



Commentary

An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs)



Amy M. Avila^a, Ilona Bebenek^a, Jessica A. Bonzo^a, Todd Bourcier^a, Karen L. Davis Bruno^a, David B. Carlson^a, John Dubinion^a, Ikram Elayan^a, Wafa Harrouk^a, Shwu-Luan Lee^a, Donna L. Mendrick^b, Jill C. Merrill^a, Jackye Peretz^a, Emily Place^a, Muriel Saulnier^a, Ronald L. Wange^a, Jia Yao^a, Dong Zhao^a, Paul C. Brown^{a,*}

^a Center for Drug Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD, 20903, United States

^b National Center for Toxicological Research, US Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD, 20903, United States

ARTICLE INFO

Keywords:

FDA
Regulatory
Alternative
Methodology
Approach
In vitro
Drug development

ABSTRACT

Nonclinical testing of human pharmaceuticals is conducted to assess the safety of compounds to be studied in human clinical trials and for marketing of new drugs. Although there is no exact number and type of nonclinical studies required for safety assessments, as there is inherent flexibility for each new compound, the traditional approach is outlined in various FDA and ICH guidance documents and involves a combination of in vitro assays and whole animal testing methods. Recent advances in science have led to the emergence of numerous new approach methodologies (NAMs) for nonclinical testing that are currently being used in various aspects of drug development. Traditional nonclinical testing methods can predict clinical outcomes, although improvements in these methods that can increase predictivity of clinical outcomes are encouraged and needed. This paper discusses FDA/CDER's view on the opportunities and challenges of using NAMs in drug development especially for regulatory purposes, and also includes examples where NAMs are currently being used in nonclinical safety assessments and where they may supplement and/or enhance current testing methods. FDA/CDER also encourages communication with stakeholders regarding NAMs and is committed to exploring the use of NAMs to improve regulatory efficiency and potentially expedite drug development.

1. Introduction

CDER's mission is to ensure the availability of safe and effective drugs to improve the health of people in the United States. To help achieve that mission, applicants of new drug and biologic products are required to provide the pharmacology and toxicology (nonclinical) information from which they have concluded that it is reasonably safe to conduct clinical trials (21CFR312) and ultimately to support marketing (21CFR312; 21CFR314; 21CFR601). This includes data about the drug or biologic product's pharmacology and disposition (pharmacological effects and mechanism(s) of action, ADME), and toxicology (acute, subacute and chronic, developmental and reproductive toxicology,

carcinogenicity, and 'special' toxicology, as appropriate). This non-clinical information informs pharmaceutical development programs at key points by addressing critical issues of importance to drug developers and FDA reviewers alike (Table 1).

Federal regulations note that these data can come from studies conducted in animals or in vitro, and are inherently flexible in that exact study types are not specified. The studies considered appropriate to address these issues are largely described in FDA and International Council for Harmonization (ICH) guidances. These guidances are also flexible as they are nonbinding and typically note that alternative approaches can be used if they satisfy the applicable statutes and regulations. Nevertheless, a relatively standard set of studies has evolved

Abbreviations: (ADME), Absorption, Distribution, Metabolism, and Excretion; (AOP), adverse outcome pathway; (BSEP), bile salt export pump; (CDER), Center for Drug Evaluation and Research; (NCTR), National Center for Toxicological Research; (NAMs), new approach methodologies; (FIH), first in human; (IVIVE), in vitro in vivo extrapolation; (ICH), International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use; (FOB), functional observational battery; (CiPA), Comprehensive In Vitro Proarrhythmia Assay; (LLNA), local lymph node assay; (NOAEL), no observed adverse effect level; (MABEL), minimal anticipated biological effect level; (MPS), microphysiological systems; (QSAR), quantitative structure activity relationship; (iPSC), induced pluripotent stem cells; (WOE), weight-of-evidence; (DART), developmental and reproductive toxicity; (OECD), Organisation for Economic Co-operation and Development

* Corresponding author.

E-mail address: Paul.Brown@fda.hhs.gov (P.C. Brown).

<https://doi.org/10.1016/j.yrtph.2020.104662>

Received 23 January 2020; Received in revised form 3 April 2020; Accepted 15 April 2020

Available online 20 April 2020

0273-2300/ © 2020 Elsevier Inc. All rights reserved.

Table 1

Key issues addressed by pharmacology and toxicology studies supporting pharmaceutical development.

-
- Describes the pharmacological effects and mechanism(s) of action of the drug in vitro and/or in vivo
 - Identifies risk attributes of drug absorption, distribution, metabolism and excretion (ADME)
 - Estimates safe “first in human” (FIH) starting dose
 - Estimates safe maximum exploratory doses in early clinical trials
 - Identifies possible consequences of chronic exposure
 - Identifies risks for special populations (e.g., pediatrics)
 - Identifies specific parameters to monitor more closely in clinical trials
 - Predicts risks that are difficult to assess or are unethical to assess in humans (e.g., carcinogenicity/developmental and reproductive toxicology)
 - Allows the mechanistic understanding of an adverse biological change observed in animals or humans
-

to address the critical issues listed in Table 1. The current approach relies on in vitro and in vivo models and has enabled the safe testing of drugs in clinical trials (Butler et al., 2017; Monticello et al., 2017). There is recognition that while the current testing paradigm performs well for some of the critical issues listed above (e.g., ‘FIH’ dose selection), performance has been questioned for other critical issues that impact prediction of human risk (e.g., carcinogenicity, immunosuppression, drug-induced liver injury).

CDER has and continues to support development of and is receptive to new approach methodologies (NAMs) that would address the key issues in Table 1 and improve predictivity of clinical outcomes. CDER considers NAMs to include a broad range of methods such as in vitro, in chemico, and in silico methods. In vivo methods can also be considered NAMs when they improve predictivity, shift studies to phylogenetically lower animals, or otherwise help replace, reduce, and refine animal use (i.e., the 3Rs) in development programs. Such NAMs can improve regulatory efficiency and potentially expedite drug development.

CDER toxicologists participated in the development of both the FDA’s Predictive Toxicology Roadmap (FDA, 2019) and the Interagency Coordinating Committee for the Validation of Alternative Methods roadmap (Interagency Coordinating Committee, 2019). Both roadmaps focus on regulatory agency needs and the development of new approaches and methodologies with contexts of use that help address these needs. CDER scientists continue to be engaged with both initiatives.

The objectives of this publication are to describe the current non-clinical approach to drug evaluation encountered in CDER’s Office of New Drugs and highlight those areas in particular need of more predictive and informative testing methodologies within a human drug development context. This paper is not intended to be an exhaustive list nor an assessment of these new approaches and does not endorse any particular approach, though opportunities and challenges with some approaches are discussed. Rather, areas are identified where the current approach is less than optimal and could benefit from supplementary or fit-for-purpose testing approaches, including non-animal methodologies. An overview of those opportunities and challenges that merit particular effort in refining new approaches is summarized in Table 2.

2. Safety pharmacology

Safety pharmacology studies are conducted to determine whether a drug causes on- or off-target serious acute effects on critical organ systems (e.g., cardiovascular, respiratory, gastrointestinal, and central nervous system). Dose responses identified in these studies serve to establish a safety margin for the first in human (FIH) dose regimen. Follow-up experiments aimed at understanding the mechanism behind an observed specific toxicity can be undertaken to address the relevance of the finding to human risk.

2.1. Current approach

2.1.1. Cardiovascular system

Whole animals are used to assess the effects of a drug on blood pressure, heart rate, and the electrocardiogram. In addition, in vivo, in vitro, and/or ex vivo methods are used to evaluate repolarization and conductance abnormalities (ICH S7A, 2001; ICH S7B, 2005).

2.1.2. Central nervous system (CNS)

ICH S7A indicates that motor activity, behavioral changes, coordination, sensory/motor reflex responses and body temperature should be evaluated. For example, a functional observation battery (FOB) (Mattsson, 1996), modified Irwin’s (Irwin, 1968), or another appropriate test (Haggerty, 1991) can be used. Whole animals have been used to assess the effects of drugs on CNS function because the endpoints of interest are activities of the entire body that are controlled and coordinated by the brain and spinal cord.

2.1.3. Respiratory system

ICH S7A stresses that respiratory rate and other measures of respiratory function (e.g., tidal volume or hemoglobin oxygen saturation) should be evaluated using appropriate methodologies to complement clinical observations. These studies include evaluations of both the total respiratory system and the mechanical properties of the lung. The use of conscious animals is most typical and is recommended; the techniques and procedures for measuring respiratory function parameters in vivo are well established (Murphy, 1994).

2.1.4. Gastrointestinal, renal and other systems

Safety pharmacology studies of other organ systems are conducted when there is cause for concern from other information. A variety of in vivo and in vitro methods are used. Drug-induced effects on some of these systems can be assessed in stand-alone safety pharmacology studies or as endpoints integrated into toxicology studies.

2.2. Statement(s) of need

Safety pharmacology studies, whether they be stand-alone studies or integrated into pivotal toxicology studies, focus on complex organ functions rather than macroscopic or microscopic tissue integrity. Some in vitro or ex vivo models have been used to address specific aspects of organ function. However, in general, identifying alternatives to the use of whole animals for these endpoints has been challenging. Rather, the existing testing approach could benefit in some areas from supplementary or ‘fit-for-purpose’ evaluations that involve non-animal methodologies, as discussed below, and from refining the approach by further integrating relevant endpoints into pivotal toxicology studies. In addition, there will always be a need for newer models that are more predictive of human adverse events.

2.2.1. Cardiovascular system

The capability of a drug being pro-arrhythmic under therapeutic conditions is recognized as a serious drug liability that needs to be identified. While the current paradigm has resulted in no approved new drugs with unrecognized risk of torsades de pointes, there are also drugs that may cause hERG block or QT prolongation that do not result in torsades de pointes (Wallis et al., 2018). Consequently, methods that can identify the proarrhythmic risk without preventing potentially useful drugs from reaching clinical testing would be beneficial.

2.2.2. Central nervous system

Tier 1 tests of CNS function employ the FOB or Irwin screen to evaluate broad neurological functions whereas in tier 2, motor, sensory, and memory functions are assessed if needed. Unfortunately, the rodent FOB and Irwin test do not reliably detect some of the most common adverse events observed in Phase I clinical trials, namely headache,

Table 2

An overview of current key challenges and needs in nonclinical safety assessments supporting development of human pharmaceuticals.

Assessment	Key Issues Addressed	Key Needs
<i>Safety Pharmacology</i>	<ul style="list-style-type: none"> o Identifies potential adverse pharmacological effects of the drug in vitro and/or in vivo o Identifies specific parameters to monitor more closely in clinical trials 	<ul style="list-style-type: none"> o Refine approach to improve identification of proarrhythmic risk while reducing attrition of potentially useful drugs from clinical testing o Refine Tier 1 neurological screens to achieve more quantitative and objective measures o Standardize Tier 1 protocols and statistical approach to enable cross-laboratory comparisons o Improve Tier 2 testing of higher cognitive functions and specificity of sensory tests
<i>General Toxicity</i>	<ul style="list-style-type: none"> o Estimates safe “first in human” (FIH) starting dose o Estimates safe maximum doses in early clinical trials o Identifies possible consequences of chronic exposure o Identifies specific parameters to monitor more closely in clinical trials 	<ul style="list-style-type: none"> o Improve risk identification for rare and idiosyncratic toxicities o Refine cardiovascular assessments to include functional endpoints o Improve approach to evaluating human relevancy of toxicity findings in animals o Identify best contexts-of-use for animal models of disease for toxicology testing o Develop more human-relevant approaches to assessing risk of immune-directed therapies
<i>Carcinogenicity</i>	<ul style="list-style-type: none"> o Predicts long-term risks that are difficult to assess or are unethical to assess in humans 	<ul style="list-style-type: none"> o Develop approaches more applicable to human biology and that provide dose response data based on anticipated human exposure and mechanism.
<i>Developmental and Reproductive Toxicity</i>	<ul style="list-style-type: none"> o Predicts risks that are difficult to assess or are unethical to assess in humans o Identifies risks for special populations 	<ul style="list-style-type: none"> o Develop sensitive alternatives to detect potential teratogenicity earlier to facilitate clinical trials. o Develop sensitive alternatives for drug classes not tolerated by pregnant animals, such that the ability to conduct in vivo developmental and reproductive toxicity studies is limited
<i>Special Toxicity</i>	<ul style="list-style-type: none"> o Identifies specific parameters to monitor more closely in clinical trials o Allows the mechanistic understanding of an adverse biological change observed in animals or humans 	<ul style="list-style-type: none"> o Develop validated non-animal methods for assessing human skin sensitization of mixtures

nausea, dizziness, fatigue/somnolence, and pain as these methods were not specifically designed to detect these findings or are not sensitive enough to detect them (Mead et al., 2016). This lack of concordance between the standard rodent neurofunctional assessment studies and adverse events in humans is a concern and highlights the need for better and more predictive preclinical tools and models. More quantitative and objective measures for tier I studies and standardization of protocols to enable cross-laboratory comparisons, including statistical analyses, are needed. There is also a need to improve the tier 2 testing of higher cognitive functions and the specificity of sensory tests (Porsolt et al., 2002).

2.3. New approach methodologies

2.3.1. Cardiovascular system

The Comprehensive In Vitro Proarrhythmia Assay (CiPA) initiative is a proposal for an innovative research paradigm designed to apply cutting edge in vitro and in silico assays for determining electrophysiological mechanisms conferring pro-arrhythmic risk to candidate drugs (Cavero and Holzgrefe, 2015; Vicente et al., 2018; Wallis et al., 2018).

Components of CiPA include in vitro assessments of drug effects on multiple ion channels and in silico computer modelling to predict risk. This could be followed with in vivo ECG biomarker assessments in phase 1 clinical trials and assessment of in vitro effects in ventricular cardiomyocytes derived from human stem cells. An expected positive outcome of the CiPA paradigm is the ability to discriminate candidate drugs into those interacting with single (e.g., blockade of IKr, INaF, IKs) or multiple (e.g., IhERG and ICaL, IKr and INaL) cardiac ion channels since this can inform the degree of proarrhythmic risk inherent in a candidate drug. Candidate drugs may also be categorized into low-, intermediate-, and high-risk proarrhythmic agents in comparison with established proarrhythmic drugs.

2.3.2. Central nervous system

In vitro models of nervous system tissue have been developed. While these may be useful to assess specific effects on neurons or other cells of the CNS, they are not able to recapitulate complex neurobehavior. These in vitro methods may be suitable for screening for specific

neurotoxicity mechanisms. In particular, ‘micro-brains’ or brain organoids are 3D cultures of primary cells or induced pluripotent stem cells directed to develop into neural tissues, and hold the potential to interrogate some aspects of drug toxicity directly on human brain tissues (Abbott, 2013; Bhatia and Ingber, 2014; Kelava and Lancaster, 2016; Pamies et al., 2017).

Assays are being developed to study chemical safety issues in zebrafish, which provide the advantage of detecting effects while maintaining the biological integrity of the test system (Kanungo et al., 2014). Zebrafish have shown potential utility in testing motor activities, convulsant and proconvulsant properties, seizure liabilities, cognitive function, and drug dependency (Barros et al., 2008; Celine de Esch, 2012).

New techniques in bioimaging have evolved to visualize a broad variety of functional CNS parameters, mapping them to anatomical brain structures that are thereby “tagged” with additional information of high biological relevance. These techniques could potentially be incorporated into safety pharmacology protocols to evaluate CNS risks and target-specific mechanisms (Borsook et al., 2013; Carmichael et al., 2018; Frank, 1999; Matthews et al., 2006). Similarly, new diagnostic methodologies using integrated video electroencephalography (EEG) technologies in non-rodents could potentially be used to address convulsion and seizure risks (Authier et al., 2009; Vite, 2005).

2.3.3. Respiratory system

Techniques are being developed to mimic in vivo lung structure and function with in vitro tissue cultures of human origin. In vitro air-liquid interface (ALI) cell culture models can potentially be used to assess some inhalation toxicology endpoints. The relevance of ALI models is currently considered between classic in vitro (i.e., submerged) and animal-based models (Lacroix, 2018). These methods are still in their infancy and will require improvements and validation prior to regulatory use (Hiemstra et al., 2018).

Methods of continuous monitoring of respiratory parameters in conscious, non-restrained small and large laboratory animals have the potential to reduce the number of animals in studies by allowing simultaneous measurement of cardiovascular and respiratory parameters (Kramer and Kinter, 2003; Niemeyer, 2016; Bailey, 2012). This, in turn, has the potential to allow for the detection of cardiovascular-respiratory

dependencies to help define mechanisms of action.

2.3.4. Gastrointestinal, renal and other systems

In vitro approaches are also being developed to recapitulate aspects of other organ systems. For example, intestinal cell organoids and in vitro kidney tubule-like structures have been created (Maass et al., 2019; Yu et al., 2017). Though the in vitro technology is not yet able to completely recreate the complexity of these organs, some toxicity can be explored. Further optimization and validation are needed before more widespread use in drug development is established (Al-Saffar et al., 2015).

3. General toxicology

General toxicology studies (e.g., single-dose and repeated-dose toxicity studies) evaluate drug safety from a systems-biology perspective, encompassing drug pharmacodynamics (primary and secondary pharmacology), pharmacokinetics, and toxicology. Data generated from general toxicology studies aid in addressing critical questions needed for regulatory decision making, including estimating maximum safe FIH starting doses, determining how fast doses may be escalated in clinical trials, and to what maximum dose. Additionally, they identify target organs of toxicity without regard to mechanism or to predictions of outcome, and can determine reversibility of identified toxicities. They can identify possible consequences of acute, sub-chronic, and chronic exposure, disclose the need for additional clinical monitoring, and predict risks that are infeasible or unethical to assess in humans (e.g., irreversible tissue damage, brain lesions).

3.1. Current approach

The current practice for general toxicology testing is conducting studies in a rodent and non-rodent species of appropriate duration to support clinical trials ((ICH M3(R2), 2010)). For biologics, if the product is only pharmacologically active in one species, then studies only in that species are needed ((ICH S6R1, 2012)). If no pharmacologically responsive test species are available, then the FIH dose estimate may be based on pharmacology data alone (i.e., a minimal anticipated biological effect level (MABEL) approach can be used). Use of a MABEL is relatively uncommon and can necessitate starting with very low doses that may not provide clinical benefit and require protracted dose escalation. Toxicity study parameters generally include clinical signs (including animal behavior), body weight and food consumption, clinical pathology, organ weights, gross pathology, histopathology, and toxicokinetics. Ideally, doses used in toxicity studies should identify a no observed adverse effect level (NOAEL) for particular toxicities and a maximum tolerated dose.

3.2. Statement(s) of need

The current approach to assess the general toxicity of pharmaceuticals have been highly successful in allowing reasonably safe clinical trials to proceed (Butler et al., 2017; Monticello et al., 2017). In addition, the general ability of animal toxicity data to identify potential adverse human effects is good (Clark and Steger-Hartmann, 2018; Olson et al., 2000), and the absence of animal toxicity is particularly good at predicting an absence of toxicities in human phase 1 clinical trials (Monticello et al., 2017). However, there are areas in which traditional general toxicology studies have been less predictive of human toxicity, and improvements in these areas are desirable (Clark, 2015; Olson et al., 2000). The following describes these notable areas of need, but is not meant to be inclusive of all attributes of general toxicology studies that could benefit from development of alternatives:

pharmaceuticals such as biotherapeutics, the pharmacologic target can be highly specific to humans such that the target does not exist in species normally used in toxicity studies. Use of human cells in vitro in NAMs together with pharmacokinetic modelling approaches might help address these challenges.

- *Rare or idiosyncratic toxicities of investigational drugs in humans are not well-identified.* Of particular note, rare forms of drug-induced liver injury (DILI) have caused the failure of drug candidates during clinical trials or marketing (Stevens and Baker, 2009). This toxicity may have genetic and immune components that are not well represented in traditional general toxicity studies, though it is clear that even registration trials in human subjects can fail to predict this adverse effect in a more broadly exposed human population. Supplementary approaches that focus on prediction of DILI based on methodologies employing human materials would be of particular interest. To this end, the NCTR developed the publicly available Liver Toxicity Knowledge Base (LTKB) to aid in the identification of DILI risk. Similar approaches may also hold promise for improving insight into other potential idiosyncratic reactions, such as cutaneous and hematological adverse reactions (Utrecht and Naisbitt, 2013).
- *Cardiovascular toxicity has been a leading cause of discontinuing drug development.* Expanding assessment of cardiovascular function beyond an electrocardiogram (e.g., echocardiograph) in the course of longer duration toxicology studies could prove informative particularly for drugs that target the cardiovascular or renal systems. Coupling such approaches to those being explored for cardiovascular safety pharmacology is a notable opportunity for refining the current overall approach.
- *Establishing human relevancy of some toxicity findings in animals remains a challenge, particularly for findings that are unexpected and poorly understood.* In many cases, if the animal toxicity is sufficiently severe and near clinically relevant exposures, the clinical dose range may be restricted or the trial halted entirely. The development of a translatable biomarker of the animal toxicity may simultaneously allow investigation of 'at risk' doses and address human relevancy over the course of clinical trials. The development and successful incorporation of kidney biomarkers of injury into development programs can serve as an example for development of other biomarkers (Chen et al., 2018). Of note, biomarkers are not limited to serum factors but can include genetic expression profiles, miRNA levels, and other modalities.
- *Toxicology studies are conducted in healthy animals and not in animal models of disease.* Models capable of assessing the interaction of drug and disease may be useful in better identifying and understanding patient-relevant effects, particularly in situations where the intended pharmacology of the investigational drug becomes dose-limiting in normal, healthy animals.
- *Consequences of immune modulation by investigational drugs in humans is not well-defined.* Modification to clinical trial designs and development of more human materials-based in vitro assays has improved assessment and management of the potential for cytokine release syndrome in response to protein therapeutics (Bonini and Rasi, 2016; Grimaldi et al., 2016). However, the multitude of variables that can impact immune outcome in human subjects, notably environmental factors and genetic diversity, limits extrapolation of immunotoxicology data derived from animals that are housed in highly controlled environments and have less genetic diversity. Fortunately, the consequences of broad immune activation or suppression in human subjects can be reasonably anticipated based on general knowledge of human immunity, but the impact of targeted immunomodulation on susceptibility to infection, viral recrudescence, autoimmunity, and cancer risk is generally not well-informed by animal studies.
- *Animals used in toxicity studies often possess limited genetic diversity.* Toxicity in humans can be substantially influenced by genetic

components. This is noted above for DILI and immune effects but is not limited to these areas. FIH and even later trials may not include sufficient genetic diversity to detect drug toxicity related to genetic factors. Understanding genetic factors that contribute to toxicity in humans and incorporating those factors into nonclinical assessments could improve predictivity.

3.3. New approach methodologies

There are currently no alternative predictive toxicological methods that readily allow identification of appropriate dose ranges for clinical exploration and of potential adverse effects over time. As such, there is a very low likelihood of replacing whole animal general toxicity studies in the drug development context. However, there are numerous alternative methods and technologies currently being studied and reported in the literature that, in part, may have utility in improving nonclinical drug development programs. Such methodologies include, but are not limited to, microphysiological systems (MPS) (e.g., organ-on-a-chip, tissue-on-a-chip), organotypic cultures (e.g., co-cultures), 3D organoids, in vitro/in silico toxicity prediction tools, quantitative structure activity relationship (QSAR) computer-based models, and the use of non-mammalian alternative species for toxicity testing. Extensive advancement in the development of MPS models has taken place over recent years, from single-organ models to interconnected multi-organ models that share a common medium (Ewart et al., 2018; Marx et al., 2016)). One of the most attractive features of MPS models is the use of human cells, which may improve assay specificity for toxicity prediction in humans by reducing reliance on non-human testing systems. With the possible benefit of better concordance to humans, the use of MPS are currently being explored for use in early drug development, prior to submission to the Agency, for candidate drug selection and toxicity screening (Cavero et al., 2019). Results of early screening assays may exclude a drug candidate from further development and thereby eliminate the need for whole animal testing of such compounds. MPS might also be used to elucidate the clinical relevance of certain animal findings and to investigate the mechanism of an observed toxicity in humans or animals in a particular organ system (e.g., liver, heart, lung, intestines, and kidney). In recent years, the development of 2D cell cultures and 3D brain organoids derived from human induced pluripotent stem cells (iPSC) of patients with various neurological or psychiatric disorders have been used to enhance the understanding of relevant connections between genetics, physiology, other individual risk variants, and disease pathology (Amin and Pasca, 2018; Di Lullo and Kriegstein, 2017; Lancaster and Knoblich, 2014). In addition, the drug effect(s) in these in vitro models derived from patients appear to correlate with patients' clinical response to therapeutics, indicating a potential for these models to predict drug response and to augment the benefits of precision medicine (Quadrato et al., 2016; Temme et al., 2016).

However, there are still technical barriers in current MPS models that limit their utility as a replacement for whole animal general toxicity testing of human pharmaceuticals for regulatory purposes. These barriers include, but are not limited to, the inability of MPS to reasonably recapitulate interactions among organ systems, the failure to appropriately mimic pharmacokinetics of the drug in vivo, the incapacity to assess consequences of long-term exposure to an investigational drug as can be done in chronic animal studies, the lack of most "organ" systems to capture all cell types present in situ, and the inability to confidently estimate human doses or exposures based on in vitro drug concentrations. Advances in physiologically based pharmacokinetic modelling and in vitro to in vivo extrapolation (IVIVE) might help address these shortcomings.

Aside from innovative cell culture models, additional advances in science and technology have contributed to the refinement, reduction, and replacement of animals. For example, the development of micro-sampling technology allows for reliable toxicokinetic (TK) sampling in

the main study animals, therefore eliminating or reducing the need of additional TK animals (ICH S3(A) Q&A, 2018). New toxicity in vitro prediction assays such as mitochondrial toxicity assays and bile salt export pump (BSEP) assays have been implemented in early stage screening for potential DILI (Shah et al., 2015; Will and Dykens, 2014). Advancements in QSAR computer-based models have led to the generation of models that may predict adverse effects of drugs in humans, including hepatobiliary, renal/bladder, cardiac, and pulmonary toxicity (Kruhlak et al., 2012). QSAR modelling and other in silico technologies are currently being used for screening potential drug candidates to help eliminate candidates with unfavorable risk-benefit profiles at an earlier stage and with better accuracy, thereby reducing animal use (Muster et al., 2008; Rognan, 2017). Similarly, integrated in vitro/in silico systems have been developed to predict off-target functional responses (Leedale et al., 2018).

Another NAM used for assessment of general toxicity of human pharmaceuticals is the use of non-mammalian alternative species, most notably zebrafish and *Caenorhabditis elegans* (*C. elegans*) (Kanungo et al., 2014). The use of zebrafish, especially embryos and larvae, has potential because their major organ systems including the nervous, cardiovascular, digestive and visual systems are similar to mammalian species at the anatomical, physiological and molecular levels. Zebrafish appear to be most useful at this time for chemical hazard identification and candidate drug screening. Zebrafish have not been used for general toxicity testing of human pharmaceuticals due to limitations of use, including ADME/pharmacokinetic differences compared to mammalian species, and the route of administration of the test article is not typically the intended human route of administration. *C. elegans* is a small nematode worm that has digestive, muscular, nervous, and reproductive systems with anatomical and physiological similarity to mammalian species. In vivo assays using *C. elegans* appear to be particularly promising for studying the developmental neurotoxic effects of chemicals (Hunt et al., 2018). *C. elegans* have not been used for general toxicity testing of human pharmaceuticals for reasons similar to the zebrafish. Although studies using non-mammalian species and other new approach methodologies may not be able to replace whole animal general toxicology studies at this time, once appropriately qualified for a particular context of use, there may be potential for their use in mechanistic follow-up toxicity studies to investigate organ/tissue-specific toxicities identified in a whole animal general toxicity study. This could reduce the subsequent use of animals in the later stage of drug development.

4. Predicting human cancer risk

Assessing cancer risk informs the safety profile of new pharmaceuticals intended to treat chronic indications. For reasons of practicality, clinical trials in non-oncologic populations lack both the size and duration needed to clearly detect an imbalance in cancer incidence from prolonged exposure to a new pharmaceutical. Thus, the carcinogenic potential of new pharmaceuticals is informed nearly exclusively by data from nonclinical studies, which ideally provide data relevant to both hazard and quantitative risk for communication in a drug label.

4.1. Current approach

The FDA follows the recommendations outlined in the ICH guidelines for assessing genotoxic risk (ICH S2(R1), 2012) and overall cancer risk for small molecules (ICH S1A, 1996; ICH S1B, 1998; ICH S1C(R2), 2008) and biotechnology-derived products (ICH S6, 1997).

- *Small molecules*: A combination of genotoxicity tests and long-term studies in two rodent species, usually the rat and the mouse, are typically conducted for assessing carcinogenic risk of small molecules. One of the long-term bioassays may be substituted for by an alternative study, typically a transgenic mouse study, that involves

less time and fewer animals than the standard mouse bioassay. Because carcinogenic risk may arise from chemical-specific attributes in addition to pharmacological properties, each small molecule has been assessed regardless of prior knowledge of the risk, or lack of risk, associated with the drug class.

- **Biologics:** A weight-of-evidence (WOE) approach is practiced for biotechnology-derived products when a carcinogenicity assessment is warranted. Assessment is tailored to the specific biologic and typically consists of an evaluation of known drug target pharmacology and compound-specific toxicology. Genotoxicity testing is generally inappropriate for biologic products and is not part of the assessment. When sufficient information is available, this type of WOE assessment can sometimes adequately address carcinogenic potential and preclude the need to conduct additional nonclinical studies, regardless of whether the biologic is active and testable in a rodent bioassay. Rodent bioassays are not warranted if the WOE clearly supports a concern regarding carcinogenic potential (e.g., immunosuppressives and growth factors). Similarly, if the WOE assessment does not suggest carcinogenic potential, a rodent bioassay is considered unlikely to add value, and no additional nonclinical testing of biologic products is recommended. Prior knowledge of target-based risks figures prominently in developing a strategy to assess carcinogenic risk of each new biologic because non-specific activity of such molecules is generally considered to be low.

4.2. Statement(s) of need

The standard strategy for characterizing carcinogenic risk for small molecule pharmaceuticals is a relatively blunt approach of counting tumors over a dose response in rodents exposed to drug for a near-lifetime. This approach is agnostic to the complexities of the pathways that can lead to cancer. This feature can be considered an advantage as any tumorigenic mechanism might be detected without requiring pre-existing knowledge of potential mechanisms. However, despite its central role in risk characterization, the predictivity of human cancer risk by the rodent bioassay has been questioned (Corvi et al., 2017; Marone et al., 2014). Predictivity and human relevance are particularly problematic for non-genotoxic pharmaceuticals. Results of rodent carcinogenicity studies are frequently positive for tumor outcomes but are often concluded to be irrelevant to human risk for reasons of dose, mode of action, or species differences (Bourcier, 2015). The value added by conducting rodent bioassays must be weighed against the extensive use of animals, resources, and time of conducting such assays. Given these considerations, the development of alternative and supplemental approaches that could improve assessment of human carcinogenic risk of new pharmaceuticals is of pressing interest.

An ideal testing strategy would identify carcinogenic hazard by methods more applicable to human biology and establish dose response data which meaningfully informs risk assessment based on exposure and mechanism. A battery of approaches that extracts relevant information from the drug's intended target, drug-specific nonclinical data, and human carcinogenomics is more likely to meet this ideal than any single approach proposed to replace or supplement rodent bioassays.

The WOE strategy recommended for therapeutic proteins does not exclude the need for rodent bioassays, but in most cases a rodent bioassay is not considered necessary to adequately assess carcinogenic risk regardless if testing in rodents is feasible or not. As discussed below in FDA's current efforts, a similar strategy is being evaluated for assessing carcinogenic risk of small molecules as well. Experience with the WOE strategy with biologics has identified a need for additional testing methods that evaluate the hazard presented by immunomodulators.

4.3. New approach methodologies

There are currently no recognized alternatives to the long-term rodent bioassay that simultaneously measures the apical endpoint of interest (i.e., tumor emergence) and provides dose-response information; both are needed for an ideal risk assessment for pharmaceuticals.

One proposed alternative strategy is leveraging certain findings in specific organs from short-term toxicology studies as having positive predictive value for long-term tumor outcome in those organs or as a general indicator of carcinogenic risk to other organs (Boobis et al., 2009; Cohen, 2010). However, alone, this poorly predicts tumor outcome on a whole-animal basis and provides limited dose-response information (Jacobs, 2005).

Mining for cancer correlations in human genomic databases (e.g., Cosmic, OMIM), correlative gene expression signatures, in silico methods, and application of the hallmarks of cancer and key characteristics of carcinogens are additional approaches under active development (Corvi et al., 2017; Hanahan and Weinberg, 2011; Jacobs et al., 2016; Smith et al., 2016). These alternative approaches have the virtue of often being based on human rather than rodent biology, are mechanism-based, and can be of particular value in identifying hazards. However, alone, these genetic and pathway signaling attributes could prove too problematic for predicting cancer risk with sufficient specificity and could result in over-prediction of hazard or risk. Similarly, a more robust understanding of the negative predictive value of these approaches is needed to appropriately interpret an absence of concerning signals using these methods. Extracting dose-response or other point-of-departure information is an additional challenge which importantly informs assessment of risk for a given exposure to a new pharmaceutical.

Another proposed alternative approach is one that emphasizes negative rather than positive predictive value for long-term tumor outcome in rodents based on the lack of certain toxicological findings in short-term studies (Sistare et al., 2011). However, the liberal definition of criteria necessary to achieve reasonable negative predictive value would often result in using findings of minimal relevance to carcinogenic risk (e.g., liver hypertrophy) as a basis for conducting long term studies in rodents. In addition, the potential of missing true rodent carcinogens of potential human relevance based on a prediction of negative outcome using this method presents implications for public safety that are not easily remedied.

The FDA is working in this area through participation at ICH. Efforts are focused on leveraging all available information on the drug target, drug-specific toxicology, and drug class in a WOE evaluation of carcinogenic risk for small molecules prior to determining the need for 2-year rodent bioassays. Of interest, in an on-going, prospective "testing" period of this approach, novel data streams not commonly seen in regulatory submissions, including genomic cancer database screens, have been submitted as part of these WOE assessments. Thus, moving toward a WOE strategy for small molecules may incentivize sponsors to develop or refine alternatives that provide more actionable data on human hazard identification or dose response. The development and vetting of such alternative data streams could support an argument for not conducting 2-year rodent bioassays with their pharmaceutical, should the ICH prospective project show this to be a feasible pathway.

5. Predicting human pregnancy/reproductive risk

5.1. Current approach

Studies that address the developmental and reproductive toxicity (DART) of a drug include assessments of fertility, embryofetal development (EFD), and pre/post-natal development (PPND), conducted using whole animals. For biologics, where nonhuman primates are the only available toxicity model species, effects on fertility can be assessed by including reproductive endpoints in chronic repeat dose toxicity

studies in sexually mature animals. An enhanced PPND can be conducted in nonhuman primates in which EFD and PPND endpoints are combined into a single study.

5.2. Statement(s) of need

The aim of DART studies is to reveal any effect of the pharmaceutical on mammalian reproduction relevant for human risk assessment. Various study designs can be conducted based on the nature of the drug and the intended clinical population. The timing of studies within the pharmaceutical development process depends on the clinical population and phase of pharmaceutical development. For example, embryofetal studies are recommended prior to the initiation of large-scale or long duration clinical trials (ICH M3(R3), 2010). All available pharmacological, toxicokinetic and toxicological data should be considered in determining the most appropriate study designs. To minimize animal use various *in vivo* approaches can be followed. In addition, it may be possible to use appropriately qualified alternative assays in certain scenarios. Adequately sensitive alternative methods to detect the potential teratogenicity of a drug when needed early in a nonclinical development program could serve to support these types of clinical development programs. Definitive developmental and reproductive toxicity studies could then be delayed to later stages of drug development. This approach can reduce the overall number of such DART studies conducted, because not all drug development programs progress to phase 3 clinical trials or marketing.

Some drug classes are not tolerated by pregnant animals such that the ability to conduct *in vivo* developmental and reproductive toxicity studies is limited; alternative assessments would be of particular value in this situation.

5.3. New approach methodologies

Several alternative developmental assays have been developed and continue to be refined. Such assays could potentially complement existing DART studies. These include embryonic stem cell assays (mouse, rat), whole embryo culture assays (rat, mouse, rabbit), studies with *C. elegans* or zebrafish, the frog embryo teratogenesis assay-Xenopus (FETAX) model, and a combination of stem cell assays with various “-omics” approaches. These assays are used in early compound lead selection by pharmaceutical companies to screen candidate drugs for potential teratogenic activity. Some of these assays are in-house while others are commercially available. Reliability and relevance of these models has been assessed to different degrees depending on the context of use. In some assays, decision criteria can be adjusted to provide a desired sensitivity and specificity. Challenges with these assays include understanding the *in vivo* relevance of sometimes relatively simple *in vitro* endpoints, such as cytotoxicity or metabolic changes. Extrapolating *in vitro* concentrations to *in vivo* pharmacokinetic parameters to calculate safety margins to clinical dose(s) is also challenging, although some methods have developed means to make such estimates. Broader discussion in the scientific and regulatory community is ongoing about whether these NAMs for developmental toxicity could add value in specific context of use scenarios beyond early drug candidate screening. Recent publication of the revised ICH S5(R3) (2020) guidance on detection of reproductive and developmental toxicity provides considerations for how to appropriately qualify alternative methods for assessing embryofetal toxicity and on potential use cases for these methods.

6. Special toxicity

The assessment of special toxicity, most commonly defined as ocular, dermal, and phototoxicity studies, has notably evolved over past years more than any other endpoint in nonclinical development programs. Evaluation of these endpoints, once addressed primarily in

whole animals, has progressed to testing paradigms that involve primarily *in vitro* and *ex vivo* assays based on human and/or animal materials.

6.1. Current approach

6.1.1. Ocular irritation

An *in vivo* rabbit eye irritation study is not recommended for topical drug product testing and the FDA recommends using appropriate *in vitro* or *ex vivo* test methods to determine the irritation potential of a drug product. Test guidelines for *in vitro* and *ex vivo* methods for ocular irritation have been developed by the Organisation for Economic Co-operation and Development (OECD) (OECD Guideline 437, 2017; OECD Guideline 438, 2018; OECD Guideline 460, 2017; OECD Guideline 491, 2018; OECD Guideline 492, 2019; OECD Guideline 494, 2019). *In vitro* use reconstructed human cornea-like epithelium is included in these test guidelines. Acceptance of these alternatives has essentially eliminated all *in vivo* eye irritation studies for human drugs for the past several years. Any *in vivo* assays submitted to the Agency today are likely old assays or ones conducted for reasons other than to support human drug development.

6.1.2. Dermal irritation

CDER does not consider an *in vivo* primary dermal irritation test to be appropriate or necessary for topical drug products. If a primary dermal irritation/corrosion assay is warranted, then a 3-dimensional reconstructed tissue model is acceptable (OECD Guideline 439, 2010). The potential for milder forms of dermal irritation can be evaluated during the repeat-dose dermal toxicology study; a standalone skin irritation study is not needed.

6.1.3. Phototoxicity

CDER recommends a phototoxic assessment for drugs which are either applied to the skin or accumulate in the skin and/or the eye. Since UV/Vis absorption initiates a photochemical reaction, a waiver for the conduct of nonclinical phototoxicity assessment may be requested for drug products that do not absorb within the range of natural sunlight (290–700 nm) (ICH S10, 2015).

Topical drug products with significant absorption in the UVB/UVA/visible spectrum (290–700 nm) should be evaluated for nonclinical photoirritation prior to clinical phototoxicity and photoallergy studies. CDER does not recommend nonclinical models of photoallergenicity.

CDER currently accepts *in vitro* and *in vivo* assays to evaluate phototoxicity of a drug. The *in vitro* 3T3 Neutral Red Uptake (NRU) phototoxicity test (OECD Guideline 432, 2019) may be used as a screen for soluble compounds prior to clinical photosafety testing. Because the *in vitro* assay lacks specificity, positive results require confirmation in a qualified photoirritation study. Reconstructed 3D human skin models have potential as a second tier in a testing strategy to confirm positive 3T3 NRU test results. Drug sponsors interested in submitting alternative data should discuss their proposal with CDER prior to data submission as part of the routine consultation process. *In vivo* phototoxicity studies can be conducted in several animal species including guinea pig, rat and mouse. Animal models with non-pigmented and pigmented skin are available. Non-pigmented skin tends to be more sensitive than pigmented skin for detecting phototoxicity. However, pigmented skin could be considered for drugs that bind significantly to melanin. The to-be-marketed clinical formulation of a topical drug product should be evaluated in an *in vivo* phototoxicity assay.

6.1.4. Skin sensitization

CDER recommends skin sensitization testing, but has no requirement for a specific test. However, the current approach to assessing skin sensitization of a drug generally relies on using studies in whole animals (e.g., the Buehler Test and the Magnusson Kligman Guinea Pig Maximization Test). The murine local lymph node assay (LLNA) is a

validated sensitization assay that reduces the number of animals used and refines their treatment. However, many dermal formulations and some dermal irritants give positive results in the LLNA that are not seen in guinea pigs or humans. Because of its higher false positive rate, CDER accepts negative LLNA data without further testing in the guinea pig test. The LLNA can also be used as part of a weight-of-evidence evaluation to discriminate between strong and weak sensitizers.

6.2. Statement(s) of need

As noted above, many of the existing approaches for assessing special toxicity endpoints are already alternatives to traditional animal models.

Protocols for in vitro ocular irritation and phototoxicity assays continue to be refined. These refinements will likely expand the applicability of the assays to additional classes of compounds and may, in some cases, increase the reliability of the assays.

Newer approaches for assessing skin sensitization have made significant progress (discussed below) and further assessment of these assays for more chemical classes and test articles is warranted. For example, CDER is not aware of any validated non-animal method for assessing human skin sensitization of chemical mixtures, and encourages the development of a screening battery of qualified in vitro assays. This is particularly important for topical drug products which are frequently complex mixtures.

6.3. New approach methodologies

Skin sensitization has been described with an adverse outcome pathway (AOP) and is the result of four key events: 1) binding of hapten to endogenous skin proteins, 2) keratinocyte activation, 3) dendritic cell activation, and 4) proliferation of antigen-specific T cells (OECD, 2014). The complexity underlying the biology of the four key events makes it unlikely that a single alternative method will be able to replace the use of animals for skin sensitization testing. Consensus among international experts suggests that the best option for replacing animal testing for skin sensitization is by combining data from several qualified non-animal test methods into an integrated testing strategy (ITS). Various ITS have been proposed, using specified combinations of non-animal methods (in silico, in chemico, in vitro methods using human-derived cells) to compensate for individual test method limitations (Strickland et al., 2017).

The currently qualified in vitro test methods, each limited to assessing only one key event, are provided below:

- Direct peptide reactivity assay (DPRA) (OECD Guideline 442C, 2019): an in chemico test that models the ability of a hapten to bind with human skin protein, forming a hapten-protein complex (key event 1). DPRA reactivity is determined by measuring the percent depletion of synthetic peptides containing either cysteine or lysine. No cells are used but the assay lacks a metabolic system to convert a pre-hapten to a hapten, potentially resulting in false negatives.
- ARE-Nrf2 luciferase test method (KeratinSens™, (OECD Guideline 442D, 2018): a measurement of human-derived keratinocyte activation of gene pathways linked to skin sensitization (e.g., anti-oxidant/electrophile response element (ARE)-dependent pathways, key event 2).
- LuSens is based on the same principle as the ARE-Nrf2 luciferase test method (i.e., is a me-too method), but measures the activation of a different pathway (Nrf2-KEAP1). LuSens is being considered by the OECD test program, but no test guideline has yet been developed.
- Human cell line activation test (h-CLAT; (OECD Guideline 442E, 2018): a measurement of activation of human dendritic cells (key event 3) by quantifying changes in the expression of cell surface markers linked to dendritic cell maturation (i.e., CD86 and CD54).

The three currently qualified in vitro test methods described above can possibly address key events 1–3 of the skin sensitization process described above. However, no in vitro test currently exists for key event 4, in which activated dendritic cells migrate to the nearest lymph node and initiate proliferation of antigen-specific T-cells. In silico read-across QSAR Toolbox software predictions have been used to cover key event 4 because in vivo data from structurally and/or mechanistically similar compounds are available to develop the read-across results (Strickland et al., 2017). The Toolbox is publicly available: <http://www.qsartoolbox.org/>. However, the in vivo data are primarily derived from the LLNA, making read-across prediction subject to the same regulatory limitations as the LLNA itself.

Recent research suggests computational methods are promising tools to identify potential human skin sensitizers without animal testing (Strickland et al., 2017). Integrated testing strategies combining in chemico, in vitro and in silico read-across predictions relying on human reference data are currently being developed (Kleinstreuer et al., 2018) and appear promising.

7. Discussion and conclusions

A desirable goal for CDER is to have reliable nonclinical testing strategies that can be used in regulatory decision making and can provide the information needed to adequately assess the safety of conducting clinical trials with human pharmaceuticals. Highlighting areas of unmet or insufficiently met needs in the current nonclinical development approach is an aim of this paper and is meant to guide stakeholders that are actively developing alternative methodologies. While we have noted some of the possible new approaches that might offer promise and their challenges, this is not a comprehensive effort at describing solutions. We anticipate further discussion and interaction with stakeholders to find acceptable paths forward. Acceptance of any new alternative method will require persuasive scientific evidence that the method improves or otherwise adds value to the current testing strategy and is fit for its intended purpose, so that regulators and the scientific community are confident in the suitability of the new method to inform the safety of human pharmaceuticals and protect public health. We believe it is important to start with the regulatory needs and the safety questions that need to be answered when considering whether a new approach methodology adds value.

Acceptance of alternative methods requires continuous dialogue and feedback between all partners from development to implementation, including review and acceptance by CDER (See FDA's Predictive Toxicology Roadmap). To that end, FDA is an active member of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) which serves to coordinate activities within the federal government relevant to new test method evaluation, acceptance, and use. The goal of ICCVAM is "to establish, wherever feasible, guidelines, recommendations, and regulations that promote the regulatory acceptance of new or revised scientifically valid toxicological tests that protect human and animal health and the environment while reducing, refining, or replacing animal tests and ensuring human safety and product effectiveness." ICCVAM also recently published a roadmap with many parallels to the concepts described in FDA's roadmap.

Context of use is an important concept in the acceptance of NAMs that has been outlined in both the FDA's Predictive Toxicology Roadmap and the ICCVAM roadmap. Context of use refers to a clearly articulated description delineating the manner and purpose of use for a particular approach. Having a clearly defined context use makes it easier for all involved to envision the potential value of an approach and helps define the amount of information needed to adequately qualify the approach. Qualification is a conclusion that the results of an assessment using the model or assay can be relied on to have a specific interpretation and application in product development and regulatory decision-making (see FDA's Predictive Toxicology Roadmap for further discussion of qualification).

The development of adverse outcome pathways for various apical toxicity outcomes can play a role in identifying possible events that can be evaluated in NAMs. As noted above, key events of the skin sensitization AOP have been used in assembling a battery of in chemico and in vitro assays for predicting in vivo sensitization potential. Other AOPs have been developed for a variety of toxicities. These AOPs have been cataloged in the literature or in resources such as the OECD AOP Knowledge Base.

Collaborations such as those with ICCVAM are also a key element of the FDA roadmap. CDER actively collaborates with a number of federal agencies and outside stakeholders in the area of alternative assays. For example, CDER collaborates with the National Center for Advancing Translational Science (NCATS) in developing programs including Tissue Chip for Drug Screening or organs-on-chips, which are 3D platforms engineered to support living human tissues and cells. CDER and other FDA center laboratories are actively evaluating such platforms. One objective is to gain experience with these technologies to understand in what areas of regulatory toxicology they may be of use. There is a need to understand the challenges and limitations of the technology and what aspects may benefit from standardization. These models may hold promise to address some of the areas of need identified above, although the limitations mentioned above must also be considered.

As new approaches and methodologies become accepted into regulatory use, incorporation of such methodologies into regulatory guidance documents can promote more widespread use. The presence of a method or alternative approach, particularly in an international guidance such as those issued by ICH, communicates to the pharmaceutical industry that such an approach will be accepted and reduces uncertainty around the use of the approach. There are several examples where a new approach was incorporated into guidance after the approach was considered acceptable. Some of these approaches are noted above for areas such as phototoxicity (ICH S10, 2015) and QSAR for potentially genotoxic impurities ((ICH M7(R1), 2018)). CDER continues to look for ways to incorporate new methodologies into CDER guidances and into ICH guidances.

CDER also notes that NAMs can potentially contribute to more efficient drug development even when data from such methods are used in the drug discovery phase, in which case regulatory acceptance is not an issue. Methods that allow more rapid and predictive selection of potentially effective and safe drug candidates in this early phase of drug development can contribute to both decreased overall development cost and advancement of the 3Rs. Better early drug selection may require fewer iterative cycles of nonclinical and clinical testing of multiple drug candidates to reach the final goal of demonstrating safety and efficacy.

7.1. Submitting NAMs to CDER

CDER encourages sponsors to submit NAMs and alternative assay studies to FDA. Drug sponsors interested in submitting alternative data should discuss their proposal with the appropriate review division prior to data submission as part of the routine consultation process. One way that sponsors can help the agency gain experience and understanding of a new approach is to submit such studies directly to review divisions concurrently with standard in vivo animal studies (parallel submission review). This enables direct comparison of the new and traditional approaches. In general, the use of benchmark controls with existing in vivo methods is optional, but CDER recommends their use for data interpretation with newer alternative test methods. CDER is exploring other pathways by which information supporting a NAM that is not associated with a specific application could be provided to the Agency for comment and feedback. CDER anticipates that if an appropriate regulatory use is identified, a package of supporting data could be submitted to the agency and a NAM could be qualified for a particular context of use if the data are found to be sufficient. The FDA Office of the Chief Scientist has also established an In Vitro Systems Working Group to facilitate cross-center discussion of in vitro approaches.

External stakeholders can contact this working group to request a webinar with FDA staff to discuss NAMs.

CDER believes that NAMs hold promise at addressing some areas of toxicology that could benefit from improved predictivity while not compromising current safety evaluations. In addition, some NAMs can contribute to the 3Rs, which CDER also supports. CDER looks forward to continued participation in the development and regulatory use of NAMs. We intend to address in future efforts issues such as what metrics can be used to measure the success of NAMs and how NAMs should be evaluated for utility in drug development.

Funding body information

None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- 21CFR312, Code of Federal Regulations. Title 21: Food and Drugs, Subchapter D: Drugs for Human Use, Part 312: Investigational New Drug Application.
- 21CFR314, Code of Federal Regulations. Title 21: Food and Drugs, Subchapter D: Drugs for Human Use, Part 314: Applications for FDA Approval to Market a New Drug.
- 21CFR601, Code of Federal Regulations. Title 21: Food and Drugs, Subchapter F: Biologics, Part 601: Licensing.
- Abbott, N.J., 2013. Blood-brain barrier structure and function and the challenges for CNS drug delivery. *J. Inherit. Metab. Dis.* 36, 437–449.
- Al-Saffar, A., et al., 2015. Gastrointestinal safety pharmacology in drug discovery and development. *Handb. Exp. Pharmacol.* 229, 291–321.
- Amin, N.D., Pasca, S.P., 2018. Building models of brain disorders with three-dimensional organoids. *Neuron* 100, 389–405.
- Authier, S., et al., 2009. Video-electroencephalography in conscious non human primate using radiotelemetry and computerized analysis: refinement of a safety pharmacology model. *J. Pharmacol. Toxicol. Methods* 60, 88–93.
- Bailey, E.L., et al., 2012. Simultaneous assessment of blood pressure and respiratory function in dogs: validation of a non-invasive telemetry method. *J. Pharmacol. Toxicol. Methods* 66, 187.
- Barros, T.P., et al., 2008. Zebrafish: an emerging technology for in vivo pharmacological assessment to identify potential safety liabilities in early drug discovery. *Br. J. Pharmacol.* 154, 1400–1413.
- Bhatia, S.N., Ingber, D.E., 2014. Microfluidic organs-on-chips. *Nat. Biotechnol.* 32, 760–772.
- Bonini, S., Rasi, G., 2016. First-in-Human clinical trials - what we can learn from tragic failures. *N. Engl. J. Med.* 375, 1788–1789.
- Boobis, A.R., et al., 2009. A data-based assessment of alternative strategies for identification of potential human cancer hazards. *Toxicol. Pathol.* 37, 714–732.
- Borsook, D., et al., 2013. Use of functional imaging across clinical phases in CNS drug development. *Transl. Psychiatry* 3, e282.
- Bourcier, T.R.D., 2015. Addressing positive findings in carcinogenicity studies. In: Graziano, M.J., Jacobson-Kram, D. (Eds.), *Genotoxicity and Carcinogenicity Testing of Pharmaceuticals*. Springer Int Pub, Switzerland, pp. 159–182.
- Butler, L.D., et al., 2017. Current nonclinical testing paradigms in support of safe clinical trials: an IQ Consortium DruSafe perspective. *Regul. Toxicol. Pharmacol.* 87 (Suppl. 3), S1–S15.
- Carmichael, O., et al., 2018. The role of fMRI in drug development. *Drug Discov. Today* 23, 333–348.
- Cavero, I., Holzgreffe, H., 2015. CiPA: ongoing testing, future qualification procedures, and pending issues. *J. Pharmacol. Toxicol. Methods* 76, 27–37.
- Cavero, I., et al., 2019. Human organotypic bioconstructs from organ-on-chip devices for human-predictive biological insights on drug candidates. *Expert Opin. Drug Saf.* 18, 651–677.
- Celine de Esch, R.S., Wolterbeek, André, Woutersen, Ruud, de Groot, Didima, 2012. Zebrafish as potential model for developmental neurotoxicity testing: a mini review. *Neurotoxicol. Teratol.* 34, 545–553.
- Chen, R., et al., 2018. Evaluating the use of KIM-1 in drug development and research following FDA qualification. *Clin. Pharmacol. Ther.* 104, 1175–1181.
- Clark, M., 2015. Prediction of clinical risks by analysis of preclinical and clinical adverse events. *J. Biomed. Inf.* 54, 167–173.
- Clark, M., Steger-Hartmann, T., 2018. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul. Toxicol. Pharmacol.* 96, 94–105.
- Cohen, S.M., 2010. Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay is no longer necessary. *Toxicol. Pathol.* 38, 487–501.

- Corvi, R., et al., 2017. Moving forward in carcinogenicity assessment: report of an EURL ECVAM/ESTIV workshop. *Toxicol In Vitro* 45, 278–286.
- Di Lullo, E., Kriegstein, A.R., 2017. The use of brain organoids to investigate neural development and disease. *Nat. Rev. Neurosci.* 18, 573–584.
- Ewart, L., et al., 2018. Application of microphysiological systems to enhance safety assessment in drug discovery. *Annu. Rev. Pharmacol. Toxicol.* 58, 65–82.
- FDA, 2019. FDA's predictive toxicology roadmap. <https://www.fda.gov/science-research/about-science-research-fda/fdas-predictive-toxicology-roadmap>, Accessed date: 22 October 2019.
- Frank, R., 1999. Nuclear biomaging in drug development and regulatory review. *J. Clin. Pharmacol.* 39, 51S–55S.
- Grimaldi, C., et al., 2016. Cytokine release: a workshop proceedings on the state-of-the-science, current challenges and future directions. *Cytokine* 85, 101–108.
- Haggerty, G.C., 1991. Strategy for and experience with neurotoxicity testing of new pharmaceuticals. *J. Am. Coll. Toxicol.* 10.
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Hiemstra, P.S., et al., 2018. Human lung epithelial cell cultures for analysis of inhaled toxicants: lessons learned and future directions. *Toxicol In Vitro* 47, 137–146.
- Hunt, P.R., et al., 2018. C. elegans Development and Activity Test detects mammalian developmental neurotoxins. *Food Chem. Toxicol.* 121, 583–592.
- ICH S6(R1), 2012. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S6 Addendum to Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals.
- ICH S7A, 2001. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S7A Safety Pharmacology Studies for Human Pharmaceuticals.
- ICH S7B, 2005. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals.
- ICH M3(R2), 2010. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: M3(R2) Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals.
- ICH M7(R1), 2018. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: M7(R1) Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk.
- ICH S1A, 1996. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S1A the Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals.
- ICH S1B, 1998. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S1B Testing for Carcinogenicity of Pharmaceuticals.
- ICH S1C(R2), 2008. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S1C(R2) Dose Selection for Carcinogenicity Studies of Pharmaceuticals.
- ICH S2(R1), 2012. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S2(R1) Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use.
- ICH S3(A), Q.A., 2018. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: Note for Guidance on Toxicokinetics: the Assessment of Systemic Exposure in Toxicity Studies: Focus on Microsampling – Questions and Answers.
- ICH S5(R3), 2020. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Harmonised Guideline Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals.
- ICH S6, 1997. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals.
- ICH S10, 2015. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guidance for Industry: S10 Photosafety Evaluation of Pharmaceuticals.
- Interagency Coordinating Committee, 2019. The validation of alternative methods roadmap. <https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy/index.html>, Accessed date: 22 October 2019.
- 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.
- Jacobs, A., 2005. Prediction of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? *Toxicol. Sci.* 88, 18–23.
- Jacobs, M.N., et al., 2016. International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. *ALTEX* 33, 359–392.
- Kanungo, J., et al., 2014. Zebrafish model in drug safety assessment. *Curr. Pharmaceut. Des.* 20, 5416–5429.
- Kelava, I., Lancaster, M.A., 2016. Dishing out mini-brains: current progress and future prospects in brain organoid research. *Dev. Biol.* 420, 199–209.
- Kleinstreuer, N.C., et al., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches (*). *Crit. Rev. Toxicol.* 48, 359–374.
- Kramer, K., Kinter, L.B., 2003. Evaluation and applications of radiotelemetry in small laboratory animals. *Physiol. Genom.* 13, 197–205.
- Kruhlak, N.L., et al., 2012. (Q)SAR modeling and safety assessment in regulatory review. *Clin. Pharmacol. Ther.* 91, 529–534.
- Lacroix, G., Koch, W., Ritter, D., Gutleb, A.C., Larsen, S.T., Lorent, T., Zanetti, F., Constant, S., Chortarea, S., Rothen-Rutishauser, B., Hiemstra, P.S., Frejafon, E., Hubert, P., Gribaldo, L., Kearns, P., Aublant, J., Diabaté, S., Weiss, C., de Groot, A., Kooter, I., 2018. Air-liquid interface in vitro models for respiratory toxicology research: consensus workshop and recommendations. 4, 91–106.
- Lancaster, M.A., Knoblich, J.A., 2014. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* 345, 1247125.
- Leedale, J., et al., 2018. A combined in vitro/in silico approach to identifying off-target receptor toxicity. *iScience* 4, 84–96.
- Maass, C., et al., 2019. Translational assessment of drug-induced proximal tubule injury using a kidney microphysiological system. *CPT Pharmacometrics Syst. Pharmacol.* 8, 316–325.
- Marone, P.A., et al., 2014. Reassessing the two-year rodent carcinogenicity bioassay: a review of the applicability to human risk and current perspectives. *Regul. Toxicol. Pharmacol.* 68, 108–118.
- Marx, U., et al., 2016. Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing. *ALTEX* 33, 272–321.
- Matthews, P.M., et al., 2006. Applications of fMRI in translational medicine and clinical practice. *Nat. Rev. Neurosci.* 7, 732–744.
- Mattsson, J.L., Spencer, P.J., Albee, R., 1996. A performance standard for clinical and functional observational battery examinations of rats. *J. Am. College Toxicol.* 15.
- Mead, A.N., et al., 2016. Assessing the predictive value of the rodent neurofunctional assessment for commonly reported adverse events in phase I clinical trials. *Regul. Toxicol. Pharmacol.* 80, 348–357.
- Monticello, T.M., et al., 2017. Current nonclinical testing paradigm enables safe entry to First-In-Human clinical trials: the IQ consortium nonclinical to clinical translational database. *Toxicol. Appl. Pharmacol.* 334, 100–109.
- Murphy, D.J., 1994. Safety pharmacology of the respiratory system. *Tech. Study Des.* 32, 237–246.
- Muster, W., et al., 2008. Computational toxicology in drug development. *Drug Discov. Today* 13, 303–310.
- Niemeyer, J.E., 2016. Telemetry for small animal physiology. *Lab. Anim. (NY)* 45, 255–257.
- OECD, 2014. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. OECD Series on Testing and Assessment No. 168.
- OECD Guideline 432, 2019. In Vitro 3T3 NRU Phototoxicity Test.
- OECD Guideline 437, 2017. Bovine Corneal Opacity and Permeability Test Method for Identifying I) Chemicals Inducing Serious Eye Damage and Ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.
- OECD Guideline 438, 2018. Isolated Chicken Eye Test Method for Identifying I) Chemicals Inducing Serious Eye Damage and Ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.
- OECD Guideline 439, 2010. In Vitro Skin Irritation.
- OECD Guideline 442C, 2019. In Chemico Skin Sensitisation: Assays Addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins.
- OECD Guideline 442D, 2018. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method.
- OECD Guideline 442E, 2018. In Vitro Skin Sensitisation: in Vitro Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation.
- OECD Guideline 460, 2017. Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants.
- OECD Guideline 491, 2018. Short Time Exposure In Vitro Test Method for Identifying I) Chemicals Inducing Serious Eye Damage and Ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.
- OECD Guideline 492, 2019. Reconstructed Human Cornea-like Epithelium (RhCE) Test Method for Identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage.
- OECD Guideline 494, 2019. Vitrigel-Eye Irritancy Test Method for Identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage.
- Olson, H., et al., 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* 32, 56–67.
- Pamies, D., et al., 2017. A human brain microphysiological system derived from induced pluripotent stem cells to study neurological diseases and toxicity. *ALTEX* 34, 362–376.
- Porsolt, R.D., et al., 2002. New perspectives in CNS safety pharmacology. *Fundam. Clin. Pharmacol.* 16, 197–207.
- Quadrato, G., et al., 2016. The promises and challenges of human brain organoids as models of neuropsychiatric disease. *Nat. Med.* 22, 1220–1228.
- Rognan, D., 2017. The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacol. Ther.* 175, 47–66.
- Shah, F., et al., 2015. Setting clinical exposure levels of concern for drug-induced liver injury (DILI) using mechanistic in vitro assays. *Toxicol. Sci.* 147, 500–514.
- Sistare, F.D., et al., 2011. An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current. *Regul. Guidel.* 39, 716–744.
- Smith, M.T., et al., 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environ. Health Perspect.* 124, 713–721.
- Stevens, J.L., Baker, T.K., 2009. The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov. Today* 14, 162–167.
- Strickland, J., et al., 2017. Multivariate models for prediction of human skin sensitization hazard. *J. Appl. Toxicol.* 37, 347–360.
- Temme, S.J., et al., 2016. Using induced pluripotent stem cells to investigate complex genetic psychiatric disorders. *Curr. Behav. Neurosci. Rep.* 3, 275–284.
- Utrecht, J., Naisbitt, D.J., 2013. Idiosyncratic adverse drug reactions: current concepts. *Pharmacol. Rev.* 65, 779–808.
- Vicente, J., et al., 2018. Mechanistic model-informed proarrhythmic risk assessment of drugs: review of the “CiPA” initiative and design of a prospective clinical validation

- study. *Clin. Pharmacol. Ther.* 103, 54–66.
- Vite, C., 2005. In: *Braund's Clinical Neurology in Small Animals: Localization, Diagnosis and Treatment*. International Veterinary Information Service. www.ivis.org.
- Wallis, R., et al., 2018. CiPA challenges and opportunities from a non-clinical, clinical and regulatory perspectives. An overview of the safety pharmacology scientific discussion. *J. Pharmacol. Toxicol. Methods* 93, 15–25.
- Will, Y., Dykens, J., 2014. Mitochondrial toxicity assessment in industry—a decade of technology development and insight. *Expet Opin. Drug Metabol. Toxicol.* 10, 1061–1067.
- Yu, H., et al., 2017. The contributions of human mini-intestines to the study of intestinal physiology and pathophysiology. *Annu. Rev. Physiol.* 79, 291–312.