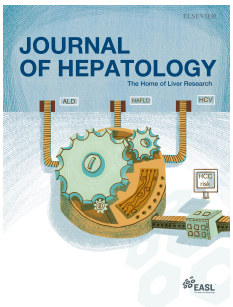


Control Compounds for Preclinical Drug-Induced Liver Injury Assessment: Consensus-driven systematic review by the ProEuroDILI Network

Antonio Segovia-Zafra, Marina Villanueva-Paz, Ana Sofia Serras, Gonzalo Matilla-Cabello, Ana Bodoque-García, Daniel Di Zeo-Sánchez, Hao Niu, Ismael Álvarez-Álvarez, Laura Sanz-Villanueva, Sergej Godec, Irina Milisav, Pierre Bagnaninchi, Raúl J. Andrade, M Isabel Lucena, José C. Fernández-Checa, Francisco Javier Cubero, Joana P. Miranda, Leonard J. Nelson

PII: S0168-8278(24)00325-8

DOI: https://doi.org/10.1016/j.jhep.2024.04.026

Reference: JHEPAT 9605

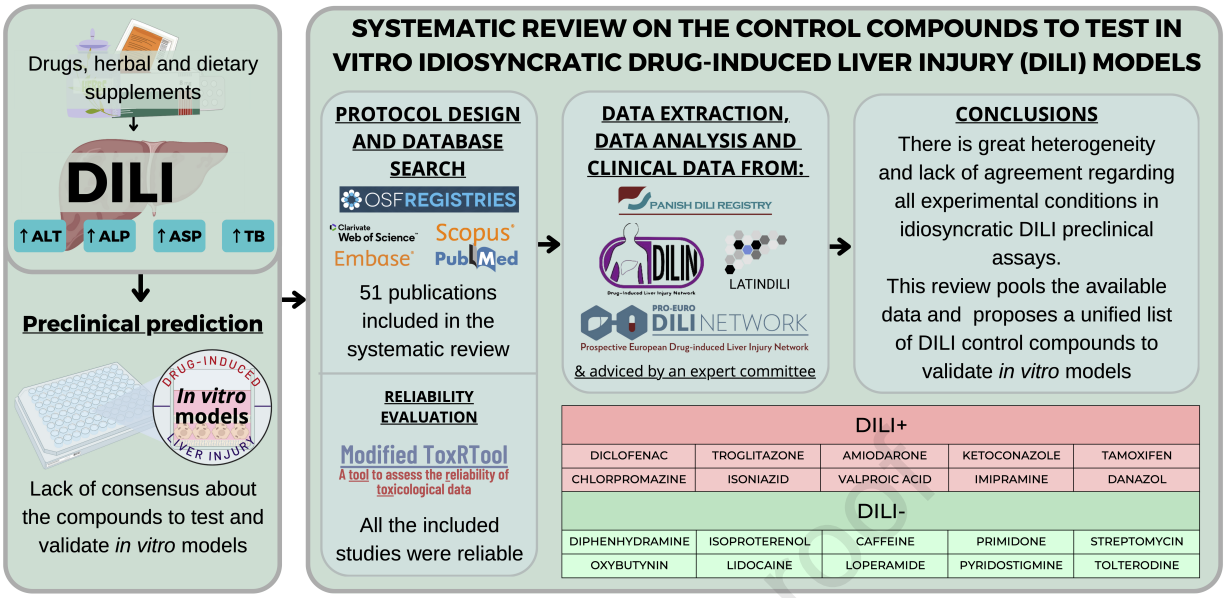
To appear in: *Journal of Hepatology*

Received Date: 14 February 2024

Revised Date: 10 April 2024

Accepted Date: 21 April 2024

Please cite this article as: Segovia-Zafra A, Villanueva-Paz M, Serras AS, Matilla-Cabello G, Bodoque-García A, Di Zeo-Sánchez D, Niu H, Álvarez-Álvarez I, Sanz-Villanueva L, Godec S, Milisav I, Bagnaninchi P, Andrade RJ, Lucena MI, Fernández-Checa JC, Cubero FJ, Miranda JP, Nelson LJ, Control Compounds for Preclinical Drug-Induced Liver Injury Assessment: Consensus-driven systematic review by the ProEuroDILI Network, *Journal of Hepatology*, https://doi.org/10.1016/j.jhep.2024.04.026.



Control Compounds for Preclinical Drug-Induced Liver Injury Assessment: Consensus-driven systematic review by the ProEuroDILI Network

Antonio Segovia-Zafra[1,2], Marina Villanueva-Paz[1,2], Ana Sofia Serras[3], Gonzalo Matilla-Cabello[1,2], Ana Bodoque-García[1], Daniel Di Zeo-Sánchez[1], Hao Niu[1], Ismael Álvarez-Álvarez[1,2], Laura Sanz-Villanueva[4,5], Sergej Godec[6,7], Irina Milisav[7,8], Pierre Bagnaninchi[9], Raúl J. Andrade[1,2,10], M Isabel Lucena[1,2,10]*, José C. Fernández-Checa[2,11,12,13,14]*, Francisco Javier Cubero[15‡], Joana P. Miranda[3‡], Leonard J. Nelson[16‡]

1 Servicios de Aparato Digestivo y Farmacología Clínica, Hospital Universitario Virgen de la Victoria, Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND, Universidad de Málaga, Málaga, Spain.

2 Centro de Investigación Biomédica en Red Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid, Spain.

3 Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal.

4 Immunology and Diabetes Unit, St Vincent's Institute, Fitzroy VIC, Australia.

5 Department of Medicine, St Vincent's Hospital, University of Melbourne, Fitzroy, VIC, Australia.

6 Department of Anaesthesiology and Surgical Intensive Care, University Medical Centre Ljubljana, Ljubljana, Slovenia

7 Institute of Pathophysiology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

8 Laboratory of oxidative stress research, Faculty of Health Sciences, University of Ljubljana, Ljubljana, Slovenia

9 Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, Scotland, United Kingdom

10 Plataforma de Investigación Clínica y Ensayos Clínicos UICEC-IBIMA, Plataforma ISCIII de Investigación Clínica, Madrid, Spain.

11 Department of Cell Death and Proliferation, Institute of Biomedical Research of Barcelona (IIBB), CSIC, Barcelona, Spain.

12 Liver Unit, Hospital Clinic I Provincial de Barcelona, Barcelona, Spain

13 Instituto de Investigaciones Biomédicas August Pi i Sunyer (IDIBAPS), Barcelona, Spain

14 Department of Medicine, Keck School of Division of Gastrointestinal and Liver disease, University of Southern California, Los Angeles, CA, United States

15 Department of Immunology, Ophthalmology and ORL, Complutense University School of Medicine, Madrid, Spain.

16 Institute for Bioengineering, School of Engineering, Faraday Building, The University of Edinburgh, Scotland, United Kingdom.

Author names in bold designate shared co-first authorship

‡ FJ.C., J.P.M. and L.J.N. equally contributed as senior authors.

* Correspondence:

M Isabel Lucena. Departamento de Farmacología, Facultad de Medicina, Universidad de Málaga, Boulevard Louis Pasteur 32, 29010, Málaga, Spain.

E-mail: lucena@uma.es

José C. Fernández-Checa. Department of Cell Death and Proliferation, Institute of Biomedical Research of Barcelona (IIBB), CSIC, 08036, Barcelona, Spain.
E-mail: checa229@yahoo.com; josecarlos.fernandezcheca@iibb.csic.es

Keywords: drug-induced liver injury; preclinical drug safety testing; validation of *in vitro* DILI models; clinical data; expert committee; panel of control drugs; control compounds.

Electronic word count: 6178 words

Abstract: 252, Manuscript: 3635, references: 2063, tables and figure legends: 228

Number of figures and tables: 1 Table, 7 Figures and 9 Supplementary Materials

Conflict of interest statement

The authors declare no competing interests.

Financial support statement

This work was supported by grants from the Instituto de Salud Carlos III and Consejería de Salud de Andalucía, cofounded by Fondo Europeo de Desarrollo Regional (FEDER) (PI21/01248, PI19-00883, PT23/00137, PEMP-0127-2020, PI-0310-2018) and Agencia Española del Medicamento y Productos Sanitarios. M.V.P and I.A.A. hold Sara Borrell contracts (CD21/00198 and CD20/00083, respectively). H.N. holds a postdoctoral contract (POSTDOC_21_00780). A.S.Z.

holds a Jaume Bosch predoctoral contract funded by CIBERehd. D.D.S. holds an iPFIS predoctoral contract funded by ISCIII (IFI21/0034). G.M.C. holds an FPU PhD fellowship from the Spanish Ministry of Science, Innovation and Universities (FPU22/03868). SCReN and CIBERehd are funded by ISCIII. A.S.S. holds a research scholarship funded by Fundação para a Ciência e a Tecnologia (2021.04902.BD). This publication is based on work from COST Action 'CA17112 Prospective European Drug-Induced Liver Injury Network' supported by COST (European Cooperation in Science and Technology); www.cost.eu. We acknowledge the support from grants PID2019-111669RB-100, PID2022-1429560B-I00 and PID2023AEP068 from Plan Nacional de I+D funded by the Agencia Estatal de Investigación (AEI), FEDER and from the CIBEREHD; as well as funding by HORIZON-HLTH-2022-STAYHLTH-02, grant number 101095679; and support from AGAUR of the Generalitat de Catalunya SGR-2021-00491, and from the Project 201916/31 funded by Fundació Marató TV3.

Authors contributions

Conceptualization & Study Design- A.S.Z., M.V.P., L.J.N., M.I.L.; Search Strategy, Study Selection and Data Extraction- A.S.Z., M.V.P., G.M.C., A.B.G, D.D.S.; Quality assessment- A.S.Z., M.V.P., A.S.S., G.M.C., A.B.G., D.D.S, I.A.A., H.N., S.G., I.M.; Data analysis- A.S.Z., M.V.P., A.S.S., G.M.C., A.B.G., D.D.S, I.A.A., H.N., L.S.-V., S.G., I.M., R.J.A; Establishment of a unified list of 10 DILI+ and 10 DILI– control compounds- A.S.Z., M.V.P, A.S.S., L.J.N., M.I.L, JC.F.-C., FJ.C., J.P.M.; Writing – Original Draft- A.S.Z., M.V.P., A.S.S. ; Writing – Review & Editing- R.J.A., M.I.L., JC.F.-C., FJ.C., J.P.M., L.J.N.

ABSTRACT

Background & Aims: Idiosyncratic drug-induced liver injury (DILI) is a complex and unpredictable event caused by drugs, herbal or dietary supplements. Early identification of human hepatotoxicity at preclinical stages remains a major challenge, in which the selection of validated *in vitro* systems and test drugs has a significant impact. This systematic review analyzed the compounds used in hepatotoxicity assays and established a list of DILI positive and negative control drugs for validation of *in vitro* models of DILI, supported by literature and clinical evidence and endorsed by an expert committee from COST Action ProEuroDILI Network (CA17112).

Methods: Following 2020 PRISMA guidelines, original research articles focusing on DILI which used *in vitro* human models and performed at least one hepatotoxicity assay with positive and negative control compounds, were included. Bias of the studies was assessed by a modified 'Toxicological Data Reliability Assessment Tool'.

Results: 51 studies (out of 2,936) met the inclusion criteria, with 30 categorized as reliable without restrictions. Although there was a broad consensus on positive compounds, the selection of negative compounds lacked clarity. 2D monoculture, short exposure times and cytotoxicity endpoints were the most tested, although there was no consensus on the drug concentrations.

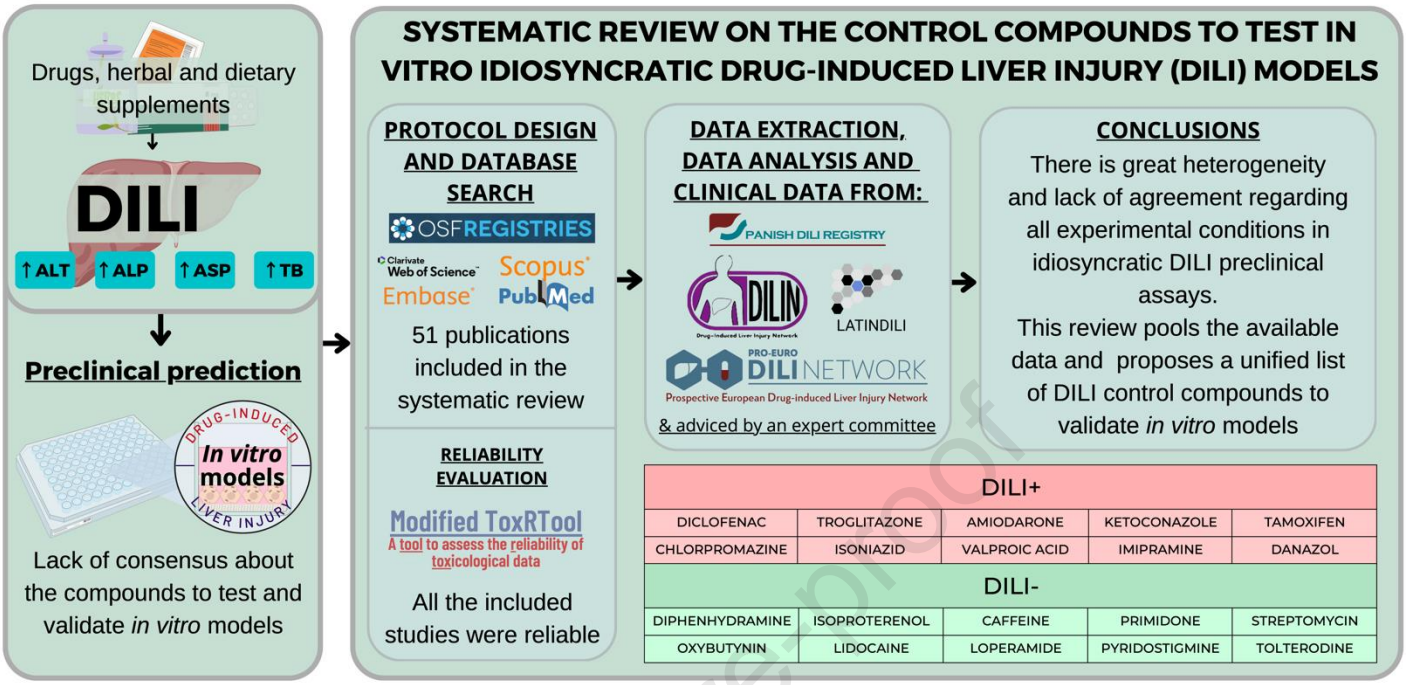
Conclusions: The extensive analysis highlighted the lack of agreement on control compounds for *in vitro* DILI assessment. Following comprehensive *in vitro* and clinical data analysis together with input from the expert committee, an evidence-based consensus-driven list of 10 positive and negative drugs is

proposed for validating *in vitro* models for improving preclinical drug safety testing regimes.

IMPACT AND IMPLICATIONS

Prediction of human toxicity early in the drug development process remains a major challenge. For this, human *in vitro* models are becoming increasingly important, however, the development of more physiologically relevant liver models and careful selection of control DILI+ and DILI- drugs are requisites to better predict DILI liability of new drug candidates. Thus, this systematic study holds critical implications for standardizing validation of new *in vitro* models for studying drug-induced liver injury (DILI). By establishing a consensus-driven list of positive and negative control drugs, the study provides a scientifically justified framework for enhancing the consistency of preclinical testing, thereby addressing a significant challenge in early hepatotoxicity identification. The results are of paramount importance to all the actors involved in the drug development process, offering a standardized approach to assess hepatotoxic risks. Practically, these findings can guide researchers in evaluating safety profiles of new drugs, refining *in vitro* models, and informing regulatory agencies on potential improvements to regulatory guidelines, ensuring a more systematic and efficient approach to drug safety assessment.

GRAPHICAL ABSTRACT



INTRODUCTION

Idiosyncratic drug-induced liver injury (DILI) encompasses liver damage caused by conventional medicines together with herbal and dietary supplements [1]. The mechanisms of toxic liver injury can be divided into at least five main categories: reactive metabolites, mitochondrial dysfunction, transporter inhibition, lysosomal impairment, and immune-mediated toxicity [2]. DILI constitutes one of the leading causes of drug attrition in clinical trials, use restriction, or withdrawal from the market [3].

Failure to predict hepatotoxicity in the drug development process is mainly due to the lack of human-relevant preclinical *in vitro* models as well as interspecies differences with animal models, resulting in poor preclinical to clinical translation. This is compounded by the multifactorial nature of DILI pathophysiology [1]. The development of more sophisticated human predictive *in vitro* models, and technologies including *in silico* approaches has thus become a priority in pharma and basic research to address hepatotoxicity risk, both in an accurate and accelerated fashion in the drug development process [4].

Predictive in vitro models for hepatotoxicity assessment must be of relevance not only at the physiological level but also of significance to pharmacological and pathological contexts [4]. A tiered approach considering not only the *in vitro* human models selected but also their phenotypic characterization, as well as pharmacological and toxicological functionality, is needed to validate new testing systems [5, 6]. Here, the lack of consensus about the selection of the most appropriate, context-specific *in vitro* human liver model, critical endpoints to analyze, number and type of control compounds to test, concentration, as well as time of exposure, contributes significantly towards the observed heterogeneity

and lack of reproducibility in results obtained in different studies [6, 7]. Several studies have also highlighted the critical importance of selecting a standardized set of prototypic hepatotoxic compounds with diverse toxicity mechanisms to validate proof-of-concept studies [8]. Moreover, pharmaceutical companies are actively engaged in applying, for example, standardized microphysiological systems for drug risk assessment [9].

When developing a new *in vitro* liver model, the study requirements determine its characteristics, in which the human cell type and tissue architecture are important initial considerations [4-6]. At a later stage, to assess the relevance of the chosen model for the study requirements, different endpoints can be applied, ranging from measurements of cell death up to more functional and mechanistic pre-cell death endpoints, reflecting the complex nature of DILI involving multiple mechanisms [10].

As for prototypic hepatotoxic compound selection, learning from drugs that either failed or were approved is an asset to test the expected hepatic response and the known mechanisms of DILI in a new *in vitro* system. Herein, we conducted a systematic review to summarize control compounds used in predictive *in vitro* human DILI models, excluding those using preclinical *in vivo* systems. We also performed a deeper analysis of the drugs most frequently used as positive and negative controls for DILI in the literature to develop a unified list of control compounds encompassing DILI-positive and DILI-negative control drugs. These findings were supported by a stringent literature-based review, using PRISMA guidelines, validated with clinical evidence and endorsed by a consensus committee of experts from the ProEuroDILI Network (CA 17112).

MATERIAL AND METHODS

Study Design and Search Strategy

This systematic review was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [11]. The protocol for the systematic review was registered in the Open Science Framework (OSF) Registries platform (osf.io/yp7g6). There were no deviations from the registered protocol.

Eligible literature published up to June 1st, 2022, was identified through a search in PubMed, Embase, Web of Science (WoS), and Scopus, with no language restrictions. The search strategy was designed based on identifying three terms: 'DILI', '*in vitro* models', and 'predictivity'. According to this strategy, the search comprised the following terms and Boolean operators: 'DRUG*' AND ('LIVER INJURY' OR 'HEPATOTOX*') AND ('PREDICT' OR 'IN VITRO' OR 'TEST*') AND ('SPECIFICITY' AND 'SENSITIVITY'). To retrieve additional studies eligible for inclusion, references cited by the included studies, narrative or systematic reviews, and meta-analyses identified throughout the literature search were manually reviewed. The retrieved literature was managed using the Rayyan online tool [12].

Inclusion and exclusion criteria

Published studies that fulfilled the following criteria were included:

1. To be a peer-reviewed original article.
2. To study the onset of DILI in preclinical stages.

3. To report at least one hepatotoxicity assay using *in vitro* human models, aiming to classify at least one drug in each of the following categories: DILI concern (DILI+) or no DILI concern (DILI-).
4. To report data about the model's predictive power, with sensitivity and specificity values (either quantitative or qualitative).

Studies conducted using *in vivo* models, and reviews, editorials, letters, commentaries, conference abstracts, and other reports with no relevant data were excluded. If the full text could not be accessed, it was searched via inter-library loan or the corresponding authors were contacted to request a copy. If the study could not be retrieved, it was finally excluded.

Study selection

The literature search was conducted by four independent researchers, who screened the title and abstract and retrieved and reviewed the full text of the relevant studies identified. Any disagreements were resolved by discussion, whilst a 5th independent researcher was consulted if a consensus needed to be reached.

Data extraction

After literature screening, the following data were extracted from each of the included studies: full name of the first author, year of publication, the model(s) used to perform the hepatotoxicity assay, drugs tested and their DILI categorization, drug concentration(s), toxicity endpoint(s) measured, and specificity and sensitivity values. For the analysis, all the drugs were classified in

the categories DILI+ and DILI-. To do so, in the studies where a binary categorization was not used, only the negative controls used for the hepatotoxicity assays and the model predictivity estimation were indexed as DILI-. The positive controls were classified as DILI+ regardless of their severity category. Corresponding authors were contacted to obtain further information if required.

Quality assessment

Based on the software-based 'Toxicological data Reliability assessment Tool' (ToxRTool) [13] and following published recommendations [14], a refined tool named 'Modified ToxRTool' was generated to assess the risk of bias in the included articles. The 'Modified ToxRTool' provides comprehensive criteria for determining the reliability of toxicological studies. Studies were evaluated in five domains: i) 'Test substance identification'; ii) 'Test system characterization'; iii) 'Study design description'; iv) 'Study results documentation'; and v) 'Plausibility of study design and data'. Each domain item was scored with 0, 0.5, or 1 point, following the recommendations of Segal D. *et al*. [14]. After the evaluation, categories of reliability proposed by Klimisch *et al.* [15], i.e., code 1 (reliable without restrictions), code 2 (reliable with restrictions), code 3 (not reliable), and code 4 (not assignable), were assigned to each domain (see **Figures 1 and 2**). Ten researchers independently conducted the quality control assessments. Three independent researchers evaluated each study and the mean score for each category was calculated based on the assessment of their assessment. Studies with a score of 15-18 points were classified as reliable without restrictions, being useful for the analysis; 11-14 points studies were classified as

reliable with restrictions, being potentially useful for the analysis; <11 points studies were classified as not reliable and were not considered further.

Data analysis

Data on the different models, drugs and conditions used to perform the toxicity assays and sensitivity and specificity values were analyzed. Other aspects of the toxicological assays were scored, such as the number of times each model and specific spatial configurations (2D, 3D) were used, how long the models were exposed to the drugs, or which concentrations were tested. However, the heterogeneity in the study design concerning cell types, model configurations, and the number of DILI+ and DILI- compounds analyzed, along with the wide range in exposure times and concentrations, prevented conducting a meta-analysis.

A full list showing all drugs and the number of articles where they appeared as positive or negative controls was created (**Supplementary Material 1**).

Drug analysis

Drugs most commonly used as positive and negative controls in the included literature were selected for examination. At least one hundred drugs from each category, DILI+ and DILI-, were extensively analyzed (**Supplementary Material 2**). If there were additional drugs with the same number of occurrences in the articles beyond the hundredth drug for positive and negative controls, such drugs were included in the analysis, expanding the list as necessary. The classification of the drugs based on the pharmacological group was extracted from the ATC/DDD index 2023 [16].

Clinical use cases | Databases

For the number of clinical cases reported per each drug, four DILI databases were analyzed: the **Spanish DILI Registry** [17] with 980 cases, the **Pro-Euro-DILI Registry** [18] with 246 cases, the DILI Network (**DILIN**) [19] with 899 cases and the **LATINDILI Network** [20] with 480 cases. The **DrugBank** database [21] was used to ascertain if a drug was withdrawn, and the reason for withdrawal. The Liver Toxicity Knowledge Base (LTKB) [22] was used to ascertain: The severity class, drug label, and DILI concern, which were then obtained from the DILIRank list [23]. The classification as DILI+ was extracted from the DILIst [24]. In addition, several of the toxicity properties (mitochondrial liability and reactive metabolite formation), pharmacokinetic properties (half-life, lipophilicity, plasma protein binding, enterohepatic circulation and hepatic metabolism), and the Biopharmaceutics Drug Disposition Classification System (BDDCS) class were also extracted from the LKTB. The DILI injury type and toxicity mechanism(s) of the drug were obtained from the Liver Toxicity (LiverTox) Database [25], whilst physicochemical properties, metabolic pathway and enzymes implied were taken from DrugBank [21].

Establishment of a unified list of 10 DILI+ and 10 DILI– control compounds

After revising the different control drugs used in the articles, clinical data from DILI cases, physicochemical, pharmacokinetic, and toxicological characteristics along with the contribution of a panel of DILI experts, a list of control DILI compounds was created. The DILI experts panel comprised members of the ProEuroDILI Network (CA 17112). To be included in the list, drugs were required to be sufficiently explored in both the selected studies and clinical databases and

represent all major types of liver injury phenotypes and drug metabolism. After the selection of a potential list of DILI+ and DILI- drugs, the DILI expert panel convened to reach a consensus on the appropriateness of these drugs, drawing on their extensive knowledge in both clinical and preclinical DILI.

RESULTS

Literature search

The search strategy retrieved 2,936 studies. Of these, 1,341 were duplicate records. After screening the title and abstract, 1,923 studies were excluded. The main reasons for exclusion were: i) The study of a condition other than DILI (75%); ii) Not studying hepatotoxicity in *in vitro* human models (17.1%), and iii) Not being an original article (7.9%). The full texts of the remaining 125 articles were assessed for eligibility, with 67 studies excluded, mainly due to lack of DILI studies in *in vitro* human models and absence of hepatotoxicity assays. Furthermore, 8 articles were excluded due to a drug bias selection. Ultimately, 51 articles [26-76] that met the stringent inclusion criteria were included for in-depth analysis and systematic review (**Figure 1**).

Study characteristics

A summary of the main characteristics of the 51 articles that fully met the inclusion criteria is shown in **Supplementary Material 3**. The relevant features assessed were the type of *in vitro* human model, the study context, the DILI control drugs used (concentrations and time of exposure), the endpoints studied and the respective predictivity of the system.

Reliability of the studies

The quality and reliability of all 51 publications assessed using the Modified ToxRTool are shown in **Figure 2**. Most studies were considered reliable without restrictions regarding test substance identification (69%), test system characterization (84%), study design description (81%), results documentation (81%), and plausibility of the study design and data (100%). Based on these parameters, most of the studies were categorized as reliable without restrictions (59%). The remaining studies were categorized as reliable with restrictions (41%), being potentially useful. Importantly, no article was classified as not reliable.

DILI categorization

Among the 51 articles included, 43 studies (84%) classified the drugs using a simple binary categorization, namely, DILI-positive (DILI+) *versus* DILI-negative (DILI-). In contrast, 7 studies (14%) used a tertiary categorization, *e.g.,* Most-, Less- or No-DILI-concern, while only 1 (2%) used further categorization (**Supplementary Material 3**).

DILI control drugs

The applicability of the *in vitro* model depends largely on the number of tested DILI+ and DILI- controls. A list of all drugs used in the 51 articles is presented in **Supplementary Material 1**.

Supplementary Material 2 summarizes the characteristics of the 104 DILI+ and 123 DILI- drugs that were most commonly used in the 51 studies analyzed. Diclofenac (45 studies; 88%) and buspirone (25 studies; 49%) were the most

widely investigated DILI-positive and DILI-negative drugs, respectively. Nevertheless, a wide heterogeneity was found in the DILI- categorization, which *a priori* stems from the fact that some articles consider Less-DILI-Concern drugs as DILI-, despite being classified as potentially hepatotoxic.

The LTKB gathers diverse datasets for DILI assessment and prediction. We therefore evaluated the distribution of DILI labelling and severity among the drugs studied (**Figure 3**). Out of 104 DILI+ control drugs analyzed, 66 drugs were categorized as Most-DILI-Concern and 29 as Less-DILI-Concern. On the other hand, within the 123 DILI- control drugs, 59 drugs were categorized as No-DILI-Concern, 33 as Less-DILI-Concern and 13 as Ambiguous-DILI-Concern. Of note, two drugs used as DILI- (2%), levofloxacin and atorvastatin, present Most-DILI-Concern and were used both as DILI+ and DILI- controls.

DILI Registries

When examining cases of hepatotoxicity for the analyzed drugs across various DILI registries, the Spanish DILI Registry had a higher number of reported cases for DILI+ drugs (53%) compared to DILI- drugs (20%). Conversely, the Pro-Euro-DILI Registry has the fewest reported cases for both DILI+ (29%) and DILI- (19%) drugs (**Supplementary Material 2)**. Nevertheless, these differences could be explained, in part, by the differences in the causative drugs among registries, being more frequent biological and immunosuppressants in recent times.

***In vitro* human models**

Factors such as the cell type(s) and the number of different model formats used in each study constitute essential features when evaluating the model's

predictivity. **Figure 4** and **Supplementary Material 4** show the different cellular, non-cellular models and culture conditions tested. The majority consisted in human primary hepatocytes (PHHs) and 2D culture configurations. A more detailed analysis of advantages, relevance and limitations of preclinical models for predicting DILI is provided in a recent review [4].

Drug concentration, time of exposure and endpoints evaluated

When deciding the optimal conditions for *in vitro* hepatotoxicity testing, important parameters to be taken into account include drug concentration (*e.g.,* multiples of C_{max}), time of exposure (*e.g.,* acute or chronic) and endpoints (*e.g.,* cytotoxicity or mechanistic endpoints). Herein, great variation in all these parameters was observed in the studies analyzed (**Figure 5, Supplementary Material 3, 5** and **6**).

Predictive capacity

Extensive variability was found when analyzing data related to the predictive capacity of the different models (**Supplementary Material 7**). For example, the number of drugs used to determine the predictive ability of a model varied among all the articles analyzed, with some using different numbers depending on the model [72], the model and the endpoint [47, 57], or the cutoff used [55].

Proposed Control Drugs: 10 DILI+ and 10 DILI– compounds

After extensive analysis of all drugs examined in this study, including their physicochemical, pharmacokinetic/pharmacodynamic characteristics, mode of action (MoA) and toxicity, a list of 10 DILI+ and 10 DILI- drugs was established

to assist in validation of *in vitro* DILI systems, applicable to both current and next-generation advanced preclinical human *in vitro* systems (**Figure 6**, **Supplementary Material 8** and **9**).

During the control compounds selection, additional features were further explored to find the most suitable ones. These included metabolic pathways, mechanisms of toxicity, and pharmacological therapeutic class (**Supplementary Material 8** and **9**).

Virtually, all phase I enzymes involved in the selected DILI+ drugs biotransformation (excluding CYP2C18, CYP2J2M, FMO1, and FMO3) are also involved in the selected DILI- drugs metabolism. However, this is not the case for phase II enzymes, since the DILI- control drugs undergo minimal phase II metabolism (*via* UGT, COMT, and GSTP enzyme families).

Moreover, different mechanisms of hepatotoxicity are represented within the DILI+ controls list, such as immune-allergic toxicity (*e.g.*, diclofenac), mitochondrial dysfunction (*e.g.*, amiodarone), cholestatic liver injury (*e.g.*, danazol), amongst others.

Regarding drug concentrations tested for each drug, the most frequently used are multiples of the maximum plasma concentration (C_{max}). Finally, given that not all studies analyzed use the same C_{max}, a simplified distribution of C_{max} concentrations was determined for these 20 drugs (**Figure 7**).

DISCUSSION

Early prediction of human toxicity in the drug development process remains a major challenge. This systematic review principally addresses the lack of a standardized panel of training compounds (DILI-positive and DILI-negative

drugs), which would allow appropriate and more robust validation of human *in vitro* liver models for hepatotoxicity studies.

The analysis performed has raised several paradoxical classifications of various drugs. Acetylsalicylic acid, a commonly used non-steroidal anti-inflammatory drug (NSAID), constitutes a drug categorized as Less-DILI-Concern by the LTKB but also presents some clinical DILI cases reported within the Spanish DILI Registry [17]. Interestingly, in the studies reviewed, this drug is widely used as both a DILI- and DILI+ control. Other drugs such as fluoxetine, warfarin, alendronic acid, entacapone, or metformin also share this confounding feature. This raises a concern regarding whether an Ambiguous-/ or Less-DILI-Concern drug should be used as a DILI- control compound. When defining a panel of control drugs, it is essential to have different categories within DILI+ drugs to cover not only severe but also mild hepatotoxicity effects, *i.e.*, to have drugs categorized as Most-DILI-Concern and Less-DILI-Concern. However, when defining true DILI- controls, the drugs should not have any propensity to cause DILI, *i.e.*, only No-DILI-Concern drugs should be included. Moreover, DILI- compounds must not bear any clinical cases within the DILI registries.

The initial aim of our review of literature-reported DILI studies was to propose a consensus list of DILI+ and DILI- drugs to validate human *in vitro* DILI models. The list evolved after conducting an extensive analysis of all the drugs and model systems examined in this study. This included data from clinical cases of DILI caused by the same drugs under analysis herein, along with their physicochemical, pharmacokinetic, and toxicological characteristics. Additionally, a panel of experts in the field of DILI (ProEuroDILI Network, CA 17112) was

convened to provide expertise and critical input on the proposed list. To our knowledge, this is the first systematic review that brings together all these data.

DILI+ Control Drugs

DILI+ drugs were the initial focus, identifying ten drugs fulfilling the established criteria and conditions. Firstly, the drugs were required to have been significantly used in both the selected studies and, importantly, also in clinical databases. Additionally, it was crucial for the drugs chosen to accurately represent all major types of liver injury sub-types (hepatocellular, cholestatic, or mixed), modes of drug metabolism and belong to either Most- or Less-DILI-Concern groups. Next, we assessed aspects of drug metabolism, mechanisms of toxicity, and pharmacological properties of the selected DILI control compounds (the list of DILI+ drugs is displayed in **Figure 6,** in red, and their main characteristics are shown in **Supplementary Material 8**).

DILI- Control Drugs

For selecting DILI- drugs, a series of pre-established criteria were followed. First, the selected drugs should be metabolized to the greatest extent by the same phase I and II enzymes as the DILI+ drugs. Additionally, these drugs should not have reported occurrences as DILI+ in the literature, be classified as severity class 0 in the LTKB, and be considered No-DILI-Concern (**Figure 6,** in green, and **Supplementary Material 9**).

Recommended concentrations to use for each drug included in the list were derived exclusively from the literature (**Table 1**). These concentrations were

selected based on the most used C_{max} value for each drug tested. In the case that the same number of articles use different concentrations for the DILI- drugs, the highest value is proposed.

Of note, as the use of *in vivo* models was part of the exclusion criteria of the present study, the suitability of these control compounds to validate DILI animal models should be further analyzed.

CONCLUSION

Given the multifactorial, complex nature of idiosyncratic DILI, no single system has yet emerged as a universal preclinical testing platform. Moreover, there is a clear and unmet need for consensus on the reference drugs to be used to validate DILI assays, recommendations about the concentrations to test and criteria for interpreting the data. This systematic study proposes an evidence-based, consensus-driven, unified list of 10 positive and 10 negative control drugs to provide benchmarking, continuity and reproducibility in the validation of human *in vitro* models for improving preclinical drug safety testing studies. From an initial corpus of nearly 3,000 literature-based studies, only 51 met the rigorous inclusion criteria to achieve this goal based on PRISMA guidelines. In addition, we report tremendous variation in the types of human *in vitro* DILI models used, for example, across cell choice and culture formats (2D monolayers to 3D multi-cellular cultures) and conditions (medium formulations; extracellular matrix configuration). Together with the lack of concordance due to interspecies differences between animal models with human DILI, a paradigm shift is required to develop models with human-relevant biological variation and complexity, which

can lead to a better understanding of mechanistic and predictive DILI signals. Therefore, cross-validation of at least 2-3 human *in vitro* models with both animal models and clinical data and integration using artificial intelligence (AI) approaches should form part of the standard operating procedure for DILI prediction in the future. Applying appropriate DILI+ and DILI- control compounds to test these systems would provide added value in terms of validation, benchmarking and improving relevance of the model to evaluate drug toxicity as well as pharmacologic effects.

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AI	Artificial intelligence
BDDCS	Biopharmaceutics Drug Disposition Classification System
CA17112	COST Action ProEuroDILI Network
C_{max}	Maximum plasma concentration
DILI	Drug-induced liver injury
DILI-	DILI negative (No DILI concern)
DILI+	DILI positive (DILI concern)
DILIN	DILI Network
LiverTox	Liver Toxicity Database
LTKB	Liver Toxicity Knowledge Base
MoA	Mode of action
NSAID	Nonsteroidal anti-inflammatory drug
OSF	Open Science Framework
PHHs	Primary human hepatocytes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ToxRTool	Toxicological data Reliability assessment Tool
WoS	Web of Science

Acknowledgements

We thank Dr. Minjun Chen for providing the LKTB list.

Materials & Correspondence

Correspondence and material requests should be addressed to:

-M Isabel Lucena. Departamento de Farmacología, Facultad de Medicina, Universidad de Málaga, Boulevard Louis Pasteur 32, 29010, Málaga, Spain.

E-mail: lucena@uma.es

-José C. Fernández-Checa. Department of Cell Death and Proliferation, Institute of Biomedical Research of Barcelona (IIBB), CSIC, 08036, Barcelona, Spain.

E-mail: checa229@yahoo.com; josecarlos.fernandezcheca@iibb.csic.es

REFERENCES

Author names in bold designate shared co-first authorship

[1] Andrade RJ, Chalasani N, Björnsson ES, et al. Drug-induced liver injury. Nature Reviews Disease Primers 2019;5:58.

[2] Walker PA, Ryder S, Lavado A, et al. The evolution of strategies to minimise the risk of human drug-induced liver injury (DILI) in drug discovery and development. Archives of Toxicology 2020;94:2559-2585.

[3] European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Drug-induced liver injury. J Hepatol 2019;70:1222-1261.

[4] Fernandez-Checa JC, Bagnaninchi P, Ye H, et al. Advanced preclinical models for evaluation of drug-induced liver injury - consensus statement by the European Drug-Induced Liver Injury Network [PRO-EURO-DILI-NET]. J Hepatol 2021;75:935-959.

[5] **Serras AS, Rodrigues JS, Cipriano M**, et al. A Critical Perspective on 3D Liver Models for Drug Metabolism and Toxicology Studies. Front Cell Dev Biol 2021;9:626805.

[6] Weaver RJ, Blomme EA, Chadwick AE, et al. Managing the challenge of drug-induced liver injury: a roadmap for the development and deployment of preclinical predictive models. Nat Rev Drug Discov 2020;19:131-148.

[7] Olson H, Betton G, Robinson D, et al. Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals. Regulatory Toxicology and Pharmacology 2000;32:56-67.

[8] Zhou Y, Shen JX, Lauschke VM. Comprehensive Evaluation of Organotypic and Microphysiological Liver Models for Prediction of Drug-Induced Liver Injury. Frontiers in Pharmacology 2019;10.

[9] Schofield CA, Walker TM, Taylor MA, et al. Evaluation of a Three-Dimensional Primary Human Hepatocyte Spheroid Model: Adoption and Industrialization for the Enhanced Detection of Drug-Induced Liver Injury. Chemical Research in Toxicology 2021;34:2485-2499.

[10] Vinken M, Hengstler JG. Characterization of hepatocyte-based in vitro systems for reliable toxicity testing. Arch Toxicol 2018;92:2981-2986.

[11] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

[12] Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews 2016;5:210.

[13] Schneider K, Schwarz M, Burkholder I, et al. “ToxRTool”, a new tool to assess the reliability of toxicological data. Toxicol Lett 2009;189:138-144.

[14] Segal D, Makris SL, Kraft AD, et al. Evaluation of the ToxRTool’s ability to rate the reliability of toxicological data for human health hazard assessments. Regulatory Toxicology and Pharmacology 2015;72:94-101.

[15] Klimisch HJ, Andreae M, Tillmann U. A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology 1997;25:1-5.

[16] World Health Organization (WHO). Guidelines for ATC classification and DDD assignment. 2010 [cited 3 May 2023]; Available from: https://www.whocc.no/atc_ddd_index/

[17] **Stephens C, Robles-Diaz M**, Medina-Caliz I, et al. Comprehensive analysis and insights gained from long-term experience of the Spanish DILI Registry. J Hepatol 2021;75:86-97.

[18] **Björnsson ES, Stephens C**, Atallah E, et al. A new framework for advancing in drug-induced liver injury research. The Prospective European DILI Registry. Liver Int 2023;43:115-126.

[19] Chalasani N, Bonkovsky HL, Fontana R, et al. Features and Outcomes of 899 Patients With Drug-Induced Liver Injury: The DILIN Prospective Study. Gastroenterology 2015;148:1340-1352.e1347.

[20] **Bessone F, Hernandez N**, Lucena MI, et al. The Latin American DILI Registry Experience: A Successful Ongoing Collaborative Strategic Initiative. Int J Mol Sci 2016;17:313.

[21] Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research 2006;34:D668-D672.

[22] **Chen M, Vijay V**, Shi Q, et al. FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discovery Today 2011;16:697-703.

[23] Chen M, Suzuki A, Thakkar S, et al. DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discovery Today 2016;21:648-653.

[24] Thakkar S, Li T, Liu Z, et al. Drug-induced liver injury severity and toxicity (DILIst): binary classification of 1279 drugs by human hepatotoxicity. Drug Discovery Today 2020;25:201-208.

[25] LiverTox. Clinical and Research Information on Drug-Induced Liver Injury. [cited 27 April 2023]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK547852/

[26] **Albrecht W, Kappenberg F, Brecklinghaus T**, et al. Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood concentrations. Arch Toxicol 2019;93:1609-1637.

[27] **Aleo MD, Shah F**, He K, et al. Evaluating the Role of Multidrug Resistance Protein 3 (MDR3) Inhibition in Predicting Drug-Induced Liver Injury Using 125 Pharmaceuticals. Chem Res Toxicol 2017;30:1219-1229.

[28] Atienzar FA, Novik EI, Gerets HH, et al. Predictivity of dog co-culture model, primary human hepatocytes and HepG2 cells for the detection of hepatotoxic drugs in humans. Toxicol Appl Pharmacol 2014;275:44-61.

[29] Basharat A, Rollison HE, Williams DP, et al. HepG2 (C3A) spheroids show higher sensitivity compared to HepaRG spheroids for drug-induced liver injury (DILI). Toxicol Appl Pharmacol 2020;408.

[30] Bell CC, Dankers ACA, Lauschke VM, et al. Comparison of Hepatic 2D Sandwich Cultures and 3D Spheroids for Long-term Toxicity Applications: A Multicenter Study. Toxicol Sci, 2 ed; 2018. p. 655-666.

[31] Boon R, Kumar M, Tricot T, et al. Amino acid levels determine metabolism and CYP450 function of hepatocytes and hepatoma cell lines. Nat Commun 2020;11:1393.

[32] Burkard A, Dähn C, Heinz S, et al. Generation of proliferating human hepatocytes using Upcyte® technology: characterisation and applications in induction and cytotoxicity assays. Xenobiotica 2012;42:939-956.

[33] Eckstrum K, Striz A, Ferguson M, et al. Evaluation of the utility of the Beta Human Liver Emulation System (BHLES) for CFSAN's regulatory toxicology program. Food Chem Toxicol 2022;161:112828.

[34] **Garside H, Marcoe KF**, Chesnut-Speelman J, et al. Evaluation of the use of imaging parameters for the detection of compound-induced hepatotoxicity in 384-well cultures of HepG2 cells and cryopreserved primary human hepatocytes. Toxicol Vitro 2014;28:171-181.

[35] Gerets HH, Tilmant K, Gerin B, et al. Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. Cell Biol Toxicol 2012;28:69-87.

[36] Harada K, Kohara H, Yukawa T, et al. Cell-based high-throughput screening for the evaluation of reactive metabolite formation potential. Toxicol Vitro 2021;74.

[37] Hirashima R, Itoh T, Tukey RH, et al. Prediction of drug-induced liver injury using keratinocytes. J Appl Toxicol 2017;37:863-872.

[38] **Hussain F, Basu S**, Heng JJH, et al. Predicting direct hepatocyte toxicity in humans by combining high-throughput imaging of HepaRG cells and machine learning-based phenotypic profiling. Arch Toxicol 2020;94:2749-2767.

[39] **Kawaguchi M, Nukaga T**, Sekine S, et al. Mechanism-based integrated assay systems for the prediction of drug-induced liver injury. Toxicol Appl Pharmacol 2020;394:114958.

[40] Khetani SR, Kanchagar C, Ukairo O, et al. Use of micropatterned cocultures to detect compounds that cause drug-induced liver injury in humans. Toxicol Sci 2013;132:107-117.

[41] Kohara H, Bajaj P, Yamanaka K, et al. High-throughput screening to evaluate inhibition of bile acid transporters using human hepatocytes isolated from chimeric mice. Toxicol Sci 2020;173:347-361.

[42] Li F, Cao L, Parikh S, Zuo R. Three-Dimensional Spheroids With Primary Human Liver Cells and Differential Roles of Kupffer Cells in Drug-Induced Liver Injury. J Pharm Sci 2020;109:1912-1923.

[43] Lin Z, Will Y. Evaluation of drugs with specific organ toxicities in organ-specific cell lines. Toxicological Sciences 2012;126:114-127.

[44] Luo Y, Rana P, Will Y. Palmitate increases the susceptibility of cells to drug-induced toxicity: An In Vitro method to identify drugs with potential contraindications in patients with metabolic disease. Toxicol Sci 2012;129:346-362.

[45] Maiuri AR, Wassink B, Turkus JD, et al. Synergistic Cytotoxicity from Drugs and Cytokines In Vitro as an Approach to Classify Drugs According to Their Potential to Cause Idiosyncratic Hepatotoxicity: A Proof-of-Concept Study. J Pharmacol Exp Ther 2017;362:459-473.

[46] Mennecozzi M, Landesmann B, Palosaari T, et al. Sex differences in liver toxicity-do female and male human primary hepatocytes react differently to toxicants in vitro? PLoS One 2015;10:e0122786.

[47] Norona LM, Fullerton A, Lawson C, et al. In vitro assessment of farnesoid X receptor antagonism to predict drug-induced liver injury risk. Arch Toxicol 2020;94:3185-3200.

[48] Novik EI, Dwyer J, Morelli JK, et al. Long-enduring primary hepatocyte-based co-cultures improve prediction of hepatotoxicity. Toxicol Appl Pharmacol. United States; 2017. p. 20-30.

[49] O'Brien PJ, Irwin W, Diaz D, et al. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. Arch Toxicol 2006;80:580-604.

[50] Oda S, Matsuo K, Nakajima A, et al. A novel cell-based assay for the evaluation of immune- and inflammatory-related gene expression as biomarkers for the risk assessment of drug-induced liver injury. Toxicol Lett 2016;241:60-70.

[51] Oda S, Uchida Y, Aleo MD, et al. An in vitro coculture system of human peripheral blood mononuclear cells with hepatocellular carcinoma-derived cells for predicting drug-induced liver injury. Arch Toxicol 2021;95:149-168.

[52] Ott LM, Ramachandran K, Stehno-Bittel L. An Automated Multiplexed Hepatotoxicity and CYP Induction Assay Using HepaRG Cells in 2D and 3D. SLAS Discov 2017;22:614-625.

[53] Persson M, Løye AF, Mow T, Hornberg JJ. A high content screening assay to predict human drug-induced liver injury during drug discovery. Journal of pharmacological and toxicological methods 2013;68:302-313.

[54] **Proctor WR, Foster AJ**, Vogt J, et al. Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury. Arch Toxicol 2017;91:2849-2863.

[55] Rana P, Aleo MD, Gosink M, et al. Evaluation of in Vitro Mitochondrial Toxicity Assays and Physicochemical Properties for Prediction of Organ Toxicity Using 228 Pharmaceutical Drugs. Chem Res Toxicol 2019;32:156-167.

[56] Rose S, Cuvellier M, Ezan F, et al. DMSO-free highly differentiated HepaRG spheroids for chronic toxicity, liver functions and genotoxicity studies. Arch Toxicol 2022;96:243-258.

[57] Schadt S, Simon S, Kustermann S, et al. Minimizing DILI risk in drug discovery - A screening tool for drug candidates. Toxicol Vitro 2015;30:429-437.

[58] Shah F, Leung L, Barton HA, et al. Setting clinical exposure levels of concern for drug-induced liver injury (DILI) using mechanistic in vitro assays. Toxicol Sci 2015;147:500-514.

[59] Shimizu Y, Sasaki T, Yonekawa E, et al. Association of CYP1A1 and CYP1B1 inhibition in in vitro assays with drug-induced liver injury. J Toxicol Sci 2021;46:167-176.

[60] **Shinozawa T, Kimura M, Cai Y**, et al. High-Fidelity Drug-Induced Liver Injury Screen Using Human Pluripotent Stem Cell-Derived Organoids. Gastroenterology 2021;160:831-846.e810.

[61] Thompson RA, Isin EM, Li Y, et al. In vitro approach to assess the potential for risk of idiosyncratic adverse reactions caused by candidate drugs. Chemical research in toxicology 2012;25:1616-1632.

[62] Tolosa L, Pinto S, Donato MT, et al. Development of a multiparametric cell-based protocol to screen and classify the hepatotoxicity potential of drugs. Toxicol Sci 2012;127:187-198.

[63] Tomida T, Okamura H, Satsukawa M, et al. Multiparametric assay using HepaRG cells for predicting drug-induced liver injury. Toxicol Lett 2015;236:16-24.

[64] Tomida T, Okamura H, Yokoi T, et al. A modified multiparametric assay using HepaRG cells for predicting the degree of drug-induced liver injury risk. J Appl Toxicol 2017;37:382-390.

[65] Vorrink SU, Zhou Y, Ingelman-Sundberg M, et al. Prediction of drug-induced hepatotoxicity using long-term stable primary hepatic 3D spheroid cultures in chemically defined conditions. Toxicol Sci 2018;163:655-665.

[66] Ware BR, Berger DR, Khetani SR. Prediction of drug-induced liver injury in micropatterned co-cultures containing iPSC-derived human hepatocytes. Toxicol Sci 2015;145:252-262.

[67] **Ware BR, Brown GE**, Soldatow VY, et al. Long-Term Engineered Cultures of Primary Mouse Hepatocytes for Strain and Species Comparison Studies During Drug Development. Gene Expr 2019;19:199-214.

[68] Ware BR, Liu JS, Monckton CP, et al. Micropatterned Coculture with 3T3-J2 Fibroblasts Enhances Hepatic Functions and Drug Screening Utility of HepaRG Cells. Toxicol Sci 2021;181:90-104.

[69] **Williams DP, Lazic SE**, Foster AJ, et al. Predicting Drug-Induced Liver Injury with Bayesian Machine Learning. Chem Res Toxicol 2020;33:239-248.

[70] Xu J, Oda S, Yokoi T. Cell-based assay using glutathione-depleted HepaRG and HepG2 human liver cells for predicting drug-induced liver injury. Toxicol In Vitro 2018;48:286-301.

[71] Xu JJ, Henstock PV, Dunn MC, et al. Cellular imaging predictions of clinical drug-induced liver injury. Toxicol Sci 2008;105:97-105.

[72] Xu Q, Liu L, Vu H, et al. Can Galactose Be Converted to Glucose in HepG2 Cells? Improving the in Vitro Mitochondrial Toxicity Assay for the Assessment of Drug Induced Liver Injury. Chem Res Toxicol 2019;32:1528-1544.

[73] Yamaoka T, Kitamura Y. Characterization of a highly sensitive and selective novel trapping reagent, stable isotope labeled glutathione ethyl ester, for the detection of reactive metabolites. J Pharmacol Toxicol Methods 2015;76:83-95.

[74] Yu KN, Nadanaciva S, Rana P, et al. Prediction of metabolism-induced hepatotoxicity on three-dimensional hepatic cell culture and enzyme microarrays. Arch Toxicol 2018;92:1295-1310.

[75] Yucha RW, He K, Shi Q, et al. In Vitro Drug-Induced Liver Injury Prediction: Criteria Optimization of Efflux Transporter IC50 and Physicochemical Properties. Toxicol Sci 2017;157:487-499.

[76] Zhang J, Doshi U, Suzuki A, et al. Evaluation of multiple mechanism-based toxicity endpoints in primary cultured human hepatocytes for the identification of drugs with clinical hepatotoxicity: Results from 152 marketed drugs with known liver injury profiles. Chem-Biol Interact 2016;255:3-11.

Tables

Table 1. Suggested C_{max} concentrations of DILI+ and DILI- control compounds.

Control Compound	Suggested Cmax (μM)
DILI+	
Diclofenac	8
Troglitazone	6.4
Amiodarone	5.3
Ketoconazole	11.3
Tamoxifen	0.2
Chlorpromazine	0.9
Isoniazid	76.6
Valproate	693.4
Imipramine	0.1
Danazol	0.1
DILI-	
Diphenhydramine	0.34
Isoproterenol	2.4
Caffeine	77.24
Primidone	4.77
Streptomycin	74.5
Oxybutynin	0.02
Lidocaine	36
Loperamide	0.08

Pyridostigmine	1.1
Tolterodine	0.04

Figure legends

Fig. 1. Flow chart of the literature review strategy.

Fig. 2. Individual (A) and overall (B) quality assessment of the 51 included publications. 1 (green), reliable without restrictions; 2 (yellow), reliable with restrictions; 3 (red), not reliable.

Fig. 3. Classification of DILI positive (A) and DILI negative (B) drugs in the 51 analyzed studies according to DILI concern, safety information in drug labelling and regulatory action.

Fig. 4. Use of different types of *in vitro* human cell models in the 51 included studies.

2D, two-dimensional; 3D, three-dimensional; HLCs, hepatocyte-like cells; iPSCs, induced pluripotent stem cells; PBMCs, peripheral blood mononuclear cells; PHHs, human primary hepatocytes; THLEs, transformed human liver epithelial cells.

Fig. 5. Different assay readouts used in the 51 included studies. (A) Percentage of studies that use each endpoint category. (B) Percentage of studies that use 1 to 8 different endpoints.

ROS, reactive oxygen species.

***Fig. 6*.** List of the 10 DILI+ and 10 DILI- drugs selected as positive (red) and negative (green) control compounds to validate new *in vitro* models.

***Fig. 7*.** Distribution of the different C_{max} used in all studies. The graphs represent the distribution range and the median of the data. Descriptive statistics are also provided for each drug. (A) shows DILI+ drugs and (B) shows DILI- drugs. C_{max}, maximum drug concentration in blood.

Figures

Figure 1:

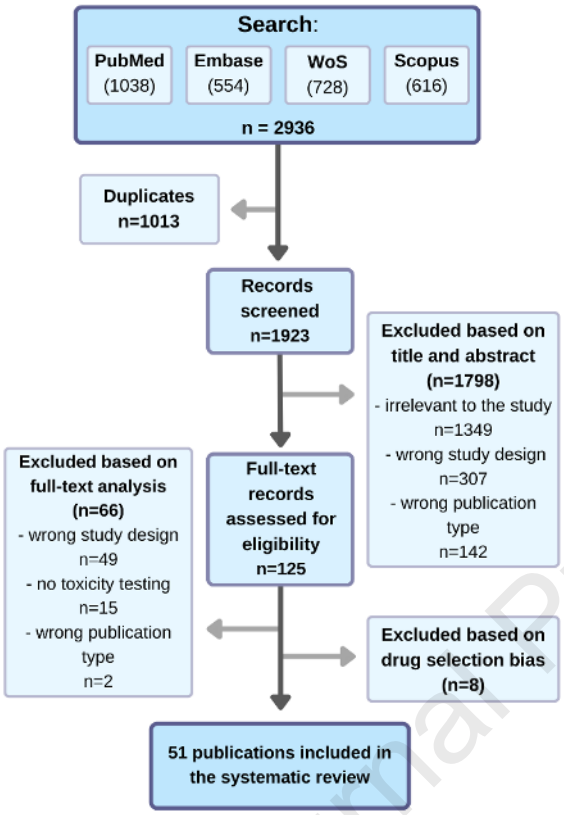


Figure 2:

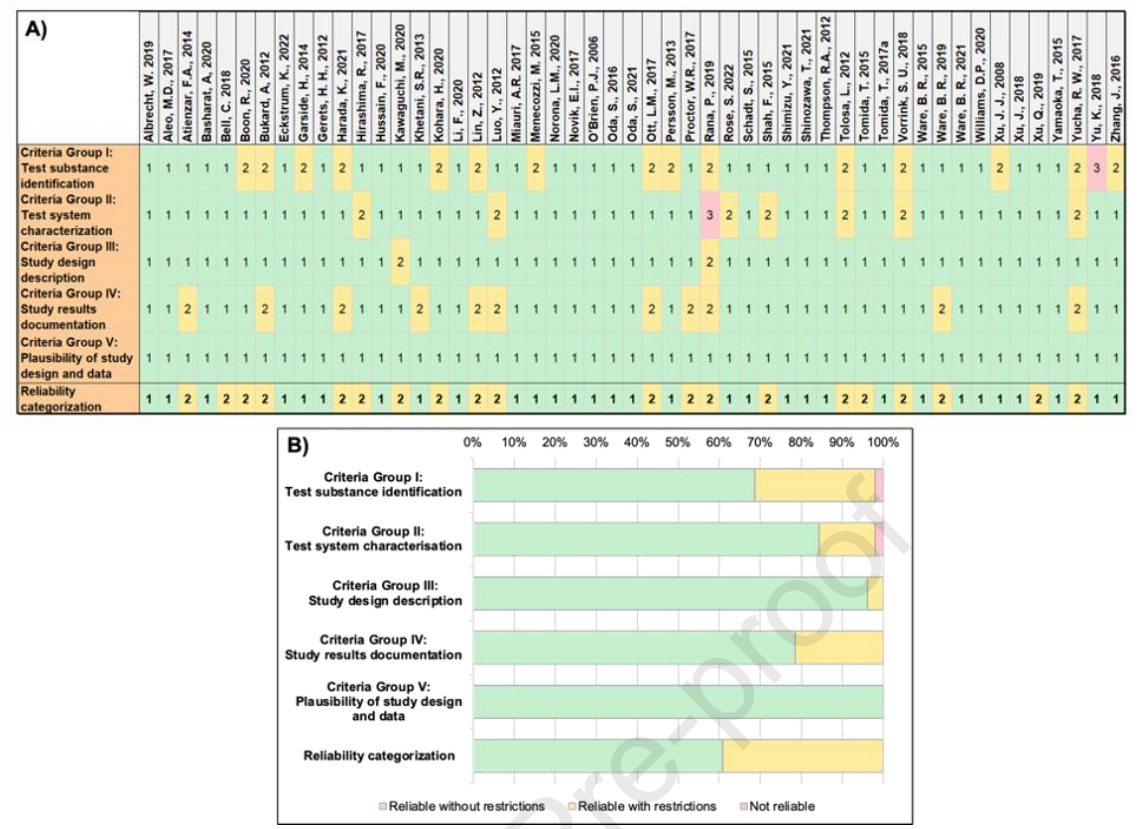


Figure 3:

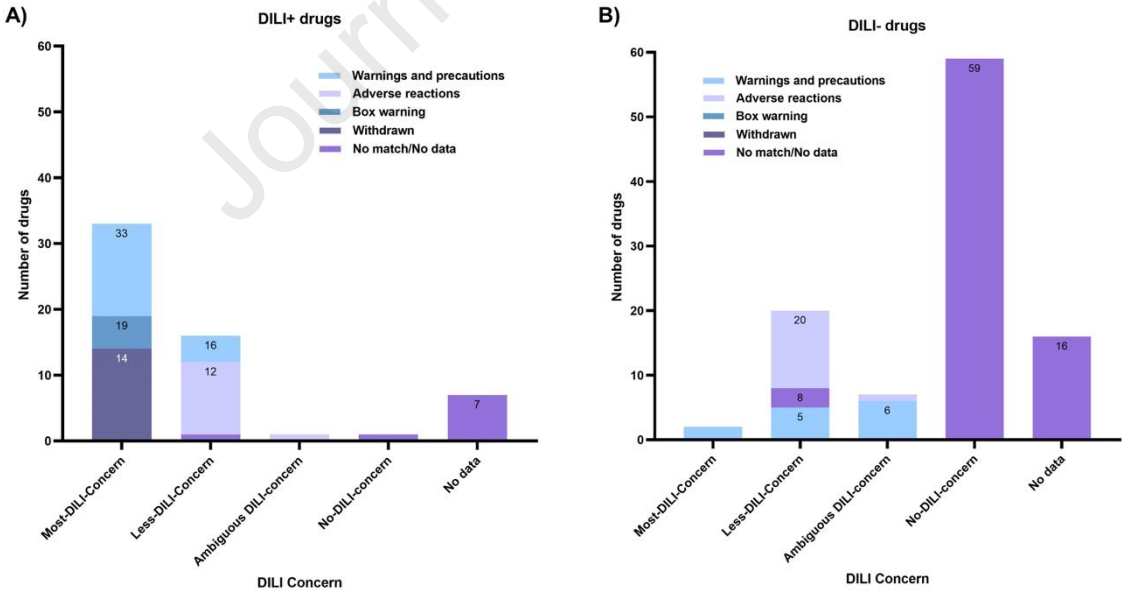


Figure 4:

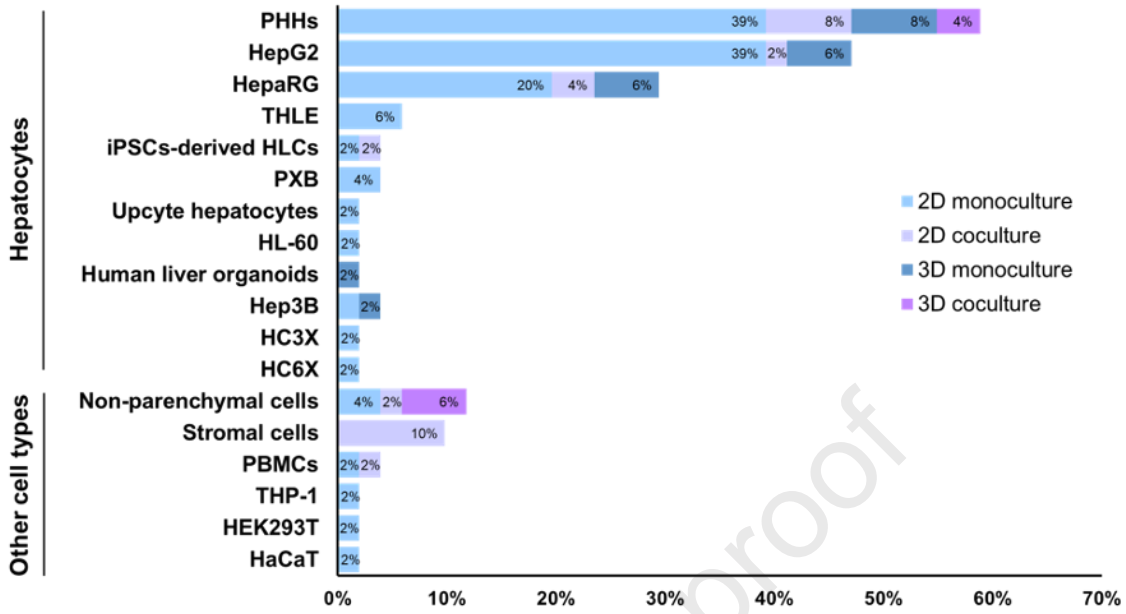


Figure 5:

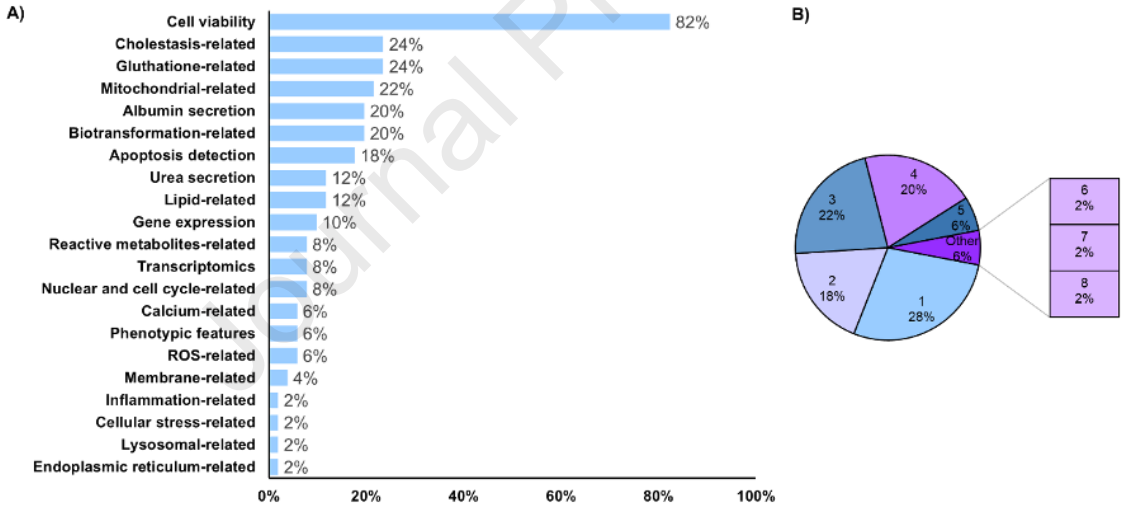
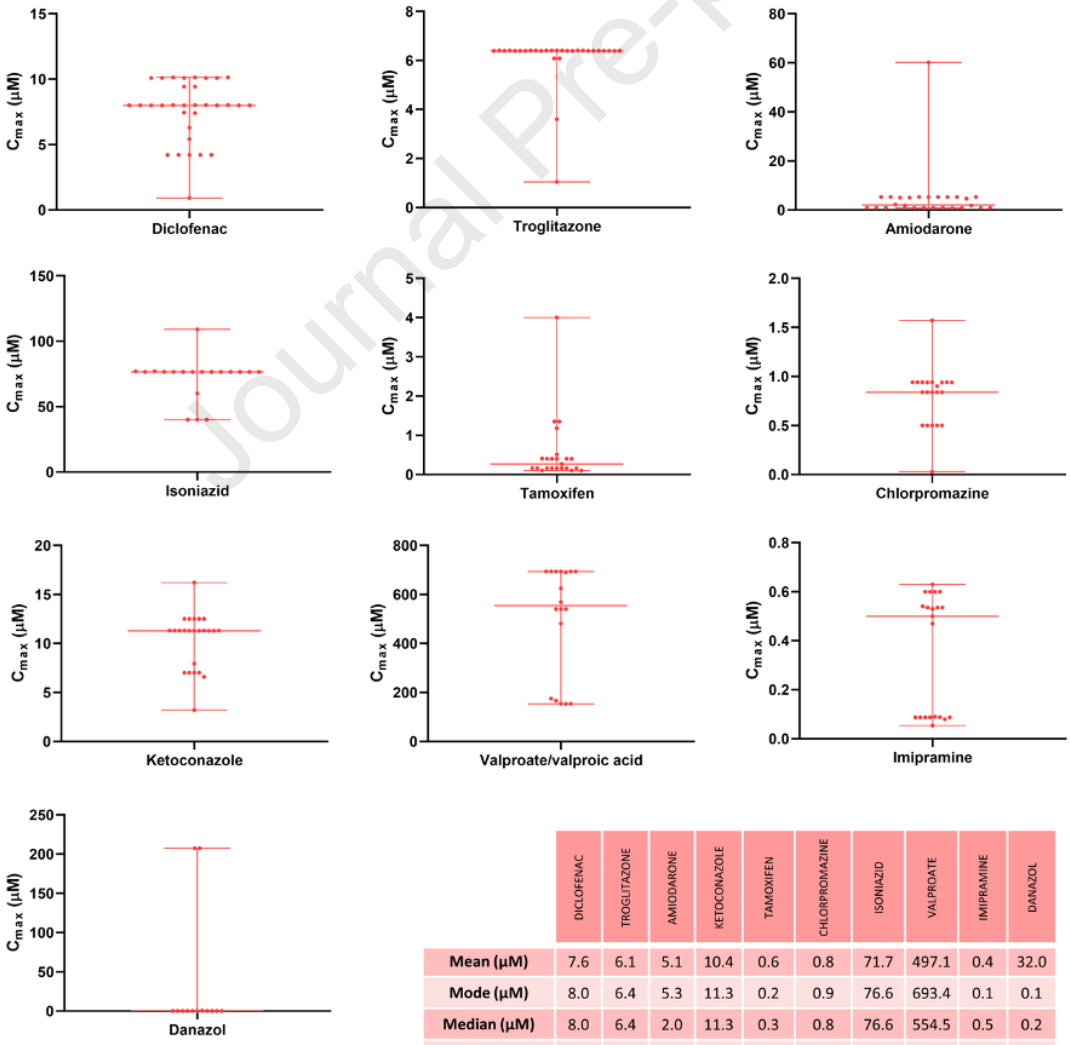


Figure 6:

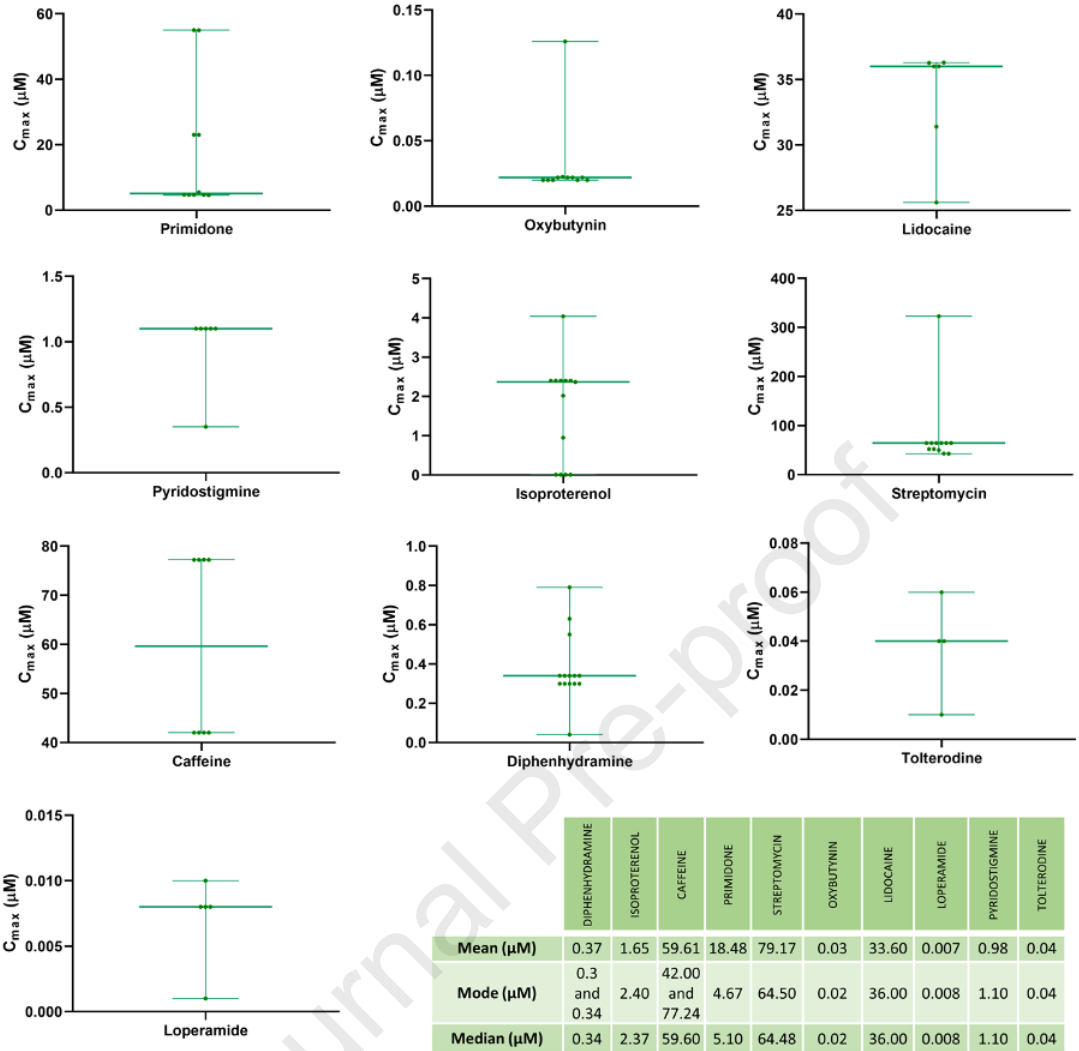
DILI+				
DICLOFENAC	TROGLITAZONE	AMIODARONE	KETOCONAZOLE	TAMOXIFEN
CHLORPROMAZINE	ISONIAZID	VALPROIC ACID	IMIPRAMINE	DANAZOL

DILI-				
DIPHENHYDRAMINE	ISOPROTERENOL	CAFFEINE	PRIMIDONE	STREPTOMYCIN
OXYBUTYNIN	LIDOCAINE	LOPERAMIDE	PYRIDOSTIGMINE	TOLTERODINE

Figure 7:



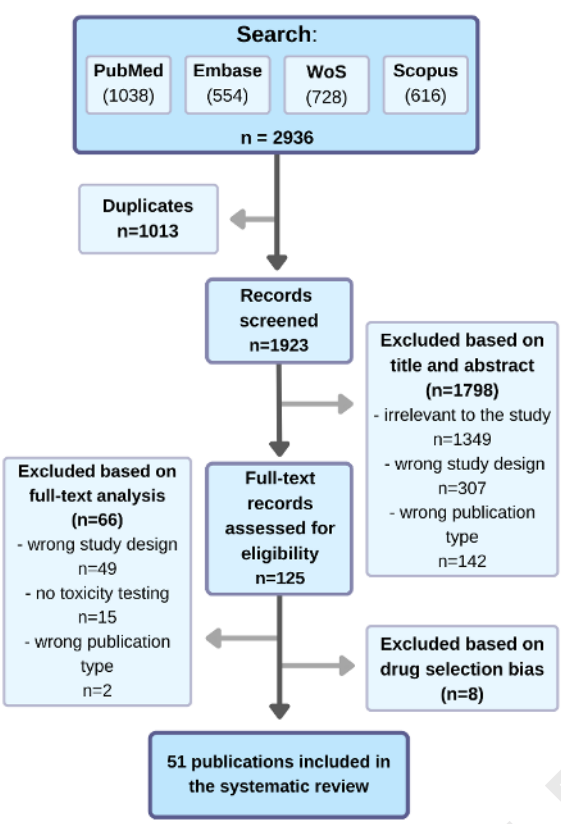
	DICLOFENAC	TROGLITAZONE	AMIODARONE	KETOCONAZOLE	TAMOXIFEN	CHLORPROMAZINE	ISONIAZID	VALPROATE	IMIPRAMINE	DANAZOL
Mean (µM)	7.6	6.1	5.1	10.4	0.6	0.8	71.7	497.1	0.4	32.0
Mode (µM)	8.0	6.4	5.3	11.3	0.2	0.9	76.6	693.4	0.1	0.1
Median (µM)	8.0	6.4	2.0	11.3	0.3	0.8	76.6	554.5	0.5	0.2
Minimum (µM)	0.9	1.0	0.8	3.2	0.1	0.0	40.0	153.9	0.1	0.1
Maximum (µM)	10.1	6.4	60.1	16.2	4.0	1.6	109.0	693.4	0.6	207.4



	DIPHENHYDRAMINE	ISOPROTERENOL	CAFFEINE	PRIMIDONE	STREPTOMYCIN	OXYBUTYNIN	LIDOCAINE	LOPERAMIDE	PYRIDOSTIGMINE	TOLTERODINE
Mean (µM)	0.37	1.65	59.61	18.48	79.17	0.03	33.60	0.007	0.98	0.04
Mode (µM)	0.3 and 0.34	2.40	42.00 and 77.24	4.67	64.50	0.02	36.00	0.008	1.10	0.04
Median (µM)	0.34	2.37	59.60	5.10	64.48	0.02	36.00	0.008	1.10	0.04
Minimum (µM)	0.04	0.01	42.00	4.58	42.99	0.02	25.60	0.001	0.35	0.01
Maximum (µM)	0.79	4.04	77.24	55.02	322.97	0.13	36.30	0.010	1.10	0.06

Table 1. Suggested C_{max} concentrations of DILI+ and DILI- control compounds.

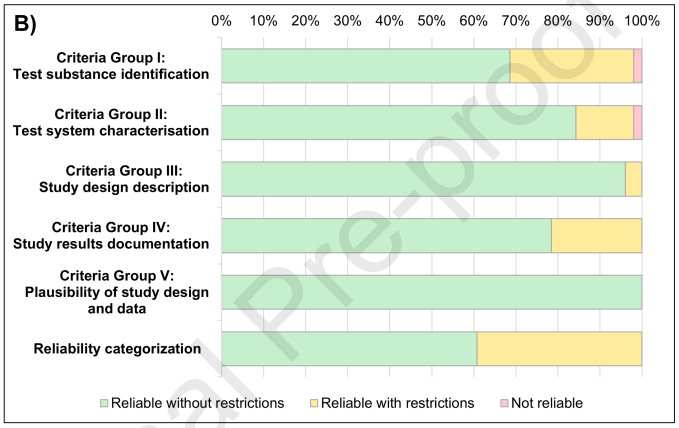
Control Compound	Suggested Cmax (μM)
Diclofenac	8
Troglitazone	6.4
Amiodarone	5.3
Ketoconazole	11.3
Tamoxifen	0.2
Chlorpromazine	0.9
Isoniazid	76.6
Valproate	693.4
Imipramine	0.1
Danazol	0.1
Diphenhydramine	0.34
Isoproterenol	2.4
Caffeine	77.24
Primidone	4.77
Streptomycin	74.5
Oxybutynin	0.02
Lidocaine	36
Pyridostigmine	1.1
Tolterodine	0.04

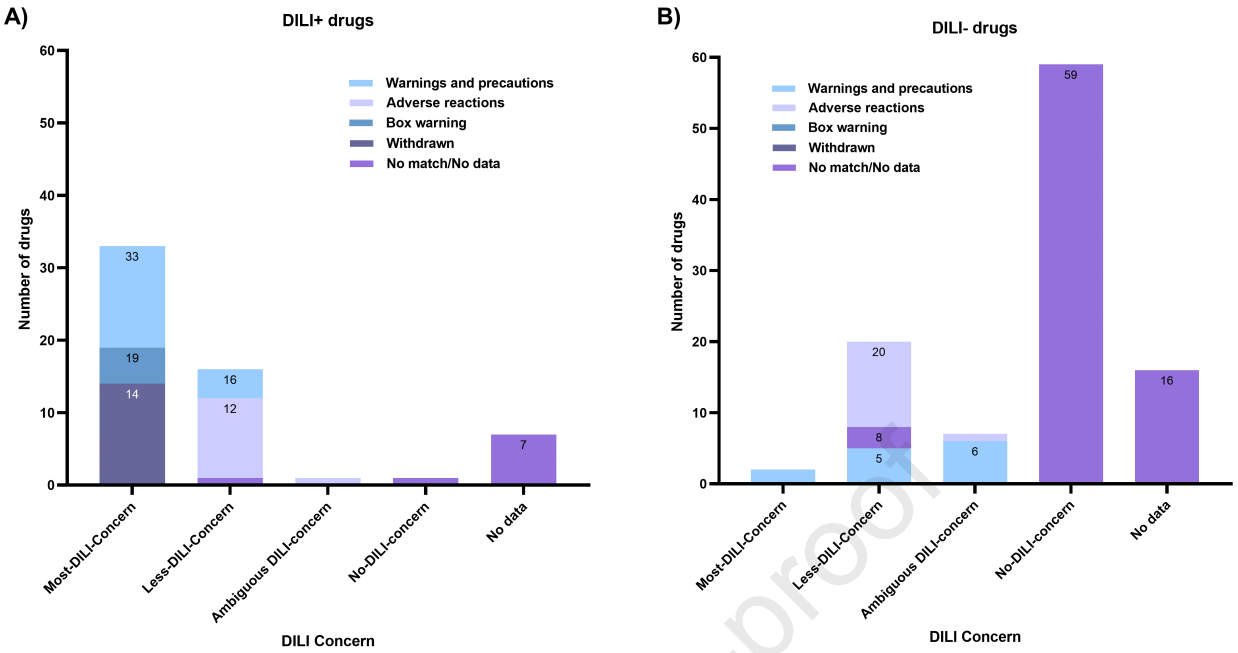


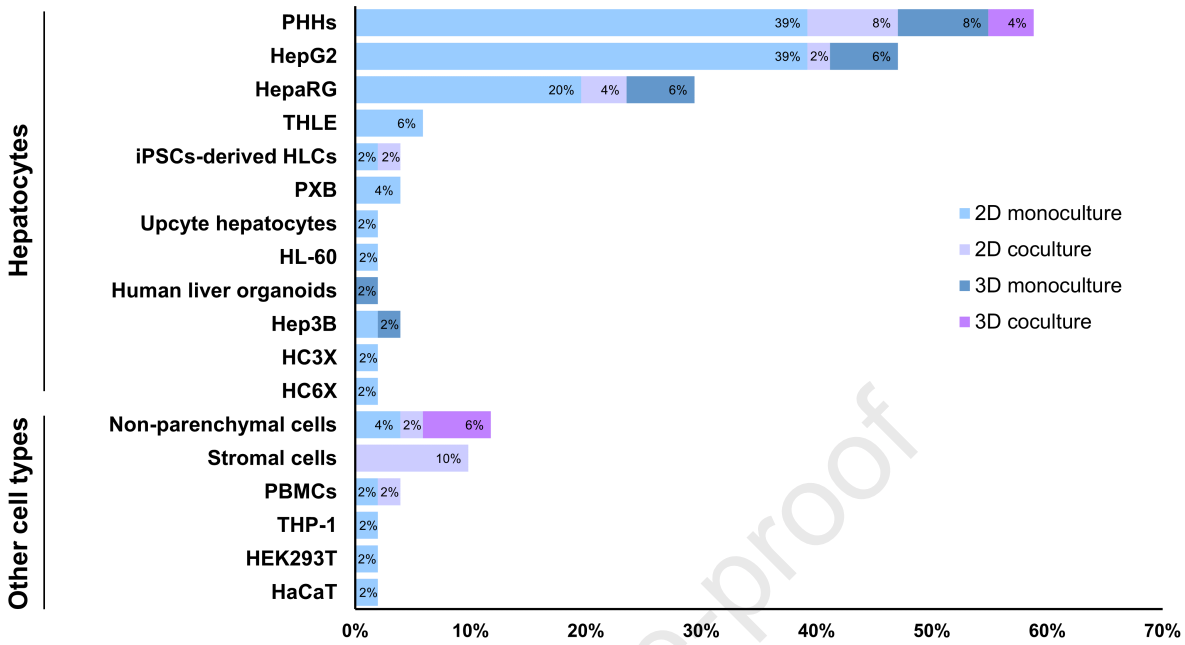
A)

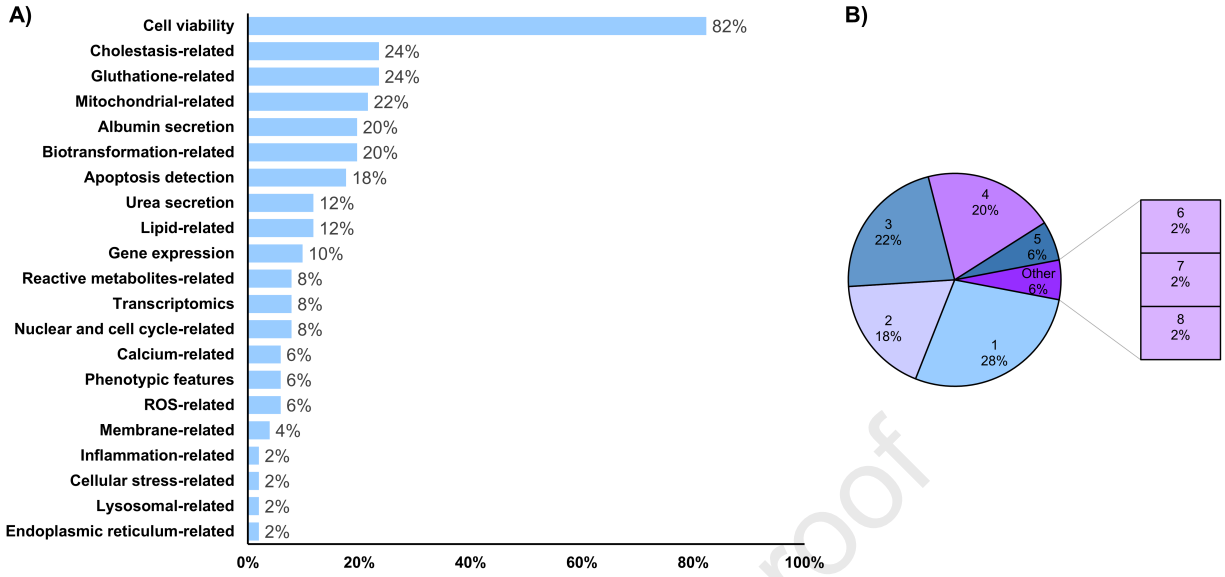
	Albrecht, W. 2019	Aleo, M.D., 2017	Atienzar, F.A., 2014	Basharat, A, 2020	Bell, C. 2018	Boon, R., 2020	Bukard, A, 2012	Eckstrum, K., 2022	Garside, H., 2014	Gerets, H. H., 2012	Harada, K., 2021	Hirashima, R., 2017	Hussain, F., 2020	Kawaguchi, M., 2020	Khetani, S.R., 2013	Kohara, H., 2020	Li, F., 2020	Lin, Z., 2012	Luo, Y., 2012	Mlauri, A.R. 2017	Menecozzi, M. 2015	Norona, L.M., 2020	Novik, E.I., 2017	O'Brien, P.J., 2006	Oda, S., 2016	Oda, S., 2021	Ott, L.M., 2017	Persson, M., 2013	Proctor, W.R., 2017	Rana, P., 2019	Rose, S. 2022	Schadt, S., 2015	Shah, F., 2015	Shimizu, Y., 2021	Shinozawa, T., 2021	Thompson, R.A., 2012	Tolosa, L., 2012	Tomida, T. 2015	Tomida, T., 2017a	Vorrink, S. U., 2018	Ware, B. R., 2015	Ware, B. R., 2019	Ware, B. R., 2021	Williams, D.P., 2020	Xu, J. J., 2008	Xu, J., 2018	Xu, Q., 2019	Yamaoka, T., 2015	Yucha, R. W., 2017	Yu, K., 2018	Zhang, J., 2016
Criteria Group I: Test substance identification	1	1	1	1	1	2	2	1	2	1	2	1	1	1	1	2	1	2	1	1	2	1	1	1	1	1	2	2	1	2	1	1	1	1	1	2	1	1	1	1	2	1	1	1	2	3	2
Criteria Group II: Test system characterization	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	3	2	1	2	1	1	1	2	1	1	2	1	1	1	1	1	1	2	1	1
Criteria Group III: Study design description	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Criteria Group IV: Study results documentation	1	1	2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	2	2	1	1	1	1	1	1	1	2	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1
Criteria Group V: Plausibility of study design and data	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Reliability categorization	**1**	**1**	**2**	**1**	**2**	**2**	**2**	**1**	**1**	**1**	**2**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**2**	**2**	**1**	**1**	**2**	**1**	**1**	**2**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**1**	**1**	**2**	**1**	**1**

B)



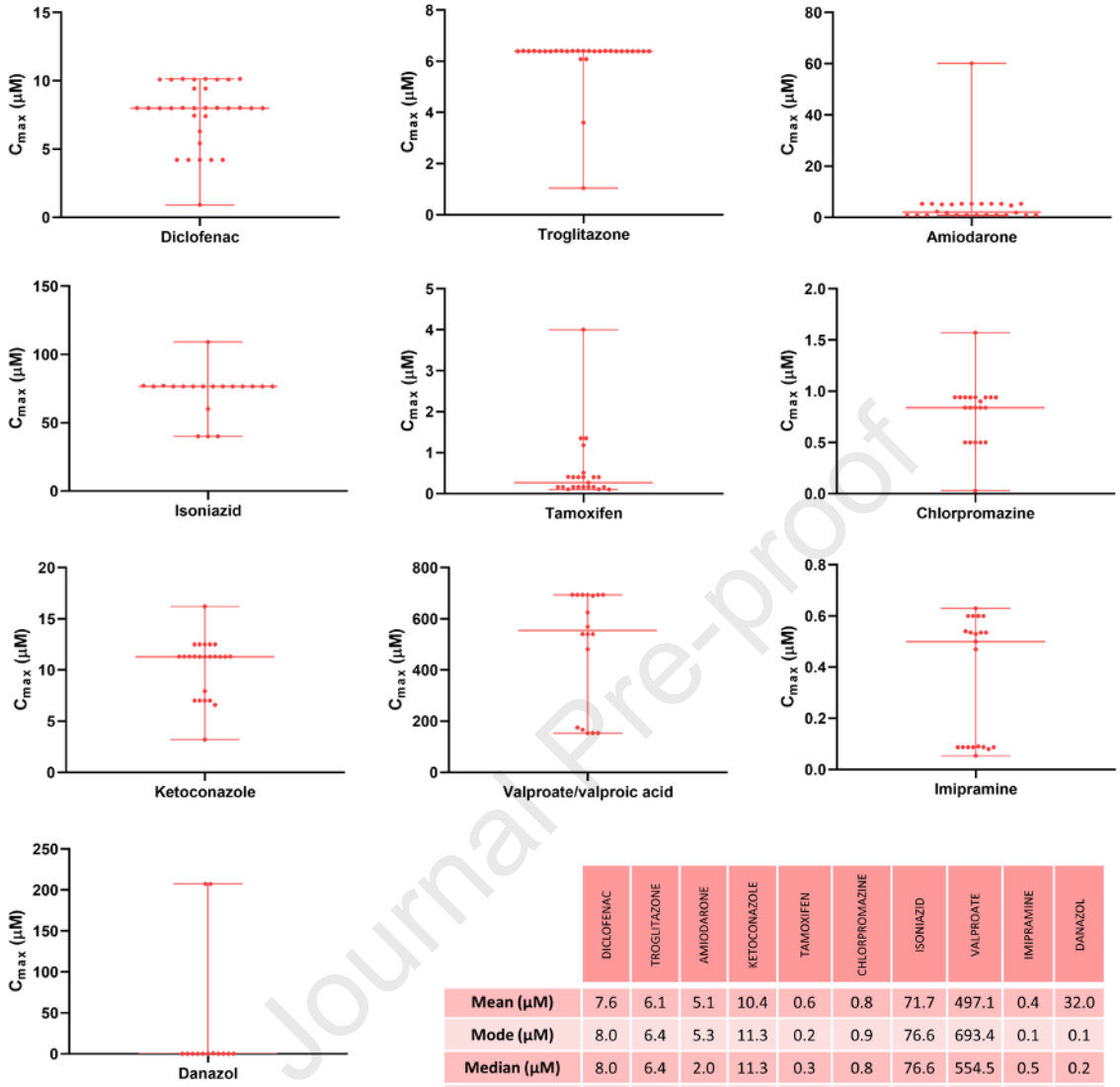




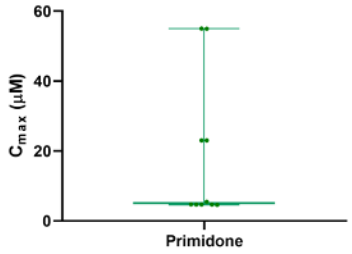


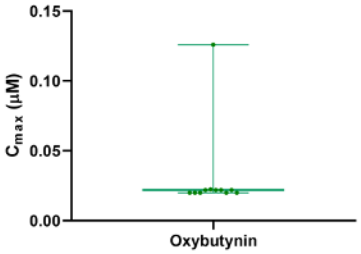
DILI+				
DICLOFENAC	TROGLITAZONE	AMIODARONE	KETOCONAZOLE	TAMOXIFEN
CHLORPROMAZINE	ISONIAZID	VALPROIC ACID	IMIPRAMINE	DANAZOL

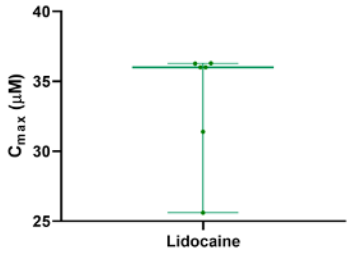
DILI-				
DIPHENHYDRAMINE	ISOPROTERENOL	CAFFEINE	PRIMIDONE	STREPTOMYCIN
OXYBUTYNIN	LIDOCAINE	LOPERAMIDE	PYRIDOSTIGMINE	TOLTERODINE

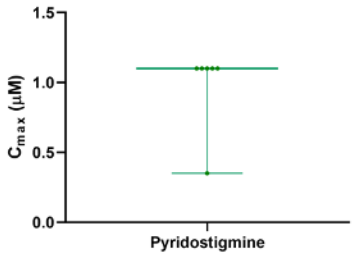


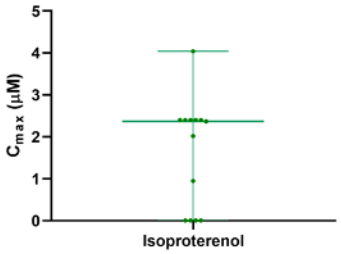
	DICLOFENAC	TROGLITAZONE	AMIODARONE	KETOCONAZOLE	TAMOXIFEN	CHLORPROMAZINE	ISONIAZID	VALPROATE	IMIPRAMIINE	DANAZOL
Mean (µM)	7.6	6.1	5.1	10.4	0.6	0.8	71.7	497.1	0.4	32.0
Mode (µM)	8.0	6.4	5.3	11.3	0.2	0.9	76.6	693.4	0.1	0.1
Median (µM)	8.0	6.4	2.0	11.3	0.3	0.8	76.6	554.5	0.5	0.2
Minimum (µM)	0.9	1.0	0.8	3.2	0.1	0.0	40.0	153.9	0.1	0.1
Maximum (µM)	10.1	6.4	60.1	16.2	4.0	1.6	109.0	693.4	0.6	207.4

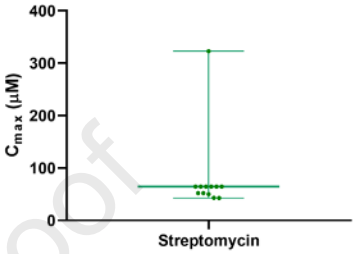


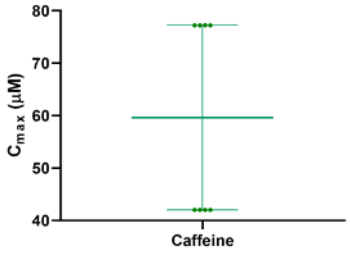


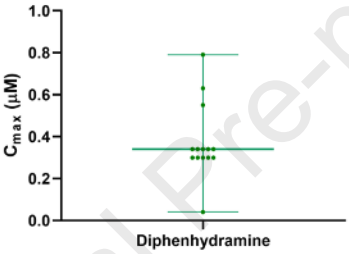


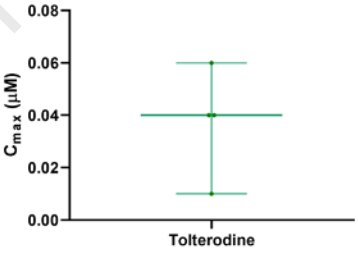


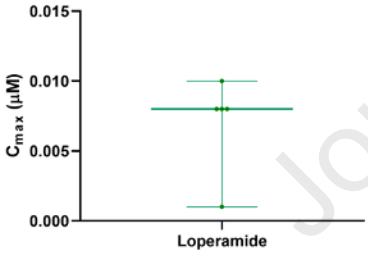












	DIPHENHYDRAMINE	ISOPROTERENOL	CAFFEINE	PRIMIDONE	STREPTOMYCIN	OXYBUTYNIN	LIDOCAINE	LOPERAMIDE	PYRIDOSTIGMINE	TOLTERODINE
Mean (µM)	0.37	1.65	59.61	18.48	79.17	0.03	33.60	0.007	0.98	0.04
Mode (µM)	0.3 and 0.34	2.40	42.00 and 77.24	4.67	64.50	0.02	36.00	0.008	1.10	0.04
Median (µM)	0.34	2.37	59.60	5.10	64.48	0.02	36.00	0.008	1.10	0.04
Minimum (µM)	0.04	0.01	42.00	4.58	42.99	0.02	25.60	0.001	0.35	0.01
Maximum (µM)	0.79	4.04	77.24	55.02	322.97	0.13	36.30	0.010	1.10	0.06

Highlights:

- Identification of DILI during preclinical stages remains challenging, underscoring the need for appropriate test drugs
- Through a systematic review, the article analyzes compounds used in *in vitro* hepatotoxicity assays
- A list of 20 control drugs, supported by literature, clinical data and an experts committee was created
- The consensus-driven list aims to enhance the validation and standardization of *in vitro* models