



Predicting the safety of medicines in pregnancy: A workshop report

J.M. Clements^a, R.G. Hawkes^{a,*}, D. Jones^a, A. Adjei^a, T. Chambers^a, L. Simon^a, H. Stemplewski^a,
N. Berry^b, S. Price^c, M. Pirmohamed^d, A.H. Piersma^e, G. Waxenecker^f, P. Barrow^g,
M.E.W. Beekhuijzen^h, A. Fowkesⁱ, H. Prior^j, F. Sewell^j

^a Medicines and Healthcare products Regulatory Agency, London, UK

^b National Institute for Biological Standards and Control, Potters Bar, UK

^c University of Surrey, UK

^d University of Liverpool, UK

^e National Institute for Public Health and the Environment (RIVM), Center for Health Protection, Bilthoven, Netherlands

^f Austrian Medicines and Medical Devices Agency, Vienna, Austria

^g Roche Pharmaceutical Research and Early Development, Basel, Switzerland

^h Charles River, 's-Hertogenbosch, Netherlands

ⁱ Lhasa Limited, Leeds, UK

^j National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), London, UK

ARTICLE INFO

Keywords:

Adverse outcome pathway
Alternative assay
Data sharing
Embryo-fetal development
In silico
Medicines regulation
Predictive models
Pregnancy

ABSTRACT

The framework for developmental toxicity testing has remained largely unchanged for over 50 years and although it remains invaluable in assessing potential risks in pregnancy, knowledge gaps exist, and some outcomes do not necessarily correlate with clinical experience. Advances in omics, *in silico* approaches and alternative assays are providing opportunities to enhance our understanding of embryo-fetal development and the prediction of potential risks associated with the use of medicines in pregnancy. A workshop organised by the Medicines and Healthcare products Regulatory Agency (MHRA), “Predicting the Safety of Medicines in Pregnancy – a New Era?”, was attended by delegates representing regulatory authorities, academia, industry, patients, funding bodies and software developers to consider how to improve the quality of and access to nonclinical developmental toxicity data and how to use this data to better predict the safety of medicines in human pregnancy. The workshop delegates concluded that based on comparative data to date alternative methodologies are currently no more predictive than conventional methods and not qualified for use in regulatory submissions. To advance the development and qualification of alternative methodologies, there is a requirement for better coordinated multidisciplinary cross-sector interactions coupled with data sharing. Furthermore, a better understanding of human developmental biology and the incorporation of this knowledge into the development of alternative methodologies is essential to enhance the prediction of adverse outcomes for human development. The output of the workshop was a series of recommendations aimed at supporting multidisciplinary efforts to develop and validate these alternative methodologies.

1. General introduction

Evidence on the safety and efficacy of medicines in pregnant women is generally lacking. Despite this, there is widespread use of medicines in pregnancy, for reasons such as ongoing treatment for pre-existing medical conditions, the development of medical conditions during pregnancy or inadvertent exposure. Current estimates suggest that the prevalence of prescription drug use during pregnancy ranges from 60 to 90% [1–5], even though only a small number of marketed products are specifically classed as safe to use in pregnancy. For many medicines,

patients and clinicians cannot make informed benefit-risk decisions because of lack of evidence, which is largely the result of ethical considerations that exclude pregnant women from clinical trials due to concerns over fetal/infant risk. Consequently, nonclinical data remains an integral part of evaluating the benefit-risk for medicines in pregnancy until it is superseded by clinical data captured through post-marketing surveillance.

Whilst the nonclinical testing of developmental toxicity in animals remain the gold standard for assessing potential risk, outcomes do not necessarily correlate with clinical experience and remain largely

* Corresponding author.

E-mail address: ross.hawkes@mhra.gov.uk (R.G. Hawkes).

<https://doi.org/10.1016/j.reprotox.2020.02.011>

Received 14 November 2019; Received in revised form 10 February 2020; Accepted 26 February 2020

Available online 29 February 2020

0890-6238/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

qualitative with limited mechanistic understanding. Such limitations have been recognised globally by researchers, industry and regulators, and initiatives developed to consider how the evidence base can be improved to strengthen our confidence in predicting and communicating potential risks associated with use of medicines in pregnancy.

In 2017, a UK Commission on Human Medicines (CHM) Expert Working Group [6] produced a set of recommendations aimed at improving the evidence base for medicines taken during pregnancy and lactation. One such recommendation was the delivery of a workshop to consider how nonclinical data could be made more predictable and accessible, as well as the feasibility of using computer modelling and molecular structural alerts to generate safety signals from *in vivo* and *in vitro* data. This workshop was held in January 2019.

Given the diverse complexities of the subject matter, multidisciplinary experts representing industry, academia, contract research organisations (CROs), regulatory bodies, patients and software development companies were invited to discuss the status of the current developmental toxicity testing paradigm and the challenges and opportunities presented when considering novel approaches for the purpose of improving the predictability of nonclinical data.

For medicines, the regulatory assessment of developmental and reproductive toxicity (DART) is routinely based on a testing paradigm involving studies in mammalian species and although this has evolved over the years, the overall base set of requirements has not changed greatly since its introduction in the 1960s. It comprises a combination of studies covering conception through embryo-fetal stages, birth and sexual maturation of test animals. The rat is usually used in all reproductive toxicity studies, with the rabbit typically used as the second species in the embryo-fetal development toxicity studies. Such studies typically require the use of between 16–20 litters per dose group for rodents and rabbits [7] are expensive, and need to have been conducted prior to certain stages of clinical development, depending on the population being exposed [8]. At least one of the test species should be pharmacologically relevant (responsive to the primary pharmacodynamic effects of the test compound). In cases where these species are considered unsuitable an alternative species may be considered, for example replacing rodents and rabbits with non-human primates (NHP) for testing biotherapeutics. The use of NHPs in these studies however has limitations.

Since the implementation of the ICH guideline on reproductive toxicology in 1993, not only has experience been gained with the testing of pharmaceuticals using the current and novel testing paradigms, but scientific, technological and regulatory knowledge has also significantly evolved, as acknowledged by the ongoing revision of the ICH S5 (R3) guideline [63]. Examples of such advancements include the development of alternative assays, *in silico* modelling and a better understanding of developmental adverse outcomes pathways (AOPs). Consequently, there are now potential opportunities to modernise the current testing paradigm, improve predictivity, re-evaluate opportunities to reduce animal use [9] and most importantly improve the safeguarding of public health through the generation of more clinically relevant data.

Given the current revision of ICH S5, the emergence of new technologies over recent years, and an opportunity to seek the opinions of multidisciplinary cross-sector experts, the workshop considered the following questions:

- 1 Can the current testing strategy for developmental toxicity be improved?
- 2 What alternative methods (assays and *in silico*) are available, how can they strengthen the current strategy and what are the challenges and opportunities?
- 3 How can stakeholders collaborate to help implement these strategies by sharing data and what are the challenges in sharing such data?

2. Limitations of current state and future opportunities

The present developmental toxicity testing models, involving the dosing of pregnant rodents and non-rodents and evaluation of fetuses and pups, have a long history of use and are considered the most definitive available. Despite this, due to species differences, there remains residual uncertainty about the interpretation of some findings for human risk assessment. These relate to a range of factors including differences in species pharmacology and physiology, translation of animal exposures to the clinical situation, differences in pharmacokinetic/pharmacodynamic (PK/PD) in pregnancy, the use of maternal rather than fetal exposure in risk assessment and that conventional studies provide limited mechanistic insight. Advances in technology mean that biological systems may be interpreted by increasingly data-rich quantitative processes requiring new multidisciplinary approaches to analyse outputs. Taking advantage of these advances presents opportunities for improving the current developmental toxicity testing protocols. It was conjectured how we might evaluate the possible effects of medicines on embryo-fetal development in humans if we made a fresh start using modern technology. Two approaches were considered, an “*evolutionary*” approach adapting existing technology or a “*revolutionary*” approach. An *evolutionary* approach might include refinements of the current paradigm to use existing information about a drug to generate hypotheses that could be tested using customised models and protocols. Modifications could include choosing a model that is pharmacodynamically like humans or adjusting the dosing regimen so that the animal exposure is more reflective of human pharmacokinetics. Alternative methods could also be introduced alongside standard protocols to complement conventional methods to develop a mechanistic understanding of adverse findings in an animal study. A *revolutionary* approach might involve replacing existing developmental toxicity testing methods with a hypothesis-driven approach where specific questions are answered with targeted experiments using a range of human-relevant assays rather than animal studies. A central theme in this approach is the knowledge that adverse events are the end result of a series of linked effects that ultimately lead to pathogenic events.

The AOP concept provides the description of an individual mechanistic pathway from the molecular initiating event (MIE), linked by all key events measured at various levels of organisation through to the resulting adverse outcome at the organism and population levels. AOPs are linear, reductive models of complex physiology and do not factor in interactions with other pathways, but they are useful for understanding how chemicals exert their toxic effects. Complex computational network modelling can be utilised to map AOPs networks and can subsequently be used for toxicity prediction purposes when used in conjunction with all relevant existing biological, toxicological and chemical knowledge. Mapping such relationships in an ontology can provide a useful tool to link molecular information to traditional toxicology outputs and human adverse outcomes. Realistically, a useful, fully AOP *in silico* driven approach for human developmental toxicity hazard assessment still requires major progress but the principles of what such a system might look like are being established. Information held within an AOP can also be used to inform and develop *in vitro/in silico*-based testing strategies based on the key events [10]. A challenge for the practical applications of AOP approaches are limitations in the current understanding of the mechanisms of developmental toxicity. The concept of defining the key characteristics of chemicals that cause reproductive toxicity as an approach for organising and evaluating mechanistic/AOP evidence is useful and could be applied to developmental toxicity [11,12].

Currently, whilst a complete *revolutionary* approach cannot be implemented straight away, advances in the field of alternative methods could be utilised now and help to contribute towards the eventual modification of the current testing paradigm. In order to progress the advancement of such innovative approaches, numerous factors need to

be considered that were acknowledged in more detail in subsequent talks and breakout sessions during the workshop and discussed below.

3. Overcoming regulatory barriers

Currently, the ‘gold standard’ for predicting effects of small molecule medicines on human reproduction and development is a package of studies in a rodent (usually rat) to evaluate effects on fertility, embryo-fetal development and postnatal development, with the rabbit usually used as the second species in embryo-fetal development studies. If needed, reassurances over the clinical relevance of findings in these studies can be further investigated by conducting additional mechanistic studies. Whilst alternative assays are typically used for screening purposes, they can also be employed for such mechanistic studies and have shown utility in characterising pharmacological targets, in applying hypothesis-driven ‘mode of action’ strategies, in testing metabolites and investigating unexpected toxic effects in animal species.

Gaining regulatory acceptance of alternative assays is challenging and to date, official validation efforts of developmental toxicity assays by European Centre for the Validation of Alternative Methods (ECVAM), Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) has been slow [13–19]. This is due in part to limited knowledge on the applicability domain of individual assays, their reductionist nature as compared to the intact organism, the classic approach towards one to one replacement of animal studies with alternative methods, and their validation against animal data as the gold standard, which can be of limited relevance to the situation in humans [20]. To reduce the uncertainties of alternative non-mammalian *in vivo* assays and cell/tissue based *in vitro* method predictions, it is necessary to apply good scientific, technical and quality practices. The Organisation for Economic Co-operation and Development (OECD) guidance document 34 [21] sets out the basic principles and processes required for the validation and acceptance of animal and non-animal test methods for regulatory hazard assessment purposes (Table 1). The European Medicines Agency (EMA) also provides guidance on the criteria for regulatory acceptance of 3Rs approaches [22].

The development and adoption of fully validated assays for regulatory purposes is clearly hindered by stringent acceptance criteria. This has subsequently led to the consideration on whether fully validated assays are required and instead whether ‘qualified’ assays can be used in specific circumstances. This concept is addressed in the latest ICH S5 revision (R3) draft document [63], which at the time of writing, provides a framework and testing scheme to facilitate the qualification of alternative assays, including a list of test compounds. The ICH Reference Compound List provides information on embryo-fetal toxicity for various reference compounds covering multiple drug classes. The applicability domain together with the intended regulatory context of use influences the factors for assay qualification and the rigor for achieving regulatory acceptance.

In incorporating new tests into regulatory risk assessment, it is

Table 1
Selected principles and criteria for test method validation from OECD Guideline 34 [21].

Criteria for test method development
The rationale for the test method should be available
The relationship between the test method's endpoint(s) and the (biological) phenomenon
A detailed protocol for the test method should be available
The intra-, and inter laboratory reproducibility of the test method
The test method's performance should be based on the testing of reference chemicals representative of the types of substances
The performance of the test method from the species of concern, and existing relevant toxicity testing data
Ideally, principles of GLP
All data should be available for expert review

essential that the method has clearly defined and scientifically valid end points. The relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose needs to be carefully considered. In integrating alternative methods, the context of use is a key consideration [23]. For example, a small molecule drug with a novel target for a chronic lifestyle indication would warrant more safety data than a therapeutic for a life-threatening indication with established developmental effects. Evidence for adequate integration of alternative assays needs to be based on scientific principles and practices such as avoiding confounding toxic effects through use of high doses/concentrations. One limitation is the current lack of human developmental toxicity information, which means *in vivo* animal data will remain for the present the best reference for qualifying alternative assays for the majority of compounds. Increased focus on obtaining high-quality human pregnancy outcome data will provide an opportunity to compare new methods to the intended species (*i.e.* human) rather than benchmark against animals.

There needs to be a consistent approach across the regulatory regions to qualifying alternative assays to avoid fragmented regional differences and thus, early interactions with the regulator are strongly encouraged. The EMA for example offers scientific advice to support the qualification of innovative methods for a specific intended use in the context of research and development into pharmaceuticals [24]. Regulatory advice regarding new technologies and methods can also be sought from individual National Competent Authorities (NCAs). The Medicines and Healthcare products Regulatory Agency (MHRA) can provide advice through either the Innovations Office or Scientific Advice Meetings, whilst the US Food and Drug Administration (FDA) can offer advice through discussions with the Center for Drug Evaluation and Research (CDER) via the Critical Path Innovation Meetings (CPIM). It is envisaged that alternative assays could be incorporated into the current paradigm and used to complement conventional models to investigate areas of residual uncertainty. For some medicines, alternative approaches could be incorporated into a tiered strategy to defer conventional studies in development and add mechanistic understanding to provide greater human relevance.

4. Alternative techniques for developmental toxicity testing

The development of alternative models for developmental toxicity testing can be seen as an attractive opportunity for the pharmaceutical industry, not only from a cost saving perspective of minimising expensive animal studies that typically take place during the latter stages of drug development, but also by speeding up drug candidate identification by allowing for better informed decisions earlier in the drug development process. However, the development of such models is a resource intensive exercise requiring clear regulatory guidance and it is for these reasons that only a handful of alternative models for developmental toxicity have been developed over the last 20 years. Some of these alternative models have been employed by the pharmaceutical industry for numerous years, but generally the data have been used for hazard identification and internal decision making only and not for regulatory submission purposes [25]. Examples of models that have been successfully developed and employed include rodent whole embryo cultures (rWEC), zebrafish embryo, mouse embryonic stem cells (mESC) and human embryonic stem cells (hESC). Advantages and disadvantages of these models are outlined in Table 2.

To counter predictivity concerns surrounding the use of non-human tissue, alternative methods can be fine-tuned to increase clinical relevance if developed from biological materials of human origin, and furthermore can provide mechanistic information that can help to better understand human hazard [26]. However, the development and validation of human-based alternative assays is hampered by the absence of sufficient data for teratogens in humans. Thus, new assays can only be qualified against animal data, meaning new models with superior predictivity for the human are penalised because of incomplete

Table 2

Advantages and disadvantages of the rWEC, zebrafish embryo, mESC and hESC alternative models.

Model	Advantages	Disadvantages
Rodent whole embryo cultures	Allows real-time monitoring of morphogenesis e.g. organ development Culture drug concentration can be equated to <i>in vivo</i> TK parameters (e.g. Cmax)	It is not animal-free i.e. requires donor females and maternal serum Does not consider maternal PK or placental influences Only considers a limited period of development i.e. late stage developmental teratogens may not be identified
Zebrafish embryo	Utilises an intact complex organism Embryonic development occurs outside of mother Visualisation of effects can be easily observed due to transparency of organism and availability of fluorescent stains Rapid development and easy breeding increases throughput Possible to automate assay Genomic profile similar to humans	Morphological and physiological differences from human Species-specific PK with some drugs Bioanalysis necessary Does not consider maternal PK or placental influences
Mouse embryonic stem cells	Animal-free Has high specificity Is high-throughput and can be automated	Sensitivity of assay may not be suitable for regulatory purposes Does not consider maternal PK or placental influences Only examines a single cell type for a short duration
Human embryonic stem cells	Animal-free Human origin Has high specificity Is high-throughput and can be automated	Sensitivity of assay may not be suitable for regulatory purposes Does not consider maternal PK or placental influences Only examines a single cell type for a short duration

Abbreviations: Cmax = maximum/peak drug concentration; PK = pharmacokinetic; TK = toxicokinetic.

concordance with existing animal models. Additionally, assessing predictive capacity is difficult for novel approaches which are based on mechanism, thus we may need to move away from validation against animal data with its potential shortcomings to defining what an ideal test would identify [27]. This approach may rely on a battery or combination of tests that cover the multiple steps in a mechanistic pathway.

A study performed by Roche aimed to assess the concordance of results from mESC, hESC and zebrafish embryo assays with *in vivo* rat/rabbit data using 20 reference compounds demonstrated that each individual alternative assay failed to detect ≥ 1 human teratogen and ≥ 1 rat/rabbit teratogen (Table 3). This limited dataset highlights the challenges of cross-species concordance and correlating *in vivo/in vitro* exposures. Another consistent problem with alternative assays, irrespective of the species, is the difficulty in extrapolating the clinical relevance of the exposure-effect relationship, which in turn makes the transition from hazard identification to risk assessment challenging. In an ideal scenario, newly developed alternative assays will have the

ability to accurately predict all known human teratogens and embryolethal agents, including maternal drug metabolites, in addition to effects on the placenta, whilst having the ability to allow for exposure comparison for risk assessment purposes.

It is generally acknowledged that the development, validation, acceptance and integration of such assays into the testing strategy will take time to achieve. In the short-term, it is more likely that alternative approaches will be used for candidate screening purposes and subsequently used to complement DART regulatory submission packages once confidence in their predictability and regulatory acceptance has been achieved. Ideally, the future of developmental toxicity testing will present itself as an integrated testing strategy whereby some currently accepted animal-based models could be replaced with a battery of alternative testing models covering a specific toxicological domain, that are animal-free, more predictive of the human situation, accepted by the health authorities as defined by regulatory guidance and are offered by several CROs.

Table 3

Roche's alternative assay comparison data.

Compound ^b	<i>In vivo</i> ^a	mESC	hESC	ZF
Phenytoin	Teratogenic	Teratogenic ⁺	Not teratogenic	Inconclusive
Levetiracetam	Inconclusive	Inconclusive	Inconclusive	Not teratogenic
Bosentan	Teratogenic	Not teratogenic	Not teratogenic	Not teratogenic
Clozapine	Not teratogenic	Inconclusive	Not teratogenic ⁺	Teratogenic
Lithium	Teratogenic	Teratogenic ⁺	Inconclusive	Not teratogenic
Ganciclovir	Teratogenic	Not teratogenic	Teratogenic ⁺	Inconclusive
Garcinol	–	Not teratogenic	–	Not teratogenic
Zebularine	–	Not teratogenic	–	Inconclusive
RO-1	Teratogenic	Teratogenic ⁺	Not teratogenic	Teratogenic ⁺
RO-2	Not teratogenic	Not teratogenic ⁺	Not teratogenic ⁺	Not teratogenic ⁺
RO-3	Teratogenic	Teratogenic ⁺	Teratogenic ⁺	Not teratogenic
RO-4	Not teratogenic	Teratogenic	Not teratogenic ⁺	Teratogenic
RO-5	Not teratogenic	Not teratogenic ⁺	Teratogenic	Not teratogenic ⁺
RO-6	Not teratogenic	Not teratogenic ⁺	Not teratogenic ⁺	Teratogenic
RO-7	Not teratogenic	Not teratogenic ⁺	Not teratogenic ⁺	Not teratogenic ⁺
RO-8	Teratogenic	Not teratogenic	Teratogenic ⁺	Not teratogenic
RO-9	Teratogenic	Teratogenic ⁺	Not teratogenic	Teratogenic ⁺
RO-10	Teratogenic	Teratogenic ⁺	Teratogenic ⁺	Not teratogenic
RO-11	Not teratogenic	Not teratogenic ⁺	Teratogenic	Teratogenic
RO-12	Not teratogenic	Not teratogenic ⁺	Not teratogenic ⁺	Teratogenic
Concordance		75 %	63 %	33 %

Abbreviations: hESC = human embryonic stem cell assay; mESC = mouse embryonic stem cell assay; ZF = zebrafish embryo assay; ⁺ = compares with *in vivo* outcome; RO- = Roche investigatory compound.

^a = in-house rat/rabbit or published human data.

^b = compounds were selected on the basis of pre-existing mEST and *in vivo* data i.e. unblinded. hEST and ZF data were blinded.

In summary, the pharmaceutical industry welcomes the development and use of alternative techniques for developmental toxicity prediction and acknowledges the potential benefits of reducing animal usage whilst increasing the efficiency of the drug discovery process. However, there is common agreement amongst industry and regulators that current alternative methods are still in their infancy and that further investment in both time and money is required along with the development of appropriate regulatory guidance to advance this field of research.

5. Adverse outcome pathways for developmental toxicity

The specificity of the manifestations of embryo-fetal toxicity and teratogenicity may vary greatly between species [28–31], which raises questions on the translatability of some findings and gives rise for a need to consider alternative, more human-relevant, end-points. The AOP concept is increasingly being recognised as a framework which can be used to develop integrated approaches to testing and safety assessment. The AOP itself is the description of the pathway characterised by the identification of a molecular initiating event followed by a series of key events (KE) that lead to the adverse outcome (AO) at the organism and population levels (Fig. 1). AOPs can be used to develop assays modelled on the key event relationships that can predict the adverse outcomes and can facilitate the translation of mechanistic data into outcomes meaningful to safety assessment.

The development of ‘omics’ technologies are essential for this framework to be successful, in order to identify critical developmental pathways and define the key components of the AOP. These ‘big data’ sets fall into various subdivisions, depending on the cellular component measured (e.g. genome, transcriptome, proteome, and metabolome). The range and sensitivity of these technologies are providing good opportunities to identify molecular perturbations leading to toxicity and will play an important role in establishing AOP-based risk assessments. These can be applied not only to tissues and cell cultures but also to single cells with technologies such as single-cell RNA sequencing. This is being used to obtain cell profiles which provides higher precision and opens new opportunities to understand cell-specific alterations and pathways [32]. These data rich technologies will require the parallel development of computational or *in silico* biology and mathematical modelling to understand the underlying biologic processes. *In silico* approaches are essential to cope with the increasing quantity and quality of biologic information, but also for modelling purposes.

In this testing paradigm, *in vitro* models will provide the main biological systems for toxicity testing instead of traditional whole-animal models. Fig. 2 describes the basic concept, in which a limited number of critical key events (represented by stars) are monitored using dedicated *in vitro* assays. The results thereof are integrated with chemical and

kinetic information into an *in silico* model, which predicts the toxicity or safety profile of the compound tested. In this approach, critical key events (the stars) are those mechanistic steps in the process that provide the thresholds that need to be overcome to result in adverse outcomes. Thus, quantitative key event relationships and *in vitro* concentration-response relationships are crucial to discriminate adaptive *versus* adverse effects in the integrated model. For risk assessment quantitative information will be critical to understand what level (threshold) of perturbation must occur to lead to the next step in the pathway [10]. Ultimately, the system could in principle include the entire adverse outcome pathway network that drives toxicity, as part of the virtual human *in silico* model.

The utilisation of an AOP approach for embryo development was considered for retinoic acid. Retinoic acid is involved for example, in cortical neurogenesis mediated by Wnt3a and in progenitor cell proliferation mediated by Fgf8 (Fig. 3). Neural tube and axial patterning are mediated by additional families of genes. Excess retinoic acid can be associated with heart defects, cleft palate, anencephaly, caudal regression, craniofacial and limb defects, mediated by down-regulation of *Dhrs3* and *Cyp26a1*, *26b1*, and *26c1*. Insufficient retinoic acid has been associated with craniofacial, cardiac, and limb malformations associated with down-regulation of *Rdh10* and *Raldh2*. These relationships can be mapped, giving a multistep, interconnected network of processes that can be evaluated in experimental systems that do not necessarily involve intact mammalian organisms [33].

Whilst retinoic acid AOPs have been identified, current knowledge of the AOP field specific to human development is limited. However, the need to enhance our understanding in this field has been acknowledged in recent years. For example, the AOP-Knowledge base open source platform is an OECD-supported web-based resource for collecting and organising AOP biological information [34,35]. At the time of the workshop there were 38 DART related AOPs being developed, out of approximately 250, demonstrating significant interest in this area. The chronic binding of N-methyl-D-aspartate (NMDA) receptors leading to neurodevelopmental effects, disruption of VEGFR signalling leading to developmental effects and COX inhibition adverse effects on reproduction are examples of only a handful developmental AOPs endorsed by the OECD. Clearly more are needed to be fully developed, upgraded to quantitative AOPs, integrated in AOP networks, represented by *in vitro* assays for critical key events, and accepted by OECD for use before an AOP testing strategy can be implemented, but they are providing initial step exemplars for what can already be achieved.

Development is a highly complex ordered process and experimental models based on less complex assays are limited by eliminating the critical cellular, temporal, and spatial interactions involved in embryo development. Although many of the critical aspects of development

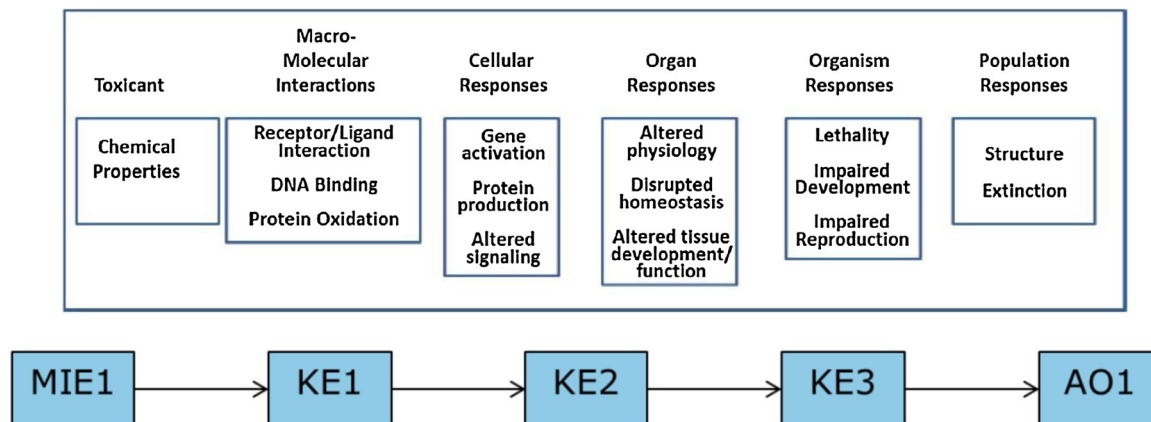


Fig. 1. AOP pathway concept [65].

Abbreviations: AOP = adverse outcome pathway; MIE = molecular initiating event; KE = key event; AO = adverse outcome.

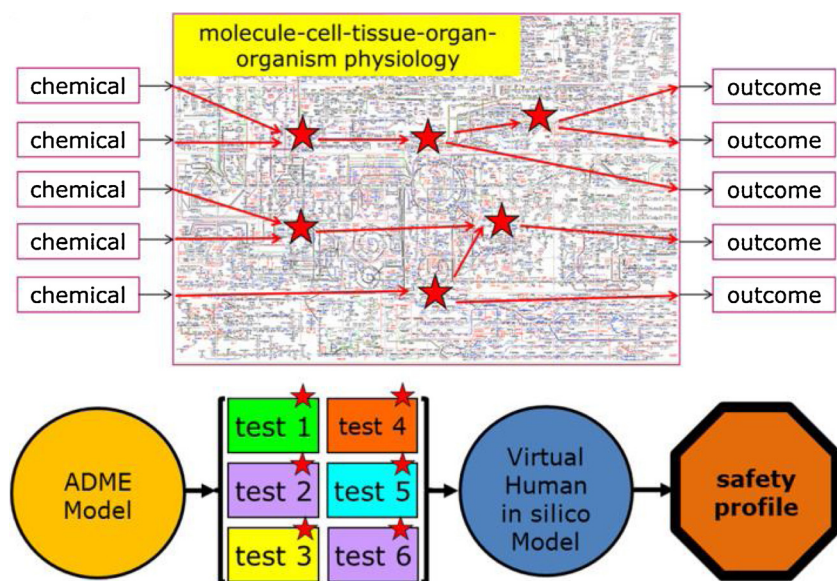


Fig. 2. AOP network concept schematic representation of the application of the virtual human in toxicological safety assessment. The virtual human (the physiological map box) challenged by chemicals will lead to adverse outcomes through a limited number of pathways, in which critical rate-limiting steps occur, indicated here graphically by red stars. Testing these aspects in quantitative *in vitro* tests combined with kinetic information provides the basis for feeding the integrated *in silico* model for predicting the resulting safety profile in humans. ADME = absorption, distribution, metabolism, excretion [66] (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

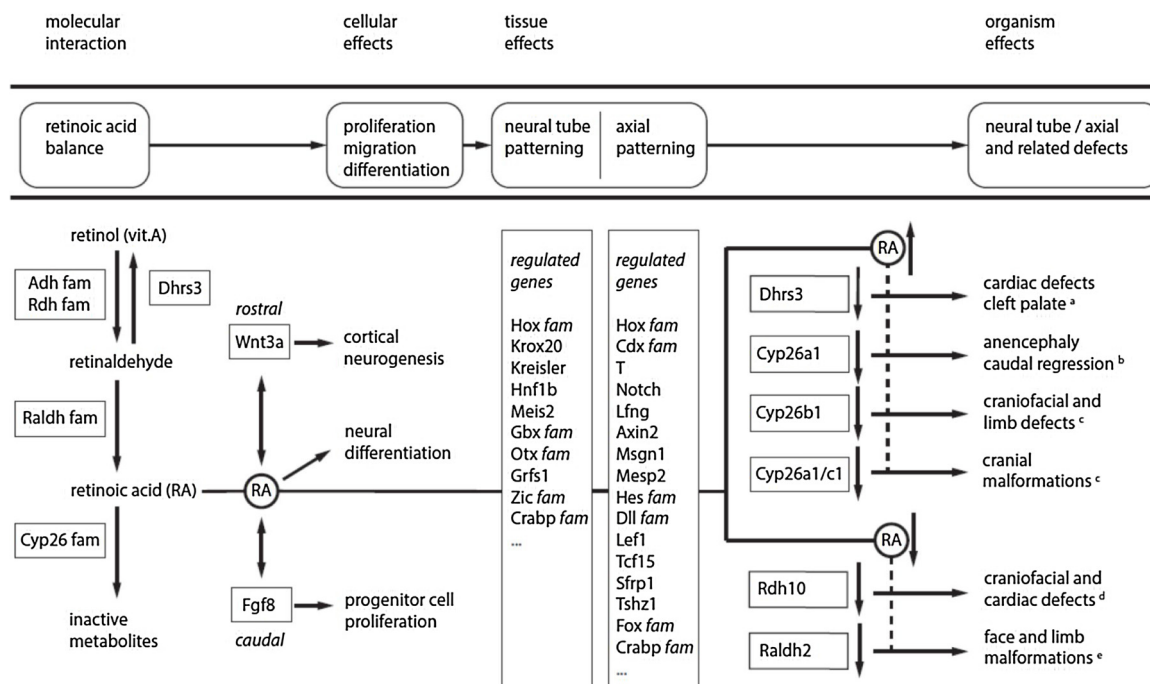


Fig. 3. Schematic representation of the proposed retinoic acid–neural tube/axial patterning adverse outcome pathway (RA–NTA AOP) framework. Reprinted from Reproductive Toxicology, Volume 55, Tonk et al. An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis. Copyright (2015) with permission Elsevier.

remain unknown, computer simulations can be used to reconstruct cellular networks and collective cell behaviour underlying particular morphogenetic events. In developmental toxicology, progress is being made by combining mechanistic data with detailed computational modelling of blood vessel development, secondary palate fusion and urethral fusion [36–38]. These approaches are promising and provide a path that may ultimately lead towards an array of systems comprising a ‘virtual embryo’ for simulation and quantitative prediction of adverse developmental outcomes. Such an approach alongside algorithms that model exposure, could lead to the development of the complete *in silico* prediction for human developmental toxicity.

6. QSAR approaches for the prediction of developmental toxicity

In silico toxicology encapsulates a wide range of computational approaches that include quantitative structure-activity relationship (QSAR)-based approaches and modelling of complex biological systems ([39]–*In silico* Toxicology, Principles and Application). These methods can be used to predict the potential toxicity of a chemical and in some situations quantitatively predict the toxic dose or potency. *In silico* approaches adhere to the 3Rs principles but are also viewed as a fast, robust and cost-effective alternative to traditional toxicity testing. However, their use by risk assessors will depend on the outputs the model can generate and the applicability domain for which the model has been proven useful.

Current off-the-shelf *in silico* approaches for predicting reproductive

toxicity are models based on QSARs and enable a prediction of hazard, which are useful in a compound prioritization setting [40,41]. QSAR models utilise experimental data and scientific knowledge to associate chemical properties with biological activity, which are then used as the basis of a prediction for a new chemical entity. Approaches include statistical-based methods which apply an algorithm to a training set of example chemicals labelled with toxicological outcomes, to generate a model which can predict a level of biological activity or potency. A complementary approach is the development of expert-rule based systems. These systems encode rules written by experts which are derived from the expert review of data generated from toxicity studies but can also consider knowledge of chemical properties and mechanisms.

QSAR models can be generated using a wide variety of methods and a number of algorithms and descriptors are available for generating a statistical model [42]. One of the challenges for the modeller is producing a model which can generate reliable predictions over a wide area of chemical space and in some cases provide a degree of transparency to the user. As confidence in the predictions tend to decrease as a model operates in areas more distant from the training set, the availability of training data is a key determinant in how a model operates throughout chemical space.

In the pharmaceutical regulatory setting, QSAR approaches are now used routinely in submissions to predict the mutagenicity of impurities in medicines [43]. For mutagenicity, the mechanisms are well understood and a vast array of data are available to support modelling across relevant chemical space. However, for developmental toxicity, QSAR approaches are currently limited to smaller and more disparate datasets, which can hinder their predictive capacity and therefore ability to support regulatory decision-making. For example, the aggregation of data from nonclinical studies involving different exposure scenarios can lead to the training of models where the clinical relevance of the predictions remains unclear.

Model developers are overcoming some of the limitations of modelling *in vivo* outcomes directly using QSARs, through the application of AOPs. Organising data around AOPs allows for QSAR models to be trained on larger *in vitro* datasets, the outputs of which can then be associated with adverse outcomes using knowledge of the pathway. This will enable QSAR models to cover larger areas of chemical and biological space relevant to an endpoint, compared to QSAR models trained solely on apical toxicity data. The collection and incorporation of a wide array of genomic, proteomic, chemical and disease-related resource data into *in silico* systems will be crucial to this approach [44]. Organisation of this ever-increasing biological data, both in terms of quantity and quality, will enable computational approaches to predict how a chemical might perturb the complex signalling networks and interactions between tissues and organs that occur during embryogenesis.

In summary, current off-the-shelf QSAR methods for developmental toxicity allow for predictions of hazard and can support hypothesis-based testing and prioritisation. Given the number of endpoints and complexity, the complete *in silico* risk assessment for human developmental toxicity where QSAR predictions act as inputs for system biology models remains a very big challenge. Greater access to data including from proprietary sources, will support the development of new *in silico* models which can operate in wider areas of relevant chemical and biological space. Anchoring these *in silico* models to AOPs will provide the necessary context required to enable risk assessors to use *in silico* approaches with increased confidence and identify when additional data should be generated.

7. Data sharing: aspirations and delivery

As highlighted, the current ICH S5 guideline on “Toxicity to Reproduction for Human Pharmaceuticals” is undergoing revision and for the first time provides guidance on the use of alternative assays in certain circumstances for DART prediction. A change of regulatory

guidance is an important step, but whether guidance change alone will stimulate the advancement in the development of such alternative assays is uncertain, especially considering that the currently accepted paradigm is well established and universally accepted, and that any change in direction needs stakeholder buy-in to invest in new processes and different data packages. One way of promoting change and building confidence is by sharing data and experiences across multiple sectors (including industry, CROs, academia and regulators), to reach consensus for evidence-based recommendations to drive change.

The advantages of data sharing projects for the purpose of developing globally accepted alternative testing strategies are numerous, benefitting not only the pharmaceutical industry, academia and regulatory authorities, but ultimately, and most importantly, patients. Firstly, and most obviously, data sharing allows for the accumulation of a larger-evidence base which can reveal patterns that are not apparent in smaller component data sets [45], thus generating improved scientific confidence in conclusions that are drawn. Furthermore, such projects can promote multidisciplinary cross-sector interactions to take place for the purpose of utilising the expertise of various disciplines to analyse and interpret the data in ways that might not have originally been envisioned, whilst having the potential to promote the harmonisation of practices and eventual global acceptance. Given that such projects can be a resource intensive exercise requiring both money, time and personnel, the chance for collaboration also opens up the possibility to share these burdens and create exploratory research opportunities that otherwise might not be commercially viable. Should data sharing projects prove successful in modifying current testing strategies the benefits could be far reaching. For the pharmaceutical industry, there are several business advantages, such as streamlining the product development lifecycle through obtaining toxicity information sooner, allowing for better informed decisions early in the drug development process and reducing the number of expensive and time-consuming animal studies. For the regulators, there are opportunities to provide scientific advice and promote the harmonisation of regulatory guidance, and from an ethical standpoint there is the opportunity to improve animal welfare and generate better quality predictive data that ultimately promotes the safeguarding of public health.

However, whilst the benefits of data sharing are clear, certain challenges exist that need to be overcome. Critically, the majority of data, which has typically been generated to regulatory standards and thus already fit for purpose, is owned by industry meaning there is a heavy reliance on the goodwill of industry to provide such data in the face of the sensitivities surrounding the sharing of confidential proprietary information that may compromise the competitive edge of the business. It is therefore important that certain assurances are provided to address the concerns of industry in this regard. In addition, pharmaceutical companies may require justification in terms of the benefits not just from a scientific perspective but also from a business perspective to encourage participation on such resource intensive exercises. There are also challenges surrounding the practicalities and processes of data sharing such as how will the data be captured, hosted, standardised and validated, and whether this will be facilitated by industry consortia, regulatory bodies or trusted independent third parties acting as “honest brokers”.

Despite such challenges there are plenty of examples where cross-company collaborations involving data sharing have succeeded, either to change regulatory guidance [46] or promote discussion and recommendations for change, including within the area of DART [47,48]. Numerous cross-company collaborations under the auspices of the Health and Environmental Sciences Institute (HESI), the NC3Rs (UK National Centre for the Replacement, Refinement and Reduction of Animals in Research), industry consortia (such as the European Federation of Pharmaceutical Industries and Associations (EFPIA) and the International Consortium for Innovation and Quality in Pharmaceutical Development [49]), and international scientific organisations such as the Society of Toxicology and Safety Pharmacology Society actively

discuss the need for new scientific approaches and work together to influence regulatory change [50].

In summary, changes to accepted practices and regulatory guidance is a process that requires a collective effort globally. Collaborative multidisciplinary cross-sector data sharing projects offer the opportunity to network to share expertise and provide a wider evidence-base to refine and improve currently accepted practices. The benefits of such projects are wide ranging and include the advancement of scientific knowledge, financial benefits for both industry and academia, harmonisation of regulatory guidance, improved animal welfare, and most importantly an opportunity to improve patient safety by delivering safer products with fewer DART concerns.

8. Discussion and consensus

Given the current revision of ICH S5, the emergence of new technologies over recent years, and an opportunity to seek the opinions of multidisciplinary cross-sector experts, the following questions were considered at the workshop.

8.1. Can the current testing strategy for developmental toxicity be improved?

In order to address whether and how the current developmental toxicity testing strategy for medicines could be improved, the status of the current testing paradigm and any apparent limitations in its predictive capabilities were discussed. It was considered that although the current regulatory testing strategy had performed well in identifying developmental toxicity hazards, the outcomes are largely qualitative, typically without mechanistic understanding and that residual uncertainty can exist concerning the human relevance of some findings. For example, the teratogenic properties of medicines are identified through the assessment of morphological and functional changes in animal models but do not provide an understanding of the events leading to these apical endpoints. There is also the potential difficulty of disentangling direct and indirect drug-induced embryo-fetal adverse effects in the presence of maternal toxicity. Providing mechanistic insight would assist in providing further clarity and reassurance in scenarios where the translatability and specificity of drug-induced adverse effects are challenged. A good example from the agrochemical sector is the replacement of the single end point uterotrophic and Hershberger *in vivo* rat assays for female and male endocrine disruption respectively with a battery of computational and *in vitro* assays in the US Environmental Protection Agency's (EPA) Endocrine Disruptor Screening Program Tier 1 battery [51].

Further limitations in the current testing paradigm were considered to include a reliance on the assessment of risk using safety margins based on maternal, rather than embryo-fetal, exposure in animals *versus* the expected clinical exposure. Furthermore, there are concerns surrounding the suitability of rats and rabbits in modelling the human placenta. Given the importance of exposure for risk assessment and the role of the placenta in pregnancy, the delegates considered future work should focus on understanding the role of the placenta in drug transport and the potential for adverse effects on placental function. The development of organ-on-chip and other *in vitro* models of placental function and the embryo implantation process were highlighted [13].

There was acknowledgement that parts of the chemicals sector are exploring moving away from traditional experimental animals to alternative strategies guided by AOPs. For example, high-throughput screening (HTS) and high content screens (HCS) are being used for *in vitro* profiling of chemicals for biological activity and potential toxicity including developmental toxicity [28]. In this context, the EPA recently announced a directive to prioritise efforts to use new technologies to reduce animal testing by calling for a 30 % reduction in mammalian studies and funding by 2025 and completely eliminating their use by 2035 [52]. Whilst the acceptance criteria for developmental toxicity

testing may differ depending on the sector, product and geographical region, alternative strategies can benefit all sectors. Cross-sector interactions could provide an opportunity to advance the field and promote harmonisation of practices.

In summary, there was a consensus that there is a need to consider how the current strategy could be improved to enhance human relevance and predictivity of embryo-fetal harm, possibly through the use of alternative methods in certain scenarios supported by regulatory guidance. Coupled with a growing interest in the development and potential utilisation of such approaches, as indicated by the current ICH S5 revision and formation of international consortia focussed on advancing this field, it was considered that there is a real opportunity to modernise the current testing paradigm to improve predictability for the ultimate purpose of safeguarding public health. Furthermore, in the interests of safeguarding public health, it was also highlighted that there was also a requirement to consider how best to communicate the outcomes of such studies to prescribers, patients and the general public.

The following recommendations from the workshop were proposed.

- More research is needed to improve current models and develop new ones to predict the potential adverse effects of medicines on the placenta during pregnancy.
- Stakeholders to determine how to communicate the outcomes of alternative testing methods to prescribers, patients and the general public.

8.2. What alternative methods (assays and *in silico*) are available, how can they strengthen the current strategy and what are the challenges and opportunities?

It was acknowledged by the delegates that within the pharmaceutical sector a limited number of *in vitro*, *ex vivo* and non-mammalian *in vivo* assays are currently being used as hazard identification screens for embryo-fetal development, to guide later *in vivo* testing strategies and to elucidate mechanisms underlying toxicity. These typically involve mESC or hESC, 'organs on a chip', rat whole embryo culture and zebrafish embryo tests. The data also have the potential to provide further mechanistic insight and clarity in scenarios where AOs have been observed, yet such data are not routinely used in regulatory submissions. This is partly due to an absence of regulatory guidance on the assay development criteria and thus confidence in the regulatory acceptability of the data. Consequently, efforts to develop alternative models have been fragmented, with no inter-laboratory sharing of methodologies or data, thus leading to a variability in outcomes and lack of harmonised protocols. For the field to move forward there is a need for regulatory authorities to provide clear guidance on steps that need to be taken to validate alternative assays.

Another barrier preventing the uptake of alternative assays was considered to be human relevance of the data generated. While the current standard animal testing approach is associated with some uncertainties and unknowns, alternative approaches based on *in vitro* assays or non-mammalian *in vivo* assays are currently considered to be associated with much higher uncertainties. Added to this, they are typically validated against imperfect mammalian *in vivo* data, thus neglecting interspecies differences, whilst also only providing limited specific information when used standalone. Additionally, for the growing number of human-only targeted therapeutics, rat/rabbit developmental toxicity testing may be considered a suboptimal approach. To address the concern of human relevance, it was acknowledged that human tissue-based assays e.g. hESC, could enhance translational relevance and that performing a battery of assays could improve the quantity and quality of alternative assay data. However, in moving forward, there is a need to consider whether a standard battery of alternatives tests should be developed for all classes of compounds or if a case-by-case approach is more appropriate.

Whilst alternative assays serve a purpose with regards to hazard

identification, it was highlighted that their predictivity could be enhanced through incorporating exposure-risk relationships into alternative models to improve risk assessment (*in vitro* to *in vivo* extrapolation). Correlating *in vitro* concentration-response with internal dose-response kinetics and understanding how *in vitro* activity from one cell type or assay links with another currently presents huge challenges. To address this, computational models that simulate kinetics (ADME) would be useful to connect *in vitro* findings to effects on development.

In considering computational approaches, it was concluded that *in silico* tools for predicting adverse effects on development were limited due to the lack of available data needed to generate predictive algorithms coupled with the challenge of modelling the complexities of embryo-fetal development. Furthermore, additional factors such as both fetal and placental changes during gestation, the level and the timing of exposure, genetic variation, and environmental factors all add to the complexity. It was considered that for developmental toxicity prediction, computational systems will for the foreseeable future be used early in development primarily as a complementary screening tool and used to generate testable hypotheses. However, there was a recognition that there have been significant advances in omics technologies coupled with bioinformatics and computational biology that mean that there are potential opportunities to utilise *in silico* approaches to advance prediction, which could be used for regulatory purposes in the future once confidence in such approaches have been established.

It was agreed that to advance the progression of alternative methods and human relevance a better understanding of human developmental biology is needed, which could be gained through identification and development of relevant AOPs. Understanding such AOPs would elucidate specific key events which could form the basis of alternative assays and *in silico* approaches could help model and network such pathways allowing for better holistic prediction. However, the delegates concluded that current knowledge of AOPs specific to human development is in its infancy and that only a handful of genes and pathways had been adequately characterised. Initial steps to promote the advancement of the AOP-based approach could be for stakeholders to agree a defined set of the key developmental pathways that could be validated and taken forward. However, identifying such pathways in the first instance is challenging. Mapping genes involved in key developmental events such as cell migration, differentiation, proliferation, cell shape and overall organ morphology is one approach that could aid the identification of such AOPs. Reverse engineering AOPs from the adverse developmental outcome back along the pathway(s) was also considered a valuable approach. Due to the complex nature of AOPs, both in terms of identifying the key events involved and mapping the associated pathways, it was concluded that the development of an AOP-based approach in the first instance requires extensive resource and investment to identify novel AOPs, from which the development of associated assays and complex *in silico* systems required to model the associated pathways will follow.

In summary, alternative methods (*in silico*, *in vitro*, tissue-based and non-mammalian *in vivo* assays) are already proving useful at least in a screening capacity to provide indications of compound effects, for prioritisation of compounds and providing mechanistic understanding. In the short-term, it was considered that alternative methods, if qualified, could be introduced in certain circumstances and potentially defer or replace conventional *in vivo* studies.

The following recommendations from the workshop were proposed.

- The pathways by which medicines can affect human development need to be mapped to enable the development of appropriate assays.
- There is a need to bridge the gap between evolving knowledge of AOPs with developing quantitative computational tools to better predict the reproductive and developmental risk of specific drug classes.
- Alternative assays need to be developed and validated in accordance

with agreed guidance.

- Alternative testing methods need to consider exposure-risk relationships to promote the translation of risk to the clinical setting.

8.3. How can stakeholders collaborate to help implement these strategies by sharing data and what are the challenges in sharing such data?

It was recognised that the development of alternative methods for better predicting adverse developmental outcomes to regulatory standards is complex and requires specific expertise. Subsequently, for such approaches to be optimally implemented, there is a requirement for multidisciplinary cross-sector interactions between toxicologists (from regulatory authorities, pharmaceutical companies, CRO's and academia), developmental biologists, computational system developers and funding bodies in order to make progress. A transition period involving case-studies whereby new approaches are compared alongside current assays would be required.

There are a number of ongoing projects that can serve as a template for developing and validating developmental toxicity prediction. For example, the IMI eTOX/eTRANSAFE project has successfully developed a drug safety database from the pharmaceutical industry toxicology reports and public toxicology data and developed novel software tools to better predict the toxicological profiles of small molecules. The US Environmental Protection Agency's (EPA) ToxCast research programme and the broader US government Tox21 consortium is using HTS-HCS (high-throughput screening-high-content screening) strategies leading to vast amounts of information on the adverse impact of environmental exposures [53]. In this project, databases with freely available HTS and HCS data on hundreds of *in vitro* screens and cellular bioactivity profiles for thousands of environmental compounds will eventually become available. The related EPA virtual embryo project is building a framework for incorporating knowledge gained from the HTS-HCS projects into *in silico* models incorporating morphogenetic programs to simulate developmental toxicity [36]. The EPA has recently proposed that the data and models underlying science that is pivotal to its regulatory decisions are publicly available in a manner sufficient for independent validation and analysis which will increase transparency to the general public [54].

To help facilitate acceptance of alternative assays in support of regulatory submissions, early engagement with the regulator in the process through early access to the data streams is warranted. It was considered that one approach to promote standardisation, validation and adoption of new assays, was for regulators and industry to identify a 'safe harbour' or 'trusted broker' for evaluation of the data from alternative assays, generating supporting evidence for new developmental toxicity testing paradigms. Consequently, an improved dialogue between regulators and industry is required which could be provided through guidance and regulatory advice meetings. A clear message was that these efforts should not be fragmented but harmonised across the ICH regions with OECD involvement for them to gain acceptance.

The challenges and opportunities provided by data sharing were also considered. There was general agreement that data sharing was valuable and had produced notable successes in other areas of regulatory toxicology, such as in the ongoing evaluation of the current requirements for dedicated carcinogenicity testing in rodents [64]. Despite researchers' enthusiasm, competitive and legal concerns over confidentiality and patent protection can lead to a reluctance of companies to share data. It was considered that the anonymisation of certain information e.g. compound structures, by an independent third-party mediator or the creation of legally binding confidentiality agreement contracts could overcome such challenges. Such approaches have been successfully implemented by current global collaborative projects.

It was also agreed that a wealth of data exists on compounds that have failed in clinical development due to adverse pathology, which can be used to generate and map AOPs and validate alternative assays.

This is potentially an important source of data that is currently being underutilised because the compounds are no longer considered commercially viable candidates and this information could be shared. However, whilst sharing data on toxic compounds is imperative, it was also considered important to share data on non-toxic compounds, especially for the purposes of developing and validating *in silico* modelling algorithms. The benefit of providing data in a standardised format to promote the harmonisation of data capture and analysis practices was also highlighted. It was considered that for new data the Standard for Exchange of Nonclinical Data (SEND) format required by the FDA for animal studies, which specifies a way to collect and present nonclinical data in a consistent format, provides a useful approach [55].

The following recommendations from the workshop were proposed.

- Greater interactions between regulatory bodies, academia, CROs, the pharmaceutical industry and funding bodies to promote the multidisciplinary collaborations required to encourage the development and utilisation of alternative and complementary methods.
- High quality developmental toxicity data with associated exposure data for selected compounds and classes tested in rodent and non-rodent models should be made available to researchers to assist with the validation of alternative testing methods.
- A “safe harbour” approach should be identified to enable the submission of alternative assay data required for comparing outcomes with currently accepted standard assays.

9. Conclusions

Advances in scientific, technological and regulatory knowledge, and the timely revision of ICH S5, are providing opportunities to consider improving nonclinical methodologies to better predict adverse developmental outcomes. This workshop explored specific opportunities that exist and the challenges that need to be overcome in order to further advance the field of predictive developmental toxicity, specifically the need to further understand human developmental biology, the requirement for multidisciplinary cross-sector interactions and a requirement for data sharing.

In light of such considerations, it is encouraging to note that the concept of needing to modernise the current DART testing paradigm is something that has gained more traction in recent years, not only from a regulatory perspective, but also from within the pharmaceutical, chemical and academic communities. Efforts within the chemical sector are already challenging the current status quo by moving away from traditional animal-based models to alternative testing strategies, which may well help pave the way for a similar approach for pharmaceutical toxicity testing. Additionally, international efforts such as NC3R's CRACK-IT challenges, HESI-DART, eTRANSAFE, IQ-DruSafe, EU-ToxRisk, Tox21 [56–60] and OECD AOP Development Programme highlight a willingness and opportunity for the whole of the scientific community to collaborate and advance the development and utilisation of better predictive methodologies. Advantages of such consortia include not only providing an opportunity for knowledge sharing through enhanced interactions of expert individuals from various disciplines, but also, and critically, the sharing of data through the provision of safe havens.

Whilst a desire and willingness to advance the field of DART testing from such communities is apparent, gaining regulatory advice during the development and validation process of new approach methodologies is critical for the acceptance and eventual routine adoption of such approaches for regulatory purposes. The provision of appropriate regulatory guidelines is one way that regulators can help support such ambitions, which is currently being addressed with the revision of ICH S5. Likewise, an open dialogue either through the potential participation of regulators on advisory boards within collaborative projects or the provision of advice through specific scientific advisory meetings are other methods by which regulators can help support.

The safety of medicines in pregnancy is a high priority issue that has gained significant regulatory interest recently, not only from the MHRA but also other global bodies such as FDA [61] and US National Institutes of Health [62]. In the current scenario where clinical evidence on the safety and efficacy of medicines in pregnant women is lacking, non-clinical data remains an integral part of evaluating the benefit-risk for medicines in pregnancy. Subsequently, opportunities to improve the current testing strategy through the utilisation of technological advancements, improved scientific knowledge and regulatory experience need to be employed.

The advantages of incorporating better predictive alternative methodologies into the current testing paradigm are wide ranging and far reaching and benefit not only industry, academia and regulators, but most importantly provides an opportunity to improve the delivery of safer medicines for pregnant women.

Declaration of Competing Interest

Paul Barrow is an employee of F. Hoffmann-La Roche Ltd. Adrian Fowkes is an employee of Lhasa Limited. Manon Beekhuijzen is an employee of Charles River Laboratories. The other authors declare no conflict of interest.

Acknowledgments

The authors would like to thank all speakers and delegates, as well as the Medicines and Healthcare products Regulatory Agency for hosting the workshop. The views expressed in this paper are those of the authors only and not of their respective institution or company.

References

- [1] M.K. Bakker, J. Jentink, F. Vroom, P.B. Van Den Berg, H.E. De Walle, L.T. De Jong-Van Den Berg, Drug prescription patterns before, during and after pregnancy for chronic, occasional and pregnancy-related drugs in the Netherlands, *BJOG* 113 (2006) 559–568, <https://doi.org/10.1111/j.1471-0528.2006.00927.x>.
- [2] A.A. Mitchell, S.M. Gilboa, M.M. Werler, K.E. Kelley, C. Louik, S. Hernández-Díaz, Medication use during pregnancy, with particular focus on prescription drugs: 1976–2008, *Am. J. Obstet. Gynecol.* 205 (2011) 51, <https://doi.org/10.1016/j.ajog.2011.02.029> e1–51.e8.
- [3] J.R. Daw, B. Mintzes, M.R. Law, G.E. Hanley, S.G. Morgan, Prescription drug use in pregnancy: a retrospective, population-based study in British Columbia, Canada (2001–2006), *Clin. Ther.* 34 (2012) 239–249, <https://doi.org/10.1016/j.clinthera.2011.11.025>.
- [4] P.G. Thorpe, S.M. Gilboa, S. Hernandez-Diaz, J. Lind, J.D. Cragan, G. Briggs, S. Kweder, J.M. Friedman, A.A. Mitchell, M.A. Honein, Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk. National Birth Defects Prevention Study, *Pharmacoepidemiol. Drug Saf.* 22 (2013) 1013–1018, <https://doi.org/10.1002/pds.3495>.
- [5] A. Engeland, T. Bjørge, K. Klungsoyr, V. Hjellvik, S. Skurtveit, Furu, Trends in prescription drug use during pregnancy and postpartum in Norway, 2005 to 2015, *Pharmacoepidemiol. Drug Saf.* 27 (2018) 995–1004, <https://doi.org/10.1002/pds.4577>.
- [6] Commission on Human Medicines' Expert Working Group on Hormone Pregnancy Test Report, (2017) <https://www.gov.uk/government/publications/report-of-the-commission-on-human-medicines-expert-working-group-on-hormone-pregnancy-tests>.
- [7] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, Detection of Toxicity to Reproduction for Medicinal Products & Toxicity To Male Fertility S5(R2), (2000) https://database.ich.org/sites/default/files/S5_R2_Guideline.pdf.
- [8] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. M3(R2), (2009) https://database.ich.org/sites/default/files/M3_R2_Guideline.pdf.
- [9] International Council for Harmonisation Final Concept Paper ICH S5 (R3), Detection of Toxicity to Reproduction for Human Pharmaceuticals, (2015) https://database.ich.org/sites/default/files/S5-R3_EWG_Concept_Paper.pdf.
- [10] F. Sewell, N. Gellatly, M. Beaumont, N. Burden, R. Currie, L. de Haan, T.H. Hutchinson, M. Jacobs, C. Mahony, I. Malcomber, J. Mehta, G. Whale, I. Kimber, The future trajectory of adverse outcome pathways: a commentary, *Arch. Toxicol.* 92 (2018) 1657–1661, <https://doi.org/10.1007/s00204-018-2183-2>.
- [11] X. Arzuaga, M.T. Smith, C.F. Gibbons, N.E. Skakkebaek, E.E. Yost, B.E.J. Beverly, A.K. Hotchkiss, R. Hauser, R.L. Pagan, S.M. Schrader, L. Zeise, G.S. Prins, Proposed key characteristics of male reproductive toxicants as an approach for organizing

- and evaluating mechanistic evidence in human health hazard assessments, *Environ. Health Perspect.* 127 (2019) 65001, <https://doi.org/10.1289/EHP5045>.
- [12] U. Luderer, B. Eskenazi, R. Hauser, K.S. Korach, C.M. McHale, F. Moran, L. Rieswijk, G. Solomon, O. Udagawa, L. Zhang, M. Zlatnik, L. Zeise, M.T. Smith, Proposed key characteristics of female reproductive toxicants as an approach for organizing and evaluating mechanistic data in hazard assessment, *Environ. Health Perspect.* 127 (2019) 75001, <https://doi.org/10.1289/EHP4971>.
 - [13] S. Bremer, E. Brittebo, L. Dencker, L.E. Knudsen, L. Mathiesen, M. Olovsson, P. Pazos, C. Pellizzer, L.R. Paulesu, W. Schaefer, M. Schwarz, F. Staud, A. Stavreus-Evers, K. Vähänkangas, *In vitro* tests for detecting chemicals affecting the embryo implantation process. The report and recommendations of ECVAM workshop 62—a strategic workshop of the EU ReProTect project, *Altern. Lab. Anim.* 35 (2007) 421–439, <https://doi.org/10.1177/026119290703500407>.
 - [14] S. Bremer, R. Cortvrindt, G. Daston, B. Elett, A. Mantovani, F. Maranghi, O. Pelkonen, I. Ruhdel, H. Spielmann, *Altern. Lab. Anim.* Reproductive and developmental toxicity, *ALTA* 33 (Suppl 1) (2005) 183–209, <https://doi.org/10.1177/026119290503301s17>.
 - [15] L. Hareng, C. Pellizzer, S. Bremer, M. Schwarz, T. Hartung, The integrated project ReProTect: a novel approach in reproductive toxicity hazard assessment, *Reprod. Toxicol.* 20 (2005) 441–452, <https://doi.org/10.1016/j.reprotox.2005.04.003>.
 - [16] S. Gibb, Toxicity testing in the 21st century: a vision and a strategy, *Reprod. Toxicol.* 25 (2008) 136–138, <https://doi.org/10.1016/j.reprotox.2007.10.013>.
 - [17] S. Hoffmann, L. Edler, I. Gardner, L. Gribaldo, T. Hartung, C. Klein, M. Liebsch, S. Sauerland, L. Schechtman, A. Stamatii, E. Nikolaidis, Points of reference in the validation process: the report and recommendations of ECVAM Workshop 66, *Altern. Lab. Anim.* 36 (2008) 343–352, <https://doi.org/10.1177/026119290803600311>.
 - [18] B. Schenk, M. Weimer, S. Bremer, B. van der Burg, R. Cortvrindt, A. Freyberger, G. Lazzari, C. Pellizzer, A. Piersma, W.R. Schäfer, A. Seiler, H. Witters, M. Schwarz, The ReProTect Feasibility Study, a novel comprehensive *in vitro* approach to detect reproductive toxicants, *Reprod. Toxicol.* 30 (2010) 200–218, <https://doi.org/10.1016/j.reprotox.2010.05.012>.
 - [19] P. Marx-Stoelting, E. Adriaens, H.J. Ahr, S. Bremer, B. Garthoff, H.P. Gelbke, A. Piersma, C. Pellizzer, U. Reuter, V. Rogiers, B. Schenk, S. Schwengberg, A. Seiler, H. Spielmann, M. Steemans, D.B. Stedman, P. Vanparys, J.A. Vericat, M. Verwei, F. van der Water, M. Weimer, M. Schwarz, A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop, *Altern. Lab. Anim.* 37 (2009) 313–328, <https://doi.org/10.1177/026119290903700314>.
 - [20] A.H. Piersma, J. Ezendam, M. Luijten, J.J. Muller, E. Rorije, L.T. van der Ven, J. van Benthem, A critical appraisal of the process of regulatory implementation of novel *in vivo* and *in vitro* methods for chemical hazard and risk assessment, *Crit. Rev. Toxicol.* 44 (2014) 876–894, <https://doi.org/10.3109/10408444.2014.940445>.
 - [21] OECD guideline No. 34, Guidance Document on The Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, (2005) [http://www.oecd.org/officialdocuments/displaydocument/?cote=en/vj/mono\(2005\)14&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocument/?cote=en/vj/mono(2005)14&doclanguage=en).
 - [22] EMA/CHMP/CVMP/JEG-3Rs/450091/2012, Guideline on the Principles of Regulatory Acceptance of 3Rs (replacement, Reduction, Refinement) Testing Approaches, (2012) https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches_en.pdf.
 - [23] C.J. Bowman, R.E. Chapin, Goldilocks' determination of what new *in vivo* data are "Just right" for different common drug development scenarios, part 1, *Birth Defects Res. B Dev. Reprod. Toxicol.* 107 (2016) 185–194, <https://doi.org/10.1002/bdrb.21184>.
 - [24] EMA/CHMP/SAWP/72894/2008, Qualification of Novel Methodologies for Drug Development: Guidance to Applicants, (2008) https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf.
 - [25] K.C. Brannen, R.E. Chapin, A.C. Jacobs, M.L. Green, Alternative models of developmental and reproductive toxicity in pharmaceutical risk assessment and the 3Rs, *ILAR J.* 57 (2016) 144–156, <https://doi.org/10.1093/ilar/ilw026>.
 - [26] A.H. Piersma, T. Burgdorf, K. Louekari, B. Desprez, R. Taalman, R. Landsiedel, J. Barroso, V. Rogiers, C. Eskes, M. Oelgeschläger, M. Whelan, A. Braeuning, A.M. Vinggaard, A. Kienhuis, J. van Benthem, J. Ezendam, Workshop on acceleration of the validation and regulatory acceptance of alternative methods and implementation of testing strategies, *Toxicol. In Vitro* 50 (2018) 62–74, <https://doi.org/10.1016/j.tiv.2018.02.018>.
 - [27] T. Hartung, S. Hoffmann, M. Stephens, Mechanistic validation, *ALTEX* 30 (2013) 119–130, <https://doi.org/10.14573/alte.2013.2.119>.
 - [28] T.B. Knudsen, M.T. Martin, R.J. Kavlock, R.S. Judson, D.J. Dix, A.V. Singh, Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB, *Reprod. Toxicol.* 28 (2009) 209–219, <https://doi.org/10.1016/j.reprotox.2009.03.016>.
 - [29] P.T. Theunissen, et al., Comparison of rat and rabbit embryo-fetal developmental toxicity data for 379 pharmaceuticals: on the nature and severity of developmental effects, *Crit. Rev. Toxicol.* 46 (2016) 900–910, <https://doi.org/10.1080/10408444.2016.1224807>.
 - [30] E. Teixidó, E. Krupp, A. Amberg, A. Czich, S. Scholz, Species-specific developmental toxicity in rats and rabbits: generation of a reference compound list for development of alternative testing approaches, *Reprod. Toxicol.* 76 (2018) 93–102, <https://doi.org/10.1016/j.reprotox.2018.01.005>.
 - [31] N.S. Sipes, M.T. Martin, D.M. Reif, N.C. Kleinstreuer, R.S. Judson, A.V. Singh, K.J. Chandler, D.J. Dix, R.J. Kavlock, T.B. Knudsen, Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data, *Toxicol. Sci.* 124 (2011) 109–127, <https://doi.org/10.1093/toxsci/kfr220>.
 - [32] Q. Zhang, W.M. Caudle, J. Pi, S. Bhattacharya, M.E. Andersen, N.E. Kaminski, R.B. Conolly, Embracing systems toxicology at single-cell resolution, *Curr. Opin. Toxicol.* 16 (2019) 49–57, <https://doi.org/10.1016/j.cotox.2019.04.003>.
 - [33] E.C. Tonk, J.L. Pennings, A.H. Piersma, An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis, *Reprod. Toxicol.* 55 (2015) 104–113, <https://doi.org/10.1016/j.reprotox.2014.10.008>.
 - [34] OECD Adverse Outcome Pathway Knowledge Base, (2020) <https://aopkb.oecd.org/>.
 - [35] AOP-WiKi, (2020) <https://aopwiki.org/>.
 - [36] M.S. Hutson, M.C.K. Leung, N.C. Baker, R.M. Spencer, T.B. Knudsen, Computational model of secondary palate fusion and disruption, *Chem. Res. Toxicol.* 30 (2017) 965–979, <https://doi.org/10.1021/acs.chemrestox.6b00350>.
 - [37] N. Kleinstreuer, D. Dix, M. Rountree, N. Baker, N. Sipes, D. Reif, R. Spencer, T. Knudsen, A computational model predicting disruption of blood vessel development, *PLoS Comput. Biol.* 9 (2013) e1002996, <https://doi.org/10.1371/journal.pcbi.1002996>.
 - [38] M.C. Leung, M.S. Hutson, A.W. Seifert, R.M. Spencer, T.B. Knudsen, Computational modeling and simulation of genital tubercle development, *Reprod. Toxicol.* 64 (2016) 151–161, <https://doi.org/10.1016/j.reprotox.2016.05.005>.
 - [39] M. Cronin, J. Madden, *In Silico Toxicology: Principles and Applications*, (2020), <https://doi.org/10.1039/9781849732093>.
 - [40] S. Wu, J. Fisher, J. Naciff, M. Lauferweiler, C. Lester, G. Daston, K. Blackburn, Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants, *Chem. Res. Toxicol.* 26 (2013) 1840–1861, <https://doi.org/10.1021/tx400226u>.
 - [41] M. Marzo, S. Kulkarni, A. Manganaro, A. Roncaglioni, S. Wu, T.S. Barton-Maclaren, C. Lester, E. Benfenati, Integrating *in silico* models to enhance predictivity for developmental toxicity, *Toxicology* 370 (2016) 127–137, <https://doi.org/10.1016/j.tox.2016.09.015>.
 - [42] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57 (June (12)) (2014) 4977–5010, <https://doi.org/10.1021/jm4004285>.
 - [43] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk (M7(R1)), (2017) https://database.ich.org/sites/default/files/M7_R1_Guideline.pdf.
 - [44] N. Baker, A. Boobis, L. Burgoon, E. Carney, R. Currie, E. Fritsche, T. Knudsen, M. Laffont, A.H. Piersma, A. Poole, S. Schneider, G. Daston, Building a developmental toxicity ontology, *Birth Defects Res.* 110 (2018) 502–518, <https://doi.org/10.1002/bdrb.21189>.
 - [45] K.A. Briggs, Is preclinical data sharing the new norm? *Drug Discov. Today* 23 (2018) 499–502, <https://doi.org/10.1016/j.drudis.2016.05.003>.
 - [46] S. Robinson, J.-L. DeLongeas, E. Donald, D. Dreher, M. Festag, S. Kervyn, et al., A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development, *Regul. Toxicol. Pharmacol.* 50 (2008) 345–352, <https://doi.org/10.1016/j.yrtph.2007.11.009>.
 - [47] K. Chapman, H. Holzgreffe, L. Black, M. Brown, G. Chellman, C. Copeman, et al., Pharmaceutical toxicology: designing studies to reduce animal use, while maximizing human translation, *Regul. Toxicol. Pharmacol.* 66 (2013) 88–103, <https://doi.org/10.1016/j.yrtph.2013.03.001>.
 - [48] A. Scialli, G. Daston, C. Chen, P. Coder, S. Euling, J. Foreman, et al., Rethinking developmental toxicity testing: evolution or revolution? *Birth Defects Res.* 110 (2018) 840–850, <https://doi.org/10.1002/bdrb.21212>.
 - [49] International Consortium for Innovation and Quality in Pharmaceutical Development (IQ) DruSafe, (2020) <https://iqconsortium.org/initiatives/leadership-groups/preclinical-safety/>.
 - [50] F. Sewell, J. Edwards, H. Prior, S. Robinson, Opportunities to apply the 3Rs in safety assessment programs, *ILAR J.* 57 (2) (2016) 234–245, <https://doi.org/10.1093/ilar/ilw024>.
 - [51] Environmental Protection Agency, Endocrine Disruptor Screening Program: Use of High Throughput Assays and Computational Tools. Federal Register, 80, No. 118 (2015) <https://www.govinfo.gov/content/pkg/FR-2015-06-19/pdf/2015-15182.pdf>.
 - [52] United States Environmental Protection Agency, Memorandum, September 10 (2019) <https://www.epa.gov/sites/production/files/2019-09/documents/image2019-09-09-231249.pdf>.
 - [53] R.S. Thomas, R.S. Paules, A. Simeonov, S.C. Fitzpatrick, K.M. Crofton, W.M. Casey, D.L. Mendrick, The US Federal Tox21 Program: a strategic and operational plan for continued leadership, *ALTEX* 35 (2) (2018) 163–168, <https://doi.org/10.14573/alte.1803011>.
 - [54] Environmental Protection Agency, Strengthening Transparency in Regulatory Science A Proposed Rule. Federal Register, 83, No. 83 (2018) <https://www.govinfo.gov/content/pkg/FR-2018-04-30/pdf/2018-09078.pdf>.
 - [55] CDISC Standard Data Tabulation Model for non-clinical studies, Standard for Exchange of Nonclinical Data (SEND), (2019) <https://www.cdisc.org/standards/foundational/send>.
 - [56] NC3Rs CRACK IT Challenges, (2020) <https://nc3rs.org.uk/crackit/crack-it-challenges>.
 - [57] Health and Environmental Sciences Institute (HESI), Developmental and Reproductive Toxicology (DART) Committee, (2020) <https://hesiglobal.org/developmental-and-reproductive-toxicology-dart/>.

- [58] eTRANSAFE, (2020) <https://etransafe.eu/>.
- [59] EU-ToxRisk, (2020) <https://www.eu-toxrisk.eu/>.
- [60] National Toxicology Program - Tox21, (2020) <https://ntp.niehs.nih.gov/results/tox21/index.html>.
- [61] Federal Food and Drug Administration, Postapproval Pregnancy Safety Studies; Draft Guidance for Industry, (2019) <https://www.fda.gov/media/124746/download>.
- [62] PRGLAC Report. National Institute of Health, Task Force on Research Specific to Pregnant Women and Lactating Women, (2018) https://www.nichd.nih.gov/sites/default/files/2018-09/PRGLAC_Report.pdf.
- [63] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, Draft ICH Harmonised Guideline Detection of Toxicity to Reproduction for Human Pharmaceuticals S5(R3), (2017) https://database.ich.org/sites/default/files/S5-R3_EWG_Draft_Guideline.pdf.
- [64] International Council for Harmonisation Concept Paper: Rodent Carcinogenicity Studies for Human Pharmaceuticals, (2012) https://database.ich.org/sites/default/files/S1-R1_EWG_Concept_Paper.pdf.
- [65] OECD guideline No. 184, Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways, (2017) [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2013\)6&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en).
- [66] A.H. Piersma, J.V. Benthem, J. Ezendam, Y.C.M. Staal, A.S. Kienhuis, The virtual human in chemical safety assessment, *Curr. Opin. Toxicol.* 15 (2019) 26–32, <https://doi.org/10.1016/j.cotox.2019.03.009>.