

## REVIEW: FRONTIERS IN PHARMACOLOGY

# Principles of Safety Pharmacology

MK Pugsley<sup>1</sup>, S Authier<sup>2</sup> and MJ Curtis<sup>3</sup>

<sup>1</sup>Department of Toxicology & Pathology, Global Preclinical Development, Johnson & Johnson Pharmaceutical Research & Development, Raritan, NJ, USA; <sup>2</sup>LAB Research Inc., 445 Armand Frappier, Laval, Quebec, Canada and <sup>3</sup>Cardiovascular Division, St Thomas' Hospital, King's College London, London, UK

Safety Pharmacology is a rapidly developing discipline that uses the basic principles of pharmacology in a regulatory-driven process to generate data to inform risk/benefit assessment. The aim of Safety Pharmacology is to characterize the pharmacodynamic/pharmacokinetic (PK/PD) relationship of a drug's adverse effects using continuously evolving methodology. Unlike toxicology, Safety Pharmacology includes within its remit a regulatory requirement to predict the risk of rare lethal events. This gives Safety Pharmacology its unique character. The key issues for Safety Pharmacology are detection of an adverse effect liability, projection of the data into safety margin calculation and finally clinical safety monitoring. This article sets out to explain the drivers for Safety Pharmacology so that the wider pharmacology community is better placed to understand the discipline. It concludes with a summary of principles that may help inform future resolution of unmet needs (especially establishing model validation for accurate risk assessment). Subsequent articles in this issue of the journal address specific aspects of Safety Pharmacology to explore the issues of model choice, the burden of proof and to highlight areas of intensive activity (such as testing for drug-induced rare event liability, and the challenge of testing the safety of so-called biologics (antibodies, gene therapy and so on.).

*British Journal of Pharmacology* (2008) **154**, 1382–1399; doi:10.1038/bjp.2008.280; published online 7 July 2008

**Keywords:** drug discovery; Safety Pharmacology; toxicology; hERG; Torsades de Pointes; TRLaD

**Abbreviations:** AE, adverse effect; ADR, adverse drug reaction reports; ADME, absorption distribution metabolism elimination; ADR, annual number of adverse drug reaction reports; CERT, Arizona Center for Education and Research on Therapeutics; CDER, Center for Drug Evaluation and Research; CHMP, Committee for Medicinal Products for Human Use; CRO, Contract Research Organization; CSA, Controlled Substances Act (of the US); CSS, Controlled Substance Staff (of the FDA); DPE, Division of Pharmacovigilance and Epidemiology; DRF, dose range finding; DEA, Drug Enforcement Agency; ECG, electrocardiogram; EMEA, European Medicines Agency; EU, European Union; FIH, first in human (first time a new class of treatment is administered to humans); FDA, Food and Drug Administration of the United States; FOB, functional observational battery of safety tests; GLP, good laboratory practise; HTS, high throughput screening; hERG, human ether-a-go-go; ICH, International Conference on Harmonization;  $I_{Kr}$ , rapid delayed rectifying potassium current; IND, investigational new drug;  $I_{to}$ , transient outward potassium current; JPMA, Japanese Pharmaceutical Manufacturers Association; NCI, National Cancer Institute; NIDA, National Institute on Drug Abuse; NIH, National Institute of Health; NCE, new chemical entity; PD, pharmacodynamic; PK, pharmacokinetic; PMS, Post-Marketing Surveillance; PES, programmed electrical stimulation; RLE, rare (but potentially) lethal event; SRS, Spontaneous Reporting System; Liability, tendency to cause an adverse effect (universal jargon term in Industry); TQTS, Thorough QT Study (jargon term for QT prolongation liability testing in humans); TRLaD, triangulation, reverse use dependence and instability; TdP, Torsades de Pointes; WHO, World Health Organization

### A definition and history of Safety Pharmacology

Safety Pharmacology is the discipline that seeks to predict whether a drug (in the widest sense of the word), if administered to human (or animal) populations, is likely to be found unsafe, and its professional mandate is to prevent such an occurrence. Prior to 1990, pharmaceutical companies

conducted toxicological testing of lead compounds as part of preclinical drug discovery. However, it has become increasingly clear over several decades that drugs may progress as far as phase 3 clinical trials (that is, the intended patient population) before rare and potentially lethal adverse effects become apparent. The vigilant post-marketing surveillance (PMS) efforts by regulatory authorities necessary to confirm the existence of a rare adverse event occur after approval for human use. The Food and Drug Administration of the United States/Center for Drug Evaluation and

Correspondence: Dr MJ Curtis, Cardiovascular Division, The Rayne Institute, St Thomas' Hospital, Lambeth Palace Road, London SE17EH, UK.  
E-mail: michael.curtis@kcl.ac.uk

Received 6 May 2008; revised 9 June 2008; accepted 12 June 2008; published online 7 July 2008

Research uses tools such as drug experience reports, medical literature (clinical trial data) and multiple federal agency data sources (Drug Enforcement Agency (DEA); National Institute of Health (NIH); National Institute on Drug Abuse (NIDA)) in conjunction with the division of pharmacovigilance and epidemiology, which utilizes the spontaneous reporting system (SRS) to monitor adverse drug effect patterns potentially indicative of a public health concern (a potential 'signal'). The SRS receives adverse drug reaction reports derived from health care providers and hospitals. When an adverse effect is very rare, it may require millions of prescriptions before an awareness of its existence emerges. There are numerous examples of this in the literature (for example, Kemp, 1992); one of the best is terfenadine.

In the mid 1990s the antihistamine, terfenadine (Seldane, Marion Merrell Dow), was withdrawn following a growing awareness that the drug could evoke the potentially life threatening cardiac syndrome, torsades de pointes (TdP), in otherwise healthy patients (Monahan *et al.*, 1990; June and Nasr, 1997). Prior to this, the general perception was that only cardiac/cardiovascular compounds were considered to possess such a tendency (liability). The problem here was that terfenadine, a non-cardiovascular drug, had low efficacy to evoke TdP making it so rare an event that it required several million prescriptions before its liability became suspected. The other important consideration here is that the indication for which terfenadine was used (hayfever) is itself far from life threatening. Therefore, risk (death) clearly outweighs benefit (the amelioration of a 'runny nose'; Rosen, 1996).

This episode was of great importance to what we now call Safety Pharmacology (a discipline that did not exist at the time). This is because predicting terfenadine's TdP risk was not possible by the conventional preclinical toxicity testing methods conducted at the time. Preclinical toxicology testing, as an approach involved determining the high-dose adverse event profile of a compound given at chronic, toxic doses, but would not have detected a rare lethal event liability at therapeutic dosage. Indeed, screening for TdP liability risk in animals or in phase 1 and 2 clinical investigations (whether by evaluating QT prolongation or by exploring other putative biomarkers) was not recognized as relevant, let alone necessary, in the late 1980s and early 1990s. Moreover, the magnitude of the effect of terfenadine on QT interval is small, and peak effects may exhibit a delayed onset (Ollerstam *et al.*, 2007) and so the effect is hard to detect even if one is looking. This problem could have been avoided if, instead of routine toxicology, a programme of specific high throughput screening (HTS) for TdP liability had been utilized in early drug discovery at the time, but consideration of biomarkers for rare adverse event liability was not part of the toxicology agenda in the early 1990s.

Even though it was known that repolarization delay in the ventricles of the heart was associated with TdP occurrence, it was not until 1996 that Brown's group identified the likely mechanism of terfenadine's 'cardiotoxic' actions (Roy *et al.*, 1996) and Rosen, by bringing together for consumption in the mainstream literature (Rosen, 1996) all of the threads, helped to inform a growing awareness in industry that toxicology alone, as practised at the time, was insufficient for

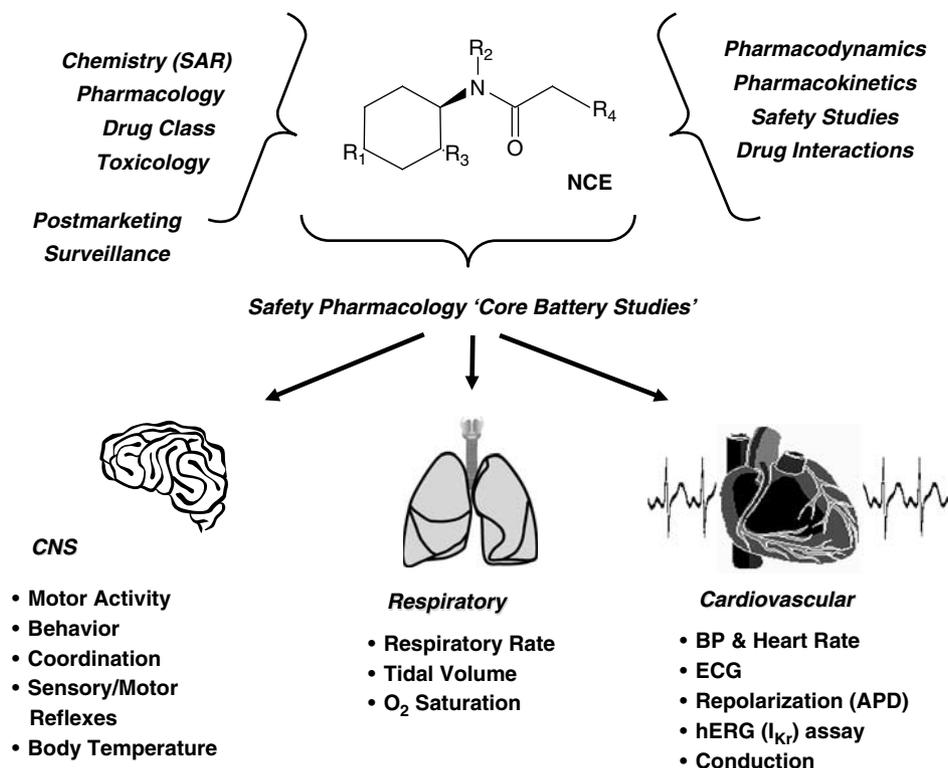
detecting rare but lethal adverse effect liability. In response to this, within 4 years, Safety Pharmacology had evolved into an industry department-based discipline designed to bridge the gap between preclinical toxicology and (preclinical and clinical) drug development (Bass *et al.*, 2004a).

The creation of Safety Pharmacology has not resolved all challenges, especially with respect to detection of rare and lethal adverse effect liability. One of the most difficult problems in Safety Pharmacology is how to conduct early HTS for adverse effect liability with precision and accuracy and in a manner that the data set for a drug deemed 'safe' by the owner of the drug can be presented in a convincing way to regulators. This is a particular problem for rare but potentially lethal adverse drug effects. Furthermore, to interject into this discourse, we noted earlier that Safety Pharmacology (as exists today) is tasked with identifying drugs as unsafe (within the therapeutic window) so, in effect, the data set the company presents to regulators is a failure to disprove that the drug is likely to be unsafe, rather than positive indication of likely safety. The extent of this difficulty becomes clear when we consider that even though more than 10 years has passed since the terfenadine episode, it is still not possible to quantify, with high certainty, a risk/benefit assessment for TdP liability for a drug about to enter phase 1 clinical studies based on preclinical (or even clinical QT test) data sets (Shah, 2008). Thus, we really remain years away from being able to take a drug's range of IC<sub>50</sub> values for different molecular targets (that is, its selectivity profile) and generate a number that reflects its risk (that is, liability to evoke TdP) that can then be balanced against a number that reflects its likely therapeutic benefit. This model applies to all and any rare, but potentially lethal, adverse effect issues.

So, how has this impacted on the unfolding (and evolving) history of Safety Pharmacology? In the absence of quantification of the predictive value of tests and programmes, industry and the regulators have attempted to accommodate one another through a series of industry- and regulatory-led initiatives. Of the latter, the most important is the International Conference on Harmonization (ICH). The ICH is a project started in 1990 that utilizes the regulatory authorities of the United States, Europe and Japan in conjunction with experts from the pharmaceutical industry (from the three regulatory regions) to discuss scientific and technical aspects of therapeutic drug registration (Bass *et al.*, 2004a). What has this to do with pharmacology? The answer is that Safety Pharmacology has been shaped in structure and function by this ongoing accommodation between pharmacologists and regulatory authorities.

## The agenda of Safety Pharmacology

Regulatory authorities (for example, the FDA (US), Health Canada (Canada), European Medicines Agency (EMA) and Japanese Pharmaceutical Manufacturers Association (JPMA)) give approval for drug use in humans. Therefore, convincing the regulators that a drug is safe and efficacious is a key part of the drug discovery/development process. Thus, it is important to consider who the regulators are and what they want to know. The structure of a Safety Pharmacology 'core



**Figure 1** An overview of the multidisciplinary integration required to evaluate the safety profile of a new chemical entity (NCE) in Safety Pharmacology. Consideration is required of the physicochemical and pharmacological nature of the compound, along with toxicological and associated ADME and pharmacokinetic findings. The lower panel of the figure depicts some of the possible non-clinical methods/parameters recommended for assessment in the safety pharmacology core battery of tests by ICH Guidelines S7A and S7B.

battery' programme (Figure 1) is to determine the potential undesirable pharmacodynamic effects of a drug on the central nervous, cardiovascular and respiratory systems, as well as to implement supplementary tests to evaluate other organ systems (Pugsley, 2004; Bass *et al.*, 2004b). Thus it is primarily designed to take account of regulatory requirements; scientific issues are secondary. Follow-up studies may be triggered if there is a need to characterize specific adverse effects found in initial Safety Pharmacology studies. Although follow-up may appear more scientifically driven than the core programme, the design of follow-up studies is nevertheless based on what is perceived by the pharmaceutical company to be the data required by the regulators. This gives a rather special flavour to Safety Pharmacology—it serves the needs of regulatory authorities primarily, and scientific proof is a secondary issue. But pharmacology is a science; this article therefore sets out to interrogate the Safety Pharmacology agenda and explore how far its mores digress from the rubric of science.

The practical agenda of Safety Pharmacology (to determine if a drug is 'unsafe', and, if this is not the case, to inform drug discovery that the drug is likely to be 'safe') is the flip side of drug discovery itself (to determine whether a drug is 'effective'). If these semantics are kept in mind, the process of Safety Pharmacology is exactly the same as that for discovery; for a drug to progress to patients, Safety Pharmacology must conclude that a drug has a sufficiently low potential to evoke adverse effects to be trialed in patients, whereas discovery must conclude that a drug

has a sufficiently high potential for benefit to be trialed in patients. Clearly, therefore, Safety and Discovery Pharmacology are interconnected, as a greater potential benefit may offset a greater potentially adverse effect liability.

Both Safety Pharmacology and discovery rely on pre-clinical (animal) research prior to phase 1 human testing. Thus both are subject to the same issues, namely concerns over whether the animal models will allow accurate and complete detection of 'hits' without false positives or negatives. Thus, both seek to identify and use full-scale clinically relevant end points (for example, detection of disease generation in Safety Pharmacology, and protection against generated disease in discovery). At the same time, for reasons of practicality, both tend to use biomarkers (surrogate end points) such as kinase inhibition as a biomarker for cancer suppression in drug discovery (Garber, 2006), and hERG block for TdP in safety (Sanguinetti and Mitcheson, 2005) to reduce the need for experimental complexity. Expert use of biomarkers is the most challenging aspect of Safety Pharmacology, and is a topic to which we will return on more than one occasion in this article, especially in the context of HTS for rare but serious adverse drug effect liability (for example, TdP liability).

One of the key roles of Safety Pharmacology is to help inform the decision to begin testing in humans. Pharmacology alone does not define the fate of a new drug. The issues determining the point at which it is ethical to proceed with clinical trials informs a risk/benefit assessment that is

weighed against clinical development costs and the potential market. Risk/benefit assessment may appear to be a rather simple process; it is not. Determining the risk/benefit ratio is especially difficult when rare, but potentially lethal events are a concern for a drug, which is intended for use against a non-life-threatening condition. The key point to emphasize again in this regard is that just as preclinical discovery studies never prove that a drug will be effective in patients, preclinical Safety Pharmacology studies never prove that a drug will be safe in patients. Thus the point at which preclinical data is sufficient to inform a decision on whether or not to proceed with a drug into clinical investigation is subjective and a matter of judgement (both for the company and for the regulators who scrutinize the application).

How is the decision to begin human testing made? In the absence of precise guidance from regulators (in some areas the ICH recommendations are vague) the decision-making process concerning when to apply for approval to proceed to phase 1 studies is difficult to understand, especially if a drug is found in preclinical tests to have a possible liability to evoke a serious adverse effect. Ultimately the regulators will decide whether to allow the drug to proceed to humans . . . or not (there is no halfway house). Nevertheless, prior to this, the regulators may put the drug 'on hold' and request more preclinical data. Each sponsor/company will make the initial decision to advance a compound into clinical testing (for example, submit an Investigational New Drug (IND) submission), and this will be based on both efficacy and safety data. Then the regulatory agency would either support an IND application, or place it 'on hold' with a request for more data.

It is important to note that both discovery and safety ought to take account of dosage to inform a likely safety margin for the drug. If animal models can be used reliably to predict the necessary dosage for benefit, and the maximum-tolerated dosage, it may be possible to calculate a projected safety margin. Of course this begins to become a challenge if one attempts to equate *in vitro* data (using drug concentrations) with *in vivo* data (and dosage).

Moreover, if biomarkers are used to substitute for real benefit or real risk in discovery or Safety Pharmacology this may lead to over- or underestimation of projected safety margins. This is a highly problematic area. Unless one proposes that both discovery efficacy and Safety Pharmacology adversity studies be conducted entirely in human volunteers (which effectively means abandoning all scientific research in medicine), a solution is required. There are two contrary possibilities. One is to minimize the use of surrogate biomarkers when estimating dose-response relationships in preclinical discovery and in preclinical Safety Pharmacology. The other is to, in effect, maximize use of surrogate biomarkers. This seems counter intuitive but it is becoming a trend in larger pharmaceutical companies. The rationale is to combine several surrogate biomarkers and conduct an 'integrated assessment' under the assumption that the predictive value of the integrated assessment is sufficiently better than the predictive value of a single surrogate biomarker to warrant this approach. Although this is pragmatic, we would recommend avoidance of surrogate

biomarkers if possible (that is, if the adverse effect itself is available as a readout in a model).

As an aside, it is worthwhile at this point to define the concept of good laboratory practise (GLP). In discovery and Safety Pharmacology there comes a stage when it is necessary to prepare IND documentation for submission for regulatory approval. Regulators take most notice of GLP studies, which use models that are formally validated (in as much as all procedures are defined, monitored and documented according to a recognized procedure, consideration of the details of which is beyond the scope of this article). GLP ensures the generation of verifiable quality data for the drug in development and, as such, defines the framework in which preclinical studies for regulatory submission must be conducted. GLP regulations encompass all components of regulated preclinical studies including the scientists involved (Study Director/Monitor), the test facility, the test system and the test article (test drug). The FDA regulates the conduct of preclinical laboratory studies under Part 58 (good laboratory practise for non-clinical laboratory studies) of Title 21 of the Code of Federal Regulations (US FDA, 2005a). We would add, as a further aside, that GLP validation is not equivalent to scientific validation; GLP is primarily a process of bookkeeping to ensure that agreed procedures have been followed. This does not mean that the agreed procedures constitute a validated method.

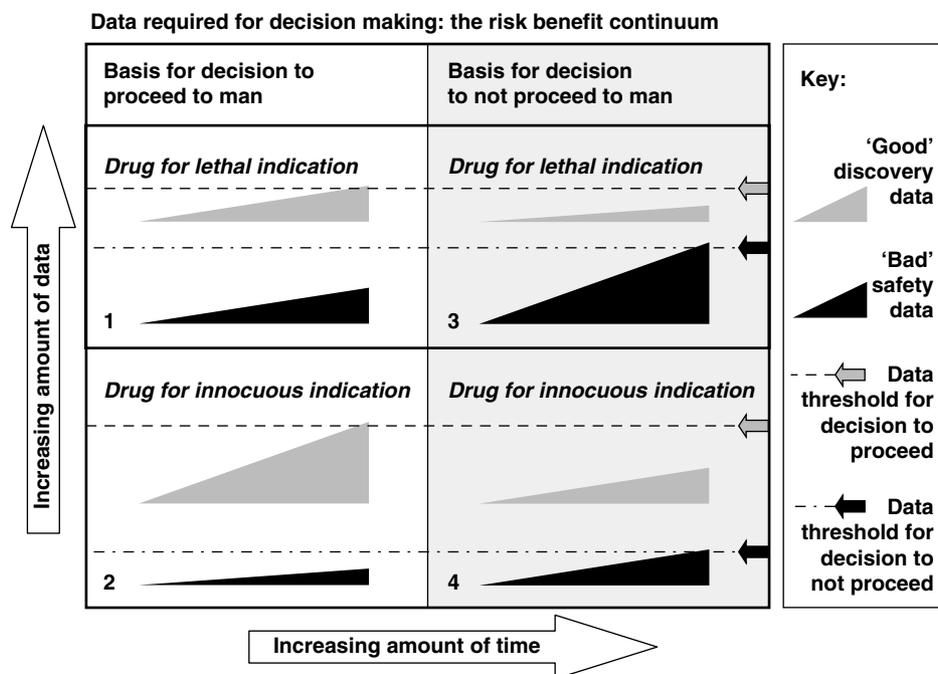
Once human testing has begun, risk/benefit assessment continues, this time in the patient population. This means taking into account the seriousness of the disease as well as the seriousness of any adverse effects. This is primarily relevant to pharmaceutical company choice making about investment (spending). Thus, a very promising cure for a rapidly progressing disease with a poor prognosis, such as pancreatic cancer, will likely be allowed to enter phase 1 clinical trials in pancreatic cancer patients with minimal preclinical Safety Pharmacology testing. In which case, the extent of Safety Pharmacology investment will be minimized. Thus, the oncology division at the FDA may not fully enforce ICH S7A (the regulatory guidance document that provides general principles and recommendations for safety pharmacology studies) depending on the seriousness of the disease and current therapy (or the absence of current therapy) in this population, in which case the company will be able to minimize their spending (by carrying out more focused and, hence, fewer Safety Pharmacology tests).

An IND is a request, under the FDA's jurisdiction, to allow initiation of clinical trials. A successful IND may be filled with an abbreviated version of the core battery investigation if the regulator deems it is worth providing this drug to patients quickly. Of course, preclinical scientists who are dealing with, for example, the oncology division of the FDA, know the requirements of this division because they will be aware of a number of IND packages that do not fully adhere to the S7A guidance, and yet were approved to allow clinical trials to proceed. This informs investment choices (spending). Subsequently, the regulatory authorities will judge if the Safety Pharmacology data is sufficient to establish that the drug does not expose patients to an unreasonable risk; with a rapidly lethal disease and a new type of treatment, time is of the essence and minor adverse effects may not be a critical

concern. Other considerations for regulators (when deciding whether to allow a drug to progress to patients) include manufacturing information documenting consistency of the drug, the proposed clinical protocol and the qualifications of the clinical investigators charged with managing the proposed studies. The flexibility of requirements associated with variations in disease severity, variations in the need for a new drug and variations in the anticipated adverse effects of the new drug should not be regarded as a charter for corner cutting as even abbreviated submissions are subjected to rigorous scrutiny.

Thus, industrial Safety Pharmacology departments seek to fulfill the requirements of the S7A core battery using different combinations of tests based on scientific judgement and the particularities of each drug candidate (that is, on a case-by-case basis). Experience is a major component in this process from both the scientist and regulator perspectives. The process is not dissimilar to jurisprudence (in the legal milieu). Once a company (or a scientist) has submitted a successful package (submission to the regulator), the company learns from this that the approach that informed generation of the package is acceptable. As an example (from personal experience) a single study with integrated telemetry recording of cardiovascular parameters, CNS neurological examination and a respiratory profile using a pneumotachometer in only  $n = 4$  dogs may be sufficient to fulfill the core battery requirement for a drug indicated for a life-threatening disease. In contrast, if the condition to be treated is not life-threatening it may be necessary to implement the full functional observational battery (FOB). The FOB is a formalized systematic evaluation of nervous system function in the rat, comprising more than 30 parameters across autonomic, neuromuscular, sensorimotor and behavioural domains in rats (Redfern *et al.*, 2005), respiratory function in a second study (Murphy and Joran, 1992) and haemodynamic telemetry in dogs (Ollerstam *et al.*, 2007). Drugs for diseases for which treatments are already available (even life-threatening diseases such as Hodgkin's lymphoma) will usually require a complete Safety Pharmacology investigation programme and a relatively favourable safety profile. Thus, there exists a risk/benefit continuum; many currently available anticancer drugs are not in any way 'safe' for healthy humans but they are considered 'safe' for cancer patients given their debilitating condition. Likewise, given the anticipated adverse effects of some anticancer drugs, no testing is needed in healthy human volunteers (see Figure 2 for details on the continuum).

The nature of the drug is also an important factor that will modulate the requirements for Safety Pharmacology testing. As an example, a monoclonal antibody (biologic) will be



**Figure 2** Risk/benefit continuum. In this figure, the wedge symbols represent the accumulation of data for positive discovery outcomes and negative safety outcomes. At some point a decision needs to be made to proceed to human studies. This decision is taken when a subjective threshold is met (indicated by arrows and dotted lines). The decision is an integrated risk assessment. The amount of time required to reach the decision is arbitrary as it is the amount of information accumulated that is paramount. The extent data (discovery and safety) necessary and sufficient for a decision is a trade off. Thus, for a drug for a lethal indication, only a moderate amount of positive discovery data is necessary for a decision to proceed provided that a sufficient amount of worrisome ('bad') safety pharmacology data has not accumulated (quadrant labelled '1'). If a threshold level of bad safety data has accumulated before the threshold amount of 'good' discovery data is reached the drug will be killed (quadrant labelled '2'). The same rules apply for a drug for an innocuous indication, except that the threshold amount of necessary positive discovery data is greater (quadrant labelled '3'), while the threshold amount of bad safety data sufficient to kill the drug is much less (quadrant 4). This figure emphasizes the role of subjective judgement in decision making, and the influence of disease severity on the risk/benefit calculation.

allowed to progress to first in human (FIH) studies with minimal investigation of, for example, TdP liability. In contrast, a small molecule first in a new drug class will require a complete Safety Pharmacology assessment before it can progress to phase 1 assessment in healthy volunteers, owing to the probability that an entirely new class of drug will have the greatest scope for unforeseen adverse effect whereas a monoclonal will have better target selectivity.

As the decision from the regulators for a given product is a risk/benefit continuum that goes beyond the ICH guidelines (US FDA, 2001), the company must choose which Safety Pharmacology tests to perform based on a subjective judgement that takes into account the need to test for 'safety preclinical signals' in the context of the potential benefit for the patient population by considering currently available drugs for the indication and their adverse effects (severity and reversibility of the adverse effects). After all, if, for example, current best therapy for a lethal indication is highly unsafe then to test a new drug for this indication for possible trivial adverse effect liability is needless, and constitutes development procrastination. Given that human lives are at stake, the drug developer is under extreme time pressure to assemble a Safety Pharmacology portfolio that will provide the regulator with data that demonstrates the risk/benefit ratio favours use of the drug in such patients. The pressure exerted by patent laws (time-limited protection) on bringing a drug as quickly as possible to market must also be acknowledged and set against the existence of regulatory guidance on what is regarded as 'reasonable promise' of expected effectiveness as well as 'reasonable expectation of safety' if a drug is allowed to pass from preclinical to clinical investigation. It is therefore clearly difficult for regulators and drug developers to know where to set the threshold for determining a judgement of 'reasonable promise', and how to weigh this against the scope for ambiguity regarding 'reasonable expectation of safety' in many areas (especially for rare but lethal adverse effect liability).

In the future, we can anticipate Safety Pharmacology studies (especially non-GLP screening) being completed earlier in the process of drug discovery and development, with preclinical Safety Pharmacology data used to inform decision making.

### **The development, validation and accreditation of preclinical Safety Pharmacology methods**

*What are the goals of the studies undertaken in safety pharmacology assessment?*

As we have explained, the primary agenda of Safety Pharmacology is to provide companies with data to discontinue development of (kill) unsafe drugs early in the preclinical development phase. The sooner a decision is made to kill a drug the sooner the company can begin to strategize on development pathways, that is, either develop another drug backup using a similar chemical scaffold or consider a dissimilar drug class or programme. As real (human) safety can usually only be decided after conduct of a meta analysis of clinical trials (a statistical approach that

evaluates the combined results of several independent clinical investigation studies, each of which has addressed a related hypothesis), which takes place after drug approval and extensive human exposure, preclinical Safety Pharmacology does not seek to ordain a drug as 'safe'. Indeed it cannot (especially for very rare but potentially lethal adverse effect liabilities, such as TdP). The best it can do is to attempt to identify a drug as potentially safe. This means that, in Safety Pharmacology, expense (of time, human resource and money) is spent in pursuit, not of bringing drugs to the market, but in stopping drugs going to market.

Understandably, therefore, it may be that companies with limited resources to develop drugs could work on the principle of 'as little Safety Pharmacology as necessary' or 'only what is required'. This may result in cutting corners to terminate a drug project, but it should not mean not cutting corners to ordain a drug as safe. In contrast, large pharmaceutical companies will commit to spend more to achieve an integrated (in this context we mean comprehensive) Safety Pharmacology programme that provides the best prediction of human response in the shortest time. Essentially, a larger, resource-rich company may be less ruthless in terminating a drug candidate early than a smaller company because a more extensive safety profile can be afforded to be developed. This ensures a reduced probability of inappropriately terminating development of what may eventually become a therapeutically useful drug. There are two ways of proceeding. First, to ensure appropriate decisions are made, a more comprehensive (and expensive) Safety Pharmacology programme may be judged to be required. To make such a comprehensive programme work, the focus is then placed upon integrating the use of time, resources and decision-making procedures. On the other hand, the company may focus on avoiding drug failures owing to adverse effects and attempt to ruthlessly weed out potential failures using approaches that may have fewer false negatives in the hope of achieving close to zero false positives. Both approaches are subjective: one seeks to avoid throwing out the baby with the bathwater whereas the other seeks to avoid leaving a piranha in the bath in place of the baby. The guiding principle, once again, is to optimize the overall preclinical Safety Pharmacology programme to minimize testing time and most importantly, to identify quickly and accurately, any 'show stopper' adverse effect liability as soon as possible. If drug development can be terminated during the preclinical testing phase instead of phase 3, this is a major resource and financial advantage.

*What guides safety pharmacology in achieving its goals?*

The guiding principle that informs the selection of what safety tests to conduct, in accordance with those outlined in guidance documents (S7A and S7B), may therefore be 'as little as is necessary and no more than is sufficient'. Safety Pharmacology is not, after all, about establishing likely therapeutic benefit, but rather is primarily about preventing further cost with uncertain benefit.

What is necessary to achieve in the laboratory becomes increasingly clear the closer one gets to a lead candidate. Thus, with a new chemical entity (NCE) no safety tests are

conducted until there is reason for considering the NCE to be a potential lead compound (that is, until there is some discovery data available). In contrast, at the other end of the discovery process, just prior to phase 1 (FIH) clinical investigations, there will have evolved a detailed portfolio of Safety Pharmacology data for the 'nominated' compound that will most typically include GLP study findings. The interesting challenge therefore is to know when to do what and how to interpret the findings, as the onus on Safety Pharmacology is to inform the risk/benefit assessment at all stages of drug discovery and development (see Figure 3).

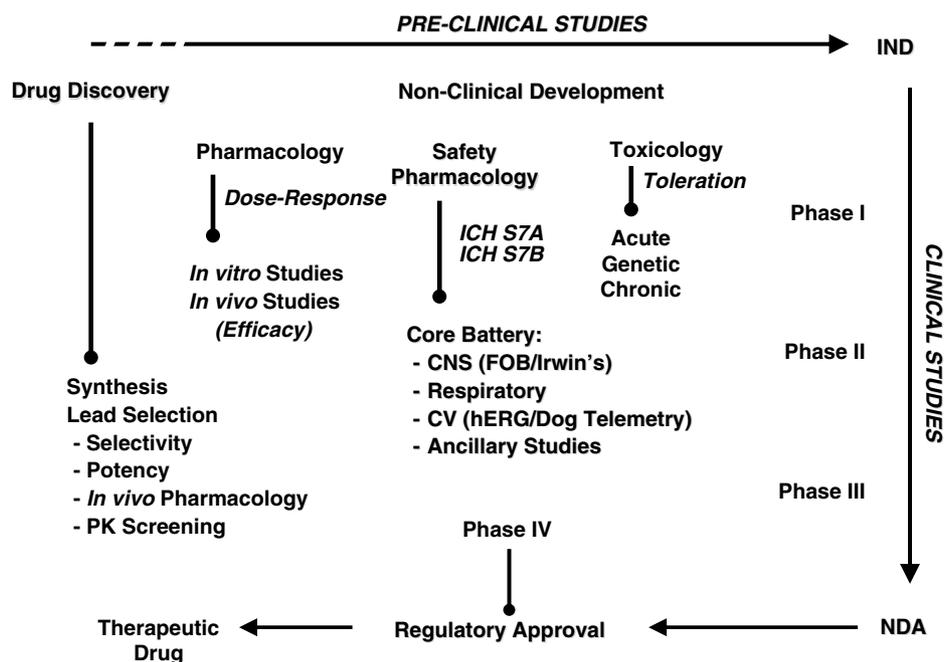
As is inevitable, a nominated compound will have a specific preclinical safety portfolio. However, an NCE that fails in preclinical safety assessment could have an incomplete portfolio (the final entry of which will be the outcome that indicated that potential risk outweighed potential benefit). For any NCE that fails it is the goal of safety assessment to inform the decision as early as possible in the discovery process for reasons of cost (animal and monetary) and time. Therefore, choice of test and timing of testing are critical.

#### *How does safety pharmacology go about achieving its goals?*

There is a core battery of CNS, respiratory and cardiovascular tests that will need to be completed if an NCE is to become a drug (see 'The agenda of safety pharmacology' section above), constructed for purposes of regulatory compliance as well as reason of good scientific practise. However, as

safety assessment progresses from no data to completion, the timing of the deployment of different safety tests (core battery and other studies conducted either in house or at a contract research organization) is a matter of subjective judgement (Table 1). This means that different companies likely conduct studies at different times (Friedrichs *et al.*, 2005; Lindgren *et al.*, 2008), choosing different non-core-battery tests from among those available according to in-house judgement and expertise. While a majority of these tests may have been validated by blinded experimentation (for example, Hamlin *et al.*, 2004; Lawrence *et al.*, 2006) others may not. For these reasons it is beyond the scope of this article to provide a logical explanation for the process of deployment of each test based on an appraisal of model validation and cost effectiveness, therefore we offer here a mere brief description of typical practise.

In some large pharmaceutical companies Safety Pharmacology may be divided into complementary phases. The initial phase is part of the process that informs lead candidate selection and optimization, and is usually not conducted under GLP compliance. For most small molecule drug candidates, this initial phase includes cardiovascular screening. This initial phase typically includes ion channel inhibition (for example,  $I_{kr}$  also known as the hERG potassium channel assay; Murphy *et al.*, 2006) and may include an isolated organ preparation (wedge preparation and/or isolated Langendorff heart; Wang *et al.*, 2008; Hondeghem *et al.*, 2003; Hamlin *et al.*, 2004), an anesthetized animal model (dogs or monkeys) using



**Figure 3** Schematic depicting the complex interaction of preclinical scientific disciplines and study models used to characterize the safety profile of a new chemical entity. A non-clinical development programme includes data from drug discovery models up through Safety Pharmacology and Toxicology where an investigational new drug application (IND) is filed for a candidate drug. The IND is the means by which a pharmaceutical company obtains regulatory permission (from the FDA) to provide drug to clinical investigators for use in phase I clinical trials. The FDA reviews the IND application for safety to assure that clinical research subjects will not be subjected to unreasonable risk. The candidate drug then proceeds through multiple clinical trials (phase I–III) after which an NDA or new drug application is made to regulatory authorities. In this document drug sponsors propose that the FDA approve a new pharmaceutical for sale and marketing. The goals of the NDA are to provide enough information to permit FDA reviewers to establish whether the drug is safe and effective for its proposed indication.

**Table 1** Non-clinical methods recommended for use in the safety pharmacology core battery of tests by ICH Guidelines S7A and S7B

Safety pharmacology core battery	Measured variables
<i>Central nervous system</i> (Modified) Irwin's test functional observation battery (FOB)	Coordination, body temperature, behavior, neuromuscular, sensorimotor, convulsions.
<i>Respiratory system</i> Plethysmography	Respiratory rate, tidal volume, airway resistance/compliance, pH, pCO <sub>2</sub> , pO <sub>2</sub>
<i>Cardiovascular system</i> QT Interval (telemetry) hERG Isolated Purkinje fibers (Langendorff Isolated Hearts) (Proarrhythmia Models)	Blood Pressure, Heart Rate, ECG, Cardiac Output, Left-Ventricular Pressure, Contractility, TRIaD, hERG IC50
<i>Supplemental systems*</i> Gastrointestinal Renal/genitourinary Blood Inflammation Immunological	Intestinal transit time, Gastric emptying and secretion, urine volume, total protein, Cl rate (GFR, Na <sup>+</sup> , K <sup>+</sup> , Cl <sup>-</sup> ) Electrolytes, BUN, platelet aggregation, bleeding time

Abbreviations: Cl rate = renal clearance rate; GFR = glomerular filtration rate. TRIaD (triangulation, reverse use dependence and instability) refers to the integrated risk assessment of Luc Hondeghem (Hondeghem *et al.*, 2003).

\*Note that there are a number of additional supplemental systems that could be interrogated, such as the immune system. This Table is not meant to be a comprehensive list. Refer to S7A and S7B guidance documents for additional study details.

continuous step infusion ( $n = 2/\text{compound}$ ) and a non-GLP conscious telemetry study (Shah, 2008). Drug class and early findings such as hERG assay results help define the next step.

The first phase of Safety Pharmacology that is usually done prior to the core battery is subjective, based on needs, findings and experience. Drug development teams will rely on the experience of their safety pharmacology team through meetings to decide on the successive steps—the process is dynamic. It is important to remember that drug discovery research (tests in disease models) occurs in parallel to safety assessment, and the outcomes of each process inform the decision making in each process (Figure 2). This is the type of decision-making process by which drugs are developed. A simple analogy is the decision making one may make about taking a swim in the sea in the UK (for non UK residents, we note that the sea in the UK can range from balmy and rewarding to chilly, turbulent and treacherous and the typical UK swimmer can range in competence from Olympic to dyslipidemic). First, we decide how strongly we desire to swim. Then we consider the weather. Then ultimately, having travelled to the coast, we dip our toes into the sea and decide (factoring in our general health and fitness) whether or not to take the plunge. This is how drugs go to market: is there a market? What conditions prevail? Do we have a launchable product?

To provide a broader assessment of the safety profile, the animal species selected for an anesthetized animal study may

often be different from that used for conscious animal telemetry. If the drug is bioavailable and has adequate PK in different species, it will be tested in a range of species. If only monkeys or dog are suitable (that is, owing to the unique expression in these species of the drug's primary molecular target for benefit and/or possible anticipated adverse effects), a single species can be used. For some drug classes, additional screening models may be added. This early phase of Safety Pharmacology may appear somewhat random but this is because it is the most difficult to design. Should every NCE be administered to monkeys? Obviously not, especially given that the goal is to find a read-out that justifies terminating drug development. It would be most ideal for a number of reasons if this assessment could be achieved using a test tube assay. However, we know that this is not achievable nor realistic despite attempts to suggest otherwise—humans are complex, integrated physiological systems so similarly complex systems are needed to evaluate the safety profile of the compound. Thus, for very novel NCEs, model and test choice may be impossible to prejudge and decision-making processes are likely to be frequently reviewed.

For the core battery of safety tests, there are regulatory guidelines that test for potential undesirable pharmacodynamic effects on physiological functions in relation to the nature of the drug exposure relevant to functions vital to life (US FDA, 2001). For the early phase Safety Pharmacology investigations, decision-making is conducted on an 'as needed basis'. As an example, cardiovascular adverse effects (for example, heart failure liability) of multi-targeted receptor tyrosine kinase inhibitors (for example, sunitinib; Kerkala *et al.*, 2006; Chu *et al.*, 2007) prompt consideration of the use of repeated dose Safety Pharmacology screening methods in conscious animals using telemetry with a focus on systemic arterial pressure and chronotropic effects for drugs in the same class (Khakoo *et al.*, 2008). For other drugs, for example, a topical acne treatment, this would be fatuous. Of course these choices must be reviewed. Overall, previous failures of the chosen Safety Pharmacology screening programme to predict human adverse effects will dictate how the programme should be revised. In drug development, companies may focus on specific therapeutic targets (for example, a specific enzyme) whereby their chemists will produce a number of iterations (backup compounds) of the parent drug. This may reduce the required extent of safety assessment.

If a company is too small to be able to afford progressing a drug into phase 3 then its ambitions may be limited to sale of their chemistry, technology and drug development programmes to a bigger company. The best time to do this is at FIH stage. Thus a smaller company will adhere to the tenet 'as little testing as needed', as described earlier. A larger company, fully intending to progress a drug to phase 3 itself is more likely to carry out a more extensive and expensive programme of preclinical Safety Pharmacology.

For drug candidates in the large molecule category, this initial screening step may not be required on the basis of the drug development team experience and the safety profile of other drugs in the same class. In these cases, the toxicology and Safety Pharmacology assessment programmes may share the same first step, known as dose range finding (DRF) studies (sometimes called toleration studies), which are

initially conducted in rodents (mice or rats) and followed by studies in a large animal species (for example, usually non-human primates). According to the ICH S6A guidance document (1997) the safety evaluation of a large molecule should include the use of relevant species defined as one in which the pharmacological activity of the large molecule (for example, protein and/or monoclonal antibody) is active because of expression of the receptor or binding epitope for that large molecule. Selected species should also demonstrate a similar immunological tissue cross-reactivity pattern to that observed in humans. The DRF/toleration study design utilizes a dose escalation paradigm to determine the dose at which adverse effects are first seen in a single or limited number of animals (somewhat rather a crude test as statistical proof cannot be part of the process with such an approach). Regardless, such studies characterize the toxicological dose–response profile (usually for the first time for a drug in development) and include cage side observations (for physical and behavioural effects), drug exposure analysis, blood chemistry, haematology, pathology and histopathology. A repeated dose administration toxicology study will often be completed in selected animal species to confirm the dose levels that will then be used in subsequent GLP toxicology and Safety Pharmacology studies.

The second part of the Safety Pharmacology programme is normally conducted in accordance with GLP guidelines for regulatory submission and includes the Safety Pharmacology core battery as defined in the ICH guideline S7A (US Food and Drug Administration, 2001) as well as a GLP hERG assay. Here decisions may be made that defer from ‘killing-the-drug’ mode to ‘presenting-the-drug-as-likely-safe’ mode. Thus there is a change in development status as the audience is no longer the company’s strategic planners; it is now the regulatory authorities who decide whether the drug is fit to be entered for human consumption.

The Safety Pharmacology core battery is typically conducted with a single administration of drug using the same administration route in conventional toxicology studies (similar to that which will be used clinically) with evaluations usually up to 24 h. Cardiovascular safety is assessed in a conscious telemetry study (for example,  $n=4$ ) usually in a Latin square or dose-escalation design with sufficient drug ‘wash out’ times between dosing. These studies usually use the same species as in the large animal toxicology studies. Respiratory Safety Pharmacology is typically evaluated in conscious rats (for example,  $n=8$  given the greater variability of respiratory parameters) but large animals such as dogs and monkeys may also be used when rodents are not suitable (for example, if target is absent in rodents or absorption distribution metabolism elimination (ADME) profile is not adequate). Neurological safety is usually evaluated using a modified Irwin test in rats (Irwin, 1968; Mattsson *et al.*, 1996) where qualitative evaluations are conducted by an evaluator blinded to study treatments (for example,  $n=10$  per group). Neurological evaluations may also be performed in other species (for example, mice, dogs, minipigs or non human primates; Moscardo *et al.*, 2007; Tontodonati *et al.*, 2007) as for respiratory Safety Pharmacology. Beyond routine CNS Safety Pharmacology evaluations, some models are developed to characterize

specific neurological adverse effects with the use of EEG monitoring by telemetry (Durmuller *et al.*, 2007). A trend to integrate some components of the Safety Pharmacology evaluations such as respiratory, CNS and ECG study end points into toxicology studies is currently noted (Luft and Bode, 2002). Development of non-invasive methodologies such as ECG monitoring (along with respiration, temperature and animal activity) using jacketed external telemetry systems has significantly contributed to this emerging practise (Morton *et al.*, 2003). Among the advantages of Safety Pharmacology assessments in toxicology studies, we have an increased sensitivity (for example, increased statistical power) based on the relatively large number of animals in toxicology studies, reduction of the number of animals required for overall safety evaluations, an integration of Safety Pharmacology end points with histopathological and hematological/clinical chemistry data and potential cost reduction (for example, when including FOB and respiratory assessments in toxicology studies).

#### *The challenge of validation of safety pharmacology approaches*

The key question about the core battery tests (as far as the regulators are concerned) is: are they validated? In other words, does the chosen model accurately identify the safety liability of the drug candidate? Validation of Safety Pharmacology test systems for GLP compliance is achieved at each test site using positive control drugs with currently accepted models (Hauser *et al.*, 2005; Chaves *et al.*, 2006, 2007; Authier *et al.*, 2007, 2008). At a higher level, some initiatives such as the QT-PRODACT project have helped characterize the sensitivity of the methodologies and inter-facility variability (Ando *et al.*, 2005; Miyazaki *et al.*, 2005; Omata *et al.*, 2005; Sasaki *et al.*, 2005; Tashibu *et al.*, 2005; Toyoshima *et al.*, 2005). These results have contributed to the increasing harmonization of industry practises, making it easier for regulators to make judgements based on retrospective comparison considerations (precedents).

Although test system validation for regulatory purposes appears to evolve within an accepted reference frame, does this mean that regulatory authorities will accept as ‘validated’ a method that has not actually been scientifically validated? From experience with regulatory audits and IND package submissions, regulatory authorities will accept models that have been demonstrated as reasonably valid in the public domain (that is, used, and the data published). Accuracy, reliability, use of standard agents as reference and security of the systems are major elements in GLP validations.

True pharmacological validation remains a vexing issue in Safety Pharmacology in exact mirror image of the issue of validation of disease models in drug discovery. It is important to emphasize that models and biomarkers are ‘valid’ only when they detect all and only those drugs that have the same effectiveness and safety in the human. There is a major paradox inherent in this requirement, one that is not well recognized and one that is a fundamental problem for the newest most potentially revolutionary drugs. Thus, because new drugs are new by definition (FIH for an untreated condition, NCE, new mechanism of action), the disease for which the drug is intended may have no presently

available treatment. Clearly without a positive control to provide a template response profile, this means there can be no validated preclinical model for discovery. Thus the models used to identify the new drug are not validated, and will not be validated until the identified drug is shown to be effective in humans. Likewise, in Safety Pharmacology, no model is validated until a range of positive and negative controls have been shown to produce the same outcome in the model as occurs in humans. This sounds simple; however, it is a huge problem for certain types of adverse effects. Thus to validate a model that is to be used for detecting a liability for a drug to evoke a very rare (but potentially) lethal event (RLE) requires precise and accurate human data on the liability of a range of drugs to evoke the RLE (the 'gold standard').

Again, drug-induced TdP liability testing provides a good example of the problems here. One of the most well known TdP-causing drugs is terfenadine. One would imagine (on the basis of foregoing considerations) that terfenadine would be one of the first drugs to be chosen to be used in validating any new TdP liability-testing model. However, one would need to know the exact rate of occurrence in humans of TdP with terfenadine to make use of terfenadine to validate a model. The rate of occurrence could then be ranked against a range of other drugs to generate a clinical ranking order. This would then serve as the template to validate the model (by generating a comparator ranking order for the model). There are several problems with this, unfortunately. For TdP there exists no reliable clinical ranking order for drugs. This is because when events (such as TdP) are rare the calculation of their preponderance cannot be made with acceptable precision. Thus, for terfenadine, the rate of occurrence of adverse cardiovascular events was 83 cases reported from time of approval in Europe (1981) until 1992 after use in millions of patients (Kemp, 1992; Schiefe and Cramer, 1996; Yap and Camm, 1999). The drug manufacturer reported the 'events' (cases presenting as TdP, QT prolongation, ventricular tachycardia, flutter and fibrillation, cardiac arrest and sudden death) to the FDA, which issued a 'black box' label warning of cardiovascular risk for the drug alone as well as when prescribed in combination with macrolide antibiotics and azole antifungal drugs (Morris and Carlson, 1998). On the basis of these findings the FDA issued a proposal to withdraw terfenadine from the market in 1998 and the manufacturer complied. Interestingly, in a cohort study comparing terfenadine to clemastine, non-prescription antihistamines and ibuprofen, life-threatening ventricular arrhythmias (used in lieu of categorical confirmed TdP incidence) occurred in less than 0.063% (or 317 out of >500 000) of patients in a Medicaid recipient database (Pratt *et al.*, 1994). The authors concluded that terfenadine users were no more likely to develop arrhythmias than those on ibuprofen or clemastine (Tavist) (Pratt *et al.*, 1994). Darpo (2001) reviewed the annual number of adverse drug reaction (ADR) reports of TdP submitted to the World Health Organization drug monitoring centre over a 16-year period (1983–1999). Of the 761 cases described 34 (or 4%) were fatal. Of the 20 most commonly reported drugs only ~46% were cardiovascular drugs (Class I, III and IV), the remainder were non-cardiovascular (antibiotics, antihistamines, anti-

psychotics and so on; Darpo, 2001). Interestingly, over this time period 41 cases (one fatal) of terfenadine TdP reports were named from 10 047 (or 0.41%) reported ADR for this drug. So, there is no reliable 'gold standard'.

As an alternative, when events are rare, some investigators have attempted to classify large numbers of drugs into a small number (5–7) of distinct classes, reflecting a subjective ranking of perceived risk. Already one can appreciate this process is fraught with uncertainty as it lacks precision and (probably also) accuracy. Indeed, the literature carries examples of variable approaches used to rank the relative risk of different drugs. The Arizona Center for Education and Research on Therapeutics (CERT) website ([www.qt drugs.org](http://www.qt drugs.org) or [www.torsades.org](http://www.torsades.org)) provides a regularly updated list of drugs (~132 as of April 2008) that can prolong QT (a subjective biomarker for TdP liability). Risk is categorized using integration of an international medical registry of drug-induced arrhythmias, case reports, FDA drug labels and data from preclinical, clinical and epidemiological studies by an expert committee of advisors for agents known to cause TdP. Drugs are then assigned as those with a possible risk, those to be avoided in congenital long QT syndrome patients and drugs unlikely to cause TdP unless other risk factors are present. This is not precise. Thus, to use even this 'template' to attempt to validate a model is hazardous.

#### *Recognizing and dealing with the validation gap*

To conclude, when an adverse event is a concern (because it is potentially lethal or debilitating), even if its occurrence is rare and when the drugs known to have a liability for the event have only a *low* liability, then it is almost impossible to validate any preclinical test (owing to a lack of precision and accuracy concerning the rank order of liability of the template drugs).

To make matters worse, if a potentially lethal adverse event is actually rather rare, even among the drugs with a known liability, then the putative test model will either have a similar low rate of adverse event making it impossible to use (if the event rate in humans is 1 in 1000 it means that many thousands of tests would be required to detect the adverse event liability in the model), or it will need to be 'modified' to exaggerate the drug's adverse event rate. The obvious drawback here is that if the model exaggerates adverse event liability then how reliable (precise and accurate) would be the rank order of liabilities of the range of positive and negative controls (the template drugs) in a validation test? Moreover, in an exaggerated liability model there are likely to be false positives.

Clearly there is no ideal approach to safety testing for rare but serious events. This area remains the most problematic in Safety Pharmacology and has generated numerous publications in recent years (Yamaguchi *et al.*, 2003; Hamlin *et al.*, 2004; Thomsen *et al.*, 2004; Valentin *et al.*, 2004; Lawrence *et al.*, 2006; Liu *et al.*, 2006; Kagstrom *et al.*, 2007). In this context, validation of the predictive value of Safety Pharmacology models is an evolving understanding of the relationship between clinical adverse effects and our integrated preclinical screening tools.

*Refining the practise of safety pharmacology*

How might we move forward? Let us again consider drug discovery. The best models in discovery are validated retrospectively when they can be shown to have had a key role in advancing a drug into clinical use. In discovery a model can be perfectly valid even if it provides only partial information (that is, accurately detects one class of effective drug but not another—for example, rat heart arrhythmia bioassays are effective for detecting class I antiarrhythmic activity (Farkas and Curtis, 2002) but not class III antiarrhythmic activity (Rees and Curtis, 1996)). Thus, in safety we cannot necessarily expect validation of the core battery tests. However, safety and discovery differ with respect to the perceived need for validation of models, and in terms of the influence of badly validated models on drug development. Thus, in discovery, if a model gives regular false positives it is quickly abandoned once the first products of the model fail in man because of the lack of effectiveness. If, on the other hand, the model regularly gives false negatives then there will be no products to test in man, and the model will eventually be abandoned and superseded by another. Model competition is inherent in academia and industry; for example, in antiarrhythmic drug discovery researchers have had species issues (dog vs cat vs rat vs pig); conscious vs anesthetized vs isolated heart preparations; myocardial ischemia vs infarction vs reperfusion vs programmed electrical stimulation (PES), with continuous re-appraisal of models (Johnston *et al.*, 1983; Botting *et al.*, 1985; Chung *et al.*, 1993; Bellemin-Baureau *et al.*, 1994; Curtis, 1998; Billman, 2006; Hamlin, 2007). As yet there has been little equivalent assessment in Safety Pharmacology. It is critical that the core battery tests be properly validated by showing they are the best among the possible options in terms of avoidance of false positives and false negatives.

What are the processes that drive new Safety Pharmacology model development and the processes used to 'accept' a model? The answer to this question is not satisfactory in terms of science, since the overriding guiding principle is pragmatism. This is exemplified (once again) by reference to TdP liability testing. HTS screening for drug block of the potassium current (so-called 'hERG screening' as the gene, hERG, encodes the channel mediating  $I_{Kr}$ ) is commonplace in industry even though it is known to generate false positives, false negatives and highly variable potency values ( $IC_{50}$ ) compared with voltage-clamp methods for channel inhibition (Zheng *et al.*, 2004; Sorota *et al.*, 2005; Murphy *et al.*, 2006; Slack *et al.*, 2006). It is used because it is quick and relatively inexpensive, once established in-house. Its use is justified because the frequency of false negatives compared with compound throughput in this crude HTS is considered to be as low as inconsequential. Moreover, on the other hand, the possibility of a false positive in an inexpensive crude HTS screen is a trivial concern in comparison with not using the screen and ending up with a candidate that has a real adverse effect liability, which is detected only much later in drug development when, for example (TdP again), *in vivo* dog telemetry studies are (now routinely) undertaken. Indeed, companies today are likely to avoid what they regard as an  $I_{Kr}$ -binding pharmacophore in early stage synthesis of new chemical entities. This

pragmatic approach is understandable, but it should be remembered that this is a gamble and is not validated in that we do not know how many potentially useful drugs are lost by this process. It has been argued by Hondeghem that crude screens for  $I_{Kr}$  block may result in truly valuable agents being lost to medicine (Shah and Hondeghem, 2005). Certainly if a drug is intended for life extension in aggressive carcinoma then it would seem ludicrously inappropriate to discard a potentially useful drug just because it blocks  $I_{Kr}$ .

The decision on whether a drug candidate should progress to the next level in the discovery process is one of the most important in Industry. Progression is driven by discovery outcomes (that is, outcomes in studies focused on potential therapeutic effectiveness) and is halted either because there is a loss of signal (lack of benefit in a disease model, for example, meaning that progression stops owing to efficacy issues) or because of the emergence of an adverse effect signal (meaning that progression may stop owing to safety issues) (see Figure 2). An adverse effect signal may not necessarily end progression, but it will certainly slow it. For a drug that is late in the candidate selection process there is reluctance for an adverse effect signal to be used to terminate progression without a proper scientific interrogation of the signal. This means a safety signal for a 'mature' drug in preclinical development is likely to trigger new mechanistic-based studies. These are likely to involve a step-up in perceived clinical relevance (for example, if the signal was in anaesthetized acutely prepared rodents the follow up studies may be conducted in conscious canines or primates with telemetry). This means a step-up in cost and time. The integrated risk assessment (or evaluation of all non-clinical study results from the core battery studies including findings obtained from follow-up studies as well as other relevant information including pharmacodynamics, tissue distribution and drug interaction studies) will therefore take into account money already spent and the likely return if the NCE becomes a drug (that is, whether or not the follow up studies fail to reiterate the adverse effect signal).

If a surrogate end point is used as the decision-making safety signal, false-positive and false-negative results may inform false decision-making. For example, using the QT interval in the ECG as a surrogate biomarker gives a false-positive for ranolazine and a false-negative for disopyramide (Shah, 2008). There is no absolute threshold for decision making with any of the available surrogate biomarkers. The burden is on the development team (discovery and safety pharmacologists working together) to establish criteria for decision making (see Figure 2). However, if the emphasis is placed on safety signals over effectiveness signals (as is inevitable), errors are unavoidable. For example, while a drug candidate in the small molecule category with  $IC_{50}$  for blocking  $I_{Kr}$  above  $1\ \mu\text{M}$  can usually be considered safe to pursue in early stage preclinical development, application of this pragmatic threshold would have led to development discontinuation of valuable candidates such as amiodarone (Lin *et al.*, 2005). The overall cost of preclinical development for a single drug (US\$2 to US\$10 millions) is trivial compared with clinical trial costs (US\$100 to US\$800 million; NCI 2007) but a large number of candidates that enter preclinical screening programmes will be 'killed' prior to reaching the

FIH milestone. This situation is increasing the weight on resource allocation at an early stage to maximize the output of the drug development pipeline.

Molecule size is an important determinant of the type of safety signal likely to be detected. Thus, drug-induced acute QT prolongation is much more common among small molecules owing to the QT-prolonging mechanism, which commonly requires drug access to discrete molecular targets, specific amino-acid residues that form the drug-binding site located within the central cavity of the hERG channel (Sanguinetti and Mitcheson, 2005; Kamiya *et al.*, 2006). In contrast, large molecules such as proteins, peptides or monoclonal antibodies (biologics) that may be too large to affect ion channels (Vargas *et al.*, 2008) may evoke adverse haemodynamic effects owing to actions on more readily accessible targets in the vascular space. A trend for the 'pipeline' to contain an increasing number of large molecules has been seen in the past few years (Marafino and Pugsley, 2003) leading to an adjustment of drug screening paradigms and Safety Pharmacology studies based on the type of molecule. Thus, although a respiratory Safety Pharmacology 'hit' is rarely the signal for small molecule drug discontinuation, it is increasingly recognized as the signal for large molecules intended for repeated administration, owing to a propensity for sensitization and allergic reactions (Murphy and Joran, 1992).

Species sensitivity should also be considered in the interpretation of Safety Pharmacology study findings. Drug or vehicle administration in dogs occasionally leads to histamine release with associated cardiovascular changes (Masini *et al.*, 1985; Eschaliier *et al.*, 1988). This phenomenon is known to be species-specific and has relatively limited clinical relevance. Pretreatment with antihistamine drugs such as diphenhydramine and cimetidine (Kien *et al.*, 1992) or measurement of histamine plasma concentration are determined in non-GLP Safety Pharmacology or mechanistic toxicology studies to confirm histamine-mediated effects.

Getting the balance right, therefore, is the big challenge, and different companies take different stances on this. Again, decision making is informed by an awareness of a paradox. Thus, in preclinical drug discovery the possibility (from studies using disease models or biomarker models of possibly dubious validity) that a drug may be of benefit informs the regulatory process to take an optimistic stance. In safety, the possibility (from preclinical studies of dubious validity) that a drug may be unsafe, informs the regulatory process to take a pessimistic stance. Although the latter might mean the delay or preclusion of useful drugs entering clinical use, it might also mean the prevention of harmful drugs doing likewise. The task for scientific Safety Pharmacology is to provide better evidence to direct the decision making. By this means the present guidelines might be expected to evolve to become less vague.

### The detail of safety pharmacology studies

From the most recent survey of industry practises (conducted by the Safety Pharmacology Society from late 2007 to early 2008; Lindgren *et al.*, 2008), we can provide the following

summary of the survey regarding issues related to the 'frontloading' of Safety Pharmacology studies. 'Frontloading' defines a safety study that is conducted with a compound prior to its selection as a drug candidate for continued development. According to the survey, ~78% of safety pharmacologists responding conduct such frontloaded studies. Both Discovery (~51%) and Drug Development (~49%) Research Centres share the responsibility for conducting these studies.

When such safety studies are partitioned and examined (Lindgren *et al.*, 2008) as to whether they are frontloaded or not, all survey responders reported frontloading cardiovascular safety studies (~69% during lead optimization prior to candidate selection). CNS studies are also almost always frontloaded (by ~63% of responders prior to candidate selection). Interestingly, but not surprisingly, frontloading of respiratory studies is low (only 28%) whereas ancillary organs (such as the gastrointestinal (GI) tract and kidneys) were generally not frontloaded (only ~21% of responders frontload such safety studies). Note that such studies are not usually conducted according to GLP standards.

Of all the studies that can be conducted during this phase of drug development, the hERG assay appears to be frontloaded by all the survey responders. Approximately 60% of respondents require mandatory hERG testing to proceed with a development candidate (Lindgren *et al.*, 2008). Of the plethora of available methodologies that can be used to determine drug effect on hERG channels the majority use the following test systems: automated HTS patch clamp (~84%); ligand-binding studies (~38%); non-automated patch clamp (~34%) and Rubidium efflux studies (~9%). Rabbit and guinea pig CV and ECG studies along with many methods used to evaluate drug effects on ventricular repolarization such as the action potential duration are frontloaded.

The FDA has recently made abuse liability assessment a mandatory part of the development phases of the submission process for all new CNS-active drug products (see comprehensive reviews by Ator and Griffiths, 2003; Balser and Bigelow, 2003). Abuse dependence liability studies are required under the USCA Title 21, Chapter 13, Controlled Substances Act, as amended 15 February 1996, §811(c), and (f) and is maintained in full accordance with the National Institutes of Mental Health's Methods and Welfare Considerations in Behavioural Research with Animals (Morrison *et al.*, 2002). A preclinical abuse liability testing guidance document was recently approved for use by the EMEA (EMEA/CHMP/SWP, 2006); however, such an equivalent guideline is only in draft stage in the US, under the auspices of the Controlled Substance Staff (CSS). The CSS provides expertise to the FDA and CDER divisions in assessing drugs for abuse liability and fulfills this unique role within the FDA under the authority of the Controlled Substances Act (CSA) of 1970. In Canada a clinical testing abuse liability guidance document is also in review (Health Canada, 2006). Many in the industry may not be fully aware of the new regulations requiring abuse liability assessments (as originally established according to the FDA Food Drug & Cosmetic Act (FD & C, 1938) and the Control Substances Act (CSA, 1970), which determines, labels and schedules abuse potential.

However, as approval by the EMEA in 2006 numerous scientific and procedural challenges have yet to be fully resolved. Sponsors will need to select appropriate preclinical *in vitro* (binding, functional) and *in vivo* (neuropharmacological, behavioural) models based upon the pharmacological nature of the test compound and the onus will be placed upon pharmaceutical companies to 'build a case' for appropriate assessment, in consultation with regional regulatory authorities. In the US, numerous agencies are involved in drug scheduling (FDA, CSS, NIDA, DEA), but as with safety studies, the FDA will require more testing rather than less and will, if directed by clinical adverse effect concerns, rely upon preclinical models for clarification of mechanisms responsible for the adverse effect. The best characterized, validated and predictable preclinical models are used for Schedule I drugs (high abuse potential drugs with no accepted medical use such as lysergic acid diethylamide) and II (high abuse potential drugs with an accepted medical use such as morphine); however, their applicability to 'weaker' compounds may not be appropriate such as drugs with Schedule 4 (low abuse potential drugs such as diazepam) or no Schedule effects. One important issue that was introduced resulting in much debate and controversy was that the EMEA regulatory authorities consider that 'behavioural pharmacology studies for investigating dependence potential...should be conducted in compliance with GLP to the greatest extent possible' (EMEA/CHMP/SWP, 2006). Thus, GLP conditions are required in Japan, expected by the EU (according to guidance) and preferred by the FDA. However, scientists in Safety/Toxicology areas tend to have limited experience with these models as preclinical testing has historically been assessed using behavioural pharmacology models (Ator and Griffiths, 2003). As only a limited number of CROs conduct preclinical abuse liability studies there is an additional concern about the potential for delay in drug package submissions; most studies are conducted at academic institutions who do not comply fully with GLP creating potential discord. Numerous other issues of concern include debate regarding choice of species (rats or non-human primates)—the FDA position is unclear (but more reliance is placed upon primate data) whereas Japanese regulators prefer non-human primates and the EU recommends avoidance of primates and advocates use of the rat. The FDA, in accordance with the EMEA, will likely suggest that abuse liability potential be characterized over a dose range, specifically up to doses that occur to several fold above the expected clinical exposure (therapeutic) range. The clinical route of drug administration is preferred as with safety/toxicology studies; however, most self administration behavioural study methods require the use of intravenous formulation necessitating development of toxicology and pharmacokinetic information before conduct of abuse liability studies. Therefore, a better integration process is needed between preclinical and clinical studies to provide an adequate 'integrated risk assessment' regarding abuse liability potential; preclinical data should be used to focus clinical investigations and aid in identification of clinical comparator compounds.

The EMEA guidance document recommends a two-tiered strategy regarding abuse liability. The first tier pharmacology

studies involve an assessment of the nature of the compound. Information regarding chemical similarity to known drugs of abuse, whether the mechanism of action is similar to compounds known to have abuse liability potential and data from receptor-binding studies, are all early signals for such a potential liability requiring subsequent evaluation. *In vitro* binding and functional cellular studies that are conducted as a part of early development can provide signals for possible dependence potential. Additional functional assays measuring neurotransmitter release and second messenger activity may also be conducted. *In vivo* neuropharmacological models including microdialysis, neurotransmitter turnover, antinociception and locomotor activity may be used (Johanson, 1990). Combined, these first tier pharmacology studies should aid in elucidation of the compound profile and mechanism of action and establish the degree of elaboration of assessment needed to establish the dependence potential.

A second tier behavioural pharmacology assessment is necessary if these initial signals suggest dependence potential and insufficient information is available to define dependence potential. Numerous animal models have been developed to assess the potential for development of drug abuse liability. Specific selection of an appropriate animal model should be based upon the pharmacological profile constructed (see 'Refining the practise of safety pharmacology' section above). A complete dose-response profile using multiple study end points (including motor and cognitive function) should be conducted and parent and metabolites considered. Clinical route of administration and appropriate animal species must be used in animal models (EMEA/CHMP/SWP, 2006; Weerts *et al.*, 2007; Feltenstein and See, 2008) that include physical dependence (drug withdrawal), reinforcing properties (self-administration), discriminative effects (drug discrimination) and tolerance.

## The emergence of scientific safety pharmacology

Although the practise of Safety Pharmacology is dictated principally by regulatory need, its development as a scientific discipline is informed by the same issues as any other biological science that requires the use of animals, namely the issues that inform appraisal of the extent to which the data sets are relevant to humans. By this we mean everything from whether the human molecular target is expressed in the chosen animal and whether the animal's basic physiology and biochemistry is sufficiently similar to that of man, through to the bioassay characteristics of the animal disease model and its cost effectiveness. However, there are distinct differences between safety and discovery pharmacology in the way these issues are treated, partly alluded to earlier. Safety Pharmacology is a discipline whose external role is simply to provide an integrated assessment of data that addresses risk and determines whether a drug will not likely be unsafe in man. In science one can never prove a negative and yet trying to prove a negative is the agenda of Safety Pharmacology. This has affected the evolution of the discipline.

The key issue to consider in this regard, given that the best test bed for human safety is a human test bed (phase 1

testing) is: how much preclinical Safety Pharmacology is necessary? Presently the guidance (ICH S7A; US FDA, 2001) is moderately explicit, but the suggestions for ancillary studies are almost open-ended. Thus, individuals conducting Safety Pharmacology studies are actually shaping the guidance in an ongoing manner by virtue of the nature of the data they generate and the nature of its relationship with the eventual clinical outcome. Thus, if the preclinical novel-type ancillary data on a new drug is accepted by regulators, its validity will presumably be assessed later by consideration of how the drug fared in man from a safety perspective. However, this process of ongoing validation will proceed only if there is scrutiny and publication of findings. This means that there is an onus on Industry and CROs to publish their Safety Pharmacology data. Indeed we hope that this will become mandatory.

In discovery, historically there has been a very meagre documentation in the literature of exactly what preclinical tests and preclinical thinking was involved in the generation (from the first idea through the preclinical screening and testing for potential effectiveness) of a commercially successful drug (note that there is an obvious reluctance to divulge thinking/serendipity because of the potential for competitors developing similar drugs). If the same holds sway with regard to Safety Pharmacology in terms of the ideas behind and the development and validation of whatever methods a company has successfully used to selectively extract potentially unsafe drugs (as is likely) then it is difficult to imagine how the process of validation will proceed in any sort of systematic fashion. Safety pharmacologists need to publish on their emerging battery of HTS safety screens, not only to reveal their validity but also to publicly display the company's safety screening prowess. This will add to credibility when presenting a drug for consideration that is claimed to be safe.

Another key issue is the relationship between preclinical Safety Pharmacology and phase 1 clinical studies. Here it is important to acknowledge the regrettable clinician concept that preclinical (animal) Safety Pharmacology studies are minimally useful or predictive. Once again, the issue of drug-induced TdP provides a good basis for elaborating this important point. For an example of this presumption, in a recent authoritative book on cardiac safety of non-cardiac drugs edited by respected clinicians (Morganroth and Gussack, 2005) only 4 out of 18 chapters focused on relevant preclinical aspects (and of those, one chapter discussed molecular aspects of ion channels and another was concerned with pharmacogenomics). The vast majority of chapters were concerned with the minutiae of recording and identifying subtle indicators of TdP liability in human subjects (from phase 1 assessment of small QT interval and QT shape changes to 'thorough QT' evaluation). This focus on determining whether a horse has bolted from the barn or is about to bolt would surely be better dealt with by ensuring the barn door had been locked in the first place (that is, by ensuring that preclinical methods are sufficient to block any progression of an unsafe drug to the clinical setting).

Nevertheless, it is undeniable (and quite proper) that once reliable human data that shows that a drug is either safe or unsafe relative to therapeutic benefit in humans has been

generated this renders most preclinical data unnecessary—however, there are times when it can provide some important information regarding mechanisms that may be responsible for adverse events that can become evident during post-marketing surveillance. Preclinical Safety Pharmacology models are constantly evolving and improving under the pressure of clinical trial findings. This is illustrated by the recent findings with sunitinib (the first multi-targeted receptor tyrosine kinase inhibitor approved simultaneously for the treatment of renal cell carcinoma and imatinib-resistant gastrointestinal stromal tumour; Demetri *et al.*, 2006) in which a cardiovascular adverse effect liability (heart failure) was not detected in initial Safety Pharmacology studies (Khakoo *et al.*, 2008).

Interestingly, preclinical model validation that is not tailored to clinical results has only a minor impact on industry practises. As an example, the lack of the transient outward potassium current ( $I_{to}$ ) in minipigs (Mow *et al.*, 2008) has not triggered major concerns for the use of this species for *in vivo* QT evaluations. The physiological and Safety Pharmacology role of  $I_{to}$  has been extensively characterized in various species including humans (Patel and Campbell, 2005) and, even though  $I_{to}$  contributes to ventricular repolarization, its block has not been associated with significant arrhythmogenic potential in humans. In other words, the fact that pigs have no  $I_{to}$  has not stopped this species from being used for QT testing, because there is no positive data to show that selective  $I_{to}$  blockers have no TdP liability. Thus although the pig will become a pariah if it fails to pick up QT widening by  $I_{to}$  block for a drug that is later found to have a TdP liability in humans, the absence of a 'hit' in this case means the pig is presently acceptable. This is pragmatism, and this is an important reality of Safety Pharmacology where the inability to predict a human response is usually needed before the industry and regulators correct the tools that are accepted. Again this is similar to jurisprudence where similar logical disposition is granted in relation to historical precedent, experience and judgement, all of which have a role to play in decision making. The problem with this is the consequence it has for preclinical Safety Pharmacology.

When a decision is to be made whether to take a new drug into clinical studies, the decision makers and clinical trial designers are required to know that the drug they hope they can show to be effective will also be safe. Thus, the clinical development attitude to Safety Pharmacology tends to reduce to a request for a simple yes/no answer. Thus, the bottom line in preclinical Safety Pharmacology is to generate a provisional integrated risk assessment that may be contemplated by individuals in charge of clinical development, and also to provide advice concerning whether the drug is likely to be sufficiently safe to warrant the start of clinical investigation. In other words: 'will it be safe?' Indeed, often the question asked by the next level of management is even more demanding: 'is it safe?' Thus, preclinical Safety Pharmacology involves an integrated risk assessment but the need for a risk/benefit calculation requires an unequivocal assessment of risk. This is very challenging.

Given these considerations it is understandable that those in charge of clinical development require clear guidance (yes

or no) and even then will wish to rely on clinical data when managing development of the drug. Indeed, (back to TdP liability once more) the concept of 'Thorough QT' assessment (TQT) in humans has emerged in recent years (US FDA, 2005a, b). 'Thorough QT' assessment derives from the ICH E14 guidance document (Darpo *et al.*, 2006) that provides recommendations concerning the design, conduct, analysis and interpretation of clinical studies to assess the potential of a drug to delay cardiac repolarization (Shah, 2005, 2008). Such a clinical trial is applicable to both new drugs with systemic bioavailability but also to approved drugs where there may be a change in dose or route of administration (resulting in an increase in exposure) or a change in patient population (US FDA, 2005b). Similarly, such a study may be triggered in response to the pharmacological class to which the drug being developed belongs where there may be an association with QT/QTc interval prolongation or TdP during post marketing surveillance (US FDA, 2005b). A 'thorough QT' study is extremely expensive to conduct (US\$3–5 million), and as it measures only QT (a putative surrogate biomarker for TdP risk) and not TdP itself, is not necessarily predictive of TdP liability. This (the cost and the uncertainty of human biomarker data sets) is one further reason why preclinical safety assessment should be inclined to take a safety-first attitude to 'hits' in safety screens.

As noted earlier, this is no different in qualitative terms from the drug discovery position, whereby clinical development will proceed on the basis of a yes/no judgement about likely effectiveness. However, once again, the stakes are different between discovery and Safety Pharmacology. If a drug fails owing to lack of effectiveness, the apparent preclinical false positive will not kill off further preclinical discovery efforts. On the other hand, when a drug fails because of adverse effects in man the consequences for preclinical development are catastrophic. In the cardiac arrhythmia suppression trial and survival with oral D-sotalol trials, two drugs intended to treat ventricular arrhythmias were found to evoke ventricular arrhythmias and kill patients (Pratt *et al.*, 1998; Weiss *et al.*, 1999). In the preclinical studies conducted at the time, although the types of proarrhythmia studies that may be used today were not undertaken, there was nevertheless an apparent failure in adequate safety testing as the proarrhythmic liability of these compounds went undetected. The consequences were that, not only did the preclinical disease models detect the potential ventricular arrhythmia effectiveness fall into disrepute, but also the world's pharmaceutical drug development programmes for treatment of ventricular arrhythmias were abandoned owing to lack of faith by the pharmaceutical industry in preclinical models in this area. This is catastrophic because 30–40% of adults today will die from ventricular arrhythmias for which there are no adequate prophylaxis (one of the largest untapped markets in the drug world). So the stakes for inadequate Safety Pharmacology are much higher than the stakes for flawed discovery (effectiveness) pharmacology. The upshot is that the scope and extent of preclinical Safety Pharmacology data sets necessary to support a claim that a drug is safe is ever growing. Moreover if, among a large set of preclinical Safety Pharmacology data that shows a lack of safety risk

there is one subset that can be interpreted as hinting at the possibility that the drug may be unsafe, a great deal of notice is taken by the regulatory authorities. The upshot is that better confidence is required for preclinical Safety Pharmacology method validity. To achieve this will require a better and more generally accepted methodology for validation of Safety Pharmacology approaches.

## The principles of safety pharmacology and the unmet needs

When there are a large number of drugs that have precise and known relative liabilities for producing common and frequent minor adverse effects it is a simple matter to validate preclinical models using the human template of responses to positive and negative controls. The challenge in Safety Pharmacology is dealing with rare events of a life threatening nature, especially for drugs aimed at treating non life-threatening diseases. Here follows a simple guide. It is not intended to be prescriptive and we invite the community to interrogate it, modify it and challenge it.

- Preclinical safety pharmacology models require better validation
- Validation requires a quantitative and accurate human template of liabilities of positive and negative controls with which to compare model data sets
- Validation is not possible for models screening for liabilities that are rare or imprecise with current drugs in humans
- Validation is also not possible for methods for evaluating human-specific biologics (that are antigenic in animals)
- When validation is not possible, especially when the liability in humans is rare but life threatening, the use of surrogate biomarkers is unavoidable
- It must be understood that interpretation of surrogate biomarker data sets is unavoidably subjective
- Preclinical safety testing in a non-validated setting must therefore be regarded as non-scientific whereby yes/no judgements will remain subjective in the absence of true validation of the models available
- Scientific validation of safety testing methods remains the goal, however, elusive this may seem
- Scientific validation requires blinded randomized testing of drugs known to have and known to not have a liability for the specific adverse effect in humans
- A rank order of liable drugs in humans ('gold standard') is the best template
- It must be acknowledged that a gold standard does not exist for most adverse effect liabilities. This poses a problem
- In the absence of validation it is better to live with false positives than risk the chance of false negatives.

## Acknowledgements

We thank Mark Holbrook (Pfizer UK) for reading the paper and providing expert advice on content.

## Conflict of interest

Two of the authors (MKP and SA) are employed in the pharmaceutical industry. One of the authors (MJC) has provided advice to various pharmaceutical companies and has received research funding from various pharmaceutical companies. The opinions presented in this article represent an objective assessment of the state of the art, and do not reflect a commercial agenda.

## References

- Ando K, Hombo T, Kanno A, Ikeda H, Imaizumi M, Shimizu N *et al.* (2005). QT PRODACT: *in vivo* QT assay with a conscious monkey for assessment of the potential for drug-induced QT interval prolongation. *J Pharmacol Sci* **99**: 487–500.
- Ator NA, Griffiths RR (2003). Principles of drug abuse liability assessment in laboratory animals. *Drug Alc Dep* **70**: S55–S72.
- Authier S, Legaspi M, Gauvin D, Chaurand F, Fournier S, Troncy E (2008). Validation of respiratory Safety Pharmacology models: conscious and anesthetized beagle dogs. *J Pharmacol Toxicol Methods* **57**: 52–60.
- Authier S, Tanguay JF, Gauvin D, Fruscia RD, Troncy E (2007). A cardiovascular monitoring system used in conscious cynomolgus monkeys for regulatory Safety Pharmacology. Part 2: pharmacological validation. *J Pharmacol Toxicol Methods* **56**: 122–130.
- Balser RL, Bigelow GE (2003). Guidelines and methodological reviews concerning drug abuse liability assessment. *Drug Alc Dep* **70**: S13–S40.
- Bass AS, Kinter L, Williams P (2004a). Origins, practices and future of safety pharmacology. *J Pharmacol Toxicol Methods* **49**: 145–151.
- Bass AS, Vargas HM, Kinter LB (2004b). Introduction to Nonclinical safety pharmacology and the safety pharmacology society. *J Pharmacol Toxicol Methods* **49**: 141–144.
- Bellemin-Baurreau J, Poizat A, Hicks PE, Armstrong JM (1994). An *in vitro* method for the evaluation of antiarrhythmic and anti-ischemic agents by using programmed electrical stimulation of rabbit heart. *J Pharmacol Toxicol Methods* **31**: 31–40.
- Billman GE (2006). A comprehensive review and analysis of 25 years of data from an *in vivo* canine model of sudden cardiac death: implications for future anti-arrhythmic drug development. *Pharmacol Ther* **111**: 808–835.
- Botting JA, Curtis MJ, Walker MJA (1985). Arrhythmias associated with myocardial ischaemia and infarction. *Mol Aspects Med* **8**: 307–422.
- Chaves AA, Keller WJ, O'Sullivan S, Williams MA, Fitzgerald LE, McPherson HE *et al.* (2006). Cardiovascular monkey telemetry: sensitivity to detect QT interval prolongation. *J Pharmacol Toxicol Methods* **54**: 150–158.
- Chaves AA, Zingaro GJ, Yordy MA, Bustard KA, O'Sullivan S, Galijatovic-Idrizbegovic A *et al.* (2007). A highly sensitive canine telemetry model for detection of QT interval prolongation: studies with moxifloxacin, haloperidol and MK-499. *J Pharmacol Toxicol Methods* **56**: 103–114.
- Chung PH, Pugsley MK, Walker MJA (1993). Arrhythmia models in the rat. *J Pharmacol Toxicol Methods* **29**: 179–184.
- Chu TF, Rupnick MA, Kerkela R, Dallabrida SM, Zurakowski D, Nguyen L *et al.* (2007). Cardiotoxicity associated with tyrosine kinase inhibitor sunitinib. *Lancet* **370**: 2011–2019.
- Curtis MJ (1998). Characterisation, utilisation and clinical relevance of isolated perfused heart models of ischaemia-induced ventricular fibrillation. *Cardiovasc Res* **39**: 194–215.
- Darpo B (2001). Spectrum of drugs prolonging QT interval and the incidence of Torsades de pointes. *Eur Heart J* **3** (Suppl. K): K70–K80.
- Darpo B, Nebout T, Sager PT (2006). Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline. *J Clin Pharmacol* **46**: 498–507.
- Demetri GD, van Oosteroum AT, Garrett CR, Blackstein ME, Shah MH, Verweij J *et al.* (2006). Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomized controlled trial. *Lancet* **368**: 1329–1338.
- Durmuller N, Guillaume P, Lacroix P, Porsolt RD, Moser P (2007). The use of the dog electroencephalogram (EEG) in Safety Pharmacology to evaluate proconvulsant risk. *J Pharmacol Toxicol Methods* **56**: 234–238.
- EMA/CPMP/SWP (2006). Guideline on the non-clinical investigation of the dependence potential of medicinal products. EMA/CPMP/SWP/94227, 23 March 2006.
- Eschaliere A, Lavarenne J, Burtin C, Renoux M, Chapuy E, Rodriguez M (1988). Study of histamine release induced by acute administration of antitumor agents in dogs. *Cancer Chemother Pharmacol* **21**: 246–250.
- Farkas A, Curtis MJ (2002). Limited effectiveness of clinically-relevant concentrations of Class I antiarrhythmics in isolated perfused rat hearts. *J Cardiovasc Pharmacol* **39**: 412–424.
- Feltenstein MW, See RE (2008). The neurocircuitry of addiction. *Br J Pharmacol* advance online publication 3 March; doi: 10.1038/bjp.2008.51.
- Friedrichs GS, Patmore L, Bass AS (2005). Non-clinical evaluation of ventricular repolarization (ICH S7B): results of an interim survey of international pharmaceutical companies. *J Pharmacol Toxicol Methods* **52**: 6–11.
- Garber K (2006). The second wave in kinase cancer drugs. *Nature Biotechnol* **24**: 127–130.
- Hamlin RL (2007). Animal models of ventricular arrhythmias. *Pharmacol Ther* **113**: 276–295.
- Hamlin RL, Cruze CA, Mittelstadt SW, Kijawornrat A, Keene BW, Roche BM *et al.* (2004). Sensitivity and specificity of isolated perfused guinea pig heart to test for drug-induced lengthening of QTc. *J Pharmacol Toxicol Methods* **49**: 15–23.
- Hauser DS, Stade M, Schmidt A, Hanauer G (2005). Cardiovascular parameters in anaesthetized guinea pigs: a Safety Pharmacology screening model. *J Pharmacol Toxicol Methods* **52**: 106–114.
- Health Canada (2006). Draft Guidance for Industry: clinical assessment of abuse liability for drugs with central nervous system activity. Health Canada Publications 02 January 2006.
- Hondeghem LM, Lu HR, van Rossem K, De Clerck F (2003). Detection of proarrhythmia in the female rabbit heart: blinded validation. *J Cardiovasc Electrophysiol* **14**: 287–294.
- Irwin S (1968). Comprehensive observational assessment: Ia. A systematic, quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia* **13**: 222–257.
- Johanson CE (1990). The evaluation of the abuse liability of drugs. *Drug Saf* **5**: 46–57.
- Johnston KM, MacLeod BA, Walker MJA (1983). Responses to ligation of a coronary artery in conscious rats and the actions of antiarrhythmics. *Can J Physiol Pharmacol* **61**: 1340–1353.
- June RA, Nasr I (1997). Torsades de pointes with terfenadine ingestion. *Am J Emerg Med* **15**: 542–543.
- Kagstrom J, Sjogren EL, Ericson AC (2007). Evaluation of the guinea pig monophasic action potential (MAP) assay in predicting drug-induced delay of ventricular repolarisation using 12 clinically documented drugs. *J Pharmacol Toxicol Methods* **56**: 186–193.
- Kamiya K, Niwa R, Mitcheson JS, Sanguinetti MC (2006). Molecular determinants of HERG channel block. *Mol Pharmacol* **69**: 1709–1716.
- Kerkala R, Grazette L, Yacobi R, Iliescu C, Patten R, Beahm C *et al.* (2006). Cardiotoxicity of the cancer therapeutic agent imatinib mesylate. *Nat Med* **12**: 908–916.
- Khakoo AY, Kassiotis CM, Tannir N, Plana JC, Halushka M, Bickford C *et al.* (2008). Heart failure associated with sunitinib malate: a multitargeted receptor tyrosine kinase inhibitor. *Cancer*; e-pub ahead of print, doi 10.1002/cncr.23460.
- Kemp JP (1992). Antihistamines—is there anything safe to prescribe? *Ann Allergy* **69**: 276–280.
- Kien ND, Quam DD, Reitan JA, White DA (1992). Mechanism of hypotension following rapid infusion of protamine sulfate in anesthetized dogs. *J Cardiothorac Vasc Anesth* **6**: 143–147.
- Lawrence CL, Bridgland-Taylor MH, Pollard CE, Hammond TG, Valentin JP (2006). A rabbit Langendorff heart proarrhythmia

- model: predictive value for clinical identification of Torsades de Pointes. *Br J Pharmacol* **149**: 845–860.
- Lin C, Ke X, Cvetanovic I, Ranade V, Somberg J (2005). The influence of extracellular acidosis on the effect of IKr blockers. *J Cardiovasc Pharmacol Ther* **10**: 67–76.
- Lindgren S, Bass AS, Briscoe R, Bruse K, Friedrichs GS, Kallman M-J *et al.* (2008). Benchmarking Safety Pharmacology Regulatory Packages and Best Practice. *J Pharmacol Toxicol Methods*.
- Liu T, Brown BS, Wu Y, Antzelevitch C, Kowey PR, Yan GX (2006). Blinded validation of the isolated arterially perfused rabbit ventricular wedge in preclinical assessment of drug-induced proarrhythmias. *Heart Rhythm* **3**: 948–956.
- Luft J, Bode G (2002). Integration of safety pharmacology endpoints into toxicology studies. *Fundam Clin Pharmacol* **16**: 91–103.
- Marafino BJ, Pugsley MK (2003). Biotechnology-derived therapeutics in cardiovascular disease. *Cardiovasc. Toxicol* **3**: 5–12.
- Masini E, Planchenault J, Pezziardi F, Gautier P, Gagnol JP (1985). Histamine-releasing properties of Polysorbate 80 *in vitro* and *in vivo*: correlation with its hypotensive action in the dog. *Agents Actions* **16**: 470–477.
- Mattsson JL, Spencer PJ, Albee RR (1996). A performance standard for clinical and functional observational battery examinations of rats. *J Am Coll Toxicol* **15**: 239.
- Miyazaki H, Watanabe H, Kitayama T, Nishida M, Nishi Y, Sekiya K *et al.* (2005). QT PRODACT: sensitivity and specificity of the canine telemetry assay for detecting drug-induced QT interval prolongation. *J Pharmacol Sci* **99**: 523–529.
- Monahan BP, Ferguson CL, Killeavy ES, Lloyd BK, Troy J, Cantilena Jr LR (1990). Torsades de pointes occurring in association with terfenadine use. *JAMA* **264**: 2788–2790.
- Morrison AR, Ator NA, Nakamura RK (eds). (2002). Report of a National Institutes of Health Workshop. NIH Publication No. 02-5083, US Government Printing Office: Washington DC.
- Mow T, Arlock P, Laursen M, Ganderup NC (2008). Major ion current except  $I_{to}$  are present in the ventricle of Gottingen minipig heart. *J. Pharmacol Toxicol Methods*, in press.
- Morganroth J, Gussack L, (eds). (2005). *Cardiac Safety of Noncardiac Drugs: Practical Guidelines for Clinical Research and Drug Development*. Humana Press: New Jersey.
- Morton DB, Hawkins P, Bevan R, Heath K, Kirkwood J *et al.* (2003). Refinements in telemetry procedures. Seventh report of the BVAWF/FRAME/RSPCA/UFaw Joint Working Group on Refinement, Part A. *Lab Animal* **37**: 261–299.
- Morris LS, Carlson A (1998). Protecting patients from harm: Terfenadine and potential drug therapy interactions. *Drug Info J* **32**: 339–345.
- Moscardo E, Maurin A, Dorigatti R, Champeroux P, Richard S. (2007). An optimised methodology for the neurobehavioural assessment in rodents. *J Pharmacol Toxicol Methods* **56**: 239–255.
- Murphy DJ, Joran ME (1992). Respiratory and cardiovascular changes associated with toxic doses of a peptide antagonist of vasopressin in the rat. *Fundam Appl Toxicol* **18**: 307–313.
- Murphy SM, Palmer M, Poole MF, Padegimas L, Hunady K, Danzig J *et al.* (2006). Evaluation of functional and binding assays in cells expressing either recombinant or endogenous hERG channel. *J Pharmacol Toxicol Methods* **54**: 42–55.
- National Cancer Institute (2007). *NCI Fact Book*. Office of Budget and Finance, National Cancer Institute: Bethesda, MD, <http://obf.cancer.gov>.
- Ollerstam A, Visser SA, Duker G, Forsberg T, Persson AH, Nilsson LB *et al.* (2007). Comparison of the QT interval response during sinus and paced rhythm in conscious and anesthetized beagle dogs. *J Pharmacol Toxicol Methods* **56**: 131–144.
- Omata T, Kasai C, Hashimoto M, Hombo T, Yamamoto K (2005). QT PRODACT: comparison of non-clinical studies for drug-induced delay in ventricular repolarization and their role in safety evaluation in humans. *J Pharmacol Sci* **99**: 531–541.
- Patel SP, Campbell DL (2005). Transient outward potassium current, 'I<sub>to</sub>', phenotypes in the mammalian left ventricle: underlying molecular, cellular and biophysical mechanisms. *J Physiol* **569**: 7–39.
- Pratt CM, Camm AJ, Cooper W, Friedman PL, MacNeil DJ, Moulton KM *et al.* (1998). Mortality in the Survival With ORal D-sotalol (SWORD) trial: why did patients die? *Am J Cardiol* **81**: 869–876.
- Pratt CM, Hertz RP, Ellis BE, Crowell SP, Louv W, Moye L. (1994). Risk of developing life-threatening ventricular arrhythmia associated with terfenadine in comparison with over-the-counter antihistamines, ibuprofen and clemastine. *Am J Cardiol* **73**: 34652.
- Pugsley MK (2004). Safety pharmacology matures into a unique pharmacological discipline. *J Pharmacol Toxicol Meth* **49**: 137–139.
- Redfern WS, Strang I, Storey S, Hays C, Barnard C, Lawton K *et al.* (2005). Spectrum of effects detected in the rat functional observational battery following oral administration of non-CNS targeted compounds. *J Pharmacol Toxicol Meth* **52**: 77–82.
- Rees SA, Curtis MJ (1996). Which cardiac potassium channel subtype is the preferred target for suppression of ventricular arrhythmias? *Pharmacol Ther* **69**: 199–217.
- Rosen MR (1996). Of oocytes and runny noses. *Circulation* **94**: 607–609.
- Roy M-L, Dumanine R, Brown AM (1996). HERG, a primary human ventricular target of the non-sedating antihistamine terfenadine. *Circulation* **94**: 817–823.
- Sanguinetti MC, Mitcheson JS (2005). Predicting drug-hERG channel interactions that cause acquired long QT syndrome. *Trends Pharmacol Sci* **26**: 119–124.
- Sasaki H, Shimizu N, Suganami H, Yamamoto K (2005). QT PRODACT: inter-facility variability in electrocardiographic and hemodynamic parameters in conscious dogs and monkeys. *J Pharmacol Sci* **99**: 513–522.
- Schiefe RT, Cramer W (1996). Sedative and Cardiovascular Profile of Newer Antihistamines. *Consult Pharm* **10**: 37–45.
- Shah RR (2005). Drugs, QTc interval prolongation and final ICH E14 guideline: an important milestone with challenges ahead. *Drug Safety* **28**: 1009–1028.
- Shah RR (2008). If a drug deemed 'safe' in nonclinical tests subsequently prolongs QT in phase 1 studies, how can its sponsor convince regulators to allow development to proceed? *Pharmacol Therap* (in press). doi: 10.1016/j.pharmthera.2008.03.002.
- Shah RR, Hondeghem LM (2005). Refining detection of drug-induced proarrhythmia: QT interval and TRIaD. *Heart Rhythm* **2**: 758–772.
- Slack M, Kirchoff C, Moller C, Winkler K, Netzer R (2006). Identification of novel Kv1.3 blockers using a fluorescent cell-based ion channel assay. *J Biomol Screen* **11**: 57–64.
- Sorota S, Zhang XS, Margulis M, Tucker K, Priestley T (2005). Characterization of a hERG screen using the IonWorks HT: comparison to a hERG rubidium efflux screen. *Assay Drug Dev Technol* **3**: 47–57.
- Tashibu H, Miyazaki H, Aoki K, Akie Y, Yamamoto K (2005). QT PRODACT: *in vivo* QT assay in anesthetized dog for detecting the potential for QT interval prolongation by human pharmaceuticals. *J Pharmacol Sci* **99**: 473–486.
- Thomsen MB, Verduyn SC, Stengl M, Beekman JD, de Pater G, van Opstal J *et al.* (2004). Increased short-term variability of repolarization predicts d-sotalol-induced torsades de pointes in dogs. *Circulation* **110**: 2453–2459.
- Tontodonati M, Faselli N, Moscardo E, Giarola A, Dorigatti R (2007). A canine model used to simultaneously assess potential neuro-behavioural and cardiovascular effects of candidate drugs. *J Pharmacol Toxicol Methods* **56**: 265–275.
- Toyoshima S, Kanno A, Kitayama T, Sekiya K, Nakai K, Haruna M *et al.* (2005). QT PRODACT: *in vivo* QT assay in the conscious dog for assessing the potential for QT interval prolongation by human pharmaceuticals. *J Pharmacol Sci* **99**: 459–471.
- US Food and Drug Administration (1997). S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals Federal Register. **62**: 61515–61519.
- US Food and Drug Administration (2001). Guidance for industry: safety pharmacology studies for human pharmaceuticals (S7A). <http://www.fda.gov/cber/gdlns/ichs7a071201.htm>.
- US Food and Drug Administration (2005a). Title 21, part 58 good laboratory practise for non clinical laboratories. **12**: 301–303.
- US Food and Drug Administration (2005b). Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs (E14). <http://www.fda.gov/cber/gdlns/iche14qtc.pdf>.
- Valentin JP, Hoffmann P, De Clerck F, Hammond TG, Hondeghem L (2004). Review of the predictive value of the Langendorff heart

- model (Screenit system) in assessing the proarrhythmic potential of drugs. *J Pharmacol Toxicol Methods* **49**: 171–181.
- Vargas HM, Bass AS, Breidenbach A, Feldman HS, Gintant GA, Harner A *et al.* (2008). Scientific review and recommendations on preclinical cardiovascular safety evaluation of biologics. *J Pharmacol Toxicol Methods* (in press) doi: 10.1016/j.vascn.2008.05.003.
- Wang D, Patel C, Cui C, Yan G-X. (2008). Preclinical assessment of drug-induced proarrhythmias: role of the arterially perfused rabbit left ventricular wedge preparation. *Pharmacology and Therapeutics* (in press) doi:10.1016/j.pharmthera.2008.02.009.
- Weerts EM, Fantegrossi WE, Goodwin AK (2007). The value of non-human primates in drug abuse research. *Clin Exp Psychopharmacol* **15**: 309–327.
- Weiss JN, Garfinkel A, Karagueuzian HS, Qu Z, Chen PS (1999). Chaos and the transition to ventricular fibrillation a new approach to antiarrhythmic drug evaluation. *Circulation* **99**: 2819–2826.
- Yamaguchi M, Shimizu M, Ino H, Terai H, Uchiyama K, Oe K *et al.* (2003). T wave peak-to-end interval and QT dispersion in acquired long QT syndrome: a new index for arrhythmogenicity. *Clin Sci (London)* **105**: 671–676.
- Yap YG, Camm AJ (1999). Arrhythmogenic mechanisms of non-sedating antihistamines. *Clin Exp Allergy* **29**: 174–181.
- Zheng W, Spencer RH, Kiss L (2004). High throughput assay technologies for ion channel drug discovery. *Assay Drug Dev Technol* **2**: 543–552.