoform confers pharmacological differences and constitutive activity. **J Pharmacol Exp Ther** 323, 888 (Dec, 2007).

P. Van Meer, J. Raber, Mouse behavioural analysis in systems biology. **Biochem J** 389, 593 (Aug 1, 2005).

DANKWOORD

Mijn eerste werkdag lijkt een eeuwigheid geleden. Toch herinner ik het me als de dag van gisteren. Ik stond rond negen uur te wachten in een slecht verlichte gang in het Went gebouw totdat ik een plekje kreeg aangewezen in de kamer van de 'terminale AIOs'. Mijn promotor en (toen nog) co- promotor waren vertrokken met vakantie, zoals dat gaat als je eerste werkdag op 1 juli is, mijn collega zou pas twee maanden later beginnen en buiten hen wist niemand echt waar ik mee bezig was. Maar, zo leerde ik, het project waar ik aan begon was toch echt wel het makkelijkste project ooit. Met het afronden van dit proefschrift gebleken dat het makkelijkste onderzoeksproject bij het departement Biofarmacie een razend snelle vier jaar duurde. De meeste andere onderzoeken duren trouwens ook ongeveer vier jaar. Dus kan de enige conclusie natuurlijk zijn dat alle onderzoeken bij dit departement eigenlijk makkelijk zijn. Of moeilijk. Daarom had ik het niet kunnen doen zonder de hulp en steun van anderen. Soms inhoudelijk, soms om me daar juist van te weerhouden.

Om te beginnen mijn promotoren, Huub Schellekens, Ellen Moors en Beatriz Silva Lima.

Huub, dank je wel voor de totale vrijheid die je mij hebt gegeven om mijn werk uit te voeren. Dat was soms best overweldigend, maar uiteindelijk zou elke andere manier voor mij waarschijnlijk aanvoelen als onderzoek doen in een dwangbuis. Het was prettig dat ik altijd even bij je kon binnenlopen met een vraag. In een paar woorden, of korter na aanschaf van een iPhone als je in het buitenland was, wist je altijd door te dringen tot de kern van het plotseling verdwijnende probleem dat ineens eigenlijk heel eenvoudig op te lossen was.

Ellen, als ik het even niet zag zitten wist je me altijd te motiveren. De werkbesprekingen, die na de noodzakelijke onderwerpen al snel veranderden in een prettige discussie over het wel en wee in de farmaceutische en academische wereld waren daarbij een constante. Jouw opgewekte stemming tijdens die besprekingen werkte besmettelijk, en mijn dank is groot dat je er ook nog eens voor waakte mijn manuscripten voor iedereen leesbaar te houden.

Dear Beatriz, your first involvement in this project was as an advisor, which grew into you being a promotor of my PhD candidature. There, your input, both on content and suggestions to improve the manuscripts, was invaluable. Not to forget your suggestion to enroll in the inspiring SafeSciMet course!

Mijn co-promotor en dagelijkse begeleider bij het CBG, Christine Gispen-de Wied. Christine, met veel plezier heb ik met je samengewerkt en vind het een voorrecht dat je mijn co-promotor wilde zijn. Je inzet voor dit project, om het in de kijker te brengen bij het grotere en regulatoire publiek en betrokkenheid om dit project een succes te maken zijn bewonderenswaardig. Onze gezellige 1-op-1 meetings waren voor mij steunend en belangrijk. Daarvoor ben ik je oprecht dankbaar. Dank je wel voor het openstellen van het CBG om daar mijn dagen achter de microfilm te kunnen slijten!

Wim, dank je dat ik mijn promotie onderzoek mocht uitvoeren binnen jouw afdeling Biofarmacie. Ik heb er een fantastische tijd gehad. Tot het volgende poker toernooi!

Ik wil nogmaals Amgen, het Nederlands Bijwerkingenfonds, het College ter Beoordeling van Geneesmiddelen, Nefarma, Stichting Proefdiervrij en het Utrecht Institute for Pharmaceutical Sciences bedanken voor hun steun bij het tot stand komen van dit boekje.

In het bijzonder wil ik mijn dank uitspreken aan een tal van collega's vrienden en familie:

Jan Willem, ik heb onze gesprekken over het nut van proefdieren erg gewaardeerd. Het sparren over en weer heeft veel nieuwe inzichten opgeleverd die dit proefschrift beter hebben gemaakt. Ook wil ik je bedanken dat je het mogelijk maakte om, in het hol van de leeuw te Münster, mijn verhaal te kunnen vertellen. Marlous, samen zijn we begonnen aan dit unieke project. Ik, toen nog in het Went bij Biofarmacie, jij in het Unnik bij Geo. Ik heb met veel plezier met je samengewerkt om van dit project een succes te maken. Dank je wel voor de stimulerende discussies, collegialiteit en de gezellige werkreizen naar Londen, Linz, Lissabon en Montreal! Hans, enorm bedankt voor de prettige samenwerking die in record tempo tot een mooie publicatie heeft geleid. De volgende komt eraan. Echt! Francisco; the one person that the entire university seems to know. I truly think that is a good thing and have enjoyed our many discussions on all topics not-science. I wish you the best of luck with your new job and am sure (and hope) you will be as much as a character there as you were here. Barbara, bedankt voor je luisterende oor en steun in de afgelopen vier jaar. Je bent terecht de spil van biofarmacie zonder wie deze machine waarschijnlijk compleet zou vastlopen. Dank je wel dat ik bij je terecht kon met allerlei gefrutslel (hoe werkt de scanner? 4x) en geregel (welke formulieren moet ik nou invullen? 8x) etc. Wanneer gaan we weer kanoën? Scan ik bij de super wel wat spulletjes voor de picknick.

Gerard, het was een genoegen om je als lid van de scientific board te hebben. Constructief en kritisch heeft je commentaar me altijd gemotiveerd. Zonder jouw idee voor de workshop was die ook nooit van de grond gekomen. Peter, hartelijk dank ook voor jouw zitting in de scientific board en de steun van Nefarma om dit project mogelijk te maken. Daan, als een van de bedenkers van dit project heb ik je interesse en betrokkenheid altijd erg gewaardeerd. Dank daarvoor.

My room mates Anastasia, Eduardo, Hans, Maryam, Meriem, and Vera; thanks for the pleasant atmosphere in the office and the off-beat discussions! Hans, bedankt voor de nodige en gezellige afleiding bij het doorspitten van meter na meter na meter microfilm! Marloes, Rob, Lonneke, Eva, bedankt voor jullie enthousiasme en inzet om al die data te verzamelen. Het was een plezier jullie te begeleiden. Marloes, ik hoop dat je nog lang met plezier in de Rijn kan dobberen! For all the fun lab activities, drinks, Christmas dinners and camaraderie, I'd like to thank all my colleagues at the department of Biopharmacy.

Robert, het was altijd prettig om oude en vooral lopende zaken met je door te nemen en te ontstressen. Eerst nog in de betere restaurants van Utrecht en Amsterdam, later via het minder culinaire en bourgondische Skype. Hopelijk kunnen we snel weer aan tafel in de toekomst. Jasper en Gaby, voor een beschaafd klein biertje/wijntje waren jullie altijd te porren. Bedankt voor de bindende muziek adviezen! Maar ondanks dat denk ik toch dat Jasper de Cylon is. Gaby en Jasper, ik ben erg dankbaar voor de fantastische lunch die jullie hebben georganiseerd, zonder jullie hulp was de start van het symposium niet zo een succes geworden. Ted, part of this Great Adventure started in Portland. Thank you for being there, by Como. I wish you and Jess all the best!

Papa, er is niets wat ik hier in een paar zinnen zou kunnen opschrijven om uit te leggen hoe dankbaar ik ben voor alles wat je voor ons hebt gedaan. Ik ben erg gelukkig dat

je in ons leven bent. Mama, dank je voor je geduld. Beiden ben ik jullie dankbaar voor alle onvoorwaardelijke steun en liefde die ik al deze jaren heb gekregen. Het heeft mij gemaakt tot de persoon die ik nu ben. Het is mij daarom een genoegen dat ik jullie straks dit proefschrift mag overhandigen. Philip en Stefanie, verbouwen is ook niet alles, maar toch leuk dat ik me kon uitleven op muren, daken en plafonds met daarna een heerlijke, gegarandeerde barbecue, als het ware. Bedankt voor de gezellige weekends!

Negar, je was er elke stap op dit hobbelige pad dat promoveren heet. Ik kan me niet voorstellen hoe ik het zonder jou had kunnen doen. Dank je wel dat je er voor me was in de tijden waarin het moeilijk was, en het plezier dat we hebben gehad in de tijden waarin het goed ging. Onze vakantie in Australië was onvergetelijk. Ik kijk uit naar onze toekomst!

ABOUT THE AUTHOR

Peter van Meer studied at the Avans polytechnic college in Etten-Leur where he obtained a bachelor's degree in organic chemistry in 2003. He also holds a master's degree in molecular pharmacology from the Vrije Universiteit in Amsterdam. He has worked as a research technician in the department of behavioral neuroscience at the Oregon Health and Science University in Portland, Oregon, USA. There, he worked with ApoE4 transgenic and knockout mice as models of Alzheimer disease to study memory and learning, and anxiety. The complexity and difficulties in working with animals triggered him to start a PhD track at the University of Utrecht in 2009 where he studied the predictive value of non-clinical animal studies in drug development. Currently, he is employed by the University of Utrecht as a post-doctoral fellow. He lives in 's Hertogenbosch.