Workshop Report

Incorporating New Approach Methodologies into Regulatory Nonclinical Pharmaceutical Safety Assessment

Abstract

New approach methodologies (NAMs) based on human biology enable the assessment of adverse biological effects of pharmaceuticals and other chemicals. Currently, however, it is unclear how NAMs should be used during drug development to improve human safety evaluation. A series of 5 workshops with 13 international experts (regulators, preclinical scientists and NAMs developers) were conducted to identify feasible NAMs and to discuss how to exploit them in specific safety assessment contexts. Participants generated four 'maps' of how NAMs can be exploited in the safety assessment of the liver, respiratory, cardiovascular, and central nervous systems. Each map shows relevant end points measured, tools used (e.g., cells, assays, platforms), and highlights gaps where further development and validation of NAMs remains necessary. Each map addresses the fundamental scientific requirements for the safety assessment of that organ system, providing users with guidance on the selection of appropriate NAMs. In addition to generating the maps, participants offered suggestions for encouraging greater NAM adoption within drug development and their inclusion in regulatory guidelines. A specific recommendation was that pharmaceutical companies should be more transparent about how they use NAMs in-house. As well as giving guidance for the four organ systems, the maps provide a template that could be used for additional organ safety testing contexts. Moreover, their conversion to an interactive format would enable users to drill down to the detail necessary to answer specific scientific and regulatory questions.

1   Introduction

Extensive nonclinical safety studies are undertaken on new pharmaceuticals prior to and alongside clinical trials. Their purpose is to identify and understand the toxic effects of the compound in order to determine whether its anticipated benefit *versus* risk profile justifies clinical evaluation and, if so, to inform the design and monitoring of clinical studies. The nonclinical safety studies are mandated by regulatory guidelines and include a variety of safety pharmacology and toxicology investigations.

Safety pharmacology studies aim to determine whether pharmaceuticals cause on- or off-target effects on biological processes which can affect the function of critical organ systems (*e.g.*, cardiovascular, respiratory, gastrointestinal, and central nervous systems) and to assess potency, which is needed to assess safety margins *versus* human clinical drug exposure. Safety pharmacology studies also help inform the selection of follow-on investigations that can aid human risk assessment and may provide insight into mechanisms which underlie any effects that arise in humans. Multiple leading pharmaceutical companies (*e.g.*, AstraZeneca, GlaxoSmithKline, Novartis, and Pfizer) have outlined the advantages provided by *in vitro* safety pharmacological profiling, including early identification of off-target interactions and the prediction of clinical side effects that may be missed in animal studies, and have highlighted that these studies enable much more cost-effective and rapid profiling of large numbers of compounds than animal procedures (Bowes et al., 2012).

Toxicology studies evaluate systemic organ toxicities, behavioral effects, reproductive and developmental toxicology, genetic toxicology, eye irritancy and dermal sensitization. They include single and repeat dose studies in rodent and non-rodent animal species, which identify target organs, assess severity and reversibility, and define dose-response and no observed adverse effect levels. These are critical parameters which are essential for regulatory decision-making on whether the compound can be progressed into clinical trials and if so, estimation of a suitable starting dose, maximum dose, dose escalation regime, and any non-standard clinical safety monitoring that may be needed. Toxicity observed in nonclinical animal safety studies is an important cause of the high attrition rate of candidate drugs prior to clinical trials that occurs in multiple pharmaceutical companies (Cook et al., 2014). However, many drugs cause clinically serious adverse effects in humans which are not detected in animals (Bailey et al., 2015). For example, human drug induced liver injury (DILI), which is not detected in animal safety studies, is an important cause of attrition late in clinical development, failed licensing and/or of restrictive drug labelling (Watkins, 2011). Attrition due to toxicity observed in animals and/or in humans is an important cause of the high failure rate of clinical drug development (Cook et al., 2014; Watkins, 2011; Thomas et al., 2021).

New approach methodologies (NAMs) include methods which predict and evaluate biological processes by which pharmaceuticals may elicit desirable pharmacological effects and/or may cause undesirable toxicity. Many different types of NAMs have been described. These include simple *in vitro* cell-based tests, more complex organotypic or microphysiological systems (MPS)/organ-on-a-chip devices, and whole human tissues maintained *ex vivo*. Interpretation of the *in vivo* relevance of the data provided by these methods is complemented by computational tools which simulate and predict *in vivo* drug disposition and kinetics, in particular physiologically based pharmacokinetic (PBPK) models. Accurate *in vitro* to *in vivo* extrapolation is further aided by human low-dose testing and microdosing studies (phase 0 testing), which provide precise data on systemic human drug exposure and kinetics *in vivo*.

The use of NAMs during drug discovery and development has the potential to reduce, and ideally eliminate, toxicities that arise in animal safety studies and in humans. Pharmaceutical companies already are using NAMs in early drug development to support efficacy testing, and as screens to deselect unsuitable molecules prior to animal safety testing, with considerable success (Morgan et al., 2018). In 2017, the US Food and Drug Administration (US FDA) outlined its aim to improve predictivity and reduce the use of animals in toxicology by promoting the development, qualification, and integration of NAMs into regulatory science (US FDA, 2017). Three years later, it announced its 'Innovative Science and Technology Approaches for New Drugs' (ISTAND) program which enables developers of NAMs to interact with the US FDA in the early stages of the development process (US FDA, 2021a). In a key 2020 paper, the US FDA gave an overview of where NAMs are currently used in preclinical safety assessment and identified areas of unmet need where NAMs might deliver more predictive and productive methodologies, potentially improving and expediting drug development (Avila et al., 2020). It recommended that regulatory requirements and unanswered safety questions should be the starting point for the development of NAMs. However, at present there is no overarching consensus on how NAMs can be incorporated within regulatory guidelines. Regulatory agencies, pharmaceutical companies and non-governmental organizations have identified numerous hurdles to doing so (Piersma et al., 2018; Butler et al., 2017; Burgdorf et al., 2019) and guidance is needed to help surmount them. Such guidance could be especially valuable to academic researchers and small biotech start-ups that drive new NAM development yet may be unfamiliar with the precise needs of regulators and other end users (ICCVAM, 2018).

The nonclinical studies used for safety assessment are identified in the International Council for Harmonization (ICH) guidance M3(R2) (non-clinical safety) (EMA, 2021) and ICH S7A (safety pharmacology) (ICH, 2000), which are nonbinding and do not specify what type of study should be conducted. The US FDA has acknowledged that a relatively standard set of studies has evolved to assess the profile and safety of a drug before it proceeds to first in human (FIH) trials, some of which (*e.g.*, FIH dose selection) arguably perform better than others (*e.g.*, predicting drug induced liver injury) (Avila et al., 2020). While most of these studies involve animal testing, federal regulations state that pharmacology and toxicology data can also come from studies conducted *ex vivo, in silico* and *in vitro* (ICH, 2000; EMA, 2018; US FDA, 2018).

The aim of this workshop was to accelerate the transition to inclusion of NAMs within an integrated, human-relevant pharmaceutical safety assessment strategy that is based on identified need, is scientifically valid and acceptable to industry and regulators. A specific objective was to propose a decision tree framework which aligns NAMs with the regulatory needs identified by the US FDA (Avila et al., 2020) and highlights where additional NAMs need to be, or currently are being, developed.

2   Methods

A number of international experts (regulators, preclinical scientists and NAM developers) were identified from the authors' professional networks, including the Alliance for Human Relevant Science[1] and were invited to participate in a series of workshops. It was stipulated that the workshops would focus on the safety testing of new pharmaceuticals only, and would not cover efficacy testing, the testing of medical devices or vaccines, chronic effects, carcinogenicity, reproductive toxicity, pediatric or special toxicity concerns. Those who accepted were asked to read several papers in advance (US FDA, 2017; Avila et al., 2020; Butler et al., 2017; EMA, 2021; Patterson et al., 2021; Andersen et al., 2019) and were directed to Table 2 in the 2020 US FDA paper (Avila et al., 2020) which presents the nonclinical safety assessment needs and was used as a starting point for discussions.

Five workshops took place in 2021, conducted online due to Covid-19 restrictions. The workshops focused on four, clinically important, organ systems concerning adverse drug reactions which are assessed routinely in nonclinical animal safety tests: cardiovascular, respiratory, liver, and central nervous system (CNS). For three of the organ systems (cardiovascular, respiratory, and CNS), adverse functional effects are assessed *via* the core battery of safety pharmacology tests described in ICH S7A. The primary focus was on *in vitro* NAMs due to the participants' expertise, although the additional value provided by computational NAM models also was recognized. Based on their knowledge, participants were specifically allocated to one of the four organ system groups. Each group considered: i) which currently available NAMs could be used to address safety pharmacology/regulatory needs, ii) what outcomes would be measured by these NAMs, iii) the strengths and weaknesses of each NAM, and iv) key gaps not addressed by the currently available NAMs.

Initially, an overall map illustrating the typical process of screening drug candidates was generated to establish the scope and context of the exercise. Participants then developed maps which summarized safety requirements for each organ system, acknowledging that these represented only a part of the overall nonclinical assessment that a drug candidate would typically undergo. Participants populated the maps with *in vitro* NAMs already in use and those that could be used with further development, while identifying areas where NAMs are still needed. Where necessary, participants continued to collaborate after the workshops to complete the maps.

3   Results

3.1 Workshops
Thirteen individuals participated in the workshops; 2 preclinical scientists, 5 NAM developers, 5 who were both preclinical scientists *and* NAMs developers (of whom 2 had extensive regulatory experience) and one regulator. Their affiliations (or previous affiliation in the case of one retiree) included the pharmaceutical industry, contract research, SME/ start-up,

---

[1] https://www.humanrelevantscience.org/

toxicology, drug safety, regulation, academia, and bioengineering. Three were female. One was from the US, two from mainland Europe and 10 from the UK.

3.2 Maps

Figure 1 provides an overview of the typical process of nonclinical safety and efficacy screening. Selected drug candidates are tested using computational methods and *in vitro* assays to address safety, kinetics (drug exposure) and efficacy. To predict efficacy and safety outcomes, effects observed in *in vitro* and *in vivo* assays are compared with *in vivo* human drug exposure which is predicted using computational PBPK tools which may include Quantitative Systems Toxicology (QST). These analyses inform decisions about whether the predicted benefit *versus* risk justifies progression to clinical trials and, if so, whether a bespoke safety monitoring plan is required.

The safety evaluation for each organ system (purple boxes) shown within the orange box in Figure 1 is detailed in Figures 2 - 5. Safety evaluation of the liver included drug-drug interaction (DDI) studies, for which the liver is the primary organ of concern.
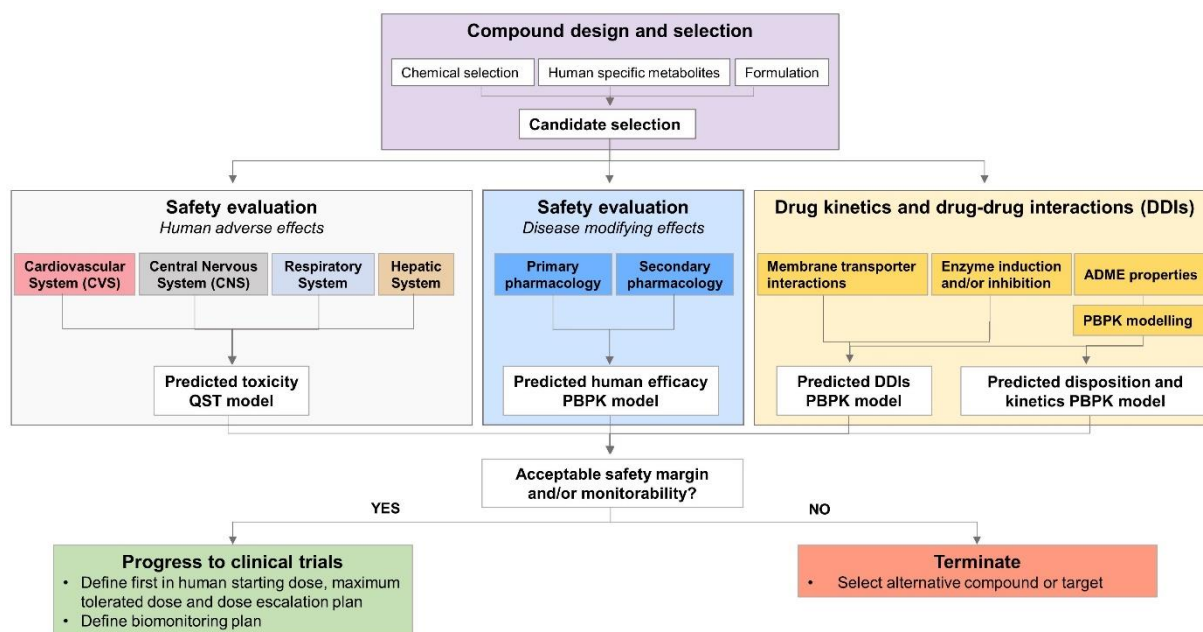

Fig. 1: Nonclinical testing overview
ADME, absorption, distribution, metabolism and excretion; PBPK, physiologically based pharmacokinetic; QST, quantitative systems toxicology

3.2.1 Central nervous system

For the central nervous system (CNS), ICH S7 specifies that the effects of a test substance on motor activity, behavioral changes, coordination, sensory/motor reflex responses and body temperature should be evaluated. Participants described NAMs which could assess motor activity, behavioral changes and blood-brain barrier integrity.

Participants noted that established models such as Ntera-2D1 cells (differentiated with post-mitotic cell product) have been used extensively in motor activity screening programs over the last 10-20 years to assess elementary functionality and network formation, so a good body of data supports their use. Consequently, they are advanced in terms of addressing regulatory requirements and display favorable characteristics (*e.g.*, viable for extended periods) which allow several measurements to be routinely gathered using cellular platforms such as high-throughput imaging as well as cell viability assays. The Ntera-2D1 platform can also generate functional astrocytes and neuronal cells which can be used in calcium imaging and electrophysiological studies to give an understanding of brain cortex impact, and therefore, give some indication of behavioral change under drug pressure (Hill et al., 2012; Woehrling et al., 2015). However, Ntera-2D1 cells typically take 6-8 weeks to grow, making them expensive to maintain, and genetic instability can also result in post-mitotic cells. Other, less sophisticated models such as SH-SY5Y cells can be differentiated relatively quickly (~ 1 week) into neuronal-like cells for functionality studies.

It was observed that blood-brain barrier (BBB) models using transformed cell lines (*e.g.*, hCMEC/D3) are well established and used to measure paracellular integrity and, therefore. BBB function. Although these cell lines exhibit only moderate paracellular tightness, they are applicable for these studies. Standard transporter assays, inhibitor studies and subcellular localization studies by immunofluorescent imaging, together with permeability measurements, provide an overall picture of BBB integrity. However, complexity was noted to be an issue with any model of the BBB, including the number and types of cells necessary. Regarding human brain endothelial cells, it was acknowledged that few cell lines with physiological paracellular tightness are available, leading to attempts to recapitulate the BBB physiology more closely using multicellular (*e.g.*, spheroids) and microfluidic (including shear stress) models. However, participants noted that these models are still under development and not yet standardized. Participants also observed that few primary human brain endothelial cell models achieve
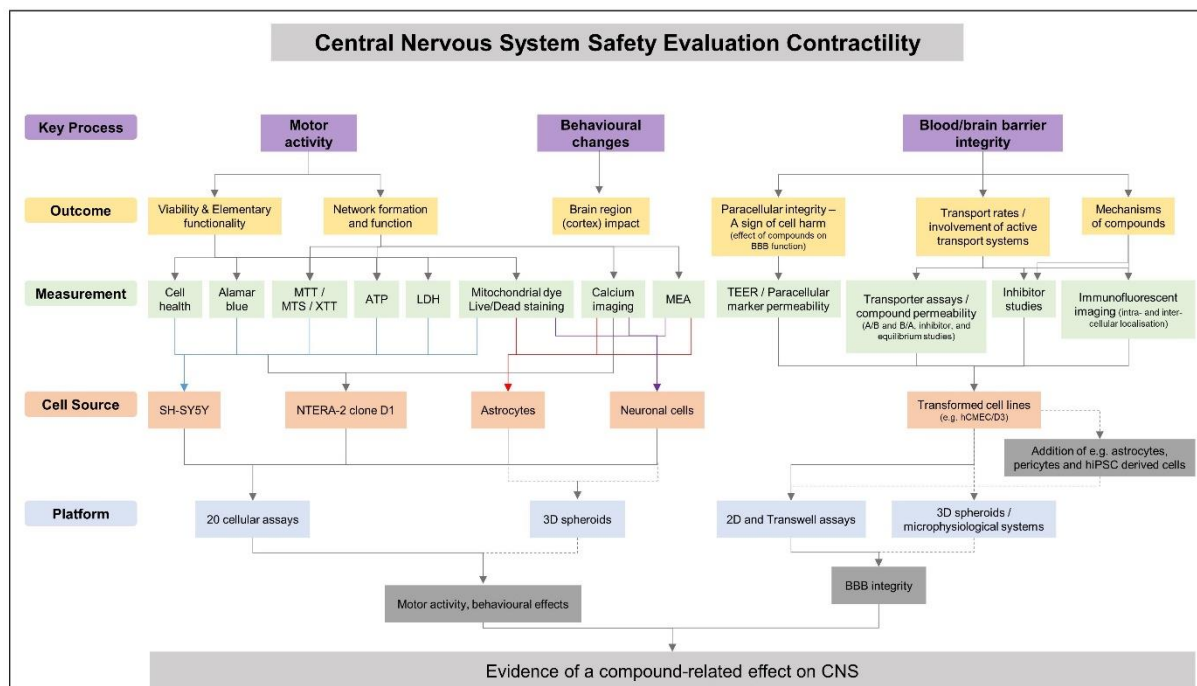
Fig. 2: Central nervous system map
Participants highlighted key functional processes (purple boxes) identified by regulators and assay outcomes (yellow boxes) which could address those processes. Assay outputs, using the NAM platforms (blue boxes), cell types (orange boxes) and measurements (green boxes), are integrated to provide evidence of the compound related effect on each system. Boxes with dotted lines indicate current/new developments in progress or required.
2D, 2-dimensional; 3D, 3-dimensional; A/B and B/A, "A-to-B" direction and "B-to-A" direction; ATP, adenosine triphosphate; BBB, blood-brain barrier; CNS, central nervous system; hCMEC, human cardiac microvascular endothelial cells; hiPSC, human induced pluripotent stem cells; LDH, lactate dehydrogenase; MEA, microelectrode/multielectrode array; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; MTT, 3:(4,5: dimethylthiazol:2:yl):2,5:diphenyltetrazolium bromide; TEER, transepithelial/transendothelial electrical resistance; XTT, (2,3:bis:(2:methoxy:4:nitro:5:sulfophenyl):2H:tetrazolium:5:carboxanilide

a paracellular tightness similar to the *in vivo* situation. Consequently, lower cost and more standardized transformed cells are currently preferred.

Participants noted that, while iPSC-derived neuronal cell types offer hope for the future of CNS studies, they currently suffer from lack of reproducibility and variable functionality depending on the differentiation protocols used. Nevertheless, it was acknowledged that they offer the possibility of long term and recovery studies since they can be retained in culture for many months and may also provide improved BBB models with superior paracellular integrity compared with current transformed cell line models.

Participants observed that the current *in vitro* methods alone cannot yet generate the 'higher level' data requested by regulators, such as sensory and motor reflex responses or whole-body temperature effects; although computational packages exploiting real world and/or clinical data, combined with NAMs data, are beginning to address these challenges.

3.2.2 Cardiovascular system

For the cardiovascular system, ICH S7 specifies that the effects of a test substance on blood pressure, heart rate, and the electrocardiogram should be evaluated. Participants identified NAMs which could address repolarization and conductance abnormalities, contractility and, to a very limited extent, vascular end points.

Participants noted that, while *in vitro* measurement of cardiac arrhythmia risk has traditionally employed hERG transfected cell lines (typically HEK293 and CHO), the development of human iPSC-derived cardiomyocytes (iPSC-CM) as well as platforms such as microelectrode arrays (Kanda et al., 2018) and automated patch clamp (Li et al., 2019) have enabled more extensive examination of multiple ion channel activities simultaneously. Co-culture of cardiomyocytes with cardiac fibroblasts or preparations of commercially available cells with mixed cell populations have also proved beneficial in many of these assays, although standardization remains a goal for many suppliers and users of these cells. It was acknowledged that reservations remain concerning the maturity and functionality of iPSC-CMs (Guo and Pu, 2020), with many cell types displaying fetal characteristics. This is already under consideration within the CiPA program[2].

The use of iPSC-derived engineered heart tissue was discussed as it has enabled more contractility applications to be developed for assessing cardiovascular effects (Lemme et al., 2018). Data from traditional platforms like Fluorescence Imaging Plate Reader (FLIPR) and high content imaging are still used along with those from MPS incorporating co-culture 3D models. It was considered that further advances with multi-organ models (*e.g.*, liver and heart (Ferrari and Rasponi, 2021)) will go some way towards creating a higher-level model of cardiovascular risk. Such *in vitro* models are also starting to reduce the need for
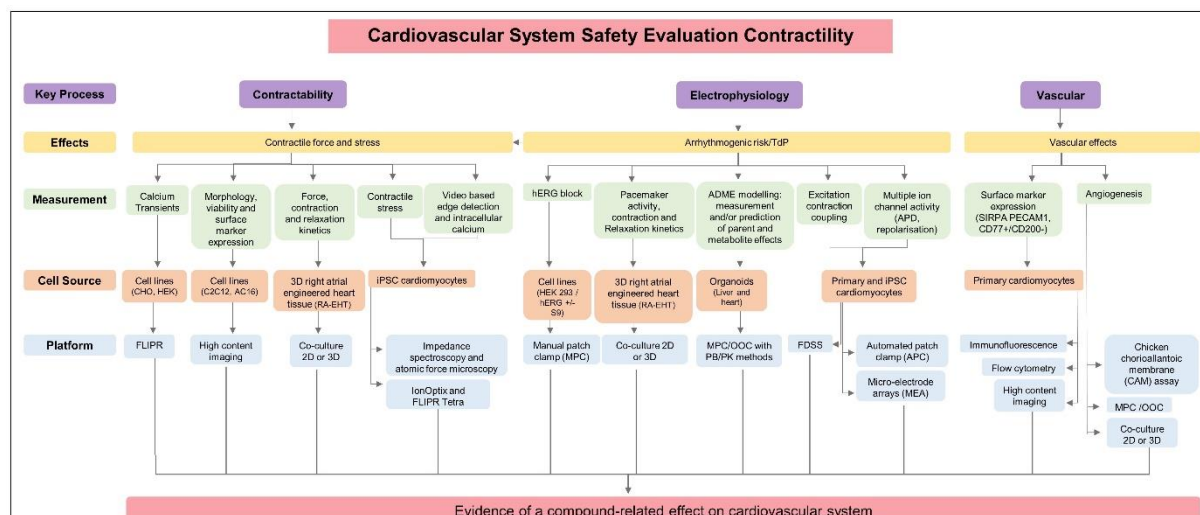
---

[2] https://cipaproject.org

Fig. 3: Cardiovascular map

Participants highlighted key functional processes (purple boxes) identified by regulators and assays that could address those processes. Assay outputs, using the NAM platforms (blue boxes), cell types (orange boxes) and measurements (green boxes), are integrated to provide evidence of the compound related effect on each system.

2D, 2-dimensional; 3D, 3-dimensional; ADME, absorption, distribution, metabolism and excretion; APD, action potential duration; CHO, Chinese hamster ovary; FDSS, functional drug screening system; FLIPR, fluorescent imaging plate reader; HEK, human embryonic kidney; hERG, human ether-à-go-go-related gene; iPSC, induced pluripotent stem cells; OOC, organ-on-a-chip; PECAM1, platelet and endothelial cell adhesion molecule 1; PBPK, physiologically based pharmacokinetic; SIRPA, signal regulatory protein alpha; TdP, torsades de pointes

clinical Thorough QT (TQT) studies and to provide insights into the monitoring required for late phase clinical trials[3] (ICH, 2018).

The absence of expertise needed to discuss i) contractility measurements with iPSC-CMs and ii) vascular effects was apparent from the limited population of these parts of the cardiovascular system map. Publications describing combined methods for measuring contraction with iPSC-CMs were highlighted (Pointon et al., 2015; Wang et al., 2020), but with the available expertise the panel members were unable to comment extensively on these approaches. Angiogenesis models were mentioned, and several technologies were suggested, although participants provided little detail. Methods and readouts for contractile force and stress, arrhythmogenic risk and direct effects on the vascular system of the heart were covered. However, it was acknowledged that further development of this map is required to provide a more comprehensive coverage of the numerous mechanisms by which drugs may adversely affect the cardiovascular system.

3.2.3 Respiratory system

For the respiratory system, ICH S7 specifies that the effects of a test substance on respiratory rate and other measures of respiratory function such as tidal volume and hemoglobin oxygen saturation should be evaluated. Participants identified NAMs that could address many aspects of respiratory function for regulatory purposes, including sensitization, inhalation toxicity, fibrosis, and irritation. Methods for determining correct exposure levels were also described.

Participants agreed that manifested physiological effects of respiratory function were often due to damage or changes at the cellular level which could be obtained as *in vitro* read-outs and when integrated, could provide an indication of overall respiratory function and/or toxicity. Notably, although effects such as CNS depression or effects on pulmonary vasculature would be absent from *in vitro* models, participants felt that the cellular endpoints described in this map and the NAMs associated with these pathways offered a more human-relevant assessment of respiratory toxicity than current *in vivo* animal measurements could provide. Nonetheless, it was acknowledged that measurements requested in current guidelines for tidal volume, respiratory rate *etc*. are only possible in whole organisms. Participants noted that *in vitro* measurement of epithelial cell health, immune response, cilia motion, -omics approaches and mucus secretion are well developed within the chemicals testing sector, where acute inhalation toxicity may be monitored using a variety of commercially available proprietary models such as SenzaGen's[4] GARD®air, Epithelix's[5] MucilAir™ and SmallAir™, MatTek's[6] EpiAirway™ and EpiAlveolar™, ImmuONE's[7] ImmuPHAGE™ and ImmuLUNG™ and Invitrolize's[8] ALISens™.

The use of mucus producing epithelial cells, or co-culture of human alveolar epithelial cells with pulmonary microvascular endothelial cells, enables cilia function, mucus production and transepithelial electrical resistance to be monitored, giving an indication of barrier function (CIPA, 2021). Moreover, it was felt that the use of patient/disease donor cells in this area (*e.g.*, from chronic obstructive pulmonary disease, asthma and cystic fibrosis patients) provides additional translational information and has enabled better characterization of these disease phenotypes and potential tailored treatments.

---

[3] https://cipaproject.org/about-cipa/#SteeringTeam

[4] https://senzagen.com/

[5] https://www.epithelix.com/

[6] https://www.mattek.com/

[7] https://immuone.com/
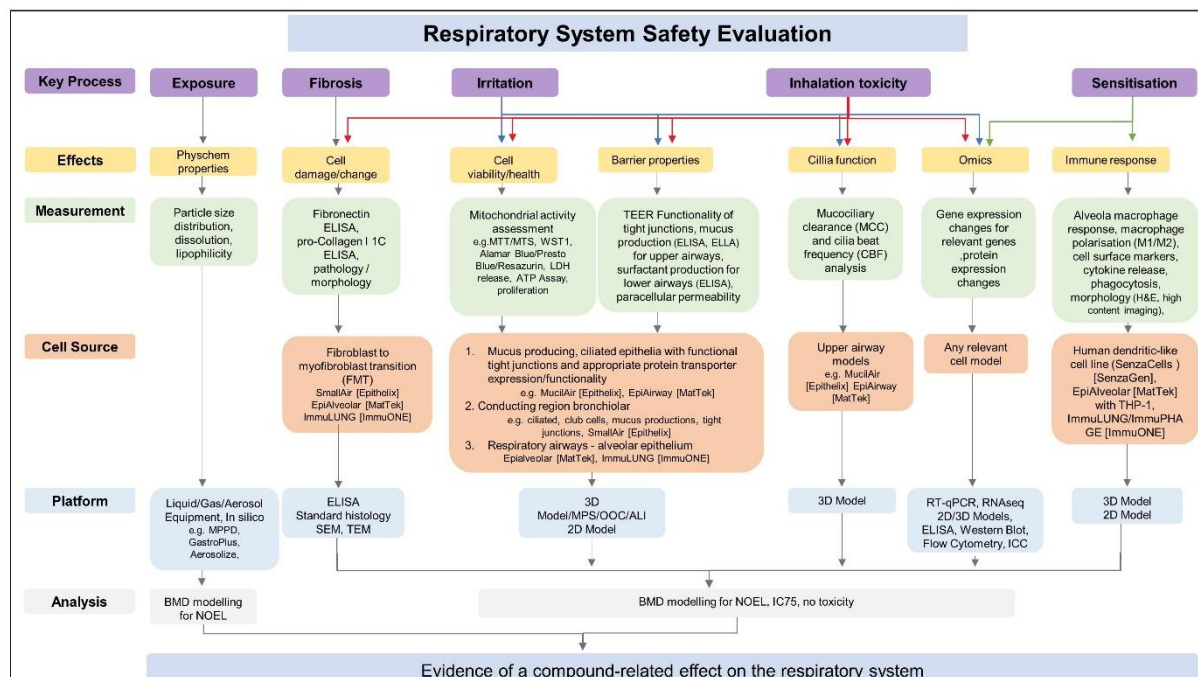
[8] https://invitrolize.com/

Fig. 4: Respiratory map
Participants highlighted key functional processes (purple boxes) identified by regulators and assays which could address those processes. Assay outputs, using the NAM platforms (blue boxes), cell types (orange boxes) and measurements (green boxes), are analyzed and integrated (white boxes) to provide evidence of the compound related effect on each system.

2D, 2-dimensional; 3D, 3-dimensional; ALI, air-liquid interface; ATP, adenosine triphosphate; BMD, benchmark dose; ELISA, enzyme-linked immunosorbent assay; ELLA, enzyme-linked lectin assay; H&E, hematoxylin and eosin; IC, inhibitory level; ICC, immunocytochemistry; LDH, lactate dehydrogenase; MPPD, multiple path particle dosimetry; MPS, microphysiological system; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; MTT, 3:(4,5: dimethylthiazol:2:yl):2,5:diphenyltetrazolium bromide; NOEL, no observable effect level; OOC, organ-on-a-chip; RT-qPCR, reverse transcription quantitative real-time polymerase chain reaction; SEM, scanning electron microscopy; TEM, transmission electron microscopy; WST1, 4-[3-(4-Iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio]-1,3-benzene disulfonate

One participant commented that the US FDA wants co-culture 3D models with a pathology output and biomarker support for safety assessment to predict an animal or clinical result (Clippinger et al., 2018). He also suggested that good markers as surrogates for the respiratory response are needed. 3D models offer many advantages over traditional 2D models because they can be maintained for months and used to study longer term exposure, multiple dosing and recovery. They also may be implemented with other tissue models in multi-organ chips, to investigate systems biology. However, as with other models that include multiple cultured cell types, it was felt that quality control and genetic stability were necessary for their widespread adoption.

A major challenge identified for respiratory safety testing was that exposure and cell response *in vivo* are affected markedly by the complexity of the lung. Although relevant cell types for co-culture are becoming more established, instrumentation to generate exposure conditions which mimic those that occur *in vivo* are still being developed and only a few are commercially available (Sadler et al., 2011; Primavessy et al., 2021). Greater use of computational models (*e.g.*, Multiple-Path Particle Dosimetry) or computational fluid dynamics was suggested (Kuprat et al., 2021; Corley et al., 2021; Su et al., 2020; Tsega, 2018) to determine the benchmark dose (BMD) or no-observed effect level (NOEL) and was highlighted as an area ripe for research. One approach accepted by the US Environmental Protection Agency (US EPA) involves computational fluid dynamics, particle size, MucilAir™ and BMD modelling, resulting in an EPA waiver for the 90-day rat inhalation toxicity test (Hardy et al., 2017; Varewyck and Verbeke, 2017; LASA and NC3Rs, 2009; US EPA, 2016).

3.2.4 Liver
ICH S7 suggests that the effects of a test substance on organ systems not investigated as part of the core battery should be assessed when there is a reason for concern. Assessment of DILI was regarded by participants as an established part of the safety testing of all drugs, with the liver being a relatively frequent toxicity target organ in *in vivo* animal safety. The liver is also the primary organ involved in induction and/or inhibition of compound metabolism, the generation of metabolites which may be toxic to the liver itself or to other organs, and any potential DDIs.

Many different liver NAMs have been described, and hepatoxicity testing was regarded by participants as an established part of the nonclinical safety testing undertaken by many pharmaceutical companies. However, currently there is no consensus on which of these should be used during drug discovery and development to reduce risk of *in vivo* liver toxicity (Fernandez-Checa et al., 2021). To address this gap, NAMs which address key biological processes by which hepatotoxicity and DDIs may arise, and which are suitable for inclusion in for regulatory safety testing, were discussed and proposed.

The most clinically concerning consequence of DILI is acute liver failure, which arises following damage to a substantial fraction of hepatocytes (Fontana, 2008). Participants felt that current human primary hepatocyte (PHH) NAMs capture most of the endpoints required to assess toxicity to hepatocytes. A particular advantage of human hepatocytes is that
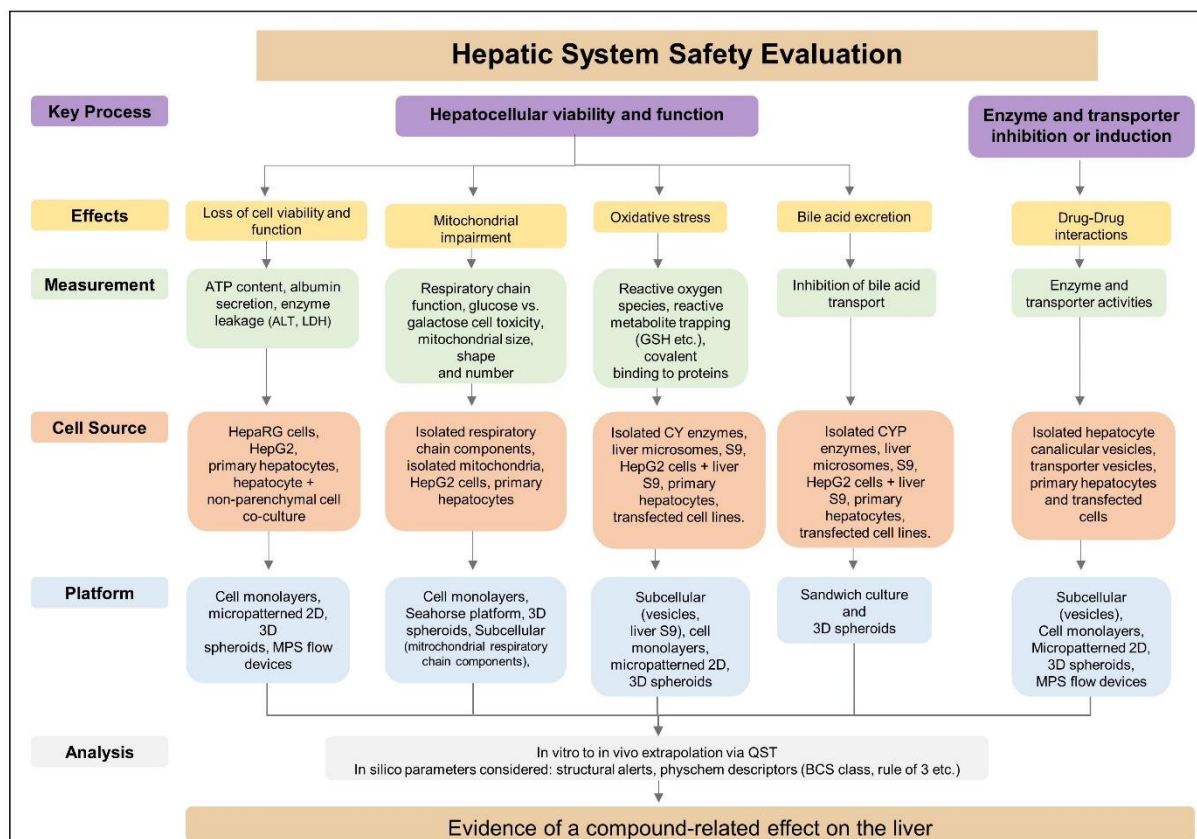
Fig. 5: Liver map
Participants highlighted key functional processes (purple boxes) identified by regulators and industry, and assays to address those processes. Assay outcomes (yellow boxes), using the NAM platforms (blue boxes), cell or preparation types (orange boxes) and measurements (green boxes), are integrated (white boxes) into QST platforms to provide evidence of the compound related effect on each system.
2D, 2-dimensional; 3D, 3-dimensional; ALT; alanine transaminase; ATP, adenosine triphosphate; BCS, biopharmaceutics classification system Seahorse™; CYP, cytochromes P450; GSH, glutathione; LDH, lactate dehydrogenase; MPS, microphysiological system; QST, quantitative systems toxicology

they avoid species variability in expression of liver drug metabolizing enzymes and transporter proteins, which is an important limitation of animal safety and DDI studies. In addition, studies undertaken using individual donor hepatocytes enable evaluation of the impact of inter-individual variability in drug metabolism and drug transporter function, which play a key role in many DDIs. However, many liver NAMs use hepatocytes pooled from multiple donors, which reduces or eliminates the impact of this variability. Due to the limited availability of PHHs, many companies instead undertake cellular toxicity testing of human liver-derived cell lines (*e.g.*, HepG2, HepaRG). When combined with information such as structural alerts and physico-chemical properties of compounds and the use computational methods such as QST to support *in vitro* to *in vivo* extrapolation, human liver cell derived NAM data have been successfully used to profile human DILI caused by numerous drugs (Smith et al., 2020). It was further noted that the use of liquid chromatography–mass spectrometry to detect and quantify metabolite formation has provided valuable insights into the role of biotransformation in DILI.

In conventional 2D culture, plated PHHs lose many of their hepatic phenotypes and functions within hours (Lauschke et al., 2016a). To overcome these limitations, various 3D culture methods, including spheroids, micropatterned co-cultures (MPCC) and liver MPS have been developed which maintain PHH viability and functionality for weeks (Smith et al., 2020; Lauschke et al., 2016b, 2019; Lin et al., 2016; Zhang et al., 2020). Participants noted that such 3D liver systems are now generally accessible and allow dose responses and some longer-term measurements to be performed. A recent non-peer reviewed report of *in vitro* toxicity testing of 27 hepatotoxic and non-hepatotoxic test drugs in a proprietary MPS platform claimed that it identified drugs that caused human DILI with very high specificity (97%) and high sensitivity (80%) (Ewart et al., 2021). It will be important to determine whether this performance can be replicated using larger numbers of drugs, and also whether similarly impressive results can be obtained using other liver MPS platforms.

In principle, liver NAMs which maintain viability and function for many weeks might be used to explore reversibility of effects. When undertaking reversibility studies, models will need to be monitored regularly to ensure that no deterioration or dramatic changes occur during the study period. Cell line proliferation in long term culture, and the conditions in which cells are maintained, were both highlighted as potential confounding variables to be considered when interpreting findings from such studies.

The level of complexity of the various liver NAMs was discussed at length. Participants commented that compared to liver cell suspensions or 2D culture models, more complex liver systems such as MPS, MPCC and spheroids have the advantage of longer viability periods (~6 weeks), enabling effects such as fibrosis and cholestasis to be observed through interactions between co-cultured non-parenchymal cells (Ware et al., 2015). Many of the MPS catalogued by the North

7

American 3Rs Collaborative[9] focus on the liver, although at the time of writing companies operating in 19 different tissue areas were identified.

It has been reported that iPSC technology enables the development of hepatocyte-like cells which can be used to explore variability between humans in susceptibility to drug induced liver injury *in vitro* (Choudhury et al., 2017; Takayama et al., 2014; Ouchi et al., 2019). However, participants noted that such individual donor cells may not be suitable for use in all platforms, potentially resulting in non-fully functioning liver models.

When designing and interpreting human liver NAM studies, participants noted that concentrations of drugs and drug metabolites which are formed within human liver *in vivo* in humans need to be considered. This is challenging since drug exposure within liver cells *in vivo* may be markedly higher than drug exposure in blood.

The need to evaluate differences between transient and persistent effects was also discussed. Participants cited troglitazone, which was withdrawn from the US market after deaths and severe liver failure in patients treated with the drug for prolonged intervals (≥ 1 month). It has been proposed that activation of adaptive immune responses could play an important role in severe liver injury caused by troglitazone, and by many other drugs. The currently available liver NAMs do not assess activation of adaptive immune responses. This gap was considered by the participants to be an additional important limitation of the proposed liver map.

3.3 Overall themes
Several themes recurred in the workshops, which are presented briefly here and summarized in the supplementary file[10].

3.3.1 Advantages of NAMs
Participants highlighted that NAMs have been used retrospectively to identify toxicity that animal studies fail to detect, citing examples of where *in vivo* toxicity was predicted *in vitro* (Dirven et al., 2021; Balogh Sivers et al., 2018)*;* but agreed that the challenge now was to apply NAMs prospectively to predict toxicity, citing an example of where this had already been achieved (Smith et al., 2020; Watkins, 2020). Several participants commented that US regulators have expressed confidence in NAMs, with many being used in Good Laboratory Practice toxicology tests and accepted by the EPA. Participants noted that the US FDA uses modelling software for evaluating drug safety, including Simcyp™ and GastroPlus®, neither of which underwent US FDA qualification, prompting them to question whether formal validation of NAMs was necessary.

Participants reported many advantages of NAMs, including their ability to highlight mechanisms of toxicity (Woodhead et al., 2017), represent human variability, provide information about the efficacy and safety of compounds (using diseased tissue models), investigate the impact of compounds over longer, chronic durations, evaluate reversibility (using complex cell models), and rapidly screen many different types of compounds for their adverse effects on target organs.

3.3.2 Factors discouraging the uptake of NAMs
Concerns were raised about the quality of cells used for *in vitro* studies, including the possibility of genetic drift, batch-to batch differences, problems with some culture media, and specific concerns relating to iPSCs. Several participants commented that NAMs are still unable to generate some information about higher level endpoints, or to represent exposure which occurs in the whole organism. However, it was felt that models such as computational fluid dynamic or PBPK computations are increasingly able to address this challenge, especially alongside the MPS revolution. Participants emphasized the need for platform designers to develop NAMs tailored to regulatory requirements.

Some participants reported that regulators were sometimes unclear about what data they required, and that it was unhelpful to be asked to submit data without being told why it was necessary up front. Others observed that, while several animal tests conducted in the early stages of drug development could now be replaced with *in vitro* methods, some companies lacked the confidence to jettison them. Conversely, it was noted that several pharmaceutical companies (*e.g.*, AstraZeneca (US FDA, 2020) now routinely use *in vitro* tests in their preliminary in-house studies to screen and triage compounds. Participants emphasized, however, that without more transparency about this, relatively few researchers and companies realize the opportunities provided by the *in vitro* methods. Some participants commented that using NAMs was risky because they could be rejected by regulators in favor of *in vivo* tests, resulting in lost revenue and delay. The cost of some NAMs, and the lack of a comprehensive database of positive and negative test compounds, were also felt to discourage their uptake.

3.3.3 Factors likely to increase uptake of NAMs
Participants highlighted the advantages of collaborating with stakeholders, citing the examples of previous genotoxicity working groups (European Environmental Mutagenesis and Genomics Society[11]; Health and Environmental Sciences Institute, 2021[12]; Gocke et al., 2000) and the Comprehensive *in vitro* Proarrhythmia Assay (CiPA) Initiative (CIPA, 2021), and suggested that talking to decision makers within regulatory agencies prior to using NAMs was advisable. A collaborative approach was also considered useful in the context of standardization and validation. Generally, it was felt that engaging with other disciplines allowed more scientists to learn about the capabilities of NAMs, and that much could be learned from other regulated sectors.

One participant commented that it would be helpful if NAMs were employed to assess developmental and reproductive toxicology (DART) prior to FIH. Application of NAMs to the assessment of DART was not discussed within the workshop and is a major focus of activity by other researchers. To increase the familiarity of regulators with NAMs, another participant suggested that it would be helpful for companies to describe in-house *in vitro* tests which they have used when making their regulatory submissions. For the same reason, participants supported the idea of submitting *in vitro* data alongside *in vivo* data. A bridging approach was mentioned in this context, whereby data are correlated between animal *in vivo* and animal

---

[9] https://www.na3rsc.org/mps-tech-hub/
[10] doi:10.14573/altex.2212081s
[11] https://www.eemgs.eu/affiliated-groups/
[12] https://hesiglobal.org/genetic-toxicology-gttc/

in vitro, between animal *in vitro* and human *in vitro*, and finally between human *in vitro* and human *in vivo* (NASEM, 2021). Uncertainty was expressed about how regulators would choose between the results of *in vitro* and *in vivo* experiments in the likely event that they produced conflicting results. A participant responded that the concern of regulators was to make a judgement about human risk, not to compare predictive ability.

Other suggestions included rewording ICH M3 to make the guidance more flexible (currently, it was felt to convey that animal studies were the default option), agreeing a basic battery of cellular tests which could quickly identify compounds with high intrinsic toxicity (although it was recognized that some companies might not welcome a prescriptive approach), and making more NAMs commercially available. Some noted that, while much time was now being spent on validating NAMs, the known strengths and weaknesses of animal tests were receiving less attention. Participants felt that there was now compelling evidence that the replacement of some poorly performing animal tests would result in fewer clinical trial failures for safety reasons.

4   Discussion

These workshops brought together a diverse group of scientists to develop maps of NAMs for four vital organ systems. Greater regulatory input would have benefited the discussions and confirmed whether the NAMs included in the maps are able to address the needs identified by the US FDA (Butler et al., 2017). Only one of ten invited regulators accepted the invitation and there was no representation from large pharmaceutical companies. Representation from the computational toxicology community was also limited and it is acknowledged that not all relevant computational applications are likely to have been included in the maps. Furthermore, it is recognised that the maps are not a complete framework for addressing all the nonclinical safety concerns for new pharmaceuticals.

The workshops identified general, dynamic issues affecting organ testing systems that are difficult to represent in a static format. First, acute and chronic toxicity may be expressed as different end points and, therefore, distinct methods may be required for each, but different time points cannot easily be mapped in a static format. Second, while a battery of assays may be used that encompasses multiple steps of a toxicity pathway to address questions typically derived from *in vivo* animal studies, it is difficult to represent each component in map form. Third, it was difficult to capture nascent technologies that are not yet commercially or widely available. These issues highlight the need for a more dynamic means of representing this information, such as an interactive decision tree that users could interrogate to answer specific questions.

Understanding exposure levels without employing *in vivo* animal methods was considered problematic. While an increasing number of computational models exist, some felt that the link between exposure and cell response was missing in many areas. Nevertheless, many computational models are now able to use *in vitro* and human clinical data to successfully determine effective dose levels and regimes for new or next in class molecules (ICH, 2000, 2009; Ferrari and Rasponi, 2021; Bai et al., 2020; International Programme on Chemical Safety and World Health, 2010; Busquet et al., 2020), making many animal studies redundant when undertaking dose prediction. Some participants observed that it was not yet possible to generate data on full system effects and higher-level responses using NAMs. It was noted that animal studies are likewise unable to provide reliable data on such responses and that such data can only be reliably obtained after Phase II trials. Furthermore, combining computational technologies with *in vitro*, existing *in vivo*, and clinical data, makes it possible to model a whole system response to replace the use of animals (US FDA, 2021b).

It was unequivocally acknowledged that a number (or panel) of integrated *in vitro* and/or computational assays will be required to address safety needs which currently are unmet by animal studies. The described NAMs may be used in this way, to measure cellular endpoints relating to the biological pathways involved in toxicity and pharmacological perturbations. This is the basis of the Adverse Outcome Pathways approach, which describes the sequence of molecular and cellular events necessary to produce a toxic response, as well as appropriate testing strategies for each event at multiple levels of biological complexity. The workshops identified gaps in assay endpoints, which when addressed should provide a better understanding of investigational drug effects.

The discussions exposed that traditional industry disciplines and their somewhat siloed nature (*e.g.*, safety pharmacology, primary pharmacology, secondary pharmacology, general toxicity *etc*.) may result in the unnecessary duplication of animal experiments. At present, no body oversees drug safety testing practices, although sponsors and regulators could identify where tests required for one discipline (*e.g.*, safety pharmacology) might also be used for another (*e.g.*, general toxicology), to prevent duplication and fulfil some of their 3Rs programmes objectives. It may be that the current classification of studies within traditional primary, secondary or safety pharmacology fields may no longer be appropriate as NAMs continue to uncover, for example, mechanistic data or mode of action information that traverse conventional disciplines.

An important insight was that NAMs are highly valued within pharmaceutical companies and often routinely used in-house in the early stages of drug development, but that this is not widely known. Greater transparency would increase awareness among the broader scientific community about the value of NAMs in this context. A requirement for companies to specify in their submissions to regulators which NAMs they use in-house would also familiarise regulators with NAMs. Commercial protection would be necessary, but agencies could omit the test compound, formulation information and other commercial details, and the sponsor could choose whether to be identified. Another reason for familiarizing regulators with NAMs is the possibility that the lack of guidance on *in vitro* methods in ICH M3 may reflect regulators' lack of expertise in these methods. If regulators are unable to recommend specific NAMs, guidelines will remain rooted in existing animal-based methodologies.

Regulator-sponsor collaboration prior to submission about the NAMs to be used would give all parties an understanding of the robustness, reproducibility, context of use and applicability of specific NAMs. The ISTAND program (Piersma et al., 2018) now makes this possible, enabling developers of NAMs to communicate with the US FDA in the early stages of the development process. A notable example is the evaluation of liver MPS and spheroid reproducibility for drug hepatotoxicity and metabolism studies (Rubiano et al., 2021). Such initiatives give regulators a greater acquaintance with

NAMs, enabling them to make upfront recommendations to scientists about the sort of *in vitro* tests that might be done, and to update guidelines where necessary. The use of facilitator groups such as the Health and Environmental Sciences Institute can also be successful in uniting stakeholders, enabling regulators to learn first-hand about the advantages, limitations, and context of use of NAMs, as with the CiPA initiative. The International Consortium for Innovation and Quality in Pharmaceutical Development has published pharmaceutical companies' preferred characteristics and features for MPS to guide system developers, regulators, and end users (Baudy et al., 2020; Fowler et al., 2020; Peterson et al., 2020; Pointon et al., 2021).

While human patient-derived cells retaining their inherent genotypic and phenotypic variability were considered invaluable for understanding heterogeneity in response to treatments, and the manifestation of and susceptibility to, disease (Chioccioli et al., 2019; Nantasanti et al., 2016; Bartfeld and Clevers, 2017; Kennedy et al., 2019; Yin et al., 2020; Liu et al., 2021), it was acknowledged that quality concerns need addressing if *in vitro* tests are to achieve wider acceptance. Guidelines such as Good Cell and Tissue Culture Practice 2 (Pamies et al., 2020). now include updated chapters on 3D culture, MPS, genetically modified cells and pluripotent stem cells. NAMs also need to be fit for purpose (Parish et al., 2020) and properly reported (Bracher et al., 2020). On a positive note, agency working groups set up to agree performance criteria for NAMs are already in place (EMA, 2021; US FDA, 2021c).

It was agreed that complex models using 3D culture may provide important information relating to system responses. However, low throughput, high cost and the ability to measure these responses are all challenges that need to be addressed before wider adoption is possible. Andersen et al., 2019 have described tiered application scenarios for NAMs in the risk assessment of chemicals and stratified them depending on their complexity and context of use, providing a potential framework for including more complex models within an overall testing strategy.

While some companies appear hesitant about investing in NAMs due to perceived financial risks, clearly others are already investing in them heavily for the early screening of compounds. The acceptance of those screens as direct replacements in the regulatory environment would make the drug discovery process cheaper, faster and more reliable. A return-on-investment model is needed to fully understand the financial opportunities offered by NAMs in drug discovery. As an example, for MPS, the reduction in total R&D costs for pharmaceutical companies is estimated to amount to 10–26% (~$700M) per new drug (Franzen et al., 2019).

5   Conclusion

In this workshop, a strategy for NAM use in 4 key areas of nonclinical pharmaceutical safety testing has been presented. Comprehensive organ system maps, such as those documented within this article, will provide stakeholders with guidance on, and confidence in, the NAMs that may be used to complement and ultimately replace *in vivo* animal methods. Converting these maps into interactive decision trees would allow users to ask specific questions and select the most appropriate NAM for their purpose. The development of a more dynamic and user-friendly version of these maps is, therefore, an important project for the future. With broader discussion and further development, our maps could provide a template for the safety testing of other organ systems, or use in other testing contexts, leading to greater adoption of NAMs, improved productivity within pharmaceutical companies and potentially, safer medicines.

The authors recommend that pharmaceutical companies are transparent about the NAMs employed in the early stages of drug development, that they should include the results from all relevant NAMs used in their submissions to regulators, and that there should be regulator-sponsor collaboration about NAMs use prior to submission. These actions would familiarize regulators with NAMs, enabling them to update guidelines, where necessary, and make upfront recommendations to nonclinical scientists about NAMs that might be employed.

In addition, the terminology of safety pharmacology and toxicity testing guidance should be updated so that appropriate language is used to describe where NAMs may be employed in the nonclinical safety assessment of drug candidates.

References

Andersen, M., McMullen, P., Phillips, M., et al. (2019). Developing context appropriate toxicity testing approaches using new alternative methods (NAMs). *ALTEX 36*, 523-534. doi:10.14573/altex.1906261

Avila, A. M., Bebenek, I., Bonzo, J. A., et al. (2020). An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs). *Regul Toxicol Pharmacol 114*, 104662. doi:10.1016/j.yrtph.2020.104662

Bai, J. P. F., Earp, J. C., Strauss, D. G. et al. (2020). A Perspective on Quantitative Systems Pharmacology Applications to Clinical Drug Development. *CPT Pharmacometrics Syst Pharmacol 9*, 675-677. doi:10.1002/psp4.12567

Bailey, J., Thew, M. and Balls, M. (2015). Predicting human drug toxicity and safety via animal tests: Can any one species predict drug toxicity in any other, and do monkeys help? *Altern to Lab Anim 43*, 393-403. doi:10.1177/026119291504300607

Balogh Sivars, K., Sivars, U., Hornberg, E., et al. (2018). A 3D Human Airway Model Enables Prediction of Respiratory Toxicity of Inhaled Drugs In Vitro. *Toxicol Sci 162*, 301-308. doi:10.1093/toxsci/kfx255

Bartfeld, S. and Clevers H. (2017). Stem cell-derived organoids and their application for medical research and patient treatment. *J Mol Biol 95*, 729-738. doi:10.1007/s00109-017-1531-7

Baudy, A. R., Otieno, M. A., Hewitt, P. et al. (2020). Liver microphysiological systems development guidelines for safety risk assessment in the pharmaceutical industry. *Lab Chip 20*, 215-225. doi:10.1039/C9LC00768G

Bowes, J., Brown, A. J., Hamon, J., et al. (2012). Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. *Nat Rev Drug Discov 11*, 909-922. doi:10.1038/nrd3845

Bracher, M., Pilkington, G. J., Hanemann, C. O. et al. (2020). A Systematic Approach to Review of in vitro Methods in Brain Tumour Research (SAToRI-BTR): Development of a Preliminary Checklist for Evaluating Quality and Human Relevance. *Front Bioeng Biotechnol 8*, 1-17. doi:10.3389/fbioe.2020.00936

Burgdorf, T., Piersma, A. H., Landsiedel, R. et al. (2019). Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology – Evolution versus revolution. *Toxicol In Vitro 9*, 1-11. doi:10.1016/j.tiv.2019.03.039

Busquet, F., Hartung, T., Pallocca, G. et al. (2020). Harnessing the power of novel animal-free test methods for the development of COVID-19 drugs and vaccines. *Arch Toxicol 94*, 2263-2272. doi:10.1007/s00204-020-02787-2

Butler, L. D., Guzzie-Peck, P., Hartke, J. et al. (2017). Current nonclinical testing paradigms in support of safe clinical trials: An IQ Consortium DruSafe perspective. *Regul Toxicol Pharmacol 87*, S1-S15. doi:10.1016/j.yrtph.2017.05.009

Chioccioli, M., Feriani, L., Kotar, J. et al. (2019). Phenotyping ciliary dynamics and coordination in response to CFTR-modulators in Cystic Fibrosis respiratory epithelial cells. *Nat Commun 10*, doi:10.1038/s41467-019-09798-3

Choudhury, Y., Toh, Y. C., Xing, J. et al. (2017). Patient-specific hepatocyte-like cells derived from induced pluripotent stem cells model pazopanib-mediated hepatotoxicity. *Sci Rep 7*, 41238. doi:10.1038/srep41238

Clippinger, A. J., Allen, D., Jarabek, A. M. et al. (2018). Alternative approaches for acute inhalation toxicity testing to address global regulatory and non-regulatory data requirements: An international workshop report. *Toxicol In Vitro 48*, 53-70. doi:10.1016/j.tiv.2017.12.011

Cook, D., Brown, D., Alexander, R. et al. (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: A five-dimensional framework. *Nat Rev Drug Discov 13*, 419-431. doi:10.1038/nrd4309

Corley, R. A., Kuprat, A. P., Suffield, S. R. et al. (2021). New Approach Methodology for Assessing Inhalation Risks of a Contact Respiratory Cytotoxicant: Computational Fluid Dynamics-Based Aerosol Dosimetry Modeling for Cross-Species and In Vitro Comparisons. *Toxicol Sci 182*, 243-259. doi:10.1093/toxsci/kfab062

Dirven, H., Vist, G. E., Bandhakavi, S. et al. (2021). Performance of preclinical models in predicting drug-induced liver injury in humans: a systematic review. *Sci Rep 11*, 6403. doi:10.1038/s41598-021-85708-2

EMA – European Medicines Agency (2018). Reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. https://www.ema.europa.eu/en/reporting-physiologically-based-pharmacokinetic-pbpk-modelling-simulation

EMA (2021). Mandate, Objectives and Rules of Procedure of the Scientific Advice Working Party (SAWP). https://www.ema.europa.eu/en/documents/other/mandate-objectives-rules-procedure-scientific-advice-working-party-sawp_en.pdf

Ewart, L., Apostolou, A., Briggs, S. A. et al. (2021). Qualifying a human Liver-Chip for predictive toxicology: Performance assessment and economic implications. *bioRxiv 2021.12.14*, 472674; doi:10.1101/2021.12.14.472674

Fernandez-Checa, J. C., Bagnaninchi, P., Ye, H. et al. (2021). Advanced preclinical models for evaluation of drug-induced liver injury - consensus statement by the European Drug-Induced Liver Injury Network [PRO-EURO-DILI-NET]. *J Hepatol 75*, 935-959. doi:10.1016/j.jhep.2021.06.021

Ferrari, E. and Rasponi, M. (2021). Liver–Heart on chip models for drug safety. *APL Bioeng 5*, 031505. doi:10.1063/5.0048986

Fontana, R. J. (2008). Acute liver failure due to drugs. *Semin Liver Dis 28*, 175-87. doi:10.1055/s-2008-107311747

Fowler, S., Chen, W. L. K., Duignan, D. B. et al. (2020). Microphysiological systems for ADME-related applications: current status and recommendations for system development and characterization. *Lab Chip 20*, 446-467. doi:10.1039/C9LC00857H

Franzen, N., van Harten, W. H., Retèl, V. P., et al. (2019). Impact of organ-on-a-chip technology on pharmaceutical R&D costs. *Drug Discov Today 24*, 1720-1724. doi:10.1016/j.drudis.2019.06.003

Guo, Y. and Pu, W. T. (2020). Cardiomyocyte Maturation. *Circ Res 126*, 1086-1106. doi:10.1161/CIRCRESAHA.119.315862

Hardy, A., Benford, D., Halldorsson, T. et al. (2017). Update: use of the benchmark dose approach in risk assessment. *EFSA J 15*. doi:10.2903/j.efsa.2017.4658

Hill, E. J., Jiménez-González, C., Tarczyluk, M. et al. (2012). NT2 Derived Neuronal and Astrocytic Network Signalling. Okazawa H, ed. *PLoS One 7*, e36098. doi:10.1371/journal.pone.0036098

ICH – International Conference on Harmonisation (2000). Guideline S7A Safety Pharmacology Studies for Human Pharmaceuticals. https://database.ich.org/sites/default/files/S7A_Guideline.pdf

ICH (2018). Final Concept Paper on ICH S7B and E14 Q&A; 15 November. https://database.ich.org/sites/default/files/E14S7B_IWG_Concept_Paper.pdf

ICH (2009). Guideline M3(R2) on non-clinical safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals Step 5. https://database.ich.org/sites/default/files/M3_R2__Guideline.pdf

ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods. (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam_strategicroadmap_january2018_document_508.pdf

International Programme on Chemical Safety & Inter-Organization Programme for the Sound Management of Chemicals. (2010). Characterization and application of physiologically based pharmacokinetic models in risk assessment. World Health Organization. https://apps.who.int/iris/handle/10665/44495

Gocke, E., Müller, L., Guzzie, P. J. et al. (2000). Considerations on photochemical genotoxicity: report of the International Workshop on Genotoxicity Test Procedures Working Group. *Environ Mol Metagen 35*, 173-184. doi:10.1002/(sici)1098-2280(2000)35:3<173::aid-em4>3.0.co;2-e

Kanda, Y., Yamazaki, D., Osada, T., et al. (2018). Development of torsadogenic risk assessment using human induced pluripotent stem cell-derived cardiomyocytes: Japan iPS Cardiac Safety Assessment (JiCSA) update. *J Pharmacol Sci 138*, 233-239. doi:10.1016/j.jphs.2018.10.010

Kennedy, R., Kuvshinov, D., Sdrolia, A. et al. (2019). A patient tumour-on-a-chip system for personalised investigation of radiotherapy based treatment regimens. *Sci Rep 9*, 6327. doi:10.1038/s41598-019-42745-2

Kuprat, A. P., Jalali, M., Jan, T. et al. (2021). Efficient bi-directional coupling of 3D computational fluid-particle dynamics and 1D Multiple Path Particle Dosimetry lung models for multiscale modeling of aerosol dosimetry. *J Aerosol Sci 151*, 105647. doi:10.1016/j.jaerosci.2020.105647

LASA, NC3Rs. (2009). Guidance on Dose Level Selection for Regulatory General Toxicology Studies for Pharmaceuticals.; 2009. http://www.lasa.co.uk/PDF/LASA-NC3RsDoseLevelSelection.pdf

Lauschke, V. M., Vorrink, S. U., Moro, S. M. L. et al. (2016a). Massive rearrangements of cellular MicroRNA signatures are key drivers of hepatocyte dedifferentiation. *Hepatology 64*, 1743-1756. doi:10.1002/hep.28780

Lauschke, V. M., Hendriks, D. F. G., Bell, C. C. et al. (2016b). Novel 3D Culture Systems for Studies of Human Liver Function and Assessments of the Hepatotoxicity of Drugs and Drug Candidates. *Chem Res Toxicol 29*, 1936-1955. doi:10.1021/acs.chemrestox.6b00150

Lauschke, V. M., Shafagh, R. Z., Hendriks, D. F. G. et al. (2019). 3D Primary Hepatocyte Culture Systems for Analyses of Liver Diseases, Drug Metabolism, and Toxicity: Emerging Culture Paradigms and Applications. *Biotechnol J 14*, 1800347. doi:10.1002/biot.201800347

Lemme, M., Ulmer, B. M., Lemoine, M. D. et al. (2018). Atrial-like Engineered Heart Tissue: An In Vitro Model of the Human Atrium. *Stem Cell Rep 11*, 1378-1390. doi:10.1016/j.stemcr.2018.10.008

Li, W., Luo, X., Ulbricht, Y. et al. (2019). Establishment of an automated patch-clamp platform for electrophysiological and pharmacological evaluation of hiPSC-CMs. *Stem Cell Rep 41*, 101662. doi:10.1016/j.scr.2019.101662

Lin, C. and Khetani, S. R. (2016). Advances in Engineered Liver Models for Investigating Drug-Induced Liver Injury. *Biomed Res Int 2016*, 1-20. doi:10.1155/2016/1829148

Liu, L., Yu, L., Li, Z. et al. (2021). Patient-derived organoid (PDO) platforms to facilitate clinical decision making. *J Transl Med 19*, 40. doi:10.1186/s12967-020-02677-2

Morgan, P., Brown, D. G., Lennard, S. et al. (2018). Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nat Rev Drug Discov 17*, 167-181. doi:10.1038/nrd.2017.244

Nantasanti, S., de Bruin, A., Rothuizen, J. et al. (2016). Concise Review: Organoids Are a Powerful Tool for the Study of Liver Disease and Personalized Treatment Design in Humans and Animals. *Stem Cells Transl Med 5*, 325-30.

NASEM – National Academies of Sciences, Engineering, and Medicine. (2021). Microphysiological Systems: Bridging Human and Animal Research - A Workshop. https://www.nationalacademies.org/event/01-19-2021/microphysiological-systems-bridging-human-and-animal-research-a-workshop

Ouchi, R., Togo, S., Kimura, M. et al. (2019). Modeling Steatohepatitis in Humans with Pluripotent Stem Cell-Derived Organoids. *Cell Metab 30*, 374-384.e6. doi:10.1016/j.cmet.2019.05.007

Pamies, D., Leist, M., Coecke, S. et al. (2020). Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0) - Draft for stakeholder discussion and call for action. *ALTEX 37*, 490-492. doi:10.14573/altex.2007091

Parish, S. T., Aschner, M., Casey, W. et al. (2020). An evaluation framework for new approach methodologies (NAMs) for human health safety assessment. *Regul Toxicol Pharmacol 112*, 104592. doi:10.1016/j.yrtph.2020.104592

Patterson, E. A., Whelan, M. P., Worth, A. P. (2021). The role of validation in establishing the scientific credibility of predictive toxicology approaches intended for regulatory application. *Comput Toxicol 17*, 100144. doi:10.1016/j.comtox.2020.100144

Peterson, N. C., Mahalingaiah, P. K., Fullerton, A. et al. (2020). Application of microphysiological systems in biopharmaceutical research and development. *Lab Chip 20*, 697-708. doi:10.1039/C9LC00962K

Piersma, A. H., Burgdorf, T., Louekari, K. et al. (2018). Workshop on acceleration of the validation and regulatory acceptance of alternative methods and implementation of testing strategies. *Toxicol In Vitro 50*, 62-74. doi:10.1016/j.tiv.2018.02.018

Pointon, A., Harmer, A. R., Dale, I. L. et al. (2015). Assessment of Cardiomyocyte Contraction in Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *Toxicol Sci 144*, 227-237. doi:10.1093/toxsci/kfu312

Pointon, A., Maher, J., Davis, M. et al. (2021). Cardiovascular microphysiological systems (CVMPS) for safety studies – a pharma perspective. *Lab Chip 21*, 458-472. doi:10.1039/D0LC01040E

Primavessy, D., Metz, J., Schnur, S. et al. (2021). Pulmonary in vitro instruments for the replacement of animal experiments. *Eur J Pharm Biopharm 168*, 62-75. doi:10.1016/j.ejpb.2021.08.005

Rubiano, A., Indapurkar, A., Yokosawa, R. et al. (2021). Characterizing the reproducibility in using a liver microphysiological system for assaying drug toxicity, metabolism, and accumulation. *Clin Transl Sci 14*, 1049-1061. doi:10.1111/cts.12969

Sadler, R. C., Prime, D., Burnell, P. K. et al. (2011). Integrated in vitro experimental modelling of inhaled drug delivery: deposition, dissolution and absorption. *J Drug Deliv Sci Technol 21*, 331-338. doi:10.1016/S1773-2247(11)50051-6

Smith, B., Rowe, J., Watkins, P. B. et al. (2020). Mechanistic Investigations Support Liver Safety of Ubrogepant. *Toxicol Sci 177*, 84-93. doi:10.1093/toxsci/kfaa093

Su, F. S., Lee, C. F., Rajendran, P. et al. (2020). Nasal Airflow Simulation on Healthy Child using Computational Fluid Dynamics. Technol Reports Kansai Univ. https://www.kansaiuniversityreports.com/article/nasal-airflow-simulation-on-healthy-child-using-computational-fluid-dynamics

Takayama, K., Morisaki, Y., Kuno, S. et al. (2014). Prediction of interindividual differences in hepatic functions and drug sensitivity by using human iPS-derived hepatocytes. *Proc Natl Acad Sci 111*, 16772-16777. doi:10.1073/pnas.1413481111

Thomas, D., Chancellor, D., Micklus, A. et al. (2021). Clinical Development Success Rates and Contributing Factors 2011–2020. https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf

Tsega, E. G. (2018). Computational Fluid Dynamics Modeling of Respiratory Airflow in Tracheobronchial Airways of Infant, Child, and Adult. *Comput Math Methods Med 2018*, 1-9. doi:10.1155/2018/9603451

US EPA – US Environmental Protection Agency. (2016). Benchmark dose software (BMDS) user manual. https://www.epa.gov/sites/default/files/2015-11/documents/bmds_manual.pdf

US FDA – US Food and Drug Administration. (2017). Predictive Toxicology Roadmap. https://www.fda.gov/files/science & research/published/FDA%27s-Predictive-Toxicology-Roadmap.pdf

US FDA (2018). Physiologically based pharmacokinetic analyses—format and content guidance for industry. https://www.fda.gov/media/101469/download

US FDA (2019). Liver Toxicity Knowledge Base (LTKB). https://www.fda.gov/science-research/bioinformatics-tools/liver-toxicity-knowledge-base-ltkb

US FDA (2021a). Innovative Science and Technology Approaches for New Drugs (ISTAND) Pilot Program. FDA Website. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program

US FDA (2021b). Advancing New Alternative Methodologies (NAMs) at FDA. https://www.fda.gov/media/144891/download

US FDA (2021c). Advancing Alternative Methods at FDA. FDA Website. https://www.fda.gov/science-research/about-science-research-fda/advancing-alternative-methods-fda

Varewyck, M. and Verbeke, T. (2017). Software for benchmark dose modelling. *EFSA Support Publ 14*. doi:10.2903/sp.efsa.2017.EN-1170

Wang, X., Wang, L., Dou, W. et al. (2020). Electrical impedance-based contractile stress measurement of human iPSC-Cardiomyocytes. *Biosens Bioelectron 166*, 112399. doi:10.1016/j.bios.2020.112399

Ware, B. R., Berger, D. R., Khetani, S. R. (2015). Prediction of Drug-Induced Liver Injury in Micropatterned Co-cultures Containing iPSC-Derived Human Hepatocytes. *Toxicol Sci 145*, 252-262. doi:10.1093/toxsci/kfv048

Watkins, P. B. (2011). Drug Safety Sciences and the Bottleneck in Drug Development. *Clin Pharmacol Ther 89*, 788-790. doi:10.1038/clpt.2011.63

Watkins, P. B. (2020). DILIsym: Quantitative systems toxicology impacting drug development. *Curr Opin Toxicol 23-24*, 67-73. doi:10.1016/j.cotox.2020.06.003

Woehrling, E. K., Parri, H. R., Tse, E. H. Y. et al. (2015). A Predictive In Vitro Model of the Impact of Drugs with Anticholinergic Properties on Human Neuronal and Astrocytic Systems. Ulrich H, ed. *PLoS One 10*, e0118786. doi:10.1371/journal.pone.0118786

Woodhead, J. L., Brock, W. J., Roth, S. E. et al. (2017). Application of a mechanistic model to evaluate putative mechanisms of tolvaptan drug-induced liver injury and identify patient susceptibility factors. *Toxicol Sci 155*, 61-74. doi:10.1093/toxsci/kfw193

Yin, S., Xi, R., Wu, A. et al. (2020). Patient-derived tumor-like cell clusters for drug testing in cancer therapy. *Sci Transl Med 12*, eaaz1723. doi:10.1126/scitranslmed.aaz1723

Zhang, X., Jiang, T., Chen, D. et al. (2020). Three-dimensional liver models: state of the art and their application for hepatotoxicity evaluation. *Crit Rev Toxicol 50*, 279-309. doi:10.1080/10408444.2020.1756219

Conflict of interest

The authors state that they have no conflict of interest.

Data availability

No original data was generated for this manuscript.

Acknowledgments

Jan Turner[1], Pandora Pound[1], Carla Owen[2], Isobel Hutchinson[2], Marina Hop[3], David Y. S. Chau[4], Lady V. Barrios Silva[4], Mike Coleman[5], Audrey Dubourg[6], Lorna W Harries[7], Victoria Hutter[8], J. Gerry Kenna[1], Volker M. Lauschke[9], Winfried Neuhaus[10], Clive Roper[11], Paul B. Watkins[12], Jonathan Welch[13], Laura Rego Alvarez[14] and Katy Taylor[14]

[1]Safer Medicines Trust, Kingsbridge, UK; [2]Animal Free Research UK, London, UK; [3]Viveo Consulting Ltd., London, UK;

[4]Division of Biomaterials and Tissue Engineering, UCL Eastman Dental Institute, London, UK; [5]College of Health and Life Sciences, Aston University, Birmingham, UK; [6]CN Bio Innovations Limited, Cambridge, UK; [7]University of Exeter Medical School, Exeter, UK; [8]ImmuONE Limited, Hatfield, UK: Centre for Topical Drug Delivery and Toxicology, University of Hertfordshire, Hatfield, UK; [9]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; Dr Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany; University of Tübingen, Tübingen, Germany; [10]Austrian Institute of Technology GmbH, Competence Unit Molecular Diagnostics, Vienna, Austria; Department of Medicine, Danube Private University, Krems, Austria; [11]Roper Toxicology Consulting Limited, Edinburgh, UK; [12]Division of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School Of Pharmacy, Chapel Hill NC, USA; [13]Newcells Biotech, Newcastle Upon Tyne, UK; [14]Cruelty Free International, London, UK

(laura.alvarez@crueltyfreeinternational.org)