



# A systematic review of *in vitro* models of drug-induced kidney injury

Alasdair R. Irvine<sup>1</sup>, Damiën van Berlo<sup>1,2</sup>,  
Rawan Shekhani<sup>1</sup> and Rosalinde Masereeuw<sup>1</sup>

## Abstract

Drug-induced nephrotoxicity is a major cause of kidney dysfunction with potentially fatal consequences and can hamper the research and development of new pharmaceuticals. This emphasises the need for new methods for earlier and more accurate diagnosis to avoid drug-induced kidney injury. Here, we present a systematic review of the available approaches to study drug-induced kidney injury, as one of the most common reasons for drug withdrawal, *in vitro*. The systematic review approach was selected to ensure that our findings are as objective and reproducible as possible. A novel study quality checklist, named validation score, was developed based on published regulatory guidance and industrial perspectives, and models returned by the search strategy were analysed as per their overall complexity and the kidney region studied. Our search strategy returned 1731 articles supplemented by 337 from secondary sources, of which 57 articles met the inclusion criteria for final analysis. Our results show that the proximal tubule dominates the field (84%), followed by the glomerulus and Bowman's capsule (7%). Of all drugs investigated, the focus was most on cisplatin ( $n = 29$ , 50.1% of final inclusions). We found that with increasing model complexity the validation score increased, reflecting the value of innovative *in vitro* models. Furthermore, although the highly diverse usage of cell lines and modelling approaches prevented a strong statistical verification through a meta-analysis, our findings show the downstream potential of such approaches in personalised medicine and for rare diseases where traditional trials are not feasible.

## Addresses

<sup>1</sup> Utrecht Institute for Pharmaceutical Sciences (UIPS), Division of Pharmacology, Utrecht University, David de Wied Building, Universiteitsweg 99, 3584 CG Utrecht, the Netherlands

<sup>2</sup> National Institute for Public Health and the Environment, Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, the Netherlands

Corresponding author: Masereeuw, Rosalinde ([R.Masereeuw@uu.nl](mailto:R.Masereeuw@uu.nl))

## Keywords

*In vitro* models, Toxicity testing, Validation strategies, Systematic review, Nephrotoxicity, Drug-Induced kidney Injury (DIKI).

## Introduction

Despite ever increasing expenditure on R&D, the drug discovery pipeline (DDP) has seen fewer products making their way into the clinic [1]. Traditionally, the decision to advance to clinical trials, where R&D expenditure is highest, is based on data obtained using well-characterised animal models for the assessment of pharmacokinetics and drug safety. Advancing through the stages of the DDP should be based on models that accurately reflect human physiology, but despite this, animal studies have long been the gold standard of preclinical research. However, there is increasing evidence of low predictivity of animal data for human effects in terms of efficacy and safety [2]. This mismatch can lead to unexpected adverse events in clinical trials or postmarketing, which contribute to the rising attrition rate and various (expensive) product recalls due to drug toxicity.

Toxicity has been estimated to be responsible for adverse events leading to attrition of up to a third [3] of drug candidates and is a major contributor to the high cost of drug development, particularly when not recognised until late in the clinical trials or postmarketing surveillance. The kidney is particularly susceptible to drug injury because of its high share of cardiac output and its role in excretion of waste compounds from the body. Drug-induced kidney injury (DIKI) is a relatively common clinical condition, particularly in critical care settings caused by acute kidney injury (AKI). Up to 30% of all critically ill patients develop AKI with nearly 6% of those diagnosed requiring subsequent kidney replacement therapy [4]. Despite this, only a fraction (2%) of drug candidates are rejected due to nephrotoxicity in early (phase I) clinical studies, and many forms of DIKI are not measurable until very late in the DDP [5]. This can be caused by species to species translation issues such as drug-metabolising enzyme (cytochrome P450; CYP) expression and individual differences in clearance performance [6]. The role of nephrotoxicity is reflected in the percentage attrition due to nephrotoxicity increasing to 19% in phase III clinical trials [4]. The difficulty of screening for DIKI in the preclinical setting

Current Opinion in Toxicology 2021, 27:18–26

This review comes from a themed issue on **Translational Toxicology**

Edited by **Roos Masereeuw** and **Damiën van Berlo**

Available online 12 June 2021

For a complete overview see the [Issue](#) and the [Editorial](#)

<https://doi.org/10.1016/j.cotox.2021.06.001>

2468-2020/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

is down to a number of factors—not least that the kidney is an incredibly complex organ composed of diverse tissue and terminally differentiated, specialised cells and often the mechanism of action of a substance is unknown, so selection of appropriate *in vitro* models to support *in vivo* findings is difficult. Clinical diagnosis is closely associated with overall excretory function using glomerular filtration rate, estimated from serum creatine and urinary albumin levels. Such outcomes have been clearly demonstrated to be imperfect indications of any form of kidney injury because of the delay from injury to measurable change [7,8]. Furthermore, there is still no universally accepted definition of DIKI as it may involve tubular injuries or glomerulopathies, resulting in AKI or chronic failure, and is often diagnosed late [9]. These limitations and an inadequate preservation of the organs' microenvironment in the models hampered currently used *in vitro* assays to adequately mimic native physiology and/or predict *in vivo* observed effects [10].

The lack of predictive models and the poor predictive aspects of animal studies for clinical trials show the clear need for better approaches to recapitulate kidney function *in vitro*. In recent years, advanced *in vitro* models emerged that integrated complex tissue cultures in microfluidic platforms to more closely mimic the human kidney and DIKI. Here, we used an adapted systematic review methodology to investigate whether currently available *in vitro* models are capable of successfully predicting safety outcomes in humans before going on to highlight new approaches that improve the preclinical processes and therefore translation rates to the clinic. Eventually, these models should improve (long-term) patient outcomes through more effective and safer novel medications entering the market. Furthermore, these models can also be used for detecting underlying causes of DIKI of existing drugs and therapies.

## Systematic review methodology to investigate current *in vitro* models

### Review protocol

The review methodology was prespecified as per a standard Cochrane review of medical interventions. The standard SYRCLE [11] protocol was modified to replace their animal model search strategy with *in vitro* models of kidney injury, for which the search terms have been adapted as presented in the protocol in [appendix 1](#) to answer the research question “can current *in vitro* models accurately provide safety information for potential extrapolation to human applications?”

### Literature search

A systematic search of the PubMed and EMBASE databases was performed, with articles up to July 2019 selected for further screening if they met inclusion criteria defined in the protocol. To ensure a complete

overview of the literature, reference lists of included studies or relevant reviews identified through the search were also reviewed. The list of relevant reviews is included in the protocol at [appendix 1](#).

Articles identified in this search were selected independently by two researchers based on title and abstract screening and as per the inclusion and exclusion criteria stated in the protocol. In case of discrepancies between the two independent reviewers, a third investigator was involved in the screening and discussion. No language restrictions were applied. If any articles meeting inclusion criteria were published in a non-native language, scientists with native language skills would have been asked to translate.

### Inclusion and exclusion criteria

Articles were included if they were primary research studies presenting unique data using an *in vitro* methodology to assess damage to kidney cells. Every effort was made to detail novel and alternative approaches such as organ-on-a-chip and predictive *in silico* modelling that may provide more human relevant responses at the preclinical stage. Relevant drugs included are presented in [Table 1](#). Screening was carried out using Rayyan QCRI [12], and citations were managed using EndNote X9 and Microsoft Excel. Data analysis was carried out in Microsoft Excel with manuscript figures prepared in GraphPad (GraphPad Prism, version 8.4.3, for Windows, Utrecht University).

### Data extraction and analysis

The following outcome measures related to safety were extracted: molecular indicators of cell damage and safety outcomes, that is cell death percentage, inflammatory markers and kidney-specific damage biomarkers such as clusterin, cystatin-C, KIM-1, N-acetyl- $\beta$ -D-glucosaminidase, neutrophil gelatinase-associated lipocalin and osteopontin (see [supplemental information](#) for details on the markers). Full outcome measures are listed in the protocol at [appendix 1](#).

Because of the high heterogeneity of the studies that met the inclusion criteria, a meta-analysis was not possible. Instead, the methodological quality of the included studies was assessed using a combination of SYRCLE's risk of bias tool [11] and a number of other factors extracted from references presented in the [supplemental table S.1](#). These factors led to the creation of a preliminary validation checklist of ten questions to assess the quality of the study and also to assess each experimental approach's potential use as an alternative to current *in vitro* models. These questions are as presented in [supplemental table S.2](#). Every study meeting the inclusion criteria in the protocol was assessed as per these questions and scored on how well the study complied. The studies were scored on a scale of 0 (no compliance) to 2 (full compliance) to give a

**Table 1** Relevant endpoints, reference drugs, measurement methods and parameters to assess organ functionality.

| Feature   | Study reference          | Description  |
|---|--------------------------|--|
| <b>Reference drugs</b>  |                          |  |
| Acetaminophen (PAP)   | [26]                     | Known to elicit PT-specific toxicity.  |
| Aristolochic acid 1 (AA-1)  | [24]                     | 2 linked MPSs to show the human hepatocyte-specific metabolism of AA-1 substantially increases its cytotoxicity.   |
| Cyclosporine A  | [20,25,27,28]            | Reversible vasoconstriction, with no relevant changes in tubular function.   |
| Doxorubicin   | [20,25,27,28]            | Induces apoptosis and necrosis in healthy tissue, causing toxicity in the brain, liver and kidney  |
| Gentamicin  | [27,29]                  | Specifically causes necrosis of cells in the kidney proximal tubule, resulting in acute tubular necrosis, followed by acute kidney failure                       |
| Ifosfamide  | [23,30]                  | Emphasising the role of metabolism to accurately modelling toxicity <i>in vitro</i>  |
| Ioxitalamate  | [20,25,27,28]            | Specifically causes necrosis of cells in the kidney proximal tubule  |
| DrugMatrix and TG-GATES databases                                     | [20,25,27,28]            | Excellent resources for known nephrotoxicants.   |
| <b>Relevant endpoints</b>   |                          |  |
| $\gamma$ H2AX   | [19,20,27],              | Acts as a measure of DNA damage to cellular cytoskeletal features  |
| Haem oxygenase 1 (HO-1)   | [25,26]                  | Induced in a variety of kidney substructures in response to injury.  |
| Interleukin 6 and interleukin 8 (IL-6 and IL-8)                       | [28,31]                  | Proinflammatory cytokines play a central role in the pathophysiology of AKI, including nephrotoxin-induced AKI   |
| Kidney injury molecule-1 (KIM-1)                                      | [17,26,31,32]            | Expression is markedly upregulated in damaged regions of the proximal tubule. It may participate in the progress of kidney injury or repair.                     |
| Neutrophil gelatinase-associated lipocalin (NGAL)                     | [26,31]                  | Easily detected in the blood and urine, NGAL has been identified as an early biomarker for prediction of AKI [33]  |
| <b>Measurement methods</b>  |                          |  |
| Lactate dehydrogenase (LDH) release                                   | [17,21]                  | A cytosolic enzyme present in many different cell types that is released into the cell culture medium on damage to the plasma membrane.                          |
| Live/death assay  | [19,20,24]               | Quick and simple three-colour assay to measure cell viability. Interestingly, this was not found to be a strong predictor for predictive <i>in silico</i> models |
| Metabolic assays  | [16,17,21,28,32]         | ATP-based and MTT-based assays are highly susceptible to metabolic interference; consequently, they may generate false positive results                          |
| TEER (transepithelial/transendothelial electrical resistance)         | [21,25]                  | A widely accepted quantitative technique to measure the integrity of tight junction dynamics in cell culture models  |
| TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labelling) | [24]                     | A method for detecting DNA fragmentation by labelling the 3'-hydroxyl termini in the double-strand DNA breaks generated during apoptosis.                        |
| <b>Parameters to assess organ functionality</b>                       |                          |  |
| OAT transporter expression  | [16,27]                  | Many drugs and toxins are organic anion transporter (OAT) substrates. Therefore, OAT expression is a key component of proximal tubule function.                  |
| Cytosolic sulfotransferase (SULT) transporter expression              | [24,34]                  | Expression required to allowing for liver metabolism of reference drugs  |
| Long-term viability   | All studies in table S.6 | All the studies presented in performed toxicity assessment in platforms capable of keeping cells cultured for >7 days.   |
| Three-dimensional architecture  | [16,21,32]               | Features such as the presence of apical to basolateral polarity, <i>in vivo</i> like flow conditions, are all vital to assess drug uptake.                       |

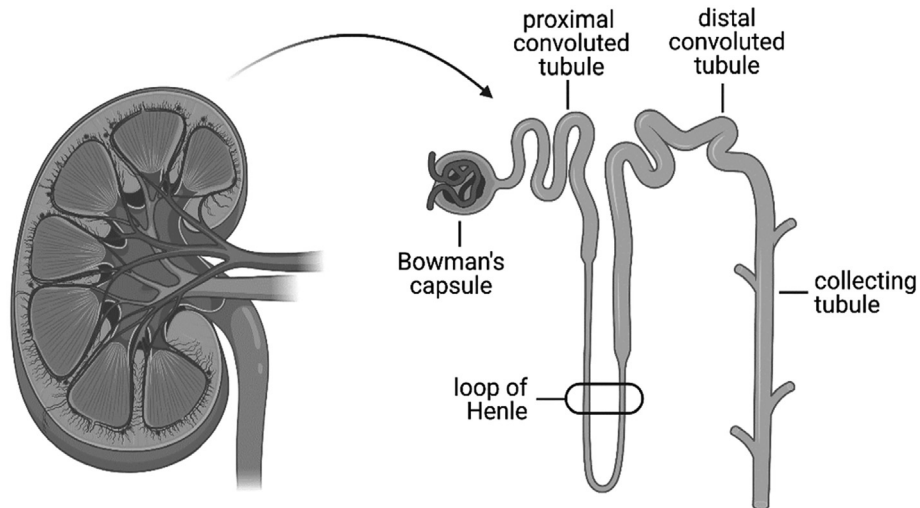
numeric and transparent assessment of each approach. Further stratification of experimental approaches was carried out as per the functional regions of the kidney nephron, separating the studies into five groups as shown in Figure 1.

### Currently used *in vitro* models systematically reviewed

The search strategy returned a total of 1731 results, and secondary sources contributed another 337 articles. After removal of duplicates ( $n = 26$ ), a total of 1824 article abstracts were screened, of which 96 met

the inclusion criteria mentioned previously and progressed to full-text screening. Full-text screening led to the exclusion of a further 39 articles because of incorrect study design (no *in vitro* data) (9), wrong publication type (17), no DIKI outcomes (12) and no full article (1), meaning 57 studies have been included for full analysis (figure S.1). Two conference abstracts, by Silva et al. (2017) and Cappadona et al. (2017), returned by the search strategy subsequently released full texts, and these were included for analysis. The latter study changed the first author and has been changed to Milansi et al. (2018) in this write-up.

Figure 1



**Kidney regions and model stratification.** Category 1 models using proximal tubule cells, category 2 models—Bowman's capsule and glomerulus, category 3—distal tubule, category 4—collecting duct and category 5—loop of Henle. The studies were also assigned a model category as described in [supplementary Table S.4](#). The type of the outcome expressed in each methodological approach was also categorised as per [supplementary Table S.5](#). This was created with BioRender.

### A systematic review is an unbiased, transparent and reproducible method for literature review, but is hampered by low standardisation across studies identified

[Table S.3](#) displays a summary of the studies and articles that fully met the criteria and were subjected to data extraction and analysis. Across all results, the overall quality was considered high (10.82). All the studies indicated the origin and type of cells and relevant control groups to their scientific hypothesis (51/57 (89.5%) having full compliance and 6 (10.5%) having medium compliance). However, there were some that did not clearly indicate whether the data were obtained from independent experiments, and the incidence of intra-laboratory or interlaboratory repetitions was very low (only 2 showing full compliance (3.5%)). Similarly, there was limited compliance with regard to drug–drug interactions and media interaction, as well as low compliance with drug stability data. No studies scored perfectly on randomisation and blinding, and only 3 (5.3%) showed partial compliance to these criteria. Interestingly, 35% of studies failed to attempt an *in vitro*–*in vivo* extrapolation [13]. Data are presented in [Figure S.2](#).

Because of the diversity of techniques and the focus on the presentation of as wide a range of novel methodologies as possible, a meta-analysis of the selected studies was not feasible. To perform such an analysis, much higher standardisation with regard to dosage, time of exposure, blinding and defined control groups would have been needed. Instead, articles meeting the inclusion criteria were stratified by the

type of model used and assessed for their potential further use in the drug development pathway, with a particular focus on their ability to screen for nephrotoxic compounds.

In [Figure S.3](#), it is demonstrated that reproducible criteria (8–10) show the lowest compliance. Building on the lack of standardisation across the studies, of particular note is the distribution of results that were generated by systematically examining the literature and manual searching using reference lists of review articles. From our systematic search, 27 articles that met the inclusion criteria were returned in comparison with 30 from the manual one, meaning our results rely on over 50% of studies that could not be located by the search strategy (after discarding of duplicates). Furthermore, when examining the validation score of the articles, our secondary sources show a higher score of 11.96 versus 9.61 for the systematic search.

### Common features and consensus to lay the foundations for validating *in vitro* models

The most commonly investigated drug was cisplatin ( $n = 29$ , 50.1% of final inclusions), which was also included in every study that used a panel of nephrotoxic compounds ( $n = 14$ ). Cisplatin (cis-diamminedichloroplatinum II) is a first-line chemotherapeutic drug used to treat diverse types of cancers. Despite being one of the most effective chemotherapeutic agents, its clinical use is limited because of the severe side effects, including AKI which can develop because of tubular cell accumulation [14]. Common drugs including acetaminophen, chemotherapeutics and environmental



toxins modelled by the studies showing the highest validation scores are presented in Table 1.

### Advanced kidney models have been developed mainly for the proximal tubule with cytotoxicity as the most common endpoint

Typically *in vitro* modelling has focussed on the culturing of singular cell types on plates or slides with toxicity of compounds being assessed via percentage of cells undergoing apoptosis under stimulus. Advanced *in vitro* models identified in this study and their characteristics are presented in Table S.6.

Figure 2 shows the vast majority of our results modelled the proximal tubule, indeed section 1 models accounted for 84% of results with an average validation score of 11. The second most well-modelled kidney region was section 2 (bowman's capsule and glomerulus) with 7% of results averaging a score of 10. Our results display one model of the distal tubule and 2 models of the loop of Henle. There was no specific modelling found for the collecting duct; however, some organoid models did model the full nephron structure.

With the proximal tubule dominating, the most common endpoints assessed were category 2, which included classical toxicity endpoints like cellular levels of ATP, and extracellular lactate dehydrogenase and MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide that can be reduced to an insoluble formazan, which has a purple color and indicative of the cells metabolic activity [15]. Further outcomes and the

modelling approaches that assess them have been summarised in Table 1.

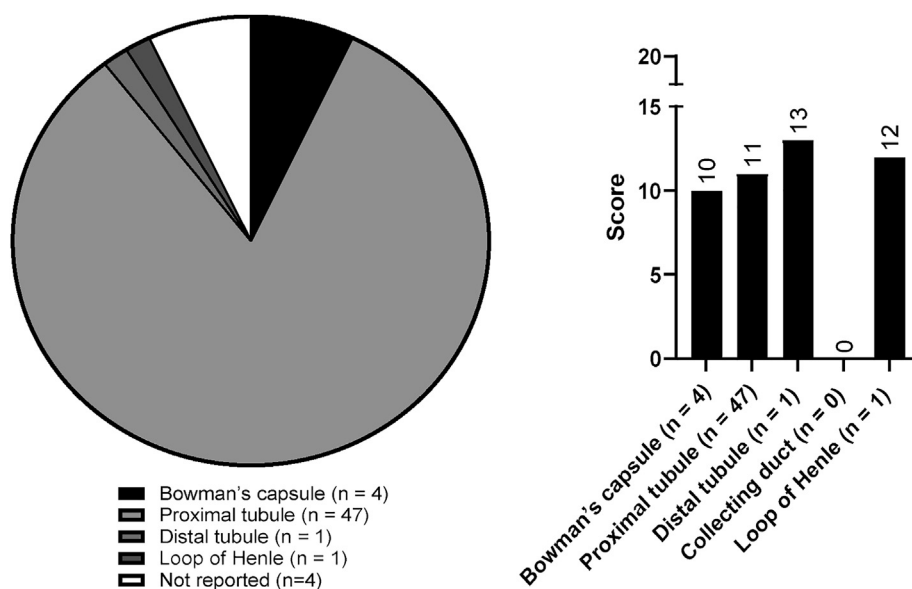
### Higher validation score observed for studies using advanced *in vitro* models

The lack of ability to perform meta-analysis on *in vitro* studies makes the synthesis and extrapolation of data to a patient population much more challenging. The 'validation score' metric developed here aims to identify techniques and models with clear applications for the identification of compounds toxic to humans—with a particular focus on the kidney.

The most common approach met the criteria for category 1—2D culture using a traditional cell line ( $n = 30$ ), followed by category 4 that includes organoids and 3D cell culture ( $n = 13$ ). In Figure 3, we see that category 4 models have the highest mean validation score (12.86,  $n = 4$ ) closely followed by category 2 models (12.67,  $n = 5$ ). In Figure 3, we can also observe a trend of a higher validation score with increasing 'complexity' of the model. There was further difficulty in comparing studies as many articles were comparing novel methods with established cell lines and toxicity, and so conformed to multiple model categories.

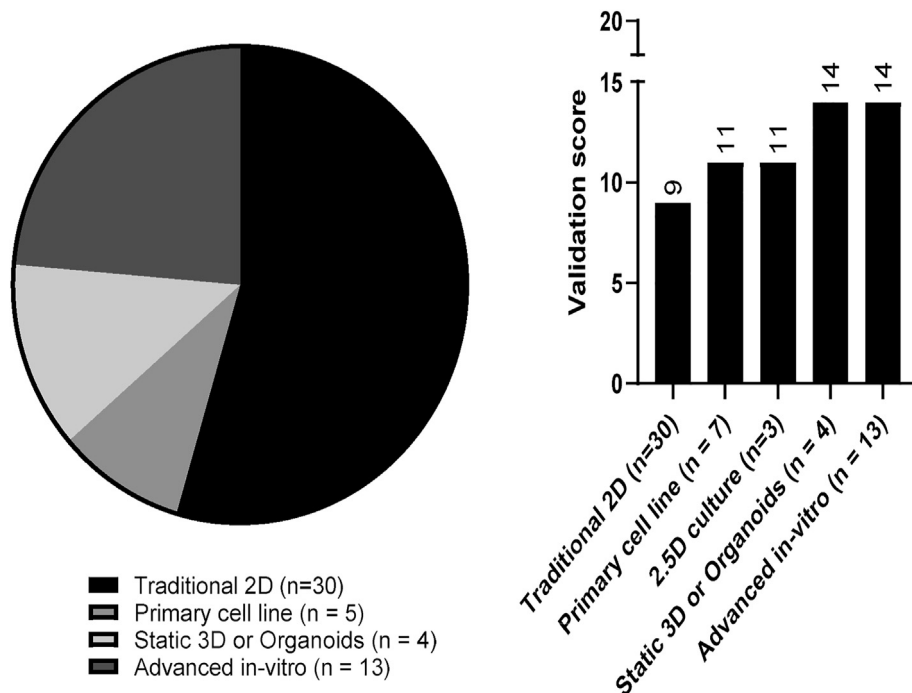
This trend is reinforced when accounting for the unequal size of the categories via the modal value with category 4 (advanced *in vitro*) showing a mode validation score of 14. Increasingly complex models that scored highly showed criteria such as increased physiological relevance (3D structure of tubules, extracellular matrix,

Figure 2



**Regions of the kidney modelled within results.** (Left) Pie chart showing the proportion of results modelling each kidney region. Proximal tubule dominates results (84%), reflecting saturation of *in vitro* modelling techniques. (Right) Mean validation scores across kidney regions. Comparison here is not valid as within our results, several studies analysed toxicity across multiple regions of the kidney.

Figure 3



**The category of *in vitro* models identified by the search strategy.** (Left) Pie chart showing proportion of identified techniques. 2D cell culture dominating with 50.6% of results reporting use. Categorical analysis (right) revealed that validation scores increased with complexity of the model. The most complex models (organoid and advanced *in vitro*) had modal validation scores of 14 in comparison with the traditional 2D approach's score of 9. Mode scores used to compensate for the uneven size of model categories shown on x-axis.

multiple compartments, presence of multiple cell types, cellular maturation, expression of relevant transporters) [16,17], inclusion of extracellular matrix components [18], compatibility with high-throughput screening and advanced imaging techniques [19,20] and ability to inform a personalised medicine approach with regard to safety testing. These common physiological features of the highest scoring models have been summarised and presented in Table 1.

The highest validation score was received by the study conducted by L Aschauer et al. [21], a score of 16 out of 20 (Table S.6). Here, the researchers used a transcriptomics approach to compare the toxicity responses of RPTEC/TERT1 cell lines cultured on filter inserts to give apical and basolateral polarisation. In addition to this, fourteen other models demonstrated a high validation score  $\geq 13$  (Table S.6). However, the Maschmeyer et al. [22] study breaks the general trend of more complex model scoring higher on the study quality checklist. The researchers here presented an interconnected four-organ-chip co-culture, the most complex model identified in our literature search. However, the study scored 8 of 20 on the study quality checklist, showing poor compliance with 'robustness' and context of use criteria on the checklist (scoring 0 on q. 1, 4 and 5). This can be attributed to the study's

overall focus more on absorption, distribution, metabolism, excretion (ADME) outcomes and proving physiological relevance and cellular health within the co-culture environment rather than a focus on clear toxicological endpoints.

Choucha et al. [23] also present a co-culture system and demonstrate the key role of liver metabolism for assessing toxicity by modelling ifosfamide toxicity via its' metabolite chloroacetaldehyde. Chang et al. [24] further reinforce the importance of accounting for hepatic metabolism showing that the biotransformation of AA-I by liver enzymes increases kidney toxicity and conversely that inclusion of an organic anion transporter inhibitor (probenecid) attenuates uptake by human cell cultures in a microphysical system. This implies that incorporation of a liver model or compartment in, for instance, an organ-on-chip or other system would greatly advance predictivity and *in vivo* relevance for humans.

### Future perspectives

Our systematic review of the literature had an overall inclusion rate of 3.3%, naturally leading to questions over the use of our methodology to identify relevant *in vitro* literature. One of the key causes of this low inclusion rate was the lack of follow-up articles to

abstracts, but it could also be explained by stringent inclusion and exclusion criteria and/or a restrictive search strategy with an opening for a significant contribution by the semantics of *in vitro* modelling versus preclinical modelling while searching the PubMed and Embase databases.

This failure of the preclinical pipeline strongly highlights the need for improved, physiologically relevant *in vitro* models that can better serve as reliable drug-screening and disease modelling tools. Here, we show the improved ability of more advanced models to predict DIKI during drug development. However, there is still a clear lack of specific indicators of DIKI that can be reliably measured *in vitro*. We identified a number of techniques, such as the use of microphysical culture systems and the integration of *in silico* modelling, that have high potential to fulfil this role using a novel evaluation metric. The models that scored the highest all used a combination of human cells that are reflective of the *in vivo* environment and multiple, diverse outcomes. This was exemplified in the work by Adler et al. [25] using both HO-1 and traditional cell death/viability assays to predict DIKI caused by nephrotoxic compounds from the DrugMatrix and TG-GATES databases. Our results clearly demonstrate the viability of expanding toxicity assays beyond single-endpoint measures of cell death. With special regard to kidney toxicity, analysis of a combination of cell health parameters and the use of a physiologically relevant cell line are key aspects for successfully developing an assay to predict drug-induced nephrotoxicity.

However, advanced *in vitro* models will have to balance physiological relevance and enhanced predictivity with factors such as ease of use and high-throughput analysis to enter into widespread use. Our results suggest that a lack of standardisation of experimental approaches coupled with an incomplete understanding of the pathology makes modelling of DIKI incredibly challenging. Furthermore, methods for and the quality of narrative reviews of *in vitro* studies are highly variable, and improved reporting standards are needed. Our validation metric attempts to do this; however, the metric itself needs further validation.

Any validation and uptake of *in vitro* models as animal replacement models takes time. Translation to the clinical setting requires both scientific excellence and quality assurance. Here, our quality assessment plays a role, and alongside the outcomes, compounds, and the physiological features in Table 1 provide the basis of a validation strategy for those seeking to develop a robust *in vitro* model of DIKI.

Of these common features, it is the integration of *in silico* approaches that is of paramount importance. Doing so allows the adoption of endpoints such as those

presented in Table 1, other parameters of cell condition and endpoints from the clinical and regulatory setting. As such, models rely on high-quality ‘known’ data points to form predictions based on, that is, characteristics, biological complexity and experimental scalability both essential for producing meaningful and robust data sets to allow the integration of *in silico* predictive models. With this in mind, there is an argument that early integration of regulatory compliance (using recently qualified biomarkers) can help drive standardisation and translation of *in vitro* data to humans. The integration of such techniques requires effective input from all stakeholders within the DDP, that is, regulators, investigators, clinicians and patients, as well as the collaboration of diverse academic disciplines to move *in vitro* innovation forward.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

R.M. reports financial support was provided by Utrecht University, by H2020-MSCA-COFUND-2017-801540 RESCUE COFUND and by ONTOX, grant number 963845 of the European Commission under the Horizon2020 research and innovation framework programme.

### Acknowledgements

This work was supported by Utrecht advanced *in vitro* models hub (U-AIM), H2020-MSCA-COFUND-2017-801540 RESCUE COFUND and by ONTOX, grant number 963845 of the European Commission under the Horizon2020 research and innovation framework programme.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cotox.2021.06.001>.

### References

Papers of particular interest, published within the period of review, have been highlighted as:

- \* of special interest
  - \*\* of outstanding interest
1. Dimasi JA, Grabowski HG, Hansen RW: **Innovation in the pharmaceutical industry: new estimates of R&D costs.** *J Health Econ* 2016, **47**:20–33.
  2. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W: **Concordance of the toxicity of pharmaceuticals in humans and in animals.** *Regul Toxicol Pharmacol* 2000, **32**:56–67.
  3. Montané E, Castells X: **Epidemiology of drug-related deaths in European hospitals: a systematic review and meta-analysis of observational studies.** *Br J Clin Pharmacol* 2021. <https://doi.org/10.1111/bcp.14799>. Online ahead of print.
  4. Naughton CA: **Drug-induced nephrotoxicity.** *Am Fam Physician* 2008, **78**:743–750.
  5. Guengerich FP: **Mechanisms of drug toxicity and relevance to pharmaceutical development.** *Drug Metabol Pharmacokin* 2011, **26**(1):3–14.
  6. Jansen K, Pou Casellas C, Groenink L, Wever KE, Masereeuw R: **Humans are animals, but are animals human enough? A systematic review and meta-analysis on interspecies**

**differences in renal drug clearance.** *Drug Discov Today* 2020, **25**:706–717.

The authors present a systematic review of the available literature of animal models used to study pharmacokinetics in the pre-clinical phase of drug discovery with a focus on the 20 renally excreted drugs. The authors demonstrated that rats despite being the most commonly used model are an inadequate species for preclinical drug clearance testing.

7. Fuchs TC, Hewitt P: **Biomarkers for drug-induced renal damage and nephrotoxicity—an overview for applied toxicology.** *AAPS J* 2011, **13**:615–631.
  8. Griffin BR, Faubel S, Edelstein CL: **Biomarkers of drug-induced kidney toxicity.** *Ther Drug Monit* 2019, **41**:213–226.
  9. Awdishu L, Mehta RL: **The 6R's of drug induced nephrotoxicity.** *BMC Nephrol* 2017, **18**.
  10. Troth SP, Simutis F, Friedman GS, Todd S, Sistare FD: **Kidney safety assessment: current practices in drug development.** In *Seminars in nephrology*. Elsevier; 2019:120–131.
  11. De Vries RBM, Hooijmans CR, Langendam MW, Van Luijk J, Leenaars M, Ritskes-Hoitinga M, Wever KE: **A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies.** *Evid Preclin Med* 2015, **2**, e00007.
  12. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A: **Rayyan—a web and mobile app for systematic reviews.** *Syst Rev* 2016, **5**:210.
  13. Hartung T: **Perspectives on in vitro to in vivo extrapolations.** *Appl in vitro Toxicol* 2018, **4**:305–316.
  14. Volarevic V, Djokovic B, Jankovic MG, Harrell CR, Fellabaum C, Djonov V, Arsenijevic N: **Molecular mechanisms of cisplatin-induced nephrotoxicity: a balance on the knife edge between renoprotection and tumor toxicity.** *J Biomed Sci* 2019, **26**:25.
  15. Arechabala B, Coiffard C, Rivalland P, Coiffard L, Roeck-Holtzhauer YD: **Comparison of cytotoxicity of various surfactants tested on normal human fibroblast cultures using the neutral red test, MTT assay and LDH release.** *J Appl Toxicol* 1999, **19**:163–165.
  16. Weber EJ, Chapron A, Chapron BD, Voellinger JL, Lidberg KA, Yeung CK, Wang Z, Yamaura Y, Hailey DW, Neumann T: **Development of a microphysiological model of human kidney proximal tubule function.** *Kidney Int* 2016, **90**:627–637.
- Here the authors show the advantages of using a MPS cultures system demonstrating full metabolic functionality, long-term viability, polarised expression and function of proteins essential for resorptive and secretory transport. The paper also shows that KIM-1, a marker of acute kidney injury, is expressed in PTECs cultured in 2D, but is absent in MPS devices. Corroborated across other papers in our results and undermining reliance on KIM-1 as a translatable biomarker for kidney toxicity.
17. Crean D, Bellwon P, Aschauer L, Limonciel A, Moenks K, Hewitt P, Schmidt T, Herrgen K, Dekant W, Lukas A: **Development of an in vitro renal epithelial disease state model for xenobiotic toxicity testing.** *Toxicol Vitro* 2015, **30**:128–137.
  18. Musah S, Mammoto A, Ferrante TC, Jeanty SSF, Hirano-Kobayashi M, Mammoto T, Roberts K, Chung S, Novak R, Ingram M: **Mature induced-pluripotent-stem-cell-derived human podocytes reconstitute kidney glomerular-capillary-wall function on a chip.** *Nature Biomed Eng* 2017, **1**:1–12.
- This paper demonstrates the feasibility of directing the differentiation of hiPS-derived cells into glomerular podocytes on-chip and the key considerations of having human based physiology at the heart of an iPSC based model which could enable the application of this organ-on-a-chip approach for patient-specific analyses in the future. The authors were able to model a drug toxicity that leads to albuminuria by integrating vascular like flow conditions and a multi compartment approach.
19. Sjögren A-K, Breitholtz K, Ahlberg E, Milton L, Forsgard M, Persson M, Stahl SH, Wilmer MJ, Hornberg JJ: **A novel multi-parametric high content screening assay in ciPTEC-OAT1 to predict drug-induced nephrotoxicity during drug discovery.** *JAT* 2018, **92**:3175–3190. <https://doi.org/10.1007/s00204-018-2284-y>.

The authors demonstrate the potential for *in silico* modelling inter-gation with *in vitro* modelling, here a support vector machine with radial basis function kernel from sci-kit learn database. They found that drug-induced changes to actin cytoskeleton organisation, mitochondria, nuclear morphology and membrane integrity were the most predictive parameters for the *in silico* components of their model.

20. Su R, Xiong S, Zink D, Loo LH: **High-throughput imaging-based nephrotoxicity prediction for xenobiotics with diverse chemical structures.** *Arch Toxicol* 2016, **90**.
- The authors quantified 129 image based phenotypic features, and identified chromatin and cytoskeletal features that can predict the human *in vivo* response to a panel of nephrotoxic compounds. The paper shows how *in vitro* data can be used to predict nephrotoxicity with high efficiency and accuracy by combining cell-based and computational methods (random forest classifier) that are suitable for automation.
21. Aschauer L, Limonciel A, Wilmes A, Stanzel S, Kopp-Schneider A, Hewitt P, Lukas A, Leonard MO, Pfaller W, Jennings P: **Application of RPTEC/TERT1 cells for investigation of repeat dose nephrotoxicity: a transcriptomic study.** *Toxicol Vitro* 2015, **30**:106–116.
- The highest scoring study in the review, the authors used micro porous supports to ensure the presence of apical to basolateral polarity in order to ensure physiological like uptake vital for accurate toxicity predication. Transcriptomic data were analysed to investigate compound-specific effects, global responses and mechanistically informative signatures and findings were confirmed based on TEER and supernatant measurements.
22. Maschmeyer I, Lorenz AK, Schimek K, Hasenberg T, Ramme AP, Hübner J, Lindner M, Drewell C, Bauer S, Thomas A: **A four-organ-chip for interconnected long-term co-culture of human intestine, liver, skin and kidney equivalents.** *Lab Chip* 2015, **15**:2688–2699.
- The researchers here presented an interconnected four-organ-chip co-culture, the most complex model identified in our literature search, interestingly study breaks the trend of validation metric and model complexity.
23. Choucha-Snouber L, Aninat C, Grsicom L, Madalinski G, Brochot C, Poleni PE, Razan F, Guillouzo CG, Legallais C, Corlu A, et al.: **Investigation of ifosfamide nephrotoxicity induced in a liver-kidney co-culture biochip.** *Biotechnol Bioeng* 2013, **110**:597–608.
- This paper is key to demonstrating the importance of liver metabolism of common clinical drugs in the assessment of kidney proximal tubule toxicity. Both mono-culture and co-culture observed, and toxicity to the proximal tubule cells was only shown when the metabolite chloroacetaldehyde was introduced via the HepaRG co-culture. Study scored less well on validation metric due to the use of tumour based (HepaRG and MDCK) cell lines and lack of reporting with regards to biomimicry of the 3D architecture of the proximal tubule.
24. Chang SY: **Human liver-kidney model elucidates the mechanisms of aristolochic acid nephrotoxicity.** *JCI Insight* 2017, **2**.
- Another example of the key role of hepatic biotransformation. The authors demonstrate that human hepatocyte-specific metabolism of AA-I substantially increases its cytotoxicity towards human kidney proximal tubular cells using an integrating multi-compartment MPS.
25. Adler M, Ramm S, Hafner M, Muhlich JL, Gottwald EM, Weber E, Jaklic A, Ajay AK, Svoboda D, Auerebach S: **A quantitative approach to screen for nephrotoxic compounds in vitro.** *J Am Soc Nephrol* 2016, **27**:1015–1028.
- The author focus on the identification of HO-1 as a translatable biomarker for *in vitro* kidney toxicity. Translatability was shown across 2D cell culture, MPS, and rat toxicogenomic databases. The authors argued that the false positive in a 2D culture against the high predictivity shown in the MPS culture shows increased need for better *in vitro* modelling. The findings also allowed the development of a novel high throughput assay based on quantifying HO-1 as a measure of kidney toxicity.
26. Astashkina AI, Mann BK, Prestwich GD, Grainger DW: **A 3-D organoid kidney culture model engineered for high-throughput nephrotoxicity assays.** *Biomaterials* 2012, **33**:4700–4711.
  27. Li Y, Kandasamy K, Chuah JKC, Lam YN, Toh WS, Oo ZY, Zink D: **Identification of nephrotoxic compounds with embryonic stem-cell-derived human renal proximal tubular-like cells.** *Mol Pharm* 2014, **11**:1982–1990.



28. Kandasamy K, Chuah JKC, Su R, Huang P, Eng KG, Xiong S, Li Y, Chia CS, Loo L-H, Zink D: **Prediction of drug-induced nephrotoxicity and injury mechanisms with human induced pluripotent stem cell-derived cells and machine learning methods.** *Sci Rep* 2015, **5**:12337.
29. Luo QH: **Evaluation of KIM-1 and NGAL as early indicators for assessment of gentamycin-induced nephrotoxicity in vivo and in vitro.** *Kidney Blood Press Res* 2016, **41**.
30. Su R, Li Y, Zink D, Loo L-H: **Supervised prediction of drug-induced nephrotoxicity based on interleukin-6 and -8 expression levels.** *BMC Bioinform* 2014, **15**:S16.
31. Li Y: **An in vitro method for the prediction of renal proximal tubular toxicity in humans.** *Toxicol Res* 2013, **2**.
32. Morizane R, Lam AQ, Freedman BS, Kishi S, Valerius MT, Bonventre JV: **Nephron organoids derived from human pluripotent stem cells model kidney development and injury** 2015, **33**:1193–1200.
33. Dieterle F, Sistare F, Goodsaid F, Papaluca M, Ozer JS, Webb CP, Baer W, Senagore A, Schipper MJ, Vonderscher J: **Renal biomarker qualification submission: a dialog between the FDA-EMA and predictive safety testing consortium.** *Nat Biotechnol* 2010, **28**:455–462.
34. Leclerc E, Hamon J, Bois FY: **Investigation of ifosfamide and chloroacetaldehyde renal toxicity through integration of in vitro liver–kidney microfluidic data and pharmacokinetic-system biology models.** *J Appl Toxicol* 2016, **36**: 330–339.