Regulatory Toxicology and Pharmacology 66 (2013) 88–103

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Pharmaceutical toxicology: Designing studies to reduce animal use, while maximizing human translation

Kathryn L. Chapman [a,*], Henry Holzgrefe [b], Lauren E. Black [c], Marilyn Brown [c], Gary Chellman [b], Christine Copeman [e], Jessica Couch [f], Stuart Creton [a], Sean Gehen [g], Alan Hoberman [d], Lewis B. Kinter [h], Stephen Madden [i], Charles Mattis [j], Hugh A. Stemple [k], Stephen Wilson [b]

[a] UK National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), Gibbs Building, 215 Euston Road, London, NW1 2BE, UK
[b] Charles River Laboratories, 6995 Longley Lane, Reno, NV 89511, USA
[c] Charles River Laboratories, P.O. Box 69, East Thetford, VT 05043, USA
[d] Charles River Laboratories, 905 Sheehy Dr., Bldg A, Horsham, PA 19044, USA
[e] Charles River Laboratories, 22022 Transcanadienne, Senneville, QC, Canada
[f] Genentech, 1 DNA Way MS59 South San Francisco, CA 94080, USA
[g] Dow AgroSciences LLC 9330 Zionsville Road, Indianapolis, IN 46268, USA
[h] AstraZeneca Pharmaceuticals LP, 1800 Concord Pike, P.O. Box 15437, Rollins 8th Floor, Wilmington, DE 19850-5437, USA
[i] Charles River Laboratories, Edinburgh EH33 2NE, UK
[j] AbbVie, Inc., 1 N Waukegan Road, Dept. R468, Bldg. AP13A, North Chicago, IL 60064-6103, USA
[k] Alberta Innovates - Technology Futures, Bag 4000, Hwy 16A & 75 Street, Vegreville, AB, Canada T9C 1T4

ARTICLE INFO

Article history:
Received 31 January 2013
Available online 21 March 2013

Keywords:
Safety
Toxicology
3Rs
In vitro methods
Safety pharmacology
Rodents
Biologicals
New chemical entities
Developmental
Reproductive

ABSTRACT

Evaluation of the safety of new chemicals and pharmaceuticals requires the combination of information from various sources (e.g. *in vitro*, *in silico* and *in vivo*) to provide an assessment of risk to human health and the environment. The authors have identified opportunities to maximize the predictivity of this information to humans while reducing animal use in four key areas; (i) accelerating the uptake of *in vitro* methods; (ii) incorporating the latest science into safety pharmacology assessments; (iii) optimizing rodent study design in biological development and (iv) consolidating approaches in developmental and reproductive toxicology. Through providing a forum for open discussion of novel proposals, reviewing current research and obtaining expert opinion in each of the four areas, the authors have developed recommendations on good practice and future strategy.

1. Introduction

Traditionally, evaluation of the safety of new chemicals and pharmaceuticals requires regulatory studies in animals to protect human health and the environment. Given their importance, the utility of animal models for prediction of human safety should be regularly reviewed as advances in both scientific understanding and technical methods evolve. This practice is essential to ensuring

Abbreviations: 3Rs, replacement, refinement and reduction of animals in research; ACSA, Agricultural Chemicals Safety Assessment; ADA, anti-drug antibody; CNS, central nervous system; DBS, dried blood spot; DART, developmental and reproductive toxicity; DRF, dose range finding; ECG, electrocardiograph; EFD, embryofetal development; EPA, environmental protection agency; ePPND, enhanced peri-postnatal toxicity study; GLP, Good Laboratory Practice; hERG, human Ether-a-Go-go Related Gene; ICATM, International Cooperation on Alternative Test Methods; ICH, International Conference for Harmonisation; ILSI-HESI, International Life Sciences Institute Health and Environmental Sciences Institute; IND, investigational new drug; JET, jacketed external telemetry; LLNA, Local Lymph Node Assay; mAb, monoclonal antibody; NHP, non-human primate; NRC, National Research Council; OECD, Organisation for Economic Co-operation and Development; PD, pharmacodynamic; PK, pharmacokinetic; PPND, peri-postnatal toxicity study; TK, toxicokinetic.

* Corresponding author. Fax: +44 (0) 7611 2260.
E-mail addresses: Kathryn.Chapman@nc3rs.org.uk (K.L. Chapman), Henry.Holzgrefe@crl.com (H. Holzgrefe), Lauren.Black@crl.com (L.E. Black), Marilyn.Brown@crl.com (M. Brown), Gary.Chellman@crl.com (G. Chellman), Christine.Copeman@crl.com (C. Copeman), couch.jessica@gene.com (J. Couch), SCGehen@dow.com (S. Gehen), Alan.Hoberman@crl.com (A. Hoberman), Lewis.Kinter@astrazeneca.com (L.B. Kinter), Stephen.Madden@crl.com (S. Madden), charles.mattis@abbott.com (C. Mattis), Hugh.Semple@albertainnovates.ca (H.A. Stemple).

0273-2300
http://dx.doi.org/10.1016/j.yrtph.2013.03.001

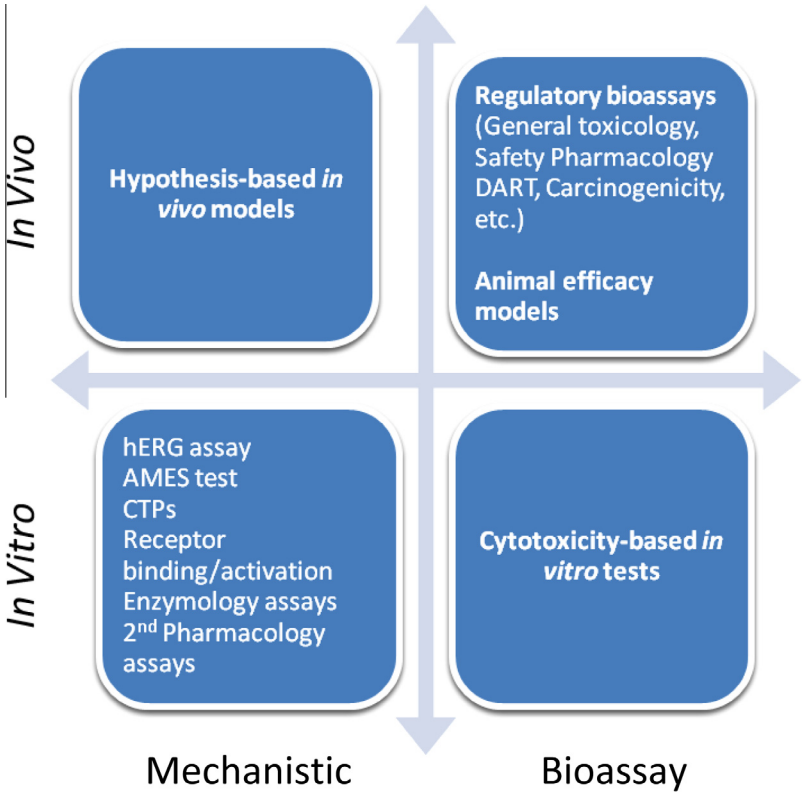

Fig. 1. Examples of *in vitro* and *in vivo* methods. Specific biological mechanisms and holistic bioassays are represented separately. hERG, human Ether-a-Go-go Related Gene; CTPs, phosphocholine cytidylyltransferase; DART, developmental and reproductive toxicity.

appropriate animal use in toxicology studies, with the continued goals of not only improving their predictive value but also reducing overall animal use and enhancing animal welfare.

Generally, risk assessment can be viewed as a process by which information from various sources (e.g. *in vitro*, *in silico*, and *in vivo* studies) is combined to characterize a particular chemical or molecular entity. Ideally, chemical and drug development would be front-loaded with experiments that can definitively select safe compounds as quickly as possible. As data accumulates to support the predictive validity of *in silico* and *in vitro* studies for human safety, these techniques will enable compounds to be deselected earlier in development, thereby limiting the need for animal testing. The replacement, refinement, and reduction of animals in research (the 3Rs) is a well-established concept, originally described in 1959 (Russell and Burch, 1959). Throughout the 1960s and 1970s, the idea that there may be alternatives to animals in research continued to increase in visibility, until finally gaining significant momentum during the 1980s when governments, academia and industry became more involved (Stephens et al., 2001). However, it is only now 50 years since the initial publication that the 3Rs are truly coming of age, with growing recognition of their benefits and widespread efforts to identify new opportunities for implementation.

In order to identify opportunities to further reduce animal use and improve efficiency in drug development, an international workshop was convened to catalyze discussion on various related themes, including: (1) accelerating the progress and uptake of *in vitro* methods, (2) incorporating the latest science into safety pharmacology assessments, (3) optimizing designs for rodent studies to support the development of biologicals, and (4) consolidating various approaches and endpoints in developmental and reproductive toxicology. Representatives from international pharmaceutical companies, contract research organizations, and regulatory agencies also discussed potential concerns around regulatory acceptance when making decisions using novel, rather than traditional approaches. In the 12 months since the workshop, drawing on the expertise of the authors and others present, we have worked towards some practical solutions to common challenges with implementing and improving 3Rs practices in these various areas. Further, expert advice on how new ideas and approaches may be effectively integrated into the constantly evolving model of drug development is discussed. Although this paper is focused on the pharmaceutical industry, participants from the agrochemical industry have also participated, and we have also drawn on their experiences to identify cross-sector parallels.

2. Predicting human toxicology using *in vitro* methods; can we accelerate progress?

There are multiple drivers for the development of new *in vitro* approaches to replace animal bioassay testing including scientific and technological advances, increased focus on animal welfare, and legislative changes. Position papers in Europe and the US (Schumann, 2002; EEC, 1986; Louhimies, 2002), European legislation for the testing of chemicals and cosmetics (EEC, 1976; REACH, 2006) and establishment of validation centers for alternative test methods illustrate the interest in this area from the international community of scientists, regulators and government agencies. Additionally, the European Medical Agency (EMA) recognized the increased use of *in vitro* methods with a recent revision of their concept paper on the replacement of animal studies with *in vitro* tests (EMA, 2012). One purpose of this paper was to more clearly define the process for regulatory acceptance of alternatives, including the need for formal validation studies on some occasions but proof of scientific validity on others.

The intended goal of this section is to provide expert opinion on the smooth integration of appropriate *in vitro* tests into current

Table 1
Examples of *in vitro* methods enhancing drug development decisions.

Subject area	*In vitro* model
Dermal Absorption	Human and animal excised skin
DMPK – metabolite profiling	Human and animal subcellular fractions and hepatocytes
DMPK – drug/drug interactions (DDI)	Human hepatic subcellular fractions, hepatocytes and transfected cell systems
Drug induced liver injury (DILI)	Human and animal subcellular fractions and hepatocytes
Pulmonary toxicity	MucillAir™ EpiAirway™
Renal toxicity	Primary human proximal tubule epithelial cells
Endocrine disruption	Recombinant hES
Cardiovascular toxicity	iCell® cardiomyocytes

processes to reduce attrition and maximize efficiency, leveraging these data to reduce animal use and refine *in vivo* studies.

2.1. Challenge: lack of regulatory acceptance of in vitro methods can limit their usefulness in replacing and reducing animal use

2.1.1. Expert solution: mechanistically-based in vitro methods can be used at all stages of pharmaceutical and chemical development to reduce and inform in vivo studies

There are two general classes of biological methods used in the discovery and development of new products: holistic bioassays and methods that evaluate specific biological mechanisms. *In vitro* and *in vivo* examples of each exist (Fig. 1). The regulatory validation process of *in vitro* methods (bioassay or mechanistic) to replace an *in vivo* regulatory bioassay may deter scientists from developing them. However, not all scientifically validated methodologies require regulatory acceptance. By focusing on the intended use, the scientific validation becomes significantly less daunting. The emergence of novel *in vitro* methods has greatly enhanced decision making in drug development, allowing the selection and progression of molecules with maximum efficacy and minimal toxicity. Some examples of such models in use within industry are shown in Table 1. Further, mechanism-based models show the potential of translation across species (including human), whereas bioassay-based models may have limited translation only within specific contexts. If *in vitro* screens are targeted towards elucidating mechanisms of action or mechanisms of toxicity and are developed based on directed hypotheses, they can be used to provide valuable safety information early in development. For example, considerable progress has recently been made in alternative assays to detect skin sensitization potential. A vision for how mechanistically-based non-animal methods may be more widely implemented in toxicity assessment was presented in a 2007 NRC report (US National Research Council, 2007). The central concept was to leverage new *in vitro* and computational tools to allow scientists to delineate 'toxicity pathways' that could then be extrapolated for use in human risk assessment.

In certain circumstances regulatory acceptance will be necessary to prevent animal studies being carried out in addition to the *in vitro* tests. In these cases, a number of principles must be met and the mechanistic relevance of the *in vitro* endpoint to the *in vivo* effect and the relationship between the *in vivo* outcome and *in vitro* test results must be demonstrated. The uses and limitations of the *in vitro* method must also be clearly defined. For full regulatory establishment and adoption of a new method, eight to nine years is typical:

- 1 year to confirm funding, design protocols and assemble the teams of scientists.
- 3 years to undertake the research.
- 4–5 years for pre-validation, validation and acceptance.

There are opportunities to streamline this process. For example, by sharing cross-company experiences, we can more readily identify the predictive assays that merit further validation and also identify those that should be dropped. Such an experience-sharing initiative may also identify gaps for future research investment. Additional information may be gained by learning how failed compounds behaved in *in vitro* tests, and through identification of the precise mechanisms that underpinned those failures (so they can be targeted and avoided in the future). We should also be encouraged by the success of test methods that have been validated and accepted (Table 2). The International Cooperation on Alternative Test Methods (ICATM) is also aimed towards supporting more rapid acceptance of new methods and may serve as a resource for validation.

2.2. Challenge: it can be difficult to put data from in vitro methods into context

2.2.1. Expert solution: increased use of technologically advanced in vitro methods will generate a rich data and experience base that will increase translational confidence

The perceived lack of confidence in the translational attributes of *in vitro* methods may be due to fear of change, limited historical data, concerns over how predictive they may be of the *in vivo* situation, and lack of clarity regarding whether the method is mechanism or bioassay based. A shift in how *in vitro* methods are viewed is needed to overcome this hurdle. For instance, an *in vitro* test may not completely replicate the equivalent *in vivo* bioassay but it may answer a specific question in a different way. Additionally, 3D cell and tissue models bring us ever closer, physiologically to the *in vivo* situation. *In vitro* bioassay methods attempting to predict *in vivo* bioassays will be susceptible to all of the criticisms and concerns associated with the latter, whereas mechanism-based assays may provide better data upon which decisions may be made.

The incentive to use novel *in vitro* methods is high due to the advantages of being able to manipulate the test system (Fig. 2) which may provide a more thorough understanding of potential mechanisms of toxicity and the human and animal response to exposure to foreign chemicals. As technically validated mechanism-focused *in vitro* methods are used more widely and more data are accrued, scientific confidence will increase. This is perhaps best exemplified by the *in vitro* hERG assay, which has become a key tool in predicting the clinical risk of QT prolongation associated with new chemical entities (Redfern et al., 2003; Moller, 2010; Piccini et al., 2009). Another is the use of mechanism-focused genetic toxicity tests (e.g. Ames Test; mouse lymphoma) to detect potential to cause genetic mutation (by direct interaction with the DNA itself or by indirect mechanisms) in drug discovery, thereby all but eliminating clinical development failures for positive carcinogenicity bioassay results related to genotoxicity.

2.3. Challenge: in vitro methods may be used for compound screening, however they have little impact on the overall number of animals used in regulatory toxicology

2.3.1. Expert solution: predictive toxicology and more accurate compound selection will avoid the use of animals for drugs destined to fail later in development

The high failure rate of candidate drugs after the decision to begin regulatory (Good Laboratory Practice (GLP)) toxicology testing arguably represents the largest opportunity to reduce overall animal use across the pharmaceutical industry. Candidate drug attrition after first in man clinical trials have been carried out is approximately 90% (Kola and Landis, 2004; Khanna, 2012). In other

Table 2
Table of validated and accepted *in vitro* test methods.

Subject area	Relevant guidance documents(s)
Genetic toxicology	OECD 471, 473, 475, 476, 487
Skin absorption	OECD 428
Skin corrosion	OECD 430, 431, 435
Phototoxicity	OECD 432
Skin irritation	OECD 439
Ocular irritation and corrosion	OECD 437, 438
Endocrine disruption	OECD 455, 456
Reproductive toxicology	OECD 414
Safety pharmacology	ICH S7B
Drug metabolism and pharmacokinetics (DMPK)	CPMP/EWP/560/95/Rev.1, FDA draft guidance on drug interaction studies, EMA draft guidance on the investigation of drug interactions

words, nine out of ten promising candidates beginning clinical phase I will not achieve marketing approval. However, much of the regulatory animal testing is front-loaded, and must be completed before clinical trials may proceed. Therefore, when candidates fail in clinical development, the animal testing associated with those candidates to their point of failure could potentially have been avoided. Reducing candidate drug attrition through better predictive *in vitro* assays, can reduce overall animal use in several ways. First, studies previously performed on compounds destined to be dropped are no longer performed because these compounds have been screened out during discovery. Second, as clinical attrition is decreased due to better screening, clinical pipelines are filled with more compounds progressing towards marketing authorizations, reducing the need for more candidates and their associated animal studies. Finally, as overall probability is increased that new candidates will achieve marketing approval, the overall size of both discovery and development pipelines is reduced, delivering further savings in animal use. All this is achievable with improved *in vitro* and *in silico* tests to select and/or drive design of compounds targeting specific mechanisms needed for therapeutic efficacy, whilst avoiding those associated with toxicity.

Currently, although *in vitro* tests are increasingly being used for compound selection, the overall use of animals has remained the same as a relatively constant number of drugs continue to be selected for *in vivo* studies (UK Home Office, 2012). However, with more predictive *in vitro* tests better decisions can be made in selection of candidate drugs, avoiding the redundant or unnecessary use of animals. Use of *in vitro* techniques in compound (de)selection will improve the quality of candidate drugs, decrease toxicological/preclinical attrition, and reduce the number of animals used in non-clinical safety assessment. For example, early detection of genotoxicity *in vitro* could preclude the further development of these new chemical entities, as is the case with European cosmetics (EEC, 1976).

Focusing on the traditional view of replacement as a 1:1 replacement of an *in vivo* test with an *in vitro* test (particularly an *in vitro* bioassay for an *in vivo* bioassay) has arguably prevented opportunities to reduce animal use. In addition, we should be more open to how *in vitro* studies can be used to avoid animal studies and/or improve animal welfare. Such a strategy is recognized in OECD Test Guideline no. 404 (Acute Dermal Irritation/Corrosion), which recommends the conduct of *in vitro* assays (OECD TG 430 and 431) to limit the severity of toxicity for compounds that progress to *in vivo* evaluation. Similarly, OECD TG 437 is designed to identify substances that are ocular corrosives or severely irritating to the eye. Substances that are negative in this test would be required to undergo further *in vivo* testing to allow accurate classification. Finally, *in vitro* studies can be used to improve the predictive value of *in vivo* toxicology studies by supporting the selection of an appropriate animal species, ensuring that testing is not conducted in species that lack human relevance.

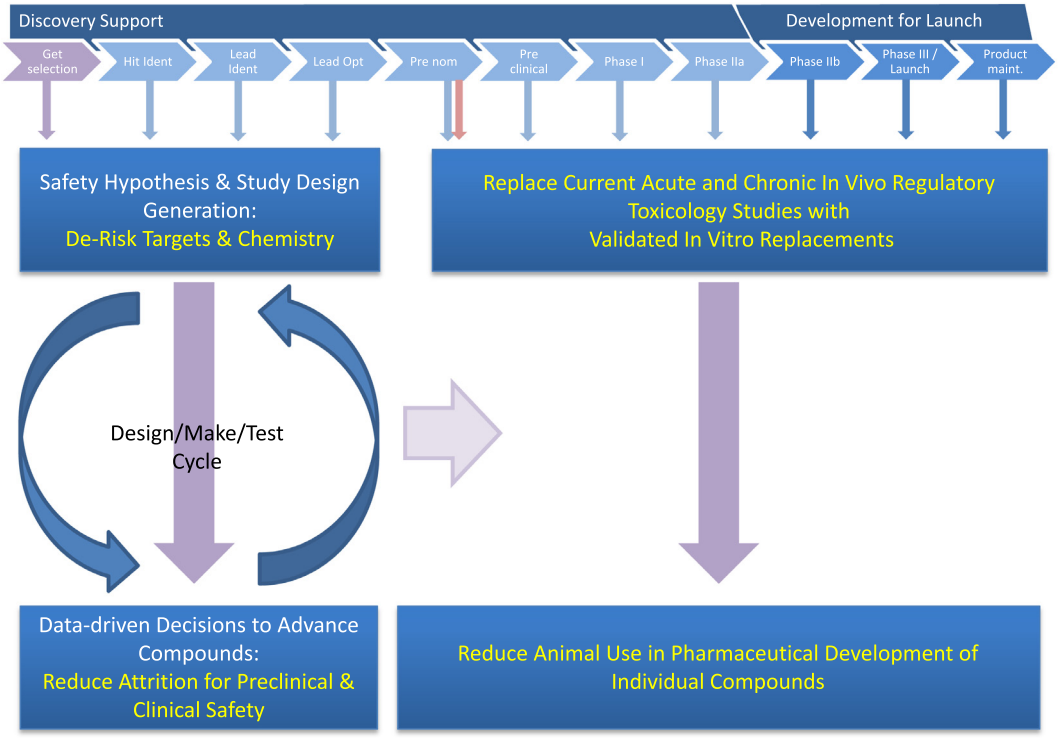

Fig. 2. How *in vitro* tests are used at the different stages of drug development to reduce animal use. Hit identification, lead identification, lead optimization, pre-nomination, product maintenance.

3. Safety pharmacology: incorporating new science into practice

The inclusion of safety pharmacology investigations in regulatory drug safety studies was first described in the International Conference for Harmonization (ICH) M3 and S6 guidance issued in 1997 (ICH, 1997a,b). While these initial documents referred to the importance of directed safety pharmacology studies, specific guidance governing the conduct of these studies was not provided until the release of ICH S7A in 2001 which provided specific recommendations for the conduct of safety pharmacology studies for human pharmaceuticals (ICH, 2001). This was followed in 2005 by the release of the S7B guidance, which addressed the nonclinical investigation of potential delayed cardiac repolarization (QT interval prolongation) by human pharmaceuticals (ICH, 2005a) and the ICH E14 guidance which addressed the clinical evaluation of QT prolongation (ICH, 2005b).

To meet current regulations, safety pharmacology studies for new chemical entities typically assess effects on the respiratory (rodent), cardiovascular (non-rodent) and central nervous system ((CNS); rodent). In order to identify potential undesirable pharmacodynamic properties that may have relevance to human safety, conscious, unrestrained animals are the preferred nonclinical model. Initial studies are typically single dose studies that include, and exceed, the expected therapeutic range of a novel pharmaceutical agent. Often, a tiered approach is used to address relevant concerns across the core battery of regulatory safety investigations. As noted in the ICH guidance, these studies may be performed as either standalone studies or incorporated into toxicity studies performed prior to 'first in man' administration.

This section investigates opportunities to improve efficiency in safety pharmacology studies by ensuring appropriate statistical analysis and state-of-the-art science are efficiently incorporated into practice. We identified some of the key challenges impacting advances in this area and discuss how these may be overcome.

3.1. Challenge: there can be significant delays between scientific and technological advances and changes in company and regulatory practice

3.1.1. Expert solution: increase information sharing between pharmaceutical, contract research organizations and regulators through an industry 'champion'

One area with the potential to reduce animal use is in combining studies, such as the incorporation of safety endpoints into toxicology studies that are already being carried out as part of an investigational new drug (IND) enabling program or clinical trial application (CTA), including dose range finding and 28-day toxicology studies. Animal use may be reduced by replacing standalone safety pharmacology studies with integrated toxicology studies, which is consistent with current ICH guidance (ICH, 2009). When studies are combined, animal use may be reduced by 20–40%. Although this is a common approach during the development of biologics and anti-cancer therapeutics, many companies are reluctant to implement this approach for other types of drugs without evidence that regulators will accept it in practice. This concern is based on the impact of possible confounding factors, such as the influence of pharmacokinetic/toxicokinetic (PK/TK) blood collections from main study animals on functional safety parameters. However, this may be overcome by multiple phase studies that first assess safety pharmacology parameters followed by other procedures (e.g. blood collection) in the same animals.

Advances in scientific knowledge also need to be efficiently incorporated into regulatory practice and the relevant information shared with the regulators to foster broad acceptance. The Safety Pharmacology Society recently released a cardiovascular "Best Practices" document which offers recommendations with the potential to further reduce animal use while improving experimental accuracy (Leishman et al., 2012). As noted in this document, our understanding of the fundamental nature of the QT interval has evolved since the release of the S7B guidelines. The QT interval is now understood to vary not only with heart rate (addressed in S7B), but also to demonstrate profound heart rate-independent variability. This paper concluded that the so-called generic QT rate-corrections such as Bazett (1920) and Fridericia (1920) were not suitable for use in studies involving nonhuman primates, recommending instead the use of individual QT rate-corrections. Importantly, current regulatory guidance does not address this. As such, regulatory safety studies are still being conducted using QT rate-corrections which have been demonstrated to be inadequate. Such studies constitute a poor use of animals and, in the worst case, may fail to accurately detect human risk. These circumstances highlight the absolute requirement for investigators and regulators to stay abreast of contemporary scientific developments that may alter the fundamental understanding of a particular model system. Such developments must then be expeditiously incorporated into the applicable regulatory guidance.

To accelerate progress in this area we suggest that an industry champion assume responsibility for identifying novel methods and approaches that have the potential to be accepted by regulators. The champion would lead a group comprised of a number of companies who are abreast of the latest science in safety pharmacology and would provide a consensus opinion on appropriate practice. This group would also provide a link between industry and regulators to enable greater interaction and data-sharing between the two groups. The champion concept is not unique to safety pharmacology and could be applied to all areas discussed in this paper.

3.2. Challenge: in order to achieve adequate statistical power in experiments, more animals must be used

3.2.1. Expert solution: technological advances combined with improved study designs enable reductions in animal use and better data

Underpowered experiments do not accurately inform us about a given risk or safety margin. When investigators fail to properly incorporate measurement error and statistical power in their study design and interpretation, experiments may need to be repeated. However, an increase in animal use (e.g. group size) is not necessary to increase statistical power in many cases. By employing more sensitive measurement techniques, such as chronic or jacketed telemetry, raw measurement error can be reduced. The corresponding improvement in statistical power reduces the number of observations necessary to establish a given effect. Similar gains have been realized as researchers advance the state-of-the-art for particular endpoints of interest (Glueck et al., 2008). For instance, recent investigations have highlighted that more frequent ECG measurements are necessary for accurate quantification of QT interval (Holzgrefe et al., 2007).

The Safety Pharmacology Society identified study power (prospective and retrospective) as a key concern during its discussions on optimizing the precision, power, design, execution, and data exploitation from safety pharmacology assays. The following example demonstrates the importance of appropriate study design and statistical analysis in a contemporary safety pharmacology study to avoid increasing the number of animals needed to get a statistically significant result. Jacketed external telemetry (JET) allows the direct incorporation of continuous ECG (noninvasive) and blood pressure (minimally invasive) monitoring in repeat-dose toxicology studies (Cavero, 2010). This technology provides the ability to obtain high fidelity continuous cardiovascular measurements in long-term repeat-dose models. However, new technology

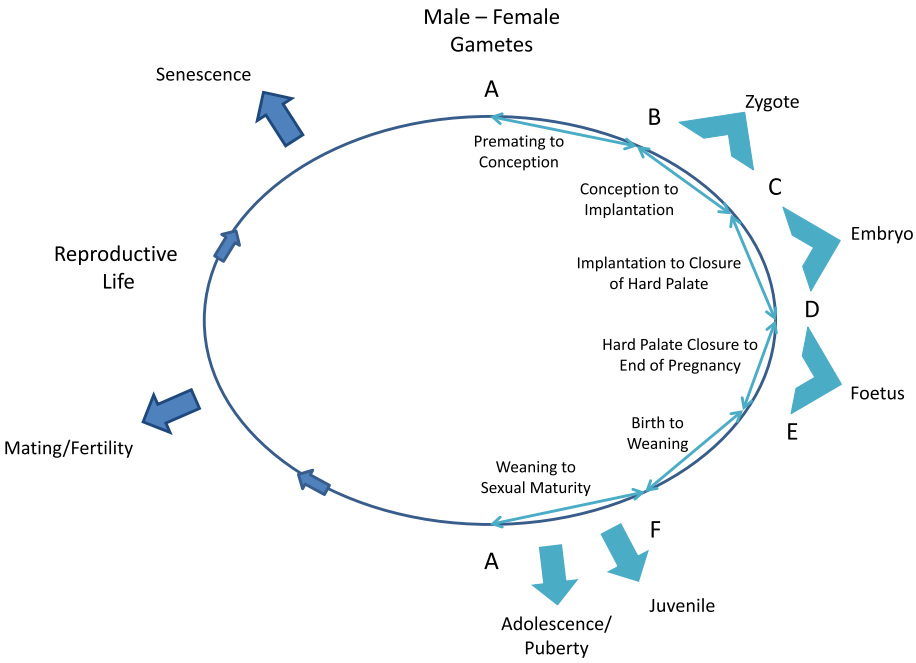

Fig. 3. The stages of the reproductive life cycle needed to assess reproductive and developmental effects of toxicity in animals.

also introduces new variables that need to be taken into account. This was highlighted by the impact of the jackets on heart rate. Each animal exhibited a unique jacket acclimatization pattern, which had to be accommodated in the experimental design and interpretation. Adequate statistical power allows the investigator to correctly eliminate false positive and false negative results, improving not only the quality of the overall safety assessment, but assuring that any unnecessary animal use is eliminated. Peers et al. argue that the systematic incorporation of improved statistical input into preclinical experiments will reduce attrition and improve translation of findings to the clinic (Peers et al., 2012).

3.3. Challenge: safety pharmacology may be perceived as a box checking exercise with a one size fits all approach

3.3.1. Expert solution: good science and therapeutic indication should drive decisions at both program selection and regulatory levels

Safety pharmacology data provide critical information in early development and discovery phases and are often used to make decisions on whether a potential drug should be discontinued from development, or not. However, the assumption that most pharmacological effects on the CNS, respiratory, or cardiovascular systems are accurately detected at the early stages of development can lead to 'box checking' in later studies, just to meet regulatory requirements.

The Animal Model Framework and other groups have been collecting data to analyze the predictivity of cardiovascular, respiratory, and CNS preclinical studies; this information will be used to improve safety pharmacology models and ensure that current animal models add value to the decision making process (Ewart et al., 2012; Valentin et al., 2009).

As safety tolerances differ for chronic and acute life-threatening indications, another approach may be to separate acute safety concerns from long-term risk by therapeutic indication. Current safety pharmacology guidance does not address this distinction. Looking to the future, it may be feasible to broadly redefine safety assessment paradigms by indication. In principle, this strategy could reduce the number of preclinical safety studies needed for a life-threatening indication while retaining the current guidance for drugs designed for chronic use. Coupled with the improved sensitivity made possible with current and emerging technologies, this further demonstrates the importance of 3Rs in human risk assessment.

4. Developmental and reproductive safety testing; consolidating opportunities in the current environment

Guidelines for developmental and reproductive toxicity (DART) studies were first issued by the US in 1966 (US FDA, 1966), accepted in the UK in 1975 (Committee on Safety of Medicines, 1975) and Japan in 1984 (Tanimura et al., 1989). The early guidelines suggested using over 10,000 (adults, juveniles and fetuses) animals per drug to assess reproductive toxicity. One driver for the establishment of the International Conferences on Harmonization (ICH) was to produce guidelines that used fewer animals while still providing an adequate evaluation of the outcome of an exposure to a drug/chemical at any point in the reproductive life cycle (Fig. 3). The ICH S5 guidance (ICH, 1994) on the testing of medicines for reproductive toxicity was the first harmonized guideline for nonclinical testing to be finalized. This harmonization halved the numbers of animals required for evaluating a DART hazard of a new pharmaceutical to approximately 5000 animals.

Since ICH S5 was first published there has been much work to further reduce animal use for DART studies (Barrow, 2009; Stewart, 2009). This section consolidates some of the opportunities to make DART studies more efficient and explores how best to implement them in current development programs.

4.1. Challenge: there may be some circumstances where DART studies do not provide additional value

4.1.1. Expert solution: timely interpretation of all available results can identify cases where DART studies can be avoided without compromising human safety

Traditionally, to assess reproductive toxicity of a new chemical entity, a rodent study for male and female fertility, studies in two species for developmental toxicity (embryo fetal development (EFD)) and a rodent study for the peri-postnatal toxicity study (PPND) are usually required. However, it may be possible to make decisions early in drug development that would prevent initiation of a full EFD study in two species. ICH S5 guidance recommends

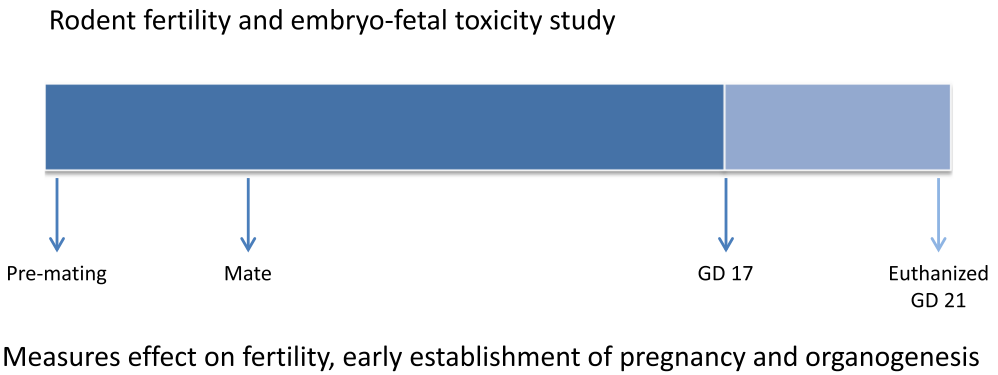

Fig. 4. Figure showing a combined male and female fertility and embryofetal development study. This can reduce animal use by 20% per compound. GD, gestational day.

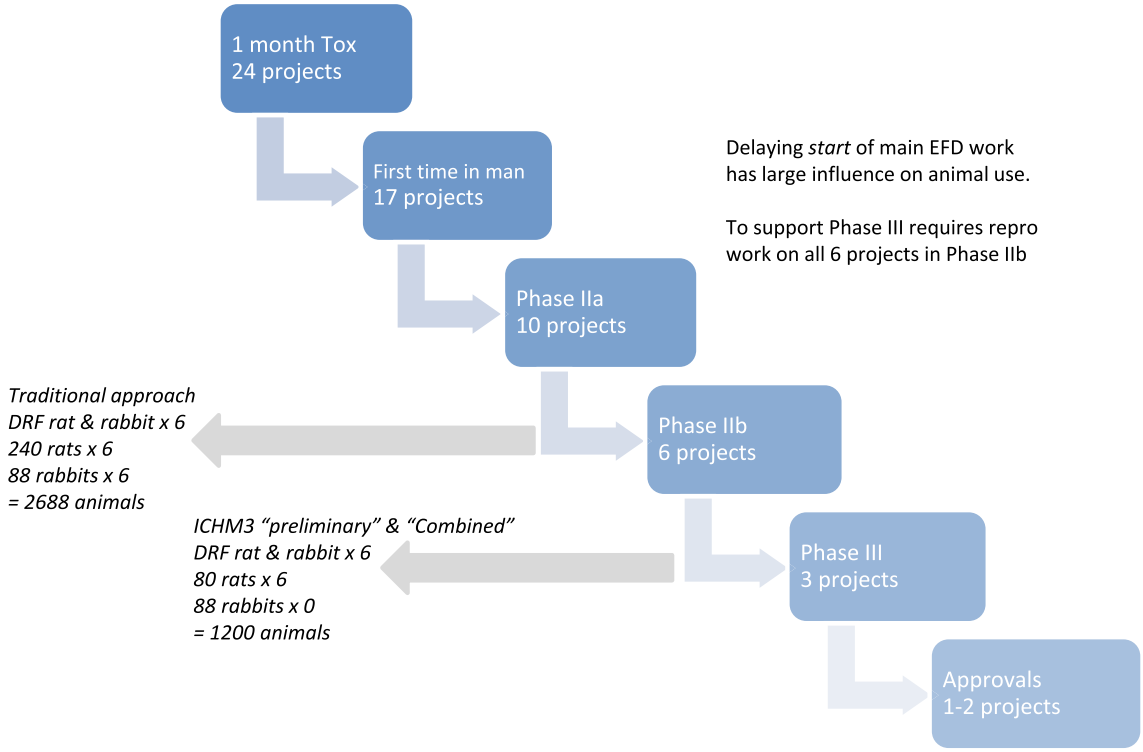

Fig. 5. Timing of studies to reduce animal use in reproductive toxicology studies due to fewer projects in later phases. DRF, dose range finding; EFD, embryofetal development.

that appropriate numbers of animals should be used to produce 20 litters per group in the EFD study. However, dose range finding (DRF) studies in pregnant animals are generally used prior to the full EFD study to ensure appropriate dose selection. If a positive result with clear evidence of developmental toxicity is found in the DRF EFD study then it is unlikely that another study will provide additional information for safety and labeling. This is especially true when TK parameters are collected in the DRF study so that exposure can be measured. In cases where sufficient data are available from general toxicity studies, a DRF study in pregnant animals may be unnecessary. For example, in non-human primates (NHPs) a DRF study is not generally conducted prior to the definitive EFD or PPND study(s). However, there are examples of low toxicity compounds which maybe very toxic to pregnant animals or to the conceptus, which makes this approach risky in terms of reducing animal use.

In vitro methods may also currently be underutilized to determine species relevance. For instance, when *in vitro* metabolism studies are being conducted for a small molecule, inclusion of rabbit microsomes along with the standard species would provide data on the appropriateness of the rabbit for an EFD study. This could lead to de-selection of the rabbit in certain circumstances.

Alternative developmental toxicity assays, including whole embryo culture, embryonic stem cells, zebrafish and several other promising assays are being used to screen compounds and study toxicity mechanisms and pathways (Chapin et al., 2008; Brannen et al., 2010). These assays are unlikely to completely replace animal EFD studies in the near future, but it is clear that the information gained is providing valuable information in predicting human hazard. Enhancements of these assays by the addition of molecular endpoints such as developmental genetic heat maps will lead to further advances.

Another approach to minimize studies on compounds that will not reach late development is to delay the timing of certain studies until phase III or even phase IV (post marketing commitment). An

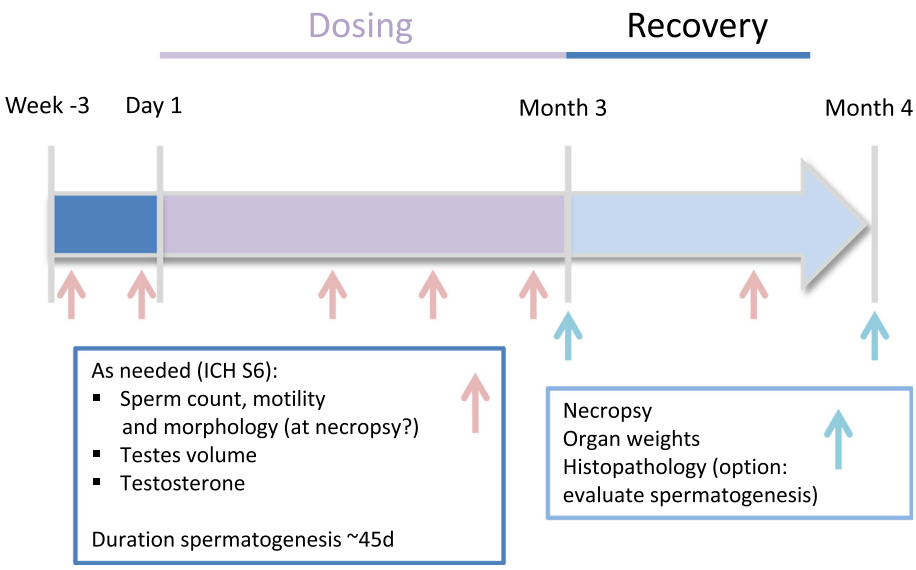

Fig. 6. Assessing effects on male reproduction in non-human primate chronic toxicity studies. The study design shown is relevant for a 3 month study but the approach is also applicable to 6 and 9 month studies. See also Table 3. d, days.

example of this is included in ICH M3 guidance (ICH, 2009) on the timing of non-clinical DART studies in relation to clinical trials. Similarly, conducting the PPND study in phase III or even as a phase IV (post marketing commitment), can further reduce animal usage simply by eliminating the need for a PPND study for a candidate molecule that fails in phase III.

4.2. Challenge: it can be difficult to estimate whether reductions in animal use are real when balancing savings in an individual program for a specific compound with overall animal use for a particular company

4.2.1. Expert solution: consideration of all available approaches within a program, while ensuring that overall animal use does not increase, offers a balanced approach

If companies were to tailor their programs to their own development needs, by focusing on studies that directly impact and guide future research directions, overall reductions in animal use would be maximized. A number of suggestions have been proposed for reducing the number of animals in individual programs, for instance by adding DART measurements to general toxicology studies or combining fertility studies with EFD studies. Combining the male and female fertility study and the EFD study can reduce animal usage for these two studies in rodents by 20% per compound. This combined design (Fig. 4) is appropriate when antigenicity is not an issue (biologicals) and when bioaccumulation of the drug in the test species is not causing unrealistically high exposure levels during the period of major organogenesis. The issue of bioaccumulation is important to ensure the EFD study is serving the purpose for which it is being performed i.e., to advise a pregnant woman exposed to a drug on the risk of that exposure. The combined fertility/EFD design has a longer dosing period and exposures during gestation may be higher than those that would be achieved from exposure only during pregnancy.

Inclusion of male and female fertility endpoints into standard 28 day or longer general toxicity studies could, in theory, eliminate the need for a separate male and female fertility study (Barrow, 2009), but timing (when the fertility evaluations need to be conducted) and power issues (number of animals required per group) make it debatable whether this will lead to an overall reduction in animal use. Potentially, many more animals will be used to test candidate drugs that will not enter phase III testing due to fertility assessment being brought forward in the development path (Fig. 5).

In certain circumstances, such as for biologicals that are not active in any other species and are intended for a patient population which includes women of child bearing age, it may be necessary to use the NHP for DART studies (Chellman et al., 2009). The relevance of the NHP as a test species is determined by tissue cross reactivity and pharmacology studies for large molecules or *in vitro* metabolism for small molecules. It is not necessary, and generally too difficult and costly, to perform mating as part of the assessment for male and female fertility. Instead, the use of sexually mature animals allows reproductive parameters to be included on chronic toxicity studies (typically the 13 week study). Histopathology and organ weights for the male and female reproductive organs from the chronic toxicology study provide an adequate assessment of male and female fertility without the need to add animals to the basic study design. Additional surrogate markers such as semen/sperm analysis, measurement of testicular volume, menstrual cycle evaluations and monitoring of reproductive hormones can be added to these studies when considered appropriate. Incorporating male and female reproductive evaluations into a chronic toxicity study reduces animal usage by 48–56%, compared with running separate stand-alone reproductive studies (Figs. 6 and 7, Table 3).

The combination of the EFD and PPND studies into an enhanced PPND (ePPND) study saves an additional 37.5–50% in NHPs per program (Fig. 8, Table 4) (Stewart, 2009). No novel data are generated by conducting a separate EFD study, since external/visceral/skeletal evaluations are incorporated into the infant evaluations on the ePPND study.

4.3. Challenge: a box checking approach to DART study design does not allow for flexibility or improved efficiency

4.3.1. Expert solution: through applied/proactive decision-making, implementation of new technologies, and differentiation of compounds by their potential to cause toxicity, fewer studies may be needed to appropriately assess risk

There are a number of suggestions for how programs could be conducted differently, rather than adhering to the status quo. These include situations where compounds have less risk, e.g. known low toxicity, use of micro-sampling techniques and/or re-use of animals where appropriate.

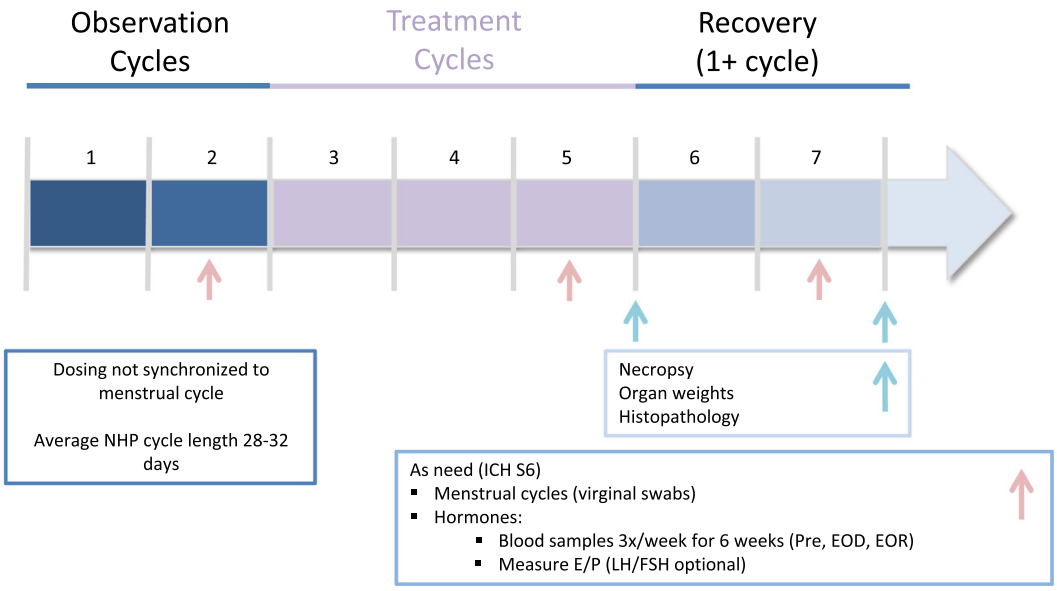

Fig. 7. Assessing effects on female reproduction in non-human primate chronic toxicity studies. The study design shown is relevant for a 3 month study but the approach is also applicable to 6 and 9 month studies. See also Table 3. NHP, non-human primate; LH, luteinizing hormone; FSH, follicle stimulating hormone; EOD, end organ damage.

If a compound is known to have low toxicity or low systemic exposure (little to no absorption), male and female fertility, EFD and PPND studies may be incorporated into a single study as covered in ICH S5. This study design would reduce animal use by about 50% (Tables 5 and 6).

As mentioned previously, DRF EFD studies in pregnant rodents and/or rabbits are used to ensure appropriate dose selection for the full EFD studies. Dosing a small number of animals per group and evaluating the uterine contents ensures that the dose levels selected produce sufficient numbers of offspring for an appropriate assessment of developmental toxicity in the full EFD study. If TK evaluations are included in the DRF-EFD studies, decisions based on blood concentration levels, which improve species extrapolation, will ensure relevant dose selection. For rabbits, conducting these DRF-EFD studies in accordance with GLP regulations can eliminate the need for a full TK profile in the EFD study, reducing the number of rabbits in each EFD study by at least 10%, with even further reductions if satellite animals are not used for blood collection in the DRF-EFD.

The use of micro-sampling and dried blood spot analysis techniques (Jonsson et al., 2012; Spooner, 2010) will enable an increased number of blood samples to be taken without an additional welfare burden on the animal, further limiting the number of animals required per study and compound. In NHP studies, these techniques would also enable hormone analysis if deemed appropriate, e.g., as part of male or female reproduction evaluations.

The reuse of animals may also provide an opportunity for overall reduction in animal use. However, this must be balanced carefully with the overall welfare burden on the individual animal that is being reused. There are a number of situations where reuse may be possible without compromising welfare. For instance, a rat PPND study generates many F1 generation animals that are not further evaluated post weaning. With 20 litters in each group and an average of 12 pups in a litter, there are approximately 200 control pups that could be used for the conduct of a DRF-juvenile toxicity study or a full juvenile toxicity study (Bailey et al., 2009).

The recent ICH S6 (R1) addendum allows for reuse of the vehicle-control treated maternal animals from PPND/ePPND studies using NHPs. Logistically, this procedure must be properly managed to avoid possible issues that may impact the results of subsequent studies. For example, maternal animals would have already delivered at least one offspring, and several factors may differ in second or third pregnancies such as abortion rate, maternal/offspring interaction, and percent infant survival. To avoid these complications, animals must be randomly assigned across dose groups in any future studies. To balance the opportunities to reduce animal use with welfare concerns of being kept on study for long periods of time, reuse would be limited to 2–3 pregnancies and/or a maximum age (e.g., 15 years) and should take into account the ability of individual animals to cope with multiple pregnancies.

5. Use of rodent models for safety assessment of biologics: how can we get the most informative data?

ICH S6 guidance covers species selection in the preclinical testing of biologics and defines a relevant animal species as one in which the test material is pharmacologically active, based on expression of the target receptor/epitope and ability to elicit anticipated biological activity. For biologics, particularly monoclonal antibodies (mAbs), the NHP has often been the only relevant, pharmacologically active animal model for nonclinical safety evaluation. However, the numbers of novel therapeutics with cross-reactivity in both rodents and NHPs is increasing; therefore, rodent use to support IND-enabling toxicity programs has become more frequent. The June 2011 ICH S6 Addendum further outlines conditions wherein a rodent model may be appropriate for safety evaluations of biologics, including:

(1) When test material is pharmacologically active in both a rodent and non-rodent species, both would be used for short term studies (<1 month); but rodent alone could be used for longer term studies if results from shorter term studies in both species are similar or findings are understood from a mechanism of action of the product, unless there is a scientific rationale to use the non-rodent species.
(2) When no pharmacologically relevant species exists, it may be feasible to use an appropriate transgenic rodent model or homologous protein with rodent cross-reactivity, if available.

Considering these recommendations and their potential implications for increased rodent studies for biologics, optimized study designs are needed to enable full evaluation of toxicity while also

Table 3
Incorporation of reproductive evaluation into chronic toxicity study vs. stand-alone studies.

Group	Treatment group	Stand-alone M or F reproduction study		Chronic toxicity study with reproduction endpoints added	
		No. main study (M or F)	No. recovery (M or F)	No. main study (M or F)	No. recovery (M or F)
1	Control	5	3	3/3 or 5/5	2/2
2	Low	5		3/3 or 5/5	2/2
3	Mid	5		3/3 or 5/5	2/2
4	High	5	3	3/3 or 5/5	2/2

Stand-alone studies use 26 animals/sex, studies can be reduced by 52 NHPs through incorporating the reproductive evaluations into a chronic toxicity study (48–56% reduction, depending on group size for chronic study).

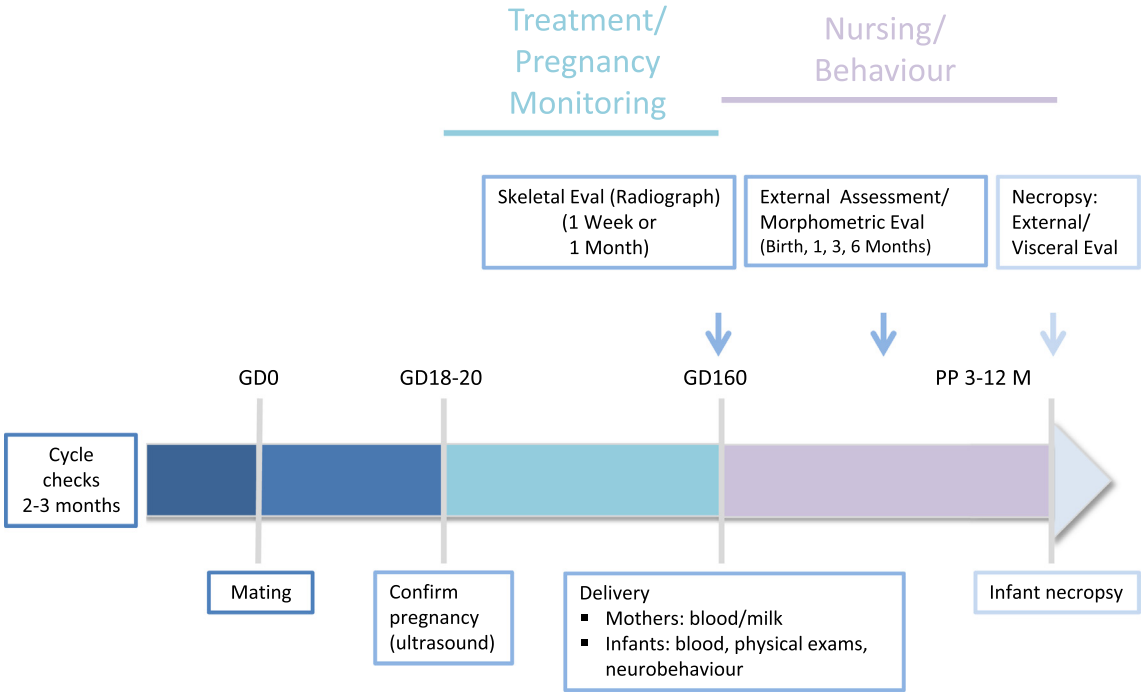

Fig. 8. Enhanced peri-postnatal toxicity study for non-human primates (Stewart, 2009). GD, gestational day; PP, post-partum; M, months; Eval, evaluation.

reducing the overall number of animals (rodents and NHPs) used. This section focuses on the safety assessment of a mAb in rat, with the assumption that the antibody showed similar cross-reactivity across humans, NHPs, and rats. The considerations proposed could also be applied to mice, with some additional species-specific considerations (i.e. difference in available blood volume).

5.1. Challenge: the number of endpoints needed to support regulatory expectations and enable robust safety evaluation of biologics in rodents increases the use of animals

5.1.1. Expert solution: investigations of these endpoints in preliminary studies can inform whether they need to be included in later studies (e.g. IND-enabling)

Standard assessments for biologics require repeated blood collections, clinical pathology, TK and anti-drug antibody (ADA) analyses to confirm systemic exposure of drug and support correlation of exposure levels with identified toxicities. According to ICH S6, both TK and ADA sampling should be incorporated into study designs to assist in the interpretation of results, due to the potential for development of immunogenicity following administration of an exogenous protein (the biologic). Incorporation of pharmacodynamic (PD) parameter(s), when available, may also be important to confirm expected on-target biological activity and assess the predictive capability of the rodent model (for example, cytokine analysis to evaluate potential immunomodulatory effects).

The multiple blood samples planned over the course of a study can, however, result in a large cumulative demand on blood volume that must be balanced with the potential impact on animal welfare and hematological data (e.g., decreased red blood cell mass and hematocrit, compensatory increases in reticulocytes), particularly in rodent species (Jain, 1987; McGuill, 1989; BVA/FRAME/RSPCA/UFAW, 1993; Van Herck et al., 1992; Hawk and Leary, 1995; Podolsky and Lukas, 1998; Diehl et al., 2001; Deng et al., 2011). Historically, this limitation on available blood volumes has been met through the use of satellite dosing groups specifically assigned for TK or PD assessment, leading to substantial increases in total animal use per study depending on the number of endpoints evaluated and blood volume required.

In order to eliminate the need for satellite TK groups in rodent GLP toxicity studies, the frequency of TK sampling and/or TK sample volumes must be reduced. A typical characteristic of the PK profile of biologics is their long half-life. Therefore, traditional sampling schedules employed in safety assessment of small molecules (i.e., several timepoints on Day 1 and repeated at the end of the dosing period) may be unnecessary. PK profiles derived from single dose PK studies can be used to optimize TK sampling and limit the number of collections in subsequent GLP toxicology studies, in which only periodic confirmation of the expected biologic drug exposure levels are needed. Using this approach, single dose PK studies would be utilized to provide full mAb exposure profiles and support human dose projections, rendering extensive TK

Table 4
Animal numbers saved by conducting an enhanced pre-postnatal development study.

Group	Treatment group	No. EFD study (pregnant F)	No. PPND study (pregnant F)	No. ePPND study (pregnant F)
1	Control	12–14	16–18	16–18
2	Low	12–14	16–18	16–18
3	High	12–14	16–18	16–18

By conducting one ePPND study instead of two separate studies (EFD and PPND) NHP use can be reduced by 36–42 animals (37.5–50% reduction).

Table 5
Rodent and rabbit developmental toxicity studies.

Study type	F0 Generation	Embryos/fetuses	F1 Generation	Embryos/fetuses
DR – RAT II	24	264	–	–
RAT I	200	1200	–	–
RAT II	100	1200	–	–
RAT III	100	1200	200	1200
DR – Rabbit II	24	200	–	–
Rabbit II	80	800	–	–
	528	4864	200	1200

Total animal use using this study design is 6792.

Table 6
Rodent and rabbit developmental toxicity studies.

Study type	F0 Generation	Embryos/fetuses	F1 Generation	Embryos/fetuses
DR – RAT II	24	264	–	–
RAT I/II	200	1200	–	–
RAT III	100	1200	200	1200
DR – Rabbit II	24	200	–	–
Rabbit II	80	800	–	–
	428	3664	200	1200

By combining segment I and II into a single study, rodent and rabbit use can be reduced by about 50%.

sampling in toxicology studies unnecessary. Limiting TK sampling demands would in turn eliminate the need for satellite animals, and ultimately lead to substantial reductions in overall animal use. Further, pilot repeat dose toxicity evaluations (e.g. 2–3 repeat doses) conducted prior to a full GLP study can serve to elucidate potential immunogenicity risk, and in some cases may support elimination of the rodent as an appropriate species for toxicity evaluation (e.g. where ADA responses substantially attenuate and/or obviate both drug exposure and PD biomarker activity).

5.2. Challenge: correlation of exposure/activity endpoints (TK and PD) to toxicity assessments

5.2.1. Expert solution: reduce the number of blood collection time points to allow sampling for exposure, functional/translational biomarkers, and toxicity endpoints to be obtained from the same animals

Satellite groups for TK and PD enable collection of these endpoints while staying within blood volume limits for any given animal, and also reduce the potential for confounding factors that may be associated with repeated blood collection on clinical pathology or other safety parameters. However, use of satellite animals can limit data interpretation, as safety and TK/PD/ADA endpoints are characterized in different sets of animals and individual correlations of toxicity/exposure relationships are not possible.

Where feasible, sampling frequencies for TK/PD/ADA endpoints should be reduced to enable integrated assessments from the main study animals and eliminate the need for satellite groups. Although relatively frequent post-dose sampling may be necessary to characterize TK for small molecules with rapid clearance, the inclusion of multiple time points over a short duration generally does not add value for biologics. For biologics with expected PK profiles (ideally characterized in a pilot non-GLP PK study, showing linear clearance and a relatively long half-life), TK sampling could be limited to 3–4 time points following the first and last dose and still enable adequate assessment of exposure, thereby reducing the need for satellite TK groups. Sampling should occur at the same intervals for all animals, including controls, to eliminate potential bias. In specific cases where additional TK and/or PD time points are considered necessary based on preliminary data to achieve study objectives, satellite groups may be used on a limited basis to support sampling needs (Table 7).

Samples for ADA analysis are typically collected after some interval following the first dose to allow adequate time for a positive ADA response to appear. Frequent ADA sampling over the dosing period is generally uninformative, as circulating drug levels are typically high and may interfere with accurate ADA detection. Therefore, a limited ADA sampling schedule is recommended. For example: once pretreatment, once or twice mid-study, at termination of the treatment period and periodic sampling during the recovery period, when circulating drug levels are decreasing.

With satellite groups, composite sampling strategies (where profiles are made up from multiple animals) generate an average TK/PD profile for a given dose level. Although this evolved as standard practice because of the known challenges with blood volume limitations in rodents, it can introduce greater variability in biomarker endpoints by pooling data and preventing direct correlation of exposure-effect relationships within individual animals. If both exposure and PD information could be obtained from the main study animals, the ability to make direct correlations with histopathology and other safety endpoints could increase the power of toxicity and translational biomarker assessments. Although satellite animals may still be required under some circumstances, integration of endpoints into the main study

population could provide a more robust study design to support interpretation of results. Other approaches that may be considered include sampling only from males where no sex difference is anticipated. Due to their larger size, fewer animals can be used for PD/toxicodynamic and other functional assays. Also, it may be more appropriate to move toxicodynamic endpoints to later in the study, to reduce blood requirements during important early sampling periods (such as the first two weeks post dose for TK), particularly since repeated or durable exposure may be necessary to elicit meaningful changes in toxicodynamic biomarkers.

Ultimately, a critical evaluation of the specific needs of each program, based on target biology and known PK/PD profiles, is necessary when designing an IND-enabling toxicity study. The solutions proposed are intended to provide opportunities to reduce unnecessary or uninformative sample collections, enable reduction or elimination of satellite groups, and obtain improved data interpretation from main study animals.

5.3. Challenge: large blood sample volume requirements during first week of dosing due to temporal clustering of early endpoints

5.3.1. Expert solution: increase age of rats at study initiation to allow larger volume of sampling at in the first week, or utilize blood microsampling techniques

One of the biggest challenges to reducing or eventually eliminating TK satellite groups is encountered during the first one to two weeks following first dose administration, where characterization of the initial exposure profile and/or PD response requires repeated blood sampling over a limited period of time. As biological differences between rats of 6, 8, or 12 weeks of age are primarily due to increases in overall size, weight and development of sexual maturity, it may be possible to eliminate satellite groups in rat toxicity studies altogether, or at least permit a reduction in the number of satellite animals required per group, by initiating studies with rats at 12 weeks instead of 6 weeks of age. At 12 weeks of age, male and female rats for most commonly used strains are approximately 225–250 g, allowing collection volumes approximately 1.5- to 2-fold that of 6 week old animals without adversely impacting welfare (Lewis et al., 2002). Therefore collection of repeated samples is possible from the main study animals if the blood volume per sample and/or sampling frequency are limited. Initiation of studies with animals at 12 weeks of age should be acceptable even in the context of chronic toxicity studies for biologics (up to 6 months in duration per ICH S6), without concerns relating to longevity or overall survival for this species. The difference in blood collection volume between 6 and 12 weeks is less evident in mice, but still merits some consideration depending on blood volume requirements defined for the study.

Finally, microsampling and dried blood spot (DBS) procedures are at the forefront of technological improvements that could result in significant reductions in the numbers of animals needed and also involve less invasive blood sampling techniques including tail snip, sublingual vein sampling, and saphenous vein sampling. These have been developed for the mouse but could equally be applied to reduce the use of rats. Both microsampling and DBS have been discussed in detail in the literature (Jonsson et al., 2012; Spooner, 2010). Overall, this is recognized as an area with great potential to reduce the number of animals needed and increase the possibility of blood sampling from main study animals. Additionally, reducing blood volume wherever possible has scientific advantages, as it minimizes the risk of confounding factors due to blood withdrawal impacting hematology results. Future work and investment is needed to validate methods to GLP standards, in the technologies used to detect and analyze mAbs (as opposed to chemicals), and in systems that can accommodate microsampling for hematology and clinical chemistry evaluation. When considering all possible study designs, a 30% reduction in study populations may be achieved while optimizing the value of the information obtained. In addition to minimizing the number of animals used, stress induced by blood sampling techniques should also be taken into consideration. Sampling methods and frequency can be selected to minimize stress responses and any related effect on study outcome, especially if sampling only main study animals (Sparrow et al., 2011).

6. Learning from experiences: a case study in the agrochemical sector

In addition to the pharmaceutical industry, other sectors are also required to conduct toxicology studies to support health protection, and are faced with similar challenges that drive the development of more efficient and less animal-dependent tools for safety testing. Sharing of experiences and practices across sectors is mutually beneficial in the identification of novel ways of improving practice. The workshop and further work presented here included the agrochemical industry perspective to facilitate cross-sector communication.

Animal use for agricultural chemical testing programs has been sizeable due to the comprehensive nature of the regulatory frameworks in place to ensure protection of human health and the environment. Prior to registration, pesticides undergo at least 120 health, safety, and environmental tests, many of which use animals (US EPA, 2007; Croplife America, 2012). In recent years, leaders in the pesticide industry as well as regulatory agencies have recognized and acted upon opportunities to provide the information needed to protect human health and the environment while also applying 3Rs principles. These opportunities have arisen from critically assessing the value of multiple studies to overall risk assessment, and removing those that are redundant or do not add significant value. In addition, it has been recognized that some study objectives can be accomplished in a more efficient manner, for example by combining endpoints into one study that have traditionally been measured in separate studies.

One noteworthy effort to improve efficiency of testing pesticides was the ILSI-HESI Agricultural Chemicals Safety Assessment (ACSA) project (Carmichael et al., 2006; Barton et al., 2006; Doe et al., 2006; Cooper et al., 2006). The objective was to provide a new testing paradigm for crop protection chemicals by developing more relevant studies that use fewer animals. Significantly, one major conclusion generated from the retrospective evaluation of best approaches for assessing systemic toxicity was that the one year dog and mouse carcinogenicity studies did not add significant value, and should no longer be required. Also, it was determined that an extended one generation reproductive toxicity study could be used in place of a multi-generation study and an appropriate test guideline was recently adopted (OECD, 2011). If fully realized, the ACSA proposals have the potential to reduce animal use for pesticide toxicity studies by up to 65% (Carmichael et al., 2006,). While some proposals have not been fully adopted, significant progress has been made with the one year dog and the extended one generation reproductive toxicity studies.

The pesticide industry is taking further steps to proactively adopt other improvements in study design intended to improve data reliability and also reduce and/or refine animal use. For example, the USEPA functional immunotoxicity study can be successfully combined with either 28-day or 90-day repeat-dose toxicity studies, thus eliminating the need for a stand-alone study (Ladics et al., 1995; Ladics et al., 1998; Terry, 2011).

Similarly, US EPA neurotoxicity requirements can also be fulfilled using a combined or integrated testing approach. In fact, both the immunotoxicity and neurotoxicity requirements can be met

Table 7
Optimization of rat study design for large molecules.

	Traditional designs – toxicity populations	Plus TK/ADA and PD satellite populations	Older animals toxicity populations – no change	Optimized approach*
Animals/group main study	10	Control = 3/sex test article groups = 12/sex	10	3/sex for TK/ADA for all groups
Number sexes	2	2	2	2
Number groups	4	4	4	4
Animal/group recovery	5 Groups 1 and 4	–	5 Groups 1 and 4	–
Total animals	100	78	100	24
Overall Total				Savings of 54 animals

* Samples for PD collection from main study animals, TK/ADA population reduced by use of older/larger animals allowing access to greater blood volumes/animal and reducing number of time points based on understanding of molecule kinetics (large molecules).

using a single 90-day study, thus reducing animal use by 55% (Table 8). The integration of TK into repeat-dose toxicity studies provides information on exposure which supports translation of animal data, used in conjunction with human biomonitoring data, to provide a more accurate and complete understanding of risk (Aylward and Hays, 2008). Also, taking steps to establish internal dosimetry can help avoid unrealistic exposure scenarios (i.e. non-linear kinetics) in animal toxicity tests, thus minimizing the potential for pain and suffering while simultaneously obtaining more relevant and reliable data (Creton et al., 2011). Other examples of proactive steps taken by the crop protection industry to apply the 3Rs include: integration of *in vivo* genotoxicity (e.g. micronucleus) into repeat-dose dietary studies, preservation of contingency tissues from repeat-dose studies for follow-up mode-of-action work (Geter et al., 2011), and utilization of dietary route of exposure for studies traditionally conducted by oral gavage (i.e. developmental toxicity) (Rasoulpour et al., 2012) (Table 9).

In addition, significant improvements have been made towards advancement of 3Rs principles and goals in the conduct of acute toxicity studies (e.g. acute oral, acute dermal, acute inhalation, skin and eye irritation, skin sensitization) for agrochemical active ingredients and formulations (Creton et al., 2010; Seidle et al., 2010; Price et al., 2011; Stallard et al., 2012). Such studies are required by regulatory agencies for the purpose of identifying acute hazards which serve to inform classification and labeling of end-use products with the ultimate goal of protecting human health. Specific advances have been accomplished by recognizing opportunities to eliminate unneeded studies through data waivers or by bridging to existing studies on similar materials. In addition, significant advances have been made through adoption of modern study guidelines (e.g. Local Lymph Node Assay (LLNA), fixed/limit dose approaches) designed to minimize animal use and improve animal welfare while providing the information needed for classification and labeling and protection of human health. In the future, many potential opportunities exist for eventual replacement of *in vivo* acute toxicity studies with *in vitro* alternatives.

One significant challenge to more systematic adoption of 3Rs principles and overall reductions in animal use in the pesticide industry is the lack of global harmonization in pesticide data requirements. For example, based on the results of a retrospective analysis, the one year dog study was recently shown to add no significant value beyond data generated from the 90-day study (Dellarco et al., 2010; Kobel et al., 2010). Accordingly, this study was removed as a core data requirement in the US and EU. However, until other regulatory agencies follow suit, registrants continue to conduct the one year study for new pesticide active ingredients intended for global markets. It was also recently shown that the mouse carcinogenicity study does not contribute significantly to the derivation of reference doses or hazard classification, and thus could also be removed as a core data requirement; however, regulatory agencies have not yet adopted proposals to remove the mouse carcinogenicity study as a core data requirement (Billington et al., 2010).

7. Discussion

Scientists face a variety of new challenges in the current economic climate that require them to achieve more with less – more predictive efficacy and safety, with less time, staff and resources. Often, the topic of reducing animal use is perceived as a European regulatory issue; however, scientists are increasingly recognizing the benefits of using the 3Rs as a framework to improve science and reduce costs. The workshop and subsequent discussions in this paper focused on areas with US strengths and interests to make drug development more efficient while minimizing animal use.

Many new technologies and approaches are being developed to advance toxicological science, but there is a time lag before these are routinely integrated into drug development. The field is moving rapidly and the most efficient way to incorporate this knowledge into industry processes is for company experience to be shared, enabling the selection of the most promising methods and rapidly identifying those that have not delivered. This would also take advantage of the current era of open innovation where companies are much more willing to sharing precompetitive knowledge. The National Centre for the 3Rs in the UK has provided 'neutral brokerage' for pre-competitive sharing of 3Rs advances in Europe and has also launched CRACK IT, an open innovation programme in the 3Rs. The recently formed International Consortium for Innovation and Quality in Pharmaceutical Development's (IQ Consortium) 3Rs Leadership Group provides a similar platform for data sharing in the US.

This is particularly pertinent in the area of *in vitro* predictive toxicology where a shift is needed in how *in vitro* methods are perceived. Consensus must be agreed on the criteria for an *in vitro* test to be considered successful, not just in replacing *in vivo* assays, but also in predicting certain mechanisms of toxicity. Once current *in vitro* methods are fully leveraged, we need to ensure that the animal models we are using are adding value to the integrated risk assessment. There are approaches described in this paper that could reduce the number of studies needed for compounds (i) with low toxicity, (ii) that are intended for acute, high risk indications or (iii) that may be dropped later in development. Additionally, there are advances that could be made in study design, for instance using new technologies to achieve adequate statistical power without increasing the number of animals, combining studies, and improving use of TK data to optimize dose selection.

What is needed to accelerate progress in this area? We believe that a three-step approach is needed. First, an industry champion working with relevant company scientists is needed to lead and build the evidence base for changes in practice. This champion would also be tasked with working with relevant trade

Table 8
Animal use in combined 90 day study.

	Traditional 90-day (OECD 408)	Stand-alone immunotoxicity	Stand-alone 90-d neurotoxicity	Integrated approach
Animal/group	10	8	10	10 (5/sex for +cont)
Number sexes	2	1	2	2
Number groups	4	5	4	4
Total animals	80	40	80	90
Overall total			80 + 40 + 80 = 200	Savings of 110 animals

Table 9
Examples of 3Rs advances and innovationsin the agriculture chemical industry.

Example	Driver	Challenges
Removal of one year dog requirement in US and EU	ACSA recommendations; further evaluation by US EPA, industry	Continues to be a requirement in some countries
Proposed removal of mouse carcinogenicity data requirement	ACASA recommendations; further evaluation by industry	Reluctance by agencies to remove requirement
Integration of US EPA – specific requirements (immunotoxicity, neurotoxicity) into 90-day	Industry innovation; desire to combine up to 3 studies into 1	Promotion of concept to other industry members and regulatory agencies
Integrated toxicokinetics	ACSA recommendations; need to establish internal dosimetry	Potentially extra cost associated, but can often prevent future cost
Dietary development toxicity	ACSA recommendations; use relevant exposure route, continuous exposure, reduced vehicle confounding	Overcoming convention (gavage)
Integrated *in vivo* genotoxicity	Relevant route (dietary), reduced animal use, more precise	Overcoming convention (stand-alone gavage)
Preservation of tissues	Frequent need to determine if toxicological MoA is relevant to humans without unnecessary animal use	Requires planning ahead and cannot always predict

associations and professional societies to increase awareness of the importance and implications of specific scientific advances, including advances in our understanding of the relative utility of specific animal models/studies to predict human safety. Second, greater global interaction with regulators on new approaches is essential to advance risk assessment in the chemical and pharmaceutical industries and achieve better harmonization. Third, a global commitment of companies, academic organizations, and regulators is needed to take 3Rs out of the 'competitive arena' and to commit to share all that they know and learn, for the benefit of the animals, as well as human patients and clients.

The authors conclude that there are many unrealized benefits and incentives that may be realized by employing a scientific approach to review and refine animal use. If the suggestions in this paper were implemented widely, there is a realistic near-term potential for significant reductions in animal use. Importantly, progressive approaches to toxicological science will lead to better prediction of human safety with reduced attrition of compounds under development.

Conflict of interest

In accordance with the ABPI Code of Practice regulating the pharmaceutical industry, ABPI and the following companies have provided financial funding to the NC3Rs; AstraZeneca plc, Covance Laboratories Ltd., GlaxoSmithKline plc, Huntingdon Life Sciences Ltd., Eli-Lilly and Company Ltd., Pfizer Ltd. and Novartis Pharmaceuticals Ltd.

Acknowledgments

The authors thank the attendees at the workshop for constructive discussion and Jane Stewart for permission to use Fig. 5.

References

Aylward, L.L., Hays, S.M., 2008. Biomonitoring Equivalents (BE) dossier for 2,4-dichlorophenoxyacetic acid (2,4-D) (CAS No. 94-75-7). Regulatory Toxicology and Pharmacology 51, S37–S48.

Bailey, G.P., Wise, L.D., Buschmann, J., Hurtt, M., Fisher, J.E., 2009. Pre- and postnatal developmental toxicity study design for pharmaceuticals. Birth Defects Research Part B: Developmental and Reproductive Toxicology 86 (6), 437–445.

Barrow, P.C., 2009. Reproductive toxicity testing for pharmaceuticals under ICH. Reproductive Toxicology 28, 172–179.

Barton, H.A., Pastoor, T.P., Baetcke, K., Chambers, J.E., Dinberto, J., Doerrer, N.G., Driver, J.H., Hastings, C.E., Iyengar, S., Krieger, R., Stahl, B., Timchalk, C., 2006. The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments. Critical Reviews in Toxicology 36, 9–35.

Bazett, H.C., 1920. An analysis of the time relations of electrocardiograms. Heart 7, 353–370.

Billington, R., Lewis, R.W., Mehta, J.M., Dewhurst, I., 2010. The mouse carcinogenicity study is no longer a scientifically justifiable core data requirement for the safety assessment of pesticides. Critical Reviews in Toxicology 40, 35–49.

Brannen, K.C., Panzica-Kelly, J.M., Danberry, T.L., Augustine-Rauch, K.A., 2010. Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. Birth Defects Research Part B: Developmental and Reproductive Toxicology 89 (1), 66–77.

BVA/FRAME/RSPCA/UFAW Joint Working Group on Refinement, 1993. Removal of blood from laboratory mammals and birds, first Report of the joint working group. Laboratory Animals 27, 1–22.

Carmichael, N.G., Barton, H.A., Boobis, A.R., Cooper, R.L., Dellarco, V.L., Doerrer, N.G., Fenner-Crisp, P.A., Doe, J.E., Lamb, J.C.T., Pastoor, T.P., 2006. Agricultural chemical safety assessment: a multisector approach to the modernization of human safety requirements. Critical Reviews in Toxicology 36, 1–7.

Cavero, I., 2010. Cardiovascular system assessment best practices: a safety pharmacology society meeting. Expert Opinion on Drug Safety 9 (5), 855–866.

Chapin, R., Augustine-Rauch, K., Beyer, B., Daston, G., Finnell, R., Flynn, T., Hunter, S., Mirkes, P., O'Shea, K.S., Piersma, A., Sandler, D., Vanparys, P., Van Maele-Fabry, G., 2008. State of the art in developmental toxicity screening methods and a way forward: a meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. Birth Defects Research Part B: Developmental and Reproductive Toxicology 83 (4), 446–456.

Chellman, G.J., Bussiere, J.L., Makori, N., Martin, P.L., Ooshima, Y., Weinbauer, G.F., 2009. Developmental and reproductive toxicology studies in nonhuman primates. Birth Defects Research (Part B) 86, 446–462.

Committee on Safety of Medicines; Guidance on Reproduction Studies. 1975. Medical Specialties register, UK.

Cooper, R.L., Lamb, J.C., Barlow, S.M., Bentley, K., Brady, A.M., Doerrer, N.G., Eisenbrandt, D.L., Fenner-Crisp, P.A., Hines, R.N., Irvine, L.F., Kimmel, C.A., Koeter, H., Li, A.A., Makris, S.L., Sheets, L.P., Speijers, G., Whitby, K.E., 2006. A tiered approach to life stages testing for agricultural chemical safety assessment. Critical Reviews in Toxicology 36, 69–98.

Creton, Dewhurst.I.C., Earl, L.K., Gehen, S.C., Guest, R.L., Hotchkiss, J.A., Indans, I., Woolhiser, M.R., Billington, R., 2010. Acute toxicity testing of chemicals opportunities to avoid redundant testing and use alternative approaches. Critical Reviews in Toxicology 40, 50–83.

Creton, S., Saghir, S.A., Bartels, M.J., Billington, R., Bus, J.S., Davies, W., Dent, M.P., Hawksworth, G.M., Parry, S., Travis, K.Z., 2011. Use of toxicokinetics to support

chemical evaluation: Informing high dose selection and study interpretation. Regulatory Toxicology and Pharmacology 62, 241–247.
Croplife America website. <http://www.croplifeamerica.org/crop-protection/pesticide-regulation> [accessed 23rd October 2012].
Dellarco, V.L., Rowland, J., May, B., 2010. A retrospective analysis of toxicity studies in dogs and impact on the chronic reference dose for conventional pesticide chemicals. Critical Reviews in Toxicology 40, 16–23.
Deng, R., Iyer, S., Theil, F.P., Mortensen, D.L., Fielder, P.J., Prabhu, S., 2011. Projecting human pharmacokinetics of therapeutic antibodies from nonclinical data. mAbs 3 (1), 61–66.
Diehl, K., Hull, R., Morton, D., et al., 2001. A good practice guide to the administration if substances and removal of blood, including routes and volumes. Journal of Applied Toxicology 21, 15–23.
Doe, J.E., Boobis, A.R., Blacjer, A., Delarco, V., Doerrer, N.G., Franklin, C., Goodman, J.I., Kronenberg, J.M., Lewis, R., McConnell, E.E., Mercier, T., Moretto, A., Nolan, C., Padilla, S., Phang, W., Solecki, R., Tilbury, L., Van Ravenzwaay, B., Wolf, D.C., 2006. A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. Critical Reviews in Toxicology 36, 37–68.
EEC, 1976. Council Directive 76/768/EEC of 27 July 1976 on the approximation of the laws of the Member States relating to cosmetic products. Official Journal of the European Communities L262, 169–172.
EEC, 1986. Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Official Journal of the European Communities L358, 1–29.
Ewart, L., Gallacher, D.J., Gintant, G., Guillon, J.M., Leishman, D., Levesque, P., McMahon, N., Mylecraine, L., Sanders, M., Suter, W., Wallis, R., Valentin, J.P., 2012. How do the top 12 pharmaceutical companies operate safety pharmacology? Journal of Pharmacology and Toxicology Methods 66 (2), 66–70.
Fridericia, L.S., 1920. Die systolendauer im elektrokardiogramm bei normalen menschen und bei herzkranken. Acta Medica Scandinavica 53, 469–486.
Geter, D., Kan, H.L., Wood, A., et al., 2011. Phenobarbital-like mode of action for liver tumours in CD1 mice and F344 rats exposed to a new developmental compound X11422208. The Toxicologist 120 (Suppl. 2 PS 223), 47.
Glueck, D.H., Muller, K.E., Karimpour-Fard, A., Hunter, L., 2008. Expected power for the false discovery rate with independence. Communications in Statistics: Theory and Methods 37 (12), 1855–1866.
Hawk, C.T., Leary, S.L., 1995. Formulary for Laboratory Animals. Iowa State University Press, Ames, Iowa, pp. 74, 79.
Holzgrefe, H.H., Cavero, I., Gleason, C.R., et al., 2007. Novel probabilistic method for precisely correcting the QT interval for heart rate in telemetered dogs and cynomolgus monkeys. Journal of Pharmacology and Toxicology Methods 55 (2), 159–175.
ICH. S6, 1997a. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S6_R1/Step4/S6_R1_Guideline.pdf>.
ICH. S6, 1997a. Preclinical safety evaluation of biotechnology-derived pharmaceuticals. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S6_R1/Step4/S6_R1_Guideline.pdf>.
ICH. M3, 1997b. Nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2__Guideline.pdf>.
ICH S7A, 2001. Safety pharmacology studies for human pharmaceuticals. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S7A/Step4/S7A_Guideline.pdf>.
ICH. S7B, 2005. The nonclinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals. 2004. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S7B/Step4/S7B_Guideline.pdf> [accessed 6 April 2011].
ICH, E14, 2005. The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for nonantiarrhythmic drugs. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E14/Step4/E14_Guideline.pdf>.
Jain, N.C., 1987. Schalm's Veterinary Hematology, 4th ed. Philadelphia Lea and Febiger.
Jonsson, O., Palma Villar, R., Nilsson, L.B., Norsten-Höög, C., Brogren, J., Eriksson, M., Königsson, K., Samuelsson, A., 2012. Capillary microsampling of 25 μl blood for the determination of toxicokinetic parameters in regulatory studies in animals. Bioanalysis 4 (6), 661–674.
Kobel, W., Fegert, I., Billington, R., Lewis, R., Bentley, K., Bomann, W., Botham, P., Stahl, B., Van Ravenzwaay, B., Spielmann, H., 2010. A 1-year toxicity study in dogs is no longer a scientifically justifiable core data requirement for the safety assessment of pesticides. Critical Reviews in Toxicology 40, 1–15.
Ladics, G.S., Smith, C., Heaps, K., Elliott, G.S., Slone, T.W., Loveless, S.E., 1995. Possible incorporation of an immunotoxicological functional assay for assessing humoral immunity for hazard identification purposes in rats on standard toxicology study. Toxicology 96, 225–238.
Ladics, G.S., Smith, C., Elliott, G.S., Slone, T.W., Loveless, S.E., 1998. Further evaluation of the incorporation of an immunotoxicological functional assay for assessing humoral immunity for hazard identification purposes in rats in a standard toxicology study. Toxicology 126, 137–152.
Leishman, D.J., Beck, T.W., Dybdal, N., et al., 2012. Best practice in the conduct of key nonclinical cardiovascular assessments in drug development: current recommendations from the safety pharmacology society. Journal of Pharmacology and Toxicology Methods 65, 93–101.
Lewis, E.M., Barnett Jr., J.F., Freshwater, L., Hoberman, A.H., Christian, M.S., 2002. Sexual maturation data for CRL Sprague–Dawley rats: criteria and confounding factors. Drug and Chemical Toxicology 25 (4), 437–458.
Louhimies, S., 2002. Directive 86/609/EEC on the protection of animals used for experimental and other scientific purposes. ATLA 30 (Suppl. 2), 217–219.
McGuill, M.W., 1989. Rowan AN (1989) biological effects of blood loss: implications for sampling volumes and techniques. Perspectives on Animal Use 31 (4), 5–18.
Moller, C., 2010. Keeping the rhythm, hERG and beyond in cardiovascular safety pharmacology. Expert Review of Clinical Pharmacology 3, 321–329.
OECD, 2011. OECD Test Guideline 433: Extended One-Generation Reproduction Toxicity Study.
Peers, I.S., Ceuppens, P.R., Harbron, C., 2012. In search of preclinical robustness. Nature Reviews Drug Discovery 11 (10), 733–734.
Piccini, J.P., Whellan, D.J., Berridge, B.R., Finkle, J.K., Pettit, S.D., Stockbridge, N., Valentin, J.P., Vargas, H.M., Krucoff, M.W., 2009. Current challenges in the evaluation of cardiac safety during drug development: translational medicine meets the critical path initiative. American Heart Journal 158, 317–326.
Podolsky, M.L., Lukas, V.S., 1998. The Care and Feeding of an IACUC. Guidelines for Blood Sample Withdrawal, CRC press, pp. 185–186.
Price, C., Stallard, N., Creton, S., Indans, I., Guest, R.L., Griffiths, D., Edwards, P., 2011. A statistical evaluation of the effects of sex differences in assessment of acute inhalation toxicity. Human and Experimental Toxicology 30, 217–238.
Rasoulpour, R.J., Ellis-Hutchings, R.G., Terry, C., Millar, N.S., Zablotny, C.L., Gibb, A., Marshall, V., Collins, T., Carney, E.W., Billington, R., 2012. A novel mode-of-action mediated by the fetal muscle nicotinic acetylcholine receptor resulting in developmental toxicity in rats. Toxicological Science 127 (2), 522–534.
Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Regulation (EC) no 1907/2006 of the European Parliament and of the Council concerning REACH establishing a European Chemicals Agency, amending Directive 1999/45/EC of the European Parliament and of the Council and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) no 1488/94, as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.
Redfern, W.S., Carlsson, L., Davis, A.S., Lynch, W.G., MacKenzie, I., Palethorpe, S., Siegl, P.K., Strang, I., Sullivan, A.T., Wallis, R., Camm, A.J., Hammond, T.G., 2003. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. Cardiovascular Research 58, 32–45.
Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique. Methuen, London.
Schumann, R., 2002. The Seventh amendment to the cosmetics directive: what does DG enterprise want from ECVAM? ATLA 30 (Suppl. 2), 213–214.
Seidle, T., Robinson, S., Holmes, T., Creton, S., Prieto, P., Scheel, J., Chlebus, M., 2010. Cross-sector review of drivers and available 3Rs approaches for acute systemic toxicity testing. Toxicological Sciences 116, 382–396.
Sparrow, S.S., Robinson, S., Bolam, S., Bruce, C., Danks, A., Everett, D., Fulcher, S., Hill, R.E., Palmer, H., Scott, E.W., Chapman, K.L., 2011. Opportunities to minimise animal use in pharmaceutical regulatory general toxicology: a cross-company review. Regulatory Toxicology and Pharmacology 61 (2), 222–229.
Spooner, N., 2010. Dried blood spot sampling for quantitative bioanalysis: time for a revolution? Bioanalysis 2 (11), 1781.
Stallard, N., Price, C., Creton, S., et al., 2012. A new sighting study for the fixed concentration procedure to allow for gender differences. Human Experimental Toxicology 30, 239–249.
Stephens, M.L., Goldberg, A.M., Rowan, A.N., 2001. The First Forty Years of Alternatives Approach: Refining, Reducing, and Replacing the Use of Laboratory Animals. In: Salem, D.J., Rowan, A.N. (Eds.), The State of Animals. Humane Society Press, Washington, DC, pp. 121–135.
Stewart, J., 2009. Developmental toxicity testing of monoclonal antibodies: an enhanced pre- and postnatal study design option. Reproductive Toxicology 28, 220–225.
Tanimura, T., Kameyama, Y., Shiota, K., Tanaka, S., Matsumoto, N., Mizutani, M., 1989. Report on the review of the guidelines for studies of the effect of drugs on reproduction. Notification No. 118, Pharmaceutical Affairs Bureau, Ministry of Health and Welfare, Japan.
Terry, C., Rasoulpour, R.J., Gollapudi, B., Billington, R., 2011. An integrated approach based on the "3" principles of animal welfare and sound science. Toxicology Letters 205S (S244).
U.S. Environmental Protection Agency, 2007. Title 40 Code of Federal Regulations Part 158: Data requirements for registration. U.S. Government Printing Office, Washington, DC.
U.S. Food and Drug Administration (FDA), 1966. Guidelines for Reproduction Studies for Safety Evaluation of Drugs for Human Use. US, FDA.
US National Research Council (NRC) Committee on Toxicity and Assessment of Environmental Agents, 2007. Toxicity testing in the 21st century: a vision and a strategy. Washington, DC. Available at: <http://dels.nas.edu/dels/rpt_briefs/Toxicity_Testing_final.pdf>.

Valentin, J.P., Bialecki, R., Ewart, L., Hammond, T., Leishmann, D., Lindgren, S., Martinez, V., Pollard, C., Redfern, W., Wallis, R., 2009. A framework to assess the translation of safety pharmacology data to humans. Journal of Pharmacology and Toxicology Methods 60 (2), 152–158.

Van Herck, H., Baumans, V., Van Der Craats, N.R., Hesp, A.P.M., Meijer, G.W., Van Tintelen, G., Walvoort, H.C., Beynen, A.C., 1992. Histological changes in the orbital region of rats after orbital puncture. Laboratory Animals 26, 53–58.