Fakultät für Mathematik und Naturwissenschaften | RWTH AACHEN UNIVERSITY

High-throughput PBK Modelling for the Prediction of Systemic Exposure and Effects of Chemicals on Humans

Von der Fakultät für Mathematik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

René Geci, Master of Science

aus

Velbert, Deutschland

Berichter: Univ.-Prof. Dr.-Ing. Lars Mathias Blank

Univ.-Prof. Dr. rer. nat. Lars Küpfer

Tag der mündlichen Prüfung: 12.03.2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek verfügbar.

Eidesstattliche Erklärung

Hiermit erkläre ich, René Geci, dass diese Dissertation und die darin dargelegten Inhalte die eigenen sind und selbstständig, als Ergebnis der eigenen originären Forschung, generiert wurden.

Hiermit erkläre ich an Eides statt:

1. Diese Arbeit wurde vollständig oder größtenteils in der Phase als Doktorand dieser Fakultät und Universität angefertigt;

2. Sofern irgendein Bestandteil dieser Dissertation zuvor für einen akademischen Abschluss oder eine andere Qualifikation an dieser oder einer anderen Institution verwendet wurde, wurde dies klar angezeigt;

3. Wenn immer andere eigene- oder Veröffentlichungen Dritter herangezogen wurden, wurden diese klar benannt;

4. Wenn aus anderen eigenen- oder Veröffentlichungen Dritter zitiert wurde, wurde stets die Quelle hierfür angegeben. Diese Dissertation ist vollständig meine eigene Arbeit, mit der Ausnahme solcher Zitate;

5. Alle wesentlichen Quellen von Unterstützung wurden benannt;

6. Wenn immer ein Teil dieser Dissertation auf der Zusammenarbeit mit anderen basiert, wurde von mir klar gekennzeichnet, was von anderen und was von mir selbst erarbeitet wurde;

7. Ein Teil oder Teile dieser Arbeit wurden zuvor veröffentlicht, und zwar in:

Geci, R. et al. *Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans.* Arch Toxicol 98, 2659–2676; 10.1007/s00204-024-03764-9 (2024).

Geci, R., Paini, A., Worth, A. P., Kuepfer, L. & Schaller, S. *Potential systemic availability classification of chemicals for safety assessment.* Environment international 202, 109636; 10.1016/j.envint.2025.109636 (2025).

Geci, R., Sayin, A. Z., Schaller, S. & Kuepfer, L. *Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury.* Arch Toxicol 100, 2029–2046; 10.1007/s00204-026-04305-2 (2026).

Edizcan, Z., Schaller, S., Kuepfer, L. & Geci, R. *High-throughput PBK modelling for dermal exposure: a pragmatic approach to predict systemic pharmacokinetics.* Arch Toxicol 100, 2421–2432; 10.1007/s00204-026-04357-4 (2026).

Aachen, der 04.12.2025

René Geci

Declaration on the Use of AI-based Tools

During the preparation of this thesis, I used the generative AI tool “ChatGPT” (GPT-5 and GPT-5.1 by OpenAI) exclusively to check and improve English phrasing, grammar, and spelling of all text sections of this thesis after I had written the original text myself. ChatGPT was not used to generate scientific content, ideas, interpretations, data analyses, figures, or code. Outputs from ChatGPT were not treated as scientific sources and are not part of the cited literature underlying this thesis.

In addition, I used the integrated spelling and grammar checker in Microsoft Word to detect and correct errors in English phrasing, grammar, and spelling.

All scientific content, arguments, and conclusions are my own, and I have carefully checked all AI-supported reformulations for correctness and consistency.

Funding

This work was performed in the context of the ONTOX project (https://ontoxproject.eu/) that has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 963845. ONTOX is part of the ASPIS project cluster (https://aspiscluster.eu/).

Acknowledgements

There are many who have walked with me on this journey and without whom this thesis would not have been possible. I am deeply grateful for every single one of them and would like to express my deep gratitude and appreciation to them.

First, I would like to thank my supervisor Prof. Lars Küpfer for his supervision, support and guidance. I especially appreciate the trust you have placed in me, giving me the freedom to follow my own passions and to explore and develop projects in directions that I found most exciting. At the same time, I always knew that your door was open whenever I needed advice, which gave me the room I needed to grow as an independent researcher.

I would like to thank Dr. Stephan Schaller without whom this project quite literally would not have existed. Your positive energy and your optimistic spirit to start new things and venture out into the unknown are truly inspiring. I am deeply grateful for the opportunity to work together with you in such a dynamic and forward-thinking environment, as well as for your trust and the freedom that you granted me.

I would also like to thank my research group at the University Hospital RWTH Aachen and everyone at ESQlabs for fostering a friendly, open, and supportive atmosphere. I appreciated the positive spirit and the welcoming environment, which made working together enjoyable and easy.

I would like to thank my family who supported me not only throughout the PhD, but long before it began. I still remember when we explored the city of Aachen on a hot summer day to see what my next chapter would hold. A lot has happened since that day, and the journey has surely not been an easy one. But change is a fundamental part of life. And as one chapter closes, a new one will begin. I am grateful that you were there with me for the one that is closing and just as thankful to know that I will have you by my side for the next one.

I would like to thank Leticia Rodríguez-Montes for her patience, companionship, relentless support, encouragement, and belief in me. Nobody was closer to all the ups and downs than you and there is no doubt in my mind that this thesis would not exist without you. I am deeply grateful for everything that you have given.

I would like to thank everyone involved in the ONTOX project for making it such an enjoyable experience. Conferences and meetings with you were always a highlight, full of friendly faces and fun conversations. Science can sometimes feel lonely, and it was good to know that I was part of a bigger community of such kind and welcoming people.

Lastly, I would like to thank Velocity Aachen and Mario from the Wok station at the university mensa. Everyday heroes like you made the routine not only manageable but genuinely enjoyable. Without e-bikes to get me to the office and a steady supply of stir-fry to keep me going this journey would at least have looked very different (and involved more bus riding and worse lunches).

“Mach da bloß keine Doktorarbeit draus.“

Kalman Geci

1962 - 2024

Table of Contents

Eidesstattliche Erklärung ..... **ii**

Declaration on the Use of AI-based Tools ..... **iii**

Funding ..... **iii**

Acknowledgements ..... **iv**

Table of Contents ..... **vi**

Abbreviations ..... **viii**

List of Figures ..... **x**

List of Tables ..... **xii**

Zusammenfassung ..... **xiii**

Abstract ..... **xiv**

General Introduction **1**

Part I: Background **5**

1. Pharmacokinetics 6
2. PBK Modelling 11
3. High-throughput PBK Modelling 14
4. New Approach Methodologies 17
5. Drug-Induced Liver Injury 19

Part II: Materials and Methods **22**

6. Methods for Intravenous and Oral PBK Modelling in Chapter 10 23
7. Methods for Dermal PBK Modelling in Chapter 11 27
8. Methods for ADME-Based Assessment of Chemical Safety in Chapter 12 29
9. Methods for Drug-Induced Liver Injury Prediction in Chapter 13 31

Part III: Results **33**

10. High-throughput PBK Prediction of Intravenous and Oral Drug Exposure 34
    - 10.1 Introduction 35
    - 10.2 Results 37
    - 10.3 Discussion 49

11. High-throughput PBK Prediction of Systemic PK after Dermal Exposure.....52
11.1 Introduction.....53
11.2 Results.....55
11.3 Discussion.....63
12. ADME-based Assessment of Chemical Safety.....65
12.1 Introduction.....66
12.2 Results.....70
12.3 Discussion.....79
13. High-throughput Prediction of Drug-Induced Liver Injury.....83
13.1 Introduction.....84
13.2 Results.....86
13.3 Discussion.....99

General Conclusions and Outlook.....102

References.....106

Appendix.....127

Curriculum Vitae.....xv

Abbreviations

ADME: Absorption, distribution, metabolism, and excretion

ALT: Alanine aminotransferase

AP: Alkaline phosphatase

AST: Aspartate aminotransferase

AI: Artificial intelligence

AUC: Area under the curve

BA: Bile acid

BER: Bioactivity exposure ratio

BSEP: Bile salt export pump

CL: Clearance

CLint: Intrinsic hepatic clearance

CLP: Classification, labelling and packaging

Cmax: Maximum concentration

DDI: Drug-drug interaction

DILI: Drug-induced liver injury

EEA: European Environment Agency

EPA: Environmental Protection Agency

EPAA: European Partnership for Alternative Approaches to Animal Testing

EU: European Union

F: Bioavailability

Fu: Fraction unbound

FIH: First-in-human

HT-PBK: High-throughput PBK

HTTK: High-throughput toxicokinetics

IC50: Half-maximal inhibitory concentration

IV: Intravenous

IVIVE: *In vitro* to *in vivo* extrapolation

LogD: Logarithm of distribution coefficient

LogMA: Logarithm of membrane affinity partition coefficient

LogP: Logarithm of octanol-water partition coefficient

MIDD: Model-informed drug development

MIDDD: Model-informed drug development and discovery

MIRA: Model-informed risk assessment

ML: Machine learning

MSAF: Mean systemic availability factor

NAM: New approach methodology

NGRA: Next generation risk assessment

NIH: National Institutes of Health

ODE: Ordinary differential equation

OSP: Open Systems Pharmacology

PBBA: Physiologically based bile acid

PBK: Physiologically based kinetic

PBPK: Physiologically based pharmacokinetic

PBT: Persistent, bioaccumulative, toxic

PBTK: Physiologically based toxicokinetic

PD: Pharmacodynamics

PFAS: Per- and polyfluoroalkyl substances

PK: Pharmacokinetics

PO: Peroral

POP: Persistent organic pollutant

PSA: Potential systemic availability

PSAF: Peak systemic availability factor

QSAR: Quantitative structure-activity relationship

QST: Quantitative systems toxicology

REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals

ROC: Receiver operating characteristic

SVHC: Substance of very high concern

TD: Toxicodynamic

Thalf: Elimination half-life

TK: Toxicokinetics

Tmax: Time to maximum concentration

Vd: Volume of distribution

vPvB: Very persistent, very bioaccumulative

List of Figures

Figure 1: Factors determining pharmacokinetics. 7

Figure 2: Key pharmacokinetic parameters derived from a concentration-time profile. 9

Figure 3: Schematic representation of a physiologically based kinetic (PBK) model structure and the PBK modelling workflow. 12

Figure 4: Schematic representation of differences between traditional and high-throughput PBK modelling. 14

Figure 5: Summary statistics of the collected IV and oral PK dataset and of various compound properties. 37

Figure 6: Conceptual overview of the simulation and analysis strategy. 38

Figure 7: Comparison of predictive performances of different physicochemical parameter sources (step 1). 40

Figure 8: Comparison of predictive performances of different fraction unbound and clearance prediction sources (step 2). 43

Figure 9: Comparison of predictive performances of different solubility and intestinal permeability prediction sources (step 3). 45

Figure 10: Predictive performance of the best fully *in silico*-based HT-PBK modelling strategy (proprietary). 48

Figure 11: Overview of dermally administered PK data and of compound properties.....56

Figure 12: Systemic PK prediction accuracy after dermal exposure across *in silico* input sources. 57

Figure 13: Predicted concentration-time profiles against observed systemic PK data. ...59

Figure 14: Correlation prediction errors and various compound properties. 60

Figure 15: Dataset-wide performance of the best dermal HT-PBK strategy and comparison with oral HT-PBK errors. 62

Figure 16: Illustration of the imbalance between the number of chemicals on the EU market and the extent of their hazard and exposure characterisation.....67

Figure 17: Data overview and general background. 69

Figure 18: Visual representation for the rationale behind distinguishing different Potential Systemic Availability concern classes. 72

Figure 19: Peak Systemic Availability Factors of the 139 EPAA Designathon Compounds for which HT-PBK predictions could be made. 75

Figure 20: Mean Systemic Availability Factors of the 139 EPAA Designathon Compounds for which HT-PBK predictions could be made. 76

Figure 21: Data collection, analysis workflow and drug properties.....87

Figure 22: Predictivity of simple DILI screening heuristics (evaluation 1). 88

Figure 23: DILI predictivity of integrated functional *in vitro* toxicity and *in vivo* Cmax data (evaluation 2). 90

Figure 24: Prospective DILI predictivity of functional *in vitro* toxicity and *in silico* Cmax predictions (evaluation 3). 93

Figure 25: DILI predictivity of *in vitro* BSEP inhibition data (evaluation 4).....95

Figure 26: Simulated effects of *in vivo* BSEP inhibition on intrahepatic bile acid levels (evaluation 5). 97

List of Tables

Table 1: Overview of different HT-PBK modelling strategies and their predictive performances.....47

Table 2: Properties of five example compounds of different Potential Systemic Availability concerns and different oral bioavailability. 71

Table 3: Comparison of classifications of 139 EPAA Designathon Compounds based on HT-PBK predictions using the Peak Systemic Availability Factor (PSAF) against expert judged classification of Potential Systemic Availability (PSA).....78

Table 4: Compounds causing meaningful intrahepatic bile acid elevations in dynamic model simulations, having Cmax to BSEP IC50 ratios larger than 0.5 and Cmax to functional *in vitro* toxicity ratios lower than 0.5, or no other *in vitro* toxicity value. 98

Zusammenfassung

Die moderne Toxikologie und Pharmakologie befinden sich in einem tiefgreifenden Wandel. Ethische Erwägungen und Fortschritte in rechnergestützten Methoden ermöglichen zunehmend die Ersetzung von Tierversuchen durch humanrelevantere, tierversuchsfreie Bewertungsmethoden. Diese Dissertation trägt zu diesem Übergang bei, indem sie einen Ansatz entwickelt und validiert, der als Hochdurchsatz-physiologisch basierte kinetische (HT-PBK) Modellierung bezeichnet wird. Diese neue Methode integriert mathematische PBK-Modelle mit modernen Werkzeugen des maschinellen Lernens (ML), welche zentrale physikochemische und Absorptions-, Distributions-, Metabolismus- und Exkretionseigenschaften (ADME) von Molekülen direkt aus ihrer chemischen Struktur vorhersagen. Dadurch ermöglicht sie eine schnelle, automatisierte Modellparametrisierung, während die mechanistische Interpretierbarkeit erhalten bleibt und Engpässe in der traditionellen PBK-Modellentwicklung beseitigt werden. Unter ausschließlicher Verwendung von *in-vitro*- und computergestützten Eingangsdaten ermöglicht die HT-PBK-Modellierung die Abschätzung der Exposition des Menschen gegenüber Chemikalien und somit die Wirkung von Chemikalien auf den Menschen. Das übergeordnete Ziel dieser Arbeit ist es, die prädiktive Leistungsfähigkeit HT-PBK-basierter pharmakokinetischer (PK) Simulationen gegenüber beim Menschen beobachteten *in-vivo*-PK-Daten zu quantifizieren und den praktischen Nutzen dieser Simulationen für Sicherheitsbewertungen aufzuzeigen. Um eine umfassende Evaluation zu ermöglichen, wurde einer der größten verfügbaren Datensätze menschlicher Plasmakonzentrations-Zeit-Profile zusammengestellt und kuratiert, der über 2000 menschliche PK-Kurven von mehr als 200 Verbindungen nach intravenöser und oraler Applikation umfasst. In diesem Datensatz reproduzierten HT-PBK-Simulationen die PK der meisten Substanzen ausreichend zuverlässig, wobei 87% der Cmax- und 84% der AUC-Werte innerhalb eines Faktors von zehn der beobachteten Daten lagen. Nach der Evaluation der intravenösen und oralen Exposition weitet die Arbeit die Analyse auf die dermale Exposition aus, einer wichtigen Expositionsroute für viele Verbraucher- und Kosmetikprodukte, für die Tierversuche regulatorisch zunehmend eingeschränkt werden. Obwohl die Vorhersage der systemischen PK nach dermaler Exposition anspruchsvoller ist, blieben 75% der Cmax- und AUC-Werte innerhalb eines Zehnfachen der beobachteten Humandaten. Anschließend wird die Evaluation auf praktische Risikobewertungsaufgaben ausgedehnt. Zuerst wird im Rahmen des European Partnership for Alternative Approaches to Animal Testing (EPAA) NAM Designathon gezeigt, wie HT-PBK-Vorhersagen allein die Klassifizierung von Chemikalien anhand ihres Potenzials zur systemischen Verfügbarkeit unterstützen können, wenn Informationen zur toxischen Potenz fehlen. Abschließend werden HT-PBK-Vorhersagen mit *in-vitro*-Toxizitätsassay-Daten kombiniert, um das Risiko zu bewerten, dass Arzneistoffe eine arzneimittelinduzierte Leberschädigung (drug-induced liver injury, DILI) verursachen. Die Integration von *in-vitro*-Assays und HT-PBK-Simulationen zeigt dabei mit einer ROC-AUC von 91% eine hohe Übereinstimmung mit klinischen DILI-Befunden. Diese Evaluation der Kombination von mechanistischen Simulationen und ML-basierter Parametervorhersage stärkt die Grundlage für toxikologische Bewertungen der nächsten Generation. Die resultierenden Werkzeuge sind transparent, unabhängig von Tierdaten und für die effiziente Bewertung großer Substanzmengen ausgelegt. Sie unterstützen somit die wachsenden wissenschaftlichen und regulatorischen Bestrebungen zur Integration ML-gestützter Modellierung in der Toxikologie und Pharmakologie.

Abstract

Modern toxicology and pharmacology are undergoing a fundamental transformation. Ethical considerations and advances in computational methods are increasingly enabling the replacement of animal testing with more human-relevant, non-animal assessment approaches. This thesis contributes to this transition by developing and validating an approach termed high-throughput physiologically based kinetic (HT-PBK) modelling. This new method integrates mathematical PBK models with modern machine learning (ML) tools that predict key physicochemical and absorption, distribution, metabolism and excretion (ADME) properties of molecules directly from their chemical structure. Thereby, it enables rapid, automated model parameterisation while preserving mechanistic interpretability and removing bottlenecks in traditional PBK model development. Relying solely on *in vitro* and computational input data, HT-PBK modelling enables the estimation of human exposure to chemicals, and consequently the effects of chemicals on humans. The overarching aim of this thesis is to quantify the predictive performance of HT-PBK-based pharmacokinetic (PK) simulations against *in vivo* PK data observed in humans, and to showcase the utility of such simulations for real-world application in safety assessment. To enable a comprehensive evaluation, the work assembled and curated one of the largest available datasets of human plasma concentration-time profiles, covering over 2000 human PK curves of more than 200 compounds following intravenous and oral administration. Across this dataset, HT-PBK simulations reproduced the PK of most compounds sufficiently reliably with 87% of Cmax and 84% of AUC values within tenfold of the observed data. After evaluating intravenous and oral exposure, the thesis extends the assessment to dermal exposure, which is a major route for many consumer and cosmetic products, and for which animal testing is increasingly restricted by regulation. Although the prediction of systemic PK after dermal exposure is more challenging, 75% of Cmax and AUC values remained within a tenfold range of observed human data. The evaluation is then extended to real-world risk assessment tasks. First, within the European Partnership for Alternative Approaches to Animal Testing (EPAA) NAM Designathon, the thesis demonstrates how HT-PBK predictions alone can support the classification of chemicals based on their potential for systemic availability when toxic potency information is unavailable. Finally, HT-PBK predictions are combined with *in vitro* toxicity assay data to assess the risk of drugs to cause drug-induced liver injury (DILI). The integration of *in vitro* assays and HT-PBK simulations yields strong alignment with clinical DILI outcomes, with a ROC AUC of 91%. By combining mechanistic simulation with ML-driven parameter prediction, this thesis advances the scientific basis for next generation risk assessment. The resulting tools are transparent, independent of animal data, and designed for efficient evaluation across large chemical sets. They support the growing scientific and regulatory efforts to integrate ML-enabled modelling into toxicology and pharmacology.

General Introduction

Over the past decades, the life sciences have undergone a profound transformation. Advances in experimental and computational methods have provided a much deeper understanding of biology, the intricate processes that underlie human life, and the ways in which exogenous substances interact with those processes. Historically, human understanding of both biology and chemical effects on health was superficial. The effects of chemicals on health, for instance, could only be assessed empirically, often through trial-and-error experimentation, and much of early biological research was descriptive in nature, limited to documenting phenomena without being able to reveal their underlying mechanisms.

This began to change with the emergence of molecular and cellular biology, catalysed by innovations such as the invention of the microscope, the deciphering of the genetic code, the sequencing of the human genome, and the rise of high-throughput omics technologies [1]. These advances have increasingly moved biology from a discipline dominated by observation toward one capable of mechanistic explanation and, crucially, prediction [2]. The ability to anticipate outcomes rather than merely to document them is now a defining feature of modern biological science.

This paradigm shift is also evident in the field of toxicology and chemical safety assessment [3,4]. Before these developments, there was no way, other than direct experimentation, to anticipate how exogenous chemicals would interact with biological processes. To determine whether a substance posed a risk to human health, this interaction had to be tested directly, most often in animal experiments. While animal testing was long considered indispensable, it has always been problematic. Beyond the ethical concerns related to this practice, its performance has historically also been inconsistent. After all, there are substantial differences between the biology of a human being and laboratory species such as mice or dogs. Hence, animal data were never perfectly predictive of human outcomes [5,6]. Despite these limitations, animal experimentation remained the dominant approach for decades, simply because no viable alternatives existed.

Today, however, the scientific community is approaching a turning point. Thanks to rapid progress in both experimental and computational methods, it is becoming possible not only to replace animal experimentation, but even to surpass it. Emerging methods increasingly allow us to assess the effects of chemicals directly based on human rather than animal biology. This transition is driven by advances in our understanding of the fundamental processes of life, as well as by methodological developments in two main areas: *in vitro* methods and computational approaches. On the experimental side, increasingly sophisticated model systems such as complex cell cultures, organoids, and organ-on-a-chip technologies enable us to gain insights into human biology outside of the body [7,8]. Many of these systems are compatible with miniaturisation, automation, and high-throughput screening, allowing thousands of tests to be performed in parallel. On the computational side, advances in computing and in methods such as data science and computational modelling enable the integration of diverse experimental outputs into coherent frameworks that holistically represent the human body, simulate biological processes, and predict outcomes of chemical exposure [9,10].

Fundamentally, two interrelated components determine how chemicals affect the body, and both need to be considered to predict those responses: pharmacokinetics (PK; or

toxicokinetics, TK, in toxicology) and pharmacodynamics (PD; or toxicodynamics, TD) [11]. In simple terms, TK/PK addresses the question of exposure, meaning how much of a chemical enters the body, where it distributes, and how long it remains in various tissues. TD/PD, by contrast, describes the biological effects a chemical can exert, and the magnitude of those effects. Both components ultimately need to be combined to understand chemical effects on human health, but this thesis focuses primarily on the prediction of the first component, the PK.

The study of PK is a long-established scientific field, and more detailed background on it is given in chapter 1. However, for a long time PK analysis was largely restricted to experimental measurements of substance concentrations in living organisms. Only in recent decades, due to the computational revolution of the late twentieth and early twenty-first centuries, did it become possible to use computational models to dynamically simulate how chemical concentrations change over time and across the entire body. A fundamental step in this development was the emergence of PBK modelling software in the early 2000s [12]. PBK models simulate the distribution of compounds in the human body by representing its anatomy and physiology as a network of connected compartments and related processes. Since their emergence, PBK models have become important tools in pharmaceutical development and chemical risk assessment, allowing researchers to estimate tissue concentrations of compounds dynamically over time [13–15]. Chapter 2 provides more detailed background on PBK modelling.

PBK models are useful tools that are routinely applied in the pharmaceutical and chemical industries to investigate *in vivo* chemical concentrations. But despite their utility, PBK models also have important limitations. For instance, their development requires significant expert knowledge, extensive data on both compound-specific properties and human physiology, and a lengthy process of calibration and validation. Traditionally, PBK models are developed for individual compounds on a case-by-case basis, often requiring iterative refinement before they achieve acceptable accuracy [16,17]. As a result, although PBK modelling is powerful and can in principle provide a key component for chemical assessment, it is not readily applicable or easily scalable to large numbers of substances.

The advent of new machine learning (ML) and artificial intelligence (AI) methods now promises to resolve these current limitations of PBK modelling and to enable rapid PK predictions of large numbers of substances without the need for human input. ML methods are increasingly capable of predicting the compound-specific properties required as inputs for PBK modelling [18]. This enables a new approach to PBK modelling, termed high-throughput PBK modelling (HT-PBK), in which predictions of chemical pharmacokinetics are generated rapidly, at scale, and with minimal human intervention [19]. HT-PBK has the potential to deliver one of the key missing pieces of modern toxicology: the ability to anticipate internal chemical concentrations in humans across large numbers of substances, without relying on animal experiments. More information on HT-PBK modelling is provided in chapter 3.

The main limitation of this new approach is that the accuracy and reliability of high-throughput PBK predictions are still uncertain, for two reasons. First, for many of the input ML methods it is not yet clear how accurate their predictions of substance properties are, especially when the substances to be evaluated are structurally different from those in their training data. Second, even if the compound properties used for modelling are accurate, the expected accuracy of the resulting PK simulations from HT-PBK models remains unclear, because such PBK models need to rely on a generic model structure to be applicable in a high-throughput manner and therefore cannot

incorporate compound-specific processes. This constraint is likely to make them inherently less accurate than traditional PBK models. These uncertainties hinder the broader acceptance of this promising approach in regulatory and industrial contexts.

The aim of this thesis is therefore twofold: first, to evaluate the accuracy and robustness of HT-PBK modelling for predicting chemical concentrations in humans, and second, to demonstrate that, when integrated with emerging TD/PD assessment methodologies, HT-PBK provides a robust foundation for predicting the outcomes of chemical exposure in humans. In doing so, this work ultimately seeks to advance the transition toward a more ethical, human-relevant, and predictive framework for chemical risk assessment that reduces reliance on animal-based testing.

The structure of the presented work is as follows:

Part I (Background) provides the basic information and context about the relevant fields and methods touched upon by this work.

Part II (Materials and Methods) describes the methodological details of the presented work. Together with the results, this material has already been published in peer-reviewed scientific journals, such as Archives of Toxicology and Environment International.

Part III then outlines the scientific results of the research performed within the context of this thesis.

- Chapter 10 presents a comprehensive validation of the accuracy of HT-PBK simulations. To this end, we assembled a large-scale PK dataset comprising over 2000 concentration-time profiles of over 200 substances following intravenous (IV) and oral (PO) administration. To our knowledge, this represents the largest human validation dataset compiled to date. Using these data, we systematically investigate which ML models provide the most suitable inputs for HT-PBK simulations, considering both freely available and proprietary ML models from industry partners. This analysis enables us to identify optimal parameterisation strategies and to quantitatively evaluate the predictive accuracy that such simulations achieve. Importantly, the study serves both as a technical optimisation effort and a validation exercise, since it characterises the expected error distributions of HT-PBK predictions. Establishing this baseline uncertainty is crucial for enabling industry and regulatory stakeholders to adopt novel HT-PBK approaches in decision making processes.

- Building upon the validation for IV and PO administration, chapter 11 extends the assessment to dermal administration. This addresses a critical need within the cosmetics and consumer products industries established by the European Union's 2013 ban on animal testing for cosmetic products [20], which has created pressing demands for alternative assessment methods among consumer goods companies such as Nestlé, Procter & Gamble, Unilever, and L'Oréal. The development of new cosmetic formulations, whether sunscreens, deodorants, or skincare products, requires robust safety assessment despite the prohibition on traditional testing approaches. Given that dermal exposure represents the primary route of contact for these products, accurate prediction of skin penetration and subsequent systemic distribution is crucial. This chapter therefore provides a targeted validation of HT-PBK predictions specifically for

transdermal absorption scenarios, establishing the reliability metrics necessary for regulatory acceptance and industrial implementation within this sector.

- Having established the predictive performance of HT-PBK across different administration routes, chapter 12 transitions from validation to practical application and demonstrates the utility of HT-PBK modelling in addressing real-world regulatory challenges. This work responds directly to a challenge set out by the European Partnership for Alternative Approaches to Animal Testing (EPAA), which seeks to promote the development of a new regulatory framework for chemical safety assessment based exclusively on New Approach Methodologies (NAMs) [21,22]. Within this initiative, we were tasked with classifying substances according to their level of concern based on their TK and ADME (absorption, distribution, metabolism, excretion) properties. The reference set comprised 150 chemicals, and through HT-PBK simulation of this chemical set, we demonstrate the feasibility of efficient, high-throughput kinetic assessment without animal data. The chapter presents the theoretical framework underlying our classification approach and introduces a novel scoring metric that integrates multiple kinetic parameters to quantify the level of concern associated with systemic exposure potential. This provides a practical tool for prioritising chemicals requiring more detailed assessment while demonstrating the readiness of HT-PBK modelling for regulatory implementation.

- The final results chapter, chapter 13, ultimately illustrates the integration of HT-PBK predictions with complementary NAMs by combining kinetic simulations with *in vitro* hepatotoxicity measurements. This integration addresses drug-induced liver injury (DILI), one of the most critical and challenging forms of drug toxicity. DILI is a leading cause of clinical trial failures and post-market drug withdrawals, and it costs the pharmaceutical industry billions of dollars annually, while being poorly predicted by conventional animal models [23–25]. Our analysis incorporates relevant *in vitro* hepatotoxicity datasets published over the past decades, demonstrating that the combination of HT-PBK-derived exposure estimates with human-relevant *in vitro* response data achieves superior concordance with clinical outcomes compared to traditional approaches. This finding establishes that existing methodologies, when properly integrated through kinetic modelling, can address even the most complex toxicological endpoints. The success of this integrated approach supports the central thesis that HT-PBK modelling, combined with other NAMs, can provide a scientifically robust and ethically superior alternative to animal testing for chemical safety assessment.

Following the described structure, the thesis provides a systematic evaluation of the accuracy of HT-PBK simulations against human *in vivo* PK data across intravenous, oral, and dermal administration, demonstrating that the method achieves sufficient predictive performance. Building on this validation, we then illustrate how HT-PBK can be applied both independently and in combination with toxicodynamic NAM data for chemical safety assessment. Using DILI as a case study, we finally validate the utility of this integrated approach by benchmarking predictions against clinical outcome data.

The thesis concludes with a synthesis of these findings in the General Conclusions and Outlook chapter, where their broader significance for the field of predictive toxicology and regulatory science is discussed.

Part I: Background

Author contributions:

René Geci wrote these chapters and created the figures and tables.

1. Pharmacokinetics

What is pharmacokinetics?

Pharmacokinetics (PK) is the study of how drugs move through the body [26]. Its main objective is to investigate how the concentrations of drugs change in different body compartments, such as the blood and the various other body organs, over time after administration. Commonly, it is defined to encompass all topics addressed by the question: “What does the body do to the drug?”.

Three broad classes of factors determine the PK of a drug [27] (Figure 1A). First, PK is influenced by properties that are inherent to the drug itself. For instance, the solubility of a given substance influences how rapidly it can be taken up by the body. Other examples include the dose at which drugs are administered, which is often directly proportional to *in vivo* concentrations, and the formulation in which drugs are provided, which influences how rapidly and to what extent substances are absorbed. The second class of properties that determine drug PK comprises factors inherent to the body. This includes the body’s anatomical and physiological characteristics: the volume of different organs, their composition, for example in terms of adipose tissue or water content, blood flow rates between organs, etc. All of these are relevant factors influencing the movement of drugs throughout the body. The third and final class includes factors that depend on both the body and the drug, and which arise from the interaction of both with each other. For instance, the rate at which enzymes metabolise a molecule depends on both the molecular structure of the drug and also on the structure of that enzyme, which in turn is determined by the genetic makeup of the body.

Another commonly used classification categorises the factors which determine PK into chronological processes [28]. In this context, they are typically summarised as the absorption, distribution, metabolism and excretion (ADME) processes (Figure 1B). Absorption refers to the rate and extent of drug uptake by the body from the local site of administration. For drugs to cause their intended effects, they usually need to enter systemic circulation via the bloodstream. The most common form of drug administration is oral administration (per os; PO), but many other routes exist, including intravenous (IV), dermal, subcutaneous, intramuscular, inhalation, ocular, and rectal administration. Distribution refers to how quickly and to what extent drugs enter different body parts. Some drugs remain largely within the blood, whereas others distribute extensively into specific organs or tissues. Certain organs, like the brain, possess specialised mechanisms for preventing drug distribution into them [29]. Metabolism describes how the enzymes of the body break down a drug into new molecules called metabolites. Metabolites are usually less biologically active and easier to dispose of than their parent compounds, although exceptions to this rule exist [30]. Excretion describes how the body ultimately eliminates drugs or their metabolites, which mostly occurs via the urine or faeces.

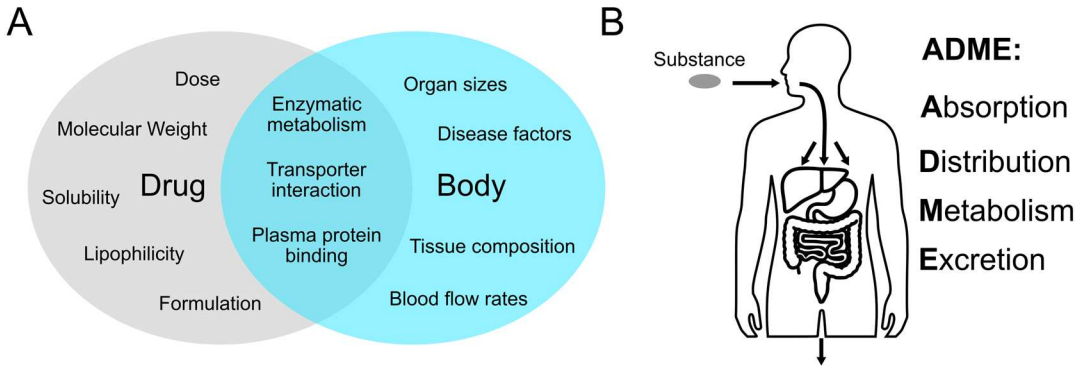


Figure 1: Factors determining pharmacokinetics. (A) Pharmacokinetic behaviour of a substance is determined by drug-specific properties (grey), body-specific properties (blue), factors arising from their interaction. Examples of each category are shown. (B) A diagram illustrating key anatomical sites involved in absorption, distribution, metabolism, and excretion (ADME) processes.

The relevance of pharmacokinetics in pharmacology and toxicology

PK is a key factor in both pharmacology and toxicology [31,32]. It is fundamental to understanding both the desired and undesired effects that substances exert on living organisms. In pharmacology, PK is essential for understanding the optimal dose at which drugs should be administered, how frequently and via which routes of administration this should be done. It provides insights into why certain subpopulations respond differently to treatments, and how other populations, such as children or the elderly, may need to be treated differently than the average adult population. It provides context for the interpretation and translation of animal study data to human outcomes. It explains why and how different drugs may interact and influence each other. For these reasons, PK is an integral part of drug development in the pharmaceutical industry, as well as of the drug approval processes in regulatory agencies.

Similarly, PK is also of crucial relevance in toxicology. Since Paracelsus stated in the 16th century “Alle Dinge sind Gift, und nichts ist ohne Gift, allein die Dosis macht dass ein Ding kein Gift ist” (commonly translated as “the dose makes the poison”) [33], it has been accepted as a central dogma of toxicology that one, if not even the most, dominating factor determining the toxicity of a substance is the extent to which organisms are exposed to it. Therefore, PK analysis is applied in toxicology in much the same way as in pharmacology. It provides context for determining safe exposure levels for a wide range of substances, including industrial chemicals, food additives, pesticides and substances present in consumer goods products. It supports the translation of toxicity outcomes from animal studies to human risk assessment and enables the evaluation of chemical effects following chronic exposure [34]. As in pharmacology, PK is also a key consideration for regulatory agencies when approving or restricting the use of chemicals.

In toxicology, the term toxicokinetics (TK) is sometimes used to distinguish its purpose, scope, and application from those in pharmacology. However, from a biological perspective, there is no fundamental difference between chemicals classified as poisons or medicines, and neither does the body distinguish between them when processing them. Although some have proposed to reconcile the two terms (PK and TK) under

biokinetics or just *kinetics* [35], pharmacokinetics remains the most widely used term and is therefore adopted throughout this thesis for clarity and accessibility.

How is PK investigated?

There are two fundamental approaches for assessing the pharmacokinetics of substances. The first is by experimental measurement and the second is by simulation. In principle, it is possible to measure the concentrations of chemicals in the body. This requires collecting samples of body tissue and then determining the quantities of molecules in those samples using analytical chemistry methods [36]. Such measurements are routinely conducted in animal studies. However, they require the corresponding animals to be sacrificed to obtain sufficient amounts of organ tissues for analysis. Consequently, this approach is unsuitable for measuring *in vivo* concentrations of substances in humans. In humans, concentration measurements are limited to easily accessible tissues that can be sampled safely without causing harm to the subjects. This is typically the case for blood and its subcomponents, as well as for urine and faeces. *In vivo* concentrations of substances in all other body sites, including specific organ concentrations, are generally not directly measurable in humans.

When *in vivo* concentrations of substances are measured directly, multiple measurements are typically taken over time to capture the dynamic changes in quantities. This provides investigators with concentration-time profiles, which are the most comprehensive representation of *in vivo* kinetics (Figure 2). For ease of comparison and analysis, these dynamic data are then often condensed into static summary parameters that are easier to interpret and compare [37]. One of the most widely used PK parameters is the maximum concentration (Cmax), which represents the peak concentration of a substance in a given tissue. Another widely used summary parameter is the area under the curve (AUC) of the concentration-time profile. It represents the total drug exposure over a given period. Cmax and AUC are the most commonly used surrogates for the strength of substance effects. Depending on the mechanism of action of a compound, either of them may be the main determining factor governing the responses that substances provoke on biological systems. Typically, it is not known *ex ante* whether Cmax or AUC is going to be the driver of biological responses of a given substance. Some notable exceptions to this general pattern exist. For example, in the case of antivirals or antibiotics, trough concentrations (minimum concentrations before the next dose) may be a more relevant parameter for their effectiveness. However, such cases represent exceptions rather than the rule.

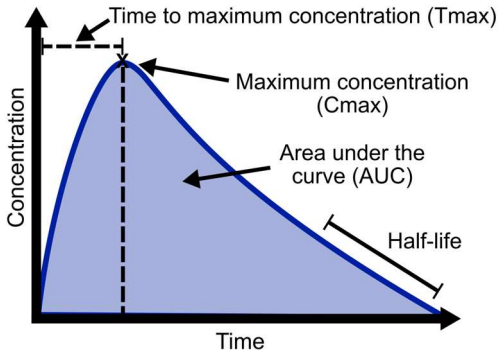


Figure 2: Key pharmacokinetic parameters derived from a concentration-time profile. The time to peak concentration (Tmax) represents the time at which the maximum observed concentration (Cmax) occurs. The area under the concentration-time curve (AUC) quantifies the total systemic exposure to the substance over the observed period. Half-life (Thalf) is the time required for concentrations to decrease by half.

In addition to Cmax and AUC, numerous other summary parameters can be calculated to characterise and summarise the PK of substances. Key additional parameters are the volume of distribution (Vd) of substances, their time to maximum concentration (Tmax), their half-life (Thalf), their bioavailability (F) and clearance (CL). Vd is the theoretical volume in which the total drug amount would need to be uniformly distributed to produce the observed blood concentration. It reflects the extent to which a drug distributes into other body tissues relative to the plasma, which is an important factor when considering dosing, as well as the potential efficacy and safety of a new drug candidate. Tmax is the time after administration at which Cmax is reached. It provides relevant insights for understanding how quickly and at what point drugs produce their effects. Thalf is the time required for drug concentrations to decrease by half, indicating the rate of drug elimination. Bioavailability is the fraction of an administered drug dose that reaches systemic circulation, reflecting both the extent of absorption and first-pass metabolism in the liver. Finally, CL is the volume of plasma from which the drug is completely removed per unit time, representing the overall efficiency of elimination via metabolism and excretion. Collectively, these parameters provide complementary insights into fundamental drug properties, are important for understanding *in vivo* effects and are frequently used in drug development. All of them can be derived from concentration-time profiles. That is why the concentration-time profile is the most informative and powerful representation of substance PK.

An alternative to direct experimental measurement is to predict PK using mathematical or computational models. There exist a range of different modelling approaches of varying complexity that can be applied for this purpose. One common classification distinguishes between empirical and mechanistic models [38]. Empirical models use observed data to describe the relationship between dosing and drug concentrations without explicitly representing the underlying biological processes determining them. This category of models includes, for instance, simple one- or two-compartment models which represent the body as a small number of discrete compartments and describe drug movement between them using simple mathematical equations. Those compartments do not necessarily match any real anatomical or physiological structures and are theoretical in nature. Mechanistic models, on the other hand, explicitly incorporate physiological,

anatomical and drug-specific properties to provide a more realistic representation of drug PK. A challenge with mechanistic models is that they are often more data-intensive and complex to parameterise compared with simpler empirical models. Physiologically based kinetic (PBK) models are examples of such mechanistic models and will be explained in detail in the next section (chapter 2).

Both empirical and mechanistic PK models require data for their parameterisation and validation. However, data requirements and the ability to incorporate different data types vary between model categories. Broadly, there are four main sources of such data: *in silico* predictions, *in vitro* measurements, animal *in vivo* PK, and human *in vivo* PK data. Human PK data can be used both to build and to validate already parameterised PK models. For example, a model built on adult PK data can be used to extrapolate PK predictions for paediatric or geriatric subjects. Similarly, animal PK data can be used to calibrate PK models for predicting substance concentrations in humans. Alternatively, *in vitro* measurements, such as the rate at which substances are metabolised by hepatocytes or specific enzymes, can also be used to parameterise PK models. This requires *in vitro* to *in vivo* extrapolation (IVIVE) to translate such measurements to the *in vivo* context, for example when translating an observed enzymatic metabolisation rate into an overall hepatic clearance rate. Finally, PK models can be informed by other computational models. For instance, machine learning models that predict the lipophilicity or water solubility of a drug can be used to provide those values as inputs for other types of modelling. However, this type of parameterisation represents a relatively recent development. The section after the next (chapter 3) will explore how such *in silico* data can be used to inform PK models. It will then be a central aim of this thesis to advance the integration of such approaches for fully *in silico*-based, high-throughput PK prediction.

2. PBK Modelling

What is PBK modelling?

Physiologically based kinetic (PBK) modelling is a computational method that enables the simulation of time courses of *in vivo* substance concentrations in organisms, using realistic representations of their anatomy and physiology [39] (Figure 3A). In pharmacology, this approach is commonly referred to as physiologically based pharmacokinetic (PBPK) modelling, whereas in toxicology it is sometimes called physiologically based toxicokinetic (PBTK) modelling. To avoid confusion and emphasise that there is no fundamental difference between these two approaches, the umbrella term PBK is used here to indicate that the discussed approach is applicable to both fields.

PBK models are mathematical models based on systems of ordinary differential equations (ODEs) which describe the body as multiple compartments (e.g., organs) interconnected by blood flow [40]. Unlike empirical models, these compartments explicitly correspond to real biological entities, such as the specific organs and tissues of which a body is composed. PBK model parameters are therefore directly linked to real and measurable physiological and anatomical properties, such as the blood flow rates between specific organs, organ volumes and tissue compositions. Although PBK models require the provision of numerous parameters, parameterisation of most properties is straightforward because many physiological and anatomical values are well established and available from standardised databases. At least, this applies to all model parameters related to the physiology and anatomy of the body. In addition, the provision of chemical-specific parameters is required to describe the PK of substances, such as their physicochemical properties or clearance rates. The provision of these chemical-specific parameters will be discussed in more detail in the next section (chapter 3).

One key reason chemical-specific parameters are required in PBK models is to describe how substances distribute from blood into different tissues. Tissue-to-plasma partition coefficients are particularly important for this, as they determine how extensively a compound distributes beyond the vascular space into organs. These coefficients quantify the equilibrium concentration ratio of a drug between tissue and plasma and depend both on physicochemical properties of a drug (e.g., lipophilicity, ionisation state) and the composition of a tissue (e.g., water, lipid, and protein content). They are typically estimated using established composition-based methods, such as the Rodgers-Rowland [41] or Poulin-Theil equations [42], which are integrated into most PBK modelling platforms. The actual entry of a compound into tissues may either be limited by the rate of blood flow to the tissue (perfusion-limited distribution) or by membrane transport processes (permeability-limited distribution), including passive diffusion and active transport, depending on the characteristics of the tissue and the compound. By dynamically integrating these data and processes, PBK models ultimately enable the simulation of substance absorption and subsequent distribution throughout the body.

In practice, PBK models are developed through an iterative, stage-wise process that can span the entire course of drug development, from preclinical species to human patients [43] (Figure 3B). Initial models are typically built using physiological data from animal models, *in vitro* measurements and chemical-specific parameters. These are used to simulate pharmacokinetics in animals and are compared against animal *in vivo* PK data to assess model performance. If models perform adequately, they are then used to extrapolate to humans, first to healthy volunteers and later to specific patient populations,

by adjusting for relevant physiological differences. Each step involves a cycle of simulation, comparison with experimental or clinical data, and model refinement. This “learn-confirm-refine” approach is fundamental to building confidence in the corresponding PBK models. It is accompanied by formal procedures of verification, validation, and uncertainty quantification. Verification ensures correct technical implementation and model structure. Validation involves comparison of simulated and observed data to assess model credibility. Uncertainty quantification evaluates the variability in parameter estimates and its impact on the confidence in model predictions. Together, these steps ensure that PBK models are not only mechanistically sound but also fit for purpose in decision making contexts.

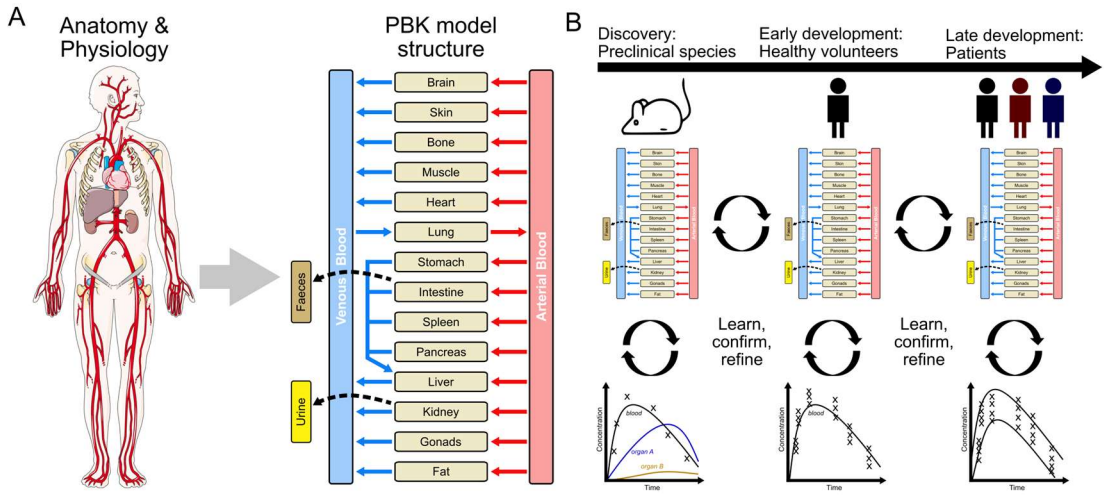


Figure 3: Schematic representation of a physiologically based kinetic (PBK) model structure and the PBK modelling workflow. (A) The model translates anatomical and physiological features of the human body (left) into interconnected compartments (right) representing individual organs and tissues, linked via arterial and venous blood flows. Excretion pathways via urine and faeces are also incorporated, enabling simulation of absorption, distribution, metabolism, and excretion processes over time. Human body schematic adapted from Servier Medical Art, https://smart.servier.com, CC BY 3.0. (B) Conceptual sketch of PBK model refinement throughout drug development stages, adapted from [43].

The use of PBK models in pharmacology and toxicology

Due to their ability to dynamically simulate substance concentrations over time, PBK models are crucial tools for analysing the PK of substances and understanding both their desired and undesired effects in pharmacology and toxicology. They bridge the gap between external exposure doses and internal concentrations and enable the prediction of tissue concentrations under scenarios that would otherwise be costly, impractical or unethical to investigate.

PBK modelling is employed for three interconnected main purposes. First, PBK models enable the prediction of PK in the absence of *in vivo* data. By integrating and translating, for instance, *in vitro* measurements of substance properties, such as their physicochemical properties or clearance rates, they provide a mechanistic basis for predicting *in vivo* PK, even in the absence of any *in vivo* data to validate those predictions. In pharmaceutical development, this can be useful for selecting safe starting

doses in first-in-human (FIH) trials, prioritising lead compounds before performing costly animal studies, and for evaluating early drug-drug interaction (DDI) risks. Even more important, however, is this use of PBK models in toxicology. Unlike in pharmacology, where preclinical animal experiments remain a standard part of the drug development process, animal studies for toxicological purposes are often infeasible, unethical or even illegal [44]. For example, in the European Union (EU), animal testing is prohibited for the development and safety assessment of cosmetic ingredients, which applies to almost all types of consumer goods products [20]. Hence, for such compounds there is no alternative for obtaining PK data other than computational prediction. Similarly, for the high-throughput screening of thousands of uncharacterised environmental chemicals, performing animal studies to determine their PK for safety assessment is at least impractical. Simulation based on *in silico* and *in vitro* parameterisation is the only feasible approach for evaluating such a large number of substances.

Second, PBK models provide a powerful framework for performing what-if scenario analyses to investigate the effects of changes in different parameters. This is, in theory, also true for many other computational modelling approaches. However, because PBK model parameters are explicitly intended to represent real biological entities, properties and mechanisms, they are especially well suited for such analyses. For instance, they can be used to investigate what effects physiological changes occurring during pregnancy or ageing may have on drug PK, how different disease states impact drug behaviour, how genetic polymorphisms cause diverging drug effects, how alternative exposure routes or formulations influence drug behaviour, or how drug PK differs between different species. Such analyses provide a mechanistic rationale for study design, dose adjustments, and for safety and risk predictions in pharmacology. In pharmacology, this analysis approach is called model-informed drug development and discovery (MIDD or MIDDD) [45]. In toxicology, it is referred to as model-informed risk assessment (MIRA) and forms a part of next generation risk assessment (NGRA) [46,47].

Third, because PBK modelling involves the simulation of substance concentrations across all body parts, it enables the investigation of *in vivo* concentrations in organs that cannot be sampled in living subjects. As outlined previously, intra-organ concentrations can be measured in animals, but are generally infeasible to obtain in humans without causing harm to subjects. This poses a particular challenge in pharmaceutical development, since the primary determinant for the *in vivo* effects of substances is their concentration at a specific, local target site. The inability to anticipate those concentrations can significantly hinder drug development. Due to its explicit representation of various body organs, PBK modelling, however, makes it possible to computationally simulate substance concentrations in such otherwise inaccessible locations. Due to the limited availability of such organ-specific concentration data, PBK modelling methods have only been developed on and evaluated with sparse animal datasets. Consequently, such organ-level predictions are not always fully reliable. They remain, fundamentally, predictions in the absence of means for measurement. Nonetheless, they are grounded in mechanistic principles and often represent the only available approach for such predictions. Together, these three capabilities make PBK modelling a cornerstone of modern drug development and chemical safety assessment.

3. High-throughput PBK Modelling

What is high-throughput PBK modelling?

High-throughput PBK (HT-PBK) modelling refers to the standardisation and automation of PBK model generation so that such models can be constructed rapidly for large numbers of substances. Traditional PBK modelling is performed for individual substances and often requires the inclusion of substance-specific processes, which may sometimes be relevant only to a single compound. In contrast, HT-PBK modelling generates models in a standardised, generic manner, without the ability to explicitly represent compound-specific processes (Figure 4). Additionally, traditional PBK modelling is typically performed iteratively by calibrating models to and validating them against various *in vivo* datasets, such as PK data from different species, populations or treatment regimens. This type of evaluation is infeasible for HT-PBK modelling because validation datasets are typically not available in a high-throughput manner, preventing model calibration and validation in the traditional sense.

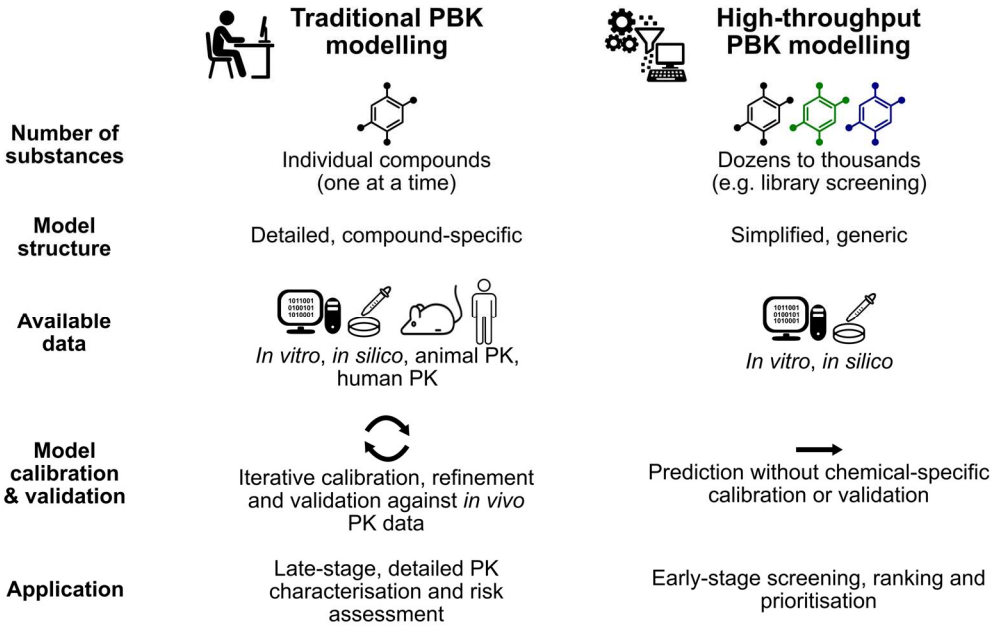


Figure 4: Schematic representation of differences between traditional and high-throughput PBK modelling. Traditional PBK modelling develops detailed, compound-specific models for individual substances, whereas high-throughput PBK modelling uses generic model templates with *in vitro* and *in silico* inputs to rapidly generate PK predictions for large numbers of compounds.

Although the technical groundwork and underlying concepts of HT-PBK modelling were already developed in previous decades, modern HT-PBK modelling started to form in the early 2010s, primarily driven by the need for more efficient PBK modelling approaches in chemical risk assessment. In 2007, the US National Research Council's "Toxicity Testing in the 21st Century" report had called for evaluating thousands of chemicals using *in vitro* and *in silico* methods to replace slow, animal-based testing approaches [3,4]. This vision led to initiatives such as the US Environmental Protection Agency's (EPA) ToxCast

programme and the federal research collaboration Tox21 [48], which began screening thousands of chemicals in biochemical and cellular assays. However, interpreting those high-throughput hazard data to prioritise chemicals by their potential risk required complementary exposure and kinetic modelling. It was necessary not only to determine the *in vitro* concentration at which chemicals exhibit activity, but also to determine whether those concentrations could be achieved *in vivo* at plausible exposure levels. This need led to the development of the concept of high-throughput toxicokinetics (HTTK) and an identically named computational tool [49].

The vision behind HTTK was to combine generic PBK models with rapidly obtainable input parameters to enable high-throughput pharmacokinetic simulations for many chemicals. Following the US EPA's development of an open-source software package, also called HTTK, the approach was then increasingly applied for toxicological assessments of substances. The demonstration of this approach's usefulness and success led to its rapid adoption in the pharmaceutical industry [50,51]. While drug discovery teams had long used PBK modelling for individual drug candidates, they subsequently began to adapt these methods for higher-throughput screening of compounds using predicted or measured ADME properties. Software providers, such as Simulations Plus with their HTPK module, and other regulatory agencies, such as the European Food Safety Authority (EFSA) with TKPlate [52], subsequently developed similar HT-PBK methods.

How is HT-PBK modelling currently used?

High-throughput PBK modelling is currently applied primarily in three contexts. First, regulatory agencies and their research branches, such as the US EPA, EFSA and Health Canada, employ it for chemical-safety prioritisation by translating high-throughput *in vitro* bioactivity data into administered-equivalent doses and for performing chemical safety assessments [53,54]. This is part of chemical prioritisation efforts under programmes such as ToxCast, Tox21, and the Canadian Chemicals Management Plan. Second, consumer and cosmetics product companies such as Unilever, Nestle, L'Oréal, and Procter & Gamble integrate HT-PBK simulations with other new approach methodologies (NAMs) to estimate the internal exposure to new substances to prove their safety without having to rely on animal testing [55,56]. These applications are aligned with legal requirements such as the EU Cosmetics Regulation and OECD case studies on next generation risk assessment [57]. Third, pharmaceutical research teams use HT-PBK approaches in early drug discovery to predict human exposure, assess interaction potential, and prioritise chemical series before initiating *in vivo* studies [51,58,59]. Across all these contexts, HT-PBK modelling provides a rapid, mechanistically informed approach to estimating internal concentrations, supporting data-driven decisions in chemical risk assessment and early-stage drug development, while reducing reliance on animal testing.

A fundamental requirement for HT-PBK modelling across all these contexts is the availability of a small set of fundamental, chemical-specific parameters that determine the PK of substances in these models. These parameters primarily comprise basic physicochemical properties such as molecular weight, lipophilicity, acid dissociation constants (pKa), and aqueous solubility. These values are often readily derived from *in silico* prediction tools based on quantitative structure-activity relationships (QSARs) or machine learning (ML) models trained on large chemical datasets [60–63]. In addition, only two further parameters relating to the distribution and excretion of compounds are

typically required as HT-PBK model inputs. These are the fraction unbound (Fu) in plasma and the clearance rate of a substance. To date, both are typically obtained through *in vitro* measurement.

Fu reflects the extent of plasma protein binding of substances and is commonly measured using standardised *in vitro* assays such as equilibrium dialysis or ultrafiltration [64]. These assays are relatively well standardised and can be applied in a high-throughput manner to new compounds. Clearance rates, on the other hand, present greater experimental and logistical challenges. They are typically determined using liver microsome or hepatocyte assays that measure metabolic depletion rates of compounds over time [65]. These assays require the development and optimisation of suitable analytical chemistry methods, such as LC-MS/MS, to quantify the parent compound and its depletion across multiple time points. This makes the determination of clearance a limiting factor for the use of HT-PBK approaches, as it is a key property determining the PK of substances but is challenging to measure.

Similar to computational QSAR or ML methods used to predict basic physicochemical compound properties, efforts have been made to develop *in silico* prediction methods for Fu and clearance rates [61,66–69]. Tools such as Simulations Plus' ADMETPredictor and GastroPlus allow the direct incorporation of such predictions for HT-PBK simulations. Similarly, different research projects have explored using predictions rather than experimental measurements of such parameters. Clearance, in particular, is difficult to predict, however. Presumably this is because it not only depends on the chemical structure of a substance, but also on the structures of dozens of enzymes and transporters, as well as multiple molecular and cellular processes which may or may not be involved in the processing of any specific molecule. By comparison, the water solubility or other physicochemical properties of substances are relatively straightforward to predict, as they only depend on simpler interactions between substances. Consequently, clearance remains a far more challenging property to predict.

Despite the availability of computational tools for the prediction of various compound properties, the accuracy of PK predictions based on their use is often uncertain. All computational tools have inherent technical errors and prediction uncertainty, which means that errors in their property predictions must propagate into mispredictions of the PK of substances. More fundamentally, it is also generally unclear how accurate HT-PBK simulations can be expected to be, even when input parameters are perfectly correct [70]. This is because PBK models must employ simplistic, generic structures to be applicable for high-throughput evaluation. For instance, they cannot represent individual transporters or their effects on drug PK. Yet such processes are routinely incorporated in traditional PBK modelling of individual compounds, and they are often essential for achieving sufficient model accuracy. These omissions and simplifications of processes therefore have to inevitably reduce the accuracy of HT-PBK predictions [71,72]. To date, the magnitude of the corresponding PK prediction errors often remains unknown. This creates uncertainty for the use of these methods in regulatory and industry decision making processes. Stakeholders in these contexts are generally accustomed to handling data uncertainty. However, the absence of information on its magnitude makes it challenging for them to rely on these methods [73].

A central aim of this work is to quantify the magnitude of errors in HT-PBK simulations based on different data types, particularly when using *in silico* predicted compound properties, in order to generate the confidence necessary for their use in pharmacological and toxicological applications.

4. New Approach Methodologies

New Approach Methodologies (NAMs; also referred to as *non-animal methods*) are methods for the assessment of chemical safety, including pharmaceuticals, that aim to partially or fully replace animal testing [74]. They encompass all types of *in silico* modelling approaches, *in chemico* assays, *in vitro* assays ranging from simple two-dimensional cell cultures to organoids and organ-on-a-chip systems, as well as *ex vivo* methods. NAMs can address a wide range of chemical effects from well-defined molecular interactions, e.g., with DNA, receptors or enzymes, to perturbations of cellular pathways or organelles, tissue- and organ-level dysfunction, and more complex, system-level outcomes such as immune, endocrine or developmental disruptions. Hence, NAMs constitute a diverse collection of methods united by the common goal of reducing or replacing animal testing.

The purpose of NAMs is to provide human-relevant evidence suitable for decision making in industrial and regulatory contexts, without necessarily replicating the exact same results obtained by animal testing. It is well recognised that, due to substantial differences in the biology of even closely related species, animal testing data are not perfectly predictive of human outcomes [75–77]. Hence, NAMs not only offer the opportunity to replace animal testing but even to overcome its current limitations by providing insights that are more relevant to human safety than those derived from conventional animal testing-based approaches. For this potential to manifest in practice, NAMs must provide information on human responses to chemical exposure at the levels and via the routes relevant to real-world use.

Over the past decades, regulators have shown an increasing willingness to rely on NAM data when the methods used are transparent, reliable and validated for their intended purpose [74,78]. International test guideline programmes now include non-animal assays for specific endpoints, and guidance documents describe how to integrate multiple lines of NAM evidence. In addition, several jurisdictions have established policies encouraging or requiring the consideration of NAM data where scientifically justified. For example, the EU fully banned animal testing for cosmetics already in 2013 [20]. In 2025, the US FDA announced plans to phase out animal testing requirements and the US NIH stated that it will no longer issue funding opportunities that exclusively rely on animal experimentation and instead prioritise human-based approaches. Although the depth of implementation varies across sectors, the overall trajectory is towards increased adoption of NAMs by both industry and regulators.

Despite progress, several challenges still hinder the replacement of animal-based legacy methods. Some of these challenges are very fundamental, and it is not entirely clear if and how they will be solved, while others are primarily technical in nature. One of the most fundamental challenges is the inability to accurately predict the *in vivo* responses of certain physiological systems using NAMs. Complex biological systems, such as the immune system, reproductive development, or neurological systems, are particularly difficult to emulate outside of the body [79,80]. In some cases, this problem stems from our incomplete understanding of these systems, while in others it arises from the complexity of the interactions between various entangled components, which makes it difficult to anticipate how chemical perturbations will propagate through these systems. Dynamic biological modelling, for example using Quantitative Systems Toxicology (QST) models, may eventually enable the prediction of such effects, but current capabilities remain far from achieving this on a comprehensive, multi-organ scale [81]. An advantage of animal testing, despite its numerous disadvantages and limitations, is the ability to

evaluate chemical effects on complex biological systems within an integrated, whole organism environment.

Nonetheless, for several endpoints, such as skin sensitisation or cardiotoxicity, robust NAMs based on well-defined molecular and cellular assays can reliably predict chemical effects by capturing fundamental toxicity mechanisms [82,83]. For such methods, the main barrier to replacing animal testing is technical rather than conceptual in nature. First, methods need to undergo rigorous technical validation, typically through extensive inter-laboratory ring trials and transferability studies [84]. It must be proven that assay results are reproducible and reliable across repetitions and different laboratories. Second, even when the predictivity and reliability of NAMs have been well established, it is often challenging to change existing legal frameworks and legal texts. Regulatory acceptance often lags scientific advances, meaning that even well-validated NAMs can remain underutilised until legislative requirements are updated [85].

Finally, many NAMs face the problem that they are only able to provide results in units of concentration. For instance, an *in vitro* assay may report half-maximal inhibitory concentration (IC50) values for receptor inhibition in molar concentrations. However, even though these data may be highly relevant for understanding toxicity, they are often not readily usable for decision making, particularly in regulatory contexts. Instead, it is typically necessary in such situations to translate given concentrations into an equivalent amount or dose, as regulatory enforcement requires actionable and easy to apply classifications. Thus, the relevant question is typically not “What concentration is safe?” but rather “What is the maximum safe dose?”. Most NAMs cannot provide this information directly, but it can be derived using PBK modelling. For this reason, PBK modelling itself may also be considered a NAM [46,86]. It allows conversion of *in vitro* measured points of departure (PODs) into specific external exposure doses, which can then be used for decision making. Because NAMs are often applied to new substances for which no *in vivo* PK data exist yet, it is particularly important that PBK simulations can be performed without such data. This is the purpose of HT-PBK modelling, which therefore plays a crucial role in replacing animal testing with NAMs.

5. Drug-Induced Liver Injury

What is drug-induced liver injury?

Drug-induced liver injury (DILI) refers to hepatic injury caused by exposure to medications [23]. Because of its proximity to the gastrointestinal tract and its central role in the metabolism of xenobiotics, the liver is particularly exposed and susceptible to the toxic effects of drugs and other foreign substances. For these reasons, DILI is one of the leading causes of adverse drug reactions, a frequent reason for drug failure in clinical studies, and a common cause for drug withdrawal from the market [87]. The economic impact of DILI on pharmaceutical companies has been estimated to be in the billions of dollars annually. Furthermore, DILI contributes to numerous patient deaths each year and is among the leading causes of acute liver failure, which can require liver transplantation or result in death [88].

DILI is an umbrella term encompassing diverse mechanistic causes and clinical phenotypes rather than a single disease. This makes its investigation and prediction challenging. In clinical practice, the diagnosis of DILI is typically one of exclusion, based on exposure timing, biochemical patterns, evolution of symptoms upon drug withdrawal or re-exposure, and the evaluation of alternative causes [89]. Standard liver biomarkers, such as the liver enzymes alanine aminotransferase (ALT), alkaline phosphatase (AP), bilirubin, and aspartate aminotransferase (AST), are routinely used to assess DILI. These markers are released upon cell damage but do not reflect specific underlying mechanisms and are also used for the evaluation of other forms of liver injury. The ratio of serum ALT to AP levels, known as the R-ratio, is used to classify the DILI patterns of drugs as hepatocellular, cholestatic, or mixed. An R-ratio greater than 5 indicates a hepatocellular pattern, values below 2 suggest a cholestatic pattern, and values between 2 and 5 are interpreted as mixed-type injury. The rationale behind this classification is that ALT is a cytosolic enzyme highly expressed in hepatocytes, whereas the AP isoenzyme is concentrated on canalicular membranes of hepatocytes. Thus, elevated ALT levels are assumed to indicate direct hepatocyte injury, while increased AP levels suggest damage to the biliary tree, as in the case of cholestasis or bile duct obstruction. However, these types of damage are not mutually exclusive, and many drugs exhibit a mixed pattern.

Additionally, the DILI potential of drugs is routinely categorised as either intrinsic or idiosyncratic. Intrinsic DILI is defined as clearly dose-dependent injury that is generally reproducible in animal models. The best known example of it is toxicity from acetaminophen (paracetamol) overdose. Idiosyncratic DILI, on the other hand, is typically dose-independent within therapeutic ranges, infrequent, and clinically heterogeneous, with variable latency and occasional immune-allergic features. Only 1 to 20 out of 100,000 patients typically exhibit DILI toxicity in response to treatment with drugs of idiosyncratic nature [90]. This low incidence makes it challenging to observe and evaluate such effects in clinical studies. To ensure patient safety and to facilitate the discovery of such DILI risks, several databases and networks have been established, including the US DILI Network, the Spanish DILI registry [91], LiverTox [92], and the FDA Liver Toxicity Knowledge Base [93]. These resources aim to identify DILI risks associated with drugs and to relate these events to patient-specific factors.

A growing body of evidence highlights that DILI risk is strongly modulated by patient-specific factors [94,95]. Genetic polymorphisms, especially in genes encoding drug-metabolising enzymes (e.g., CYP450s) or immune-related proteins (e.g., certain HLA

alleles), have been associated with increased susceptibility to DILI. Additionally, comorbidity conditions such as pre-existing liver disease, diabetes, or obesity can alter hepatic function or inflammatory status, thereby lowering the threshold for injury. Polypharmacy and drug-drug interactions further complicate the picture by affecting the metabolism or clearance of the offending drug. Age, sex, nutritional status, alcohol use, and gut microbiota composition are also increasingly recognised as contributors [96]. Together, these patient-level modifiers help explain the unpredictable and often idiosyncratic nature of DILI.

What are known mechanisms of DILI?

Despite substantial research efforts, our understanding of the mechanistic causes of the DILI risk associated with many drugs remains limited. This is likely because, in addition to drug-specific properties, individual patient factors play a crucial role in idiosyncratic DILI risk, and the relationships among these factors are not yet well understood. Nonetheless, several DILI mechanisms have been characterised with reasonable certainty, although to varying depths of mechanistic understanding [97–99].

Bioactivation: Many drugs undergo conversion by liver enzymes into reactive intermediates that covalently bind to proteins and lipids or deplete glutathione. These events can trigger endoplasmic reticulum stress, disrupt proteostasis, and lower the threshold for regulated cell death. The outcome is typically apoptosis or necrosis, depending on the context and severity of injury. In some cases, the resulting drug-protein adducts are recognised as neoantigens, thereby linking bioactivation to adaptive immune responses. Because the liver carries out the majority of xenobiotic metabolism, it is particularly vulnerable [100].

General cytotoxicity: Certain compounds cause broad cellular injury by disturbing plasma or organelle membranes, disrupting microtubules or lysosomes, or inducing marked oxidative or nitrosative stress. The resulting effects include cellular swelling, ballooning degeneration, and cell death via apoptosis or necrosis, often accompanied by secondary disruption of canalicular structure and bile flow [101].

Mitochondrial toxicity: Some drugs impair mitochondrial function by inhibiting respiratory chain complexes, uncoupling oxidative phosphorylation, or inducing opening of the permeability transition pore. Consequences include ATP depletion, increased reactive oxygen species, reduced beta-oxidation with lipid accumulation, and release of apoptotic or necrotic signals [102].

Immune mechanisms: In some cases, immune responses play a central role in DILI. Drug-protein adducts, or direct pharmacologic interactions with immune receptors, can be presented in certain HLA contexts, activating drug-reactive T cells. Damage-associated molecular patterns from stressed hepatocytes engage innate immune cells, such as Kupffer cells, and promote cytokine release. These processes may explain delayed onset, hypersensitivity features in some cases, and recurrence upon re-exposure [103].

Gene expression perturbations: Drug exposure prompts hepatocytes to adjust gene programmes that regulate chemical handling, by increasing or decreasing the expression of enzymes and transporters involved in bile or metabolite transport. These shifts may be protective, by accelerating drug clearance, or harmful, by generating more reactive metabolites or reducing bile flow. Cells also activate generic stress-response pathways to

cope with oxidative and protein stress. If these stresses persist, however, these responses may become maladaptive, worsening injury [104].

BSEP inhibition: The bile salt export pump (BSEP) mediates the transport of bile acids from hepatocytes into bile. Inhibition of BSEP by drugs leads to intrahepatic bile acid accumulation, membrane injury, and a cholestatic phenotype. The net effect depends on the activity of other transporters and the integrity of the canalicular membrane [105].

Overall, these mechanisms are not mutually exclusive and frequently co-occur in DILI cases. Other mechanisms have also been proposed in the pathogenesis of DILI, but those outlined here are among the best understood.

The current state of DILI prediction in drug development

Historically, the gold standard for predicting liver injury, similar to other toxicities, has been animal testing. However, concordance rates between animal and human hepatotoxicity outcomes are not very high. Depending on end point and study design, they are reported to be somewhere between 40% and 70% [5,106,107]. This limited concordance has driven substantial interest within the pharmaceutical industry to develop new molecular and cellular methods for DILI prediction. In toxicology, such approaches are now referred to as NAMs, yet in pharmaceutical research such methods have already been used for decades. They are now a standard component of the drug development workflow.

Building on the previously discussed understanding of DILI mechanisms, a variety of assays have been developed to screen and predict the DILI risks of potential new drug candidates. These include *in vitro* assays, often in high-throughput format, that measure compound interactions as substrates or inhibitors of well-defined enzymes or transporters, such as cytochrome P450 enzymes or BSEP. They also include mitochondrial function assays that track oxygen consumption and membrane potential, as well as general cytotoxicity assays using hepatocytes in formats ranging from simple two-dimensional monolayers to complex three-dimensional cultures. Some assays use microscopic readouts or cutting-edge, high-throughput imaging techniques. However, current preclinical strategies remain, to this day, only modestly predictive [24,25]. No assay or single approach has yet been shown to reliably and sensitively predict DILI risk, which is why drugs continue to fail in clinical studies and post-marketing due to unexpected liver toxicity findings.

Part II: Materials and Methods

Chapter 6 is adapted from:

Geci, R. et al. *Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans.* Arch Toxicol 98, 2659–2676; 10.1007/s00204-024-03764-9 (2024).

Chapter 7 is adapted from:

Edizcan, Z., Schaller, S., Kuepfer, L. & Geci, R. *High-throughput PBK modelling for dermal exposure: a pragmatic approach to predict systemic pharmacokinetics.* Arch Toxicol 100, 2421–2432; 10.1007/s00204-026-04357-4 (2026).

Chapter 8 is adapted from:

Geci, R., Paini, A., Worth, A. P., Kuepfer, L. & Schaller, S. *Potential systemic availability classification of chemicals for safety assessment.* Environment international 202, 109636; 10.1016/j.envint.2025.109636 (2025).

Chapter 9 is adapted from:

Geci, R., Sayin, A. Z., Schaller, S. & Kuepfer, L. *Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury.* Arch Toxicol 100, 2029–2046; 10.1007/s00204-026-04305-2 (2026).

Author contributions:

René Geci wrote these chapters and created the figures and tables.

6. Methods for Intravenous and Oral PBK Modelling in Chapter 10

Intravenous and oral PK data extraction

We extracted all concentration-time data from the PK-DB [108], CvT-DB [109] and OSP Observed Data Repository [110] databases and filtered for healthy, adult human subjects. We excluded all studies of special populations including paediatric (< 17 years), geriatric (> 60 years) or diseased populations, since we expected they might display altered pharmacokinetics. A few selected conditions, like “migraine”, “psychiatric disorder”, “asthma”, “diabetes” and “liver disease (minimal)” were permitted. But “liver disease (severe)” or “renal disease (end-stage)” were excluded. Filtering for profiles of administered parent compounds was performed, discarding data of metabolites, and only selecting data measured in blood, plasma or serum, and only after single dose intravenous or oral administration. We excluded studies with concomitant administrations of other drugs to avoid data from drug-drug interaction studies, however, “control” or “placebo” groups from these studies were included. Additionally, we excluded endogenous compounds like ethanol or glucose. Where information about both nominal dose and salt-free active pharmaceutical ingredient (API) was provided, we used the salt-free API value as the dose value. Where only a nominal dose was given, we assumed that this dose was corresponding to the salt-free API dose. Labels of formulations for oral administration (“tablet”, “capsule”, “solution”, etc.) were extracted as provided.

Following the retrieval of concentration-time profiles from previously mentioned databases, we further enriched the dataset by manually digitising additional concentration-time data from the public literature. We started by adding available data for compounds of relevance to the EU-funded ONTOX project [111], searching the literature for corresponding PK studies, for example using PKPDAI [112]. Next, we did the same for compounds of which we had only obtained a single PK profile before. From the retrieved literature, we manually extracted concentration-time profiles using WebPlotDigitizer [113] (version 4.6; https://automeris.io/WebPlotDigitizer). We applied the same exclusion criteria as described before, aiming for data from healthy adults. In some cases, for certain compound classes like chemotherapeutics, however, it was not possible to find PK data for healthy adults or subjects younger than 60 years. As we were able to check those studies manually to ensure that they were not performed on special populations due to assumed differences in pharmacokinetics, we occasionally decided to include them, if no other data on such compounds was available.

Finally, we standardised and curated the extracted PK data. We excluded data points with concentration values that were numerically close to zero (< 1e-7) after 4 h, because some studies had unreasonably low concentration values at the end of their concentration-time profiles. We assumed that those probably represented the lower limit of quantification or were caused by digitisation inaccuracies. We excluded all PK profiles with less than two time points. In cases where we had originally retrieved more than 30 concentration-time profiles for a single compound and a given route of administration, we excluded all profiles from individuals (where number of subjects stated as 1), as long as there were at least 20 other concentration-time profiles left. If the number of profiles was then still higher than 30, we excluded the studies with the least number of subjects until we had no more than 60 concentration-time profiles per study and route. To check the consistency of the retrieved data, we generated dose normalised plots for all compounds where multiple concentration-time curves were available to visually check that studies

were consistent with each other and that there were no extreme outliers among the data. When we found noticeable differences between concentration-time profiles for the same compound and route, we examined the corresponding original publications and either corrected or excluded those entries.

Compound property data retrieval

For the evaluation of various PBK model parameterisation strategies, we collected or predicted the necessary input properties of the compounds of which we had obtained PK data. The minimal input properties for model parameterisation in PK-Sim are a compound's lipophilicity, pKa values, plasma or hepatic intrinsic clearance and its fraction unbound (Fu) [39]. The simulation of oral administration also necessitates data for the solubility and intestinal permeability of substances.

We predicted lipophilicity using six different *in silico* tools: LogP by OCHEM (https://ochem.eu/home/show.do) and VEGA ALOGP [114,115], LogD values by ADMETLab 2.0 [116] and ADMETPredictor version 11.0.0.3 (here called "SimPlus"; https://www.simulations-plus.com/software/admetpredictor), and predictions from Bayer's inhouse models of LogD and LogMA [60]. We further converted LogP or LogD values to LogMA-type values using regression relationships taken from Yun and Edginton [117], Endo et al. [118], and Loidl-Stahlhofen et al. [119]. pKa values were predicted using ChemAxon [120] and ADMETPredictor, aqueous solubility values using four *in silico* tools named ProtoPRED (https://protopred.protoqsar.com/), OPERA [121], ADMETPredictor, and ADMETLab 2.0. Further, Fasted State Simulated Intestinal Fluid (FaSSIF) and Fed State Simulated Intestinal Fluid (FeSSIF) solubility was also predicted using ADMETPredictor.

Intestinal permeability was predicted as CACO2 values using ADMETLab 2.0, ProtoPRED and OPERA, as well as MDCK permeability using ADMETLab 2.0 and ADMETPredictor. Fu values were predicted using seven different *in silico* tools: ADMETLab 2.0, ADMETPredictor, pkCSM [68], Watanabe et al. [66], OPERA, VEGA logK and VEGA CORAL [122]. Plasma clearance values were predicted using ADMETLab 2.0, pkCSM and the ScitoVation clearance tool (https://scitovation-testing.shinyapps.io/clearancetoolgui). Hepatocyte CLint values were predicted with OPERA, ADMETPredictor and ADMET-AI [69].

For comparison, we additionally collected experimental measurements of solubility, Fu, and clearance from public sources if available. In particular, aqueous solubility (LogS) data were retrieved from Hughes et al. [123], and the CompTox Chemicals Dashboard [124]. Fu values were collected from Tonnelier et al. [125], Yamazaki and Kanaoka [126], Lombardo et al. [127], Riley et al. [128], Sohlenius-Sternbeck et al. [129], Votano et al. [130], Zhu et al. [131], from the CompTox Chemicals Dashboard and the R httk package version 2.2.1. We also retrieved intrinsic hepatocyte clearance values from the R httk package [132] (version 2.2.1), as well as fitted PK-Sim intestinal permeability values from Willmann et al. [133]. Eventually, we collected *in vivo* observed plasma clearance values from literature, preferentially from IV studies, except for compounds for which no IV studies were available. When multiple values of a property were acquired for the same compound, average values were used. When experimental values of a compound were found in the previously described *in silico* tool's training set, those values were also added to the retrieved experimental data.

High-throughput PBK simulation and evaluation

PBK model simulations were performed using the standard whole-body PBK model of PK-Sim version 11.1.137 [134] which was executed from R using the ospsuite package version 11.1.143 and R version 4.2.2 [135]. For every from the literature retrieved PK study, a corresponding PBK model simulation was run by parameterising a generic PBK model with a) all compound-specific parameter values as given by the tools of the different parameterisation strategies, b) the study-specific parameters like route of administration, dose and infusion duration (for IV) or formulation (for PO), if provided in the PK data, and c) demographic parameters like sex, age, weight, and height of subjects, if given in the PK data. When demographic parameters were not available, PK-Sim default settings (healthy adult male) were used for simulation.

Predicted or measured hepatic intrinsic clearance (CLint) values were scaled to *in vivo* liver clearance as

In vivo Intrinsic Liver Clearance $\left[\frac{\mu l}{min}\right]$ =

$$CLint\left[\frac{\mu l}{min*10^6\ cells}\right] * Hepatocellularity\ \left[\frac{cells}{g}\right] * Liver\ density\ \left[\frac{g}{ml}\right] * Liver\ volume\ [ml] \quad \text{(eq. 1)}$$

using the same values as Pearce et al. [49] with a hepatocellularity of $1.1x10^8$ cells/g [136], a liver density of 1.05 g/ml [137] and the PK-Sim default liver volume.

For oral simulations, we also set a formulation-specific dissolution parameter (80% dissolution time of Lint80 formulation as defined in PK-Sim), because some oral data was not obtained by administration of drugs in solution but, for instance, in form of capsules or tablets which are not dissolved immediately upon administration. The values used for this were 25 min for “capsule” and 40 min for “tablet” formulations.

After simulating PBK models, simulation results and observed concentration-time data of corresponding PK studies were compared against each other and different model performance metrics were calculated to quantify PBK model quality: Log2-fold changes of predicted/observed values of Cmax, Tmax, AUC 0-last calculated using the linear up-log down method of the PKNCA package version 0.10.0 [138], as well as the percentage of datapoints within different fold ranges (1.5/2/3/5/10-fold). Further, we evaluated the goodness of the entire concentration-time profile prediction against all datapoints in the *in vivo* PK profile calculating the Relative and Absolute Log2 Errors as

Relative Log2 Error =

$$\frac{1}{N} * \sum_{i=1}^{N} Log2\left(\frac{Predicted\ value_i}{Observed\ value_i}\right) \quad \text{(eq. 2)}$$

Absolute Log2 Error =

$$\frac{1}{N} * \sum_{i=1}^{N} \left|Log2\left(\frac{Predicted\ value_i}{Observed\ value_i}\right)\right| \quad \text{(eq. 3)}$$

for all N datapoints in every concentration-time profile such that the Relative Log2 Error captures biases in prediction, i.e., over- or underestimates (systematic error), while the Absolute Log2 Error quantifies overall closeness of PBK model simulations to the observed concentration-time values (random error).

In cases where we had obtained multiple concentration-time profiles of a single compound, we selected the median performance value of all studies as the representative value for the substance to ensure that individual outlier studies would not

distort our results. To later summarise the overall performance of full parameterisation strategies, we followed the same approach to integrate the results of the various compounds in our dataset, so that when we refer to the Median Log2 Error, this means the median of the Log2 Error values of all the compounds in the evaluation dataset.

7. Methods for Dermal PBK Modelling in Chapter 11

Dermal PK data and physicochemical property data retrieval

We performed a literature search to identify clinical studies in which plasma PK data was measured in healthy adult humans after single dermal administration. Studies were excluded if substances were applied to either wounds, irritated skin, or if the compounds were highly volatile or prodrugs, or if the substances were applied under repeated dosing. Only data for parent compounds was included, and only if compound concentrations were measured in plasma or serum. Concentration-time profiles were extracted manually from the retrieved literature studies and digitised with WebPlotDigitizer version 5.2 [139].

QSAR models were applied to predict physicochemical properties of all substances for which we had retrieved PK data. The minimal properties required for modelling dermal absorption using the OSP Suite model were lipophilicity, solubility, melting point, boiling point, and vapour pressure [39]. For lipophilicity predictions, we used LogD values predicted by ADMETLab 3.0 [61], ADMETPredictor (SimPlus; https://www.simulations-plus.com/software/admetpredictor), and OPERA [121], and LogP values predicted by ADMETLab 3.0, SimPlus, VEGA [115], and OPERA. Solubility values were predicted using ADMETLab 3.0, SimPlus, VEGA, and OPERA. For melting point and boiling point, predictions were made using OPERA and ADMETLab 3.0. Lastly, for vapour pressure, we obtained predicted values from OPERA and VEGA. Those PBK model parameters which were not related to dermal absorption, such as clearance and fraction unbound, were parameterised as reported previously (“Best *in silico*-free" strategy) [140] to enable comparison of results to previous reports.

Software and model simulations

To simulate compound permeation across and within different layers following dermal administration, we used the skin permeation model of the OSP Suite [141] (https://github.com/Open-Systems-Pharmacology/Skin-permeation-model). The model depicts the stratum corneum, viable epidermis, and dermis as diffusion-partition layers governed by Fick’s laws of diffusion and partition equilibria, following the biophysical framework described by Kasting and Miller [142]. This mechanistic skin model formed the basis of all HT-PBK simulations performed in this study and was coupled with the standard whole-body PBK model implemented in PK-Sim version 11.3 [134] to simulate systemic PK. For automated PBK model simulation, we executed models from R using the ospsuite package version 11.3.529 and R version 4.3.1 [143].

For each dermal PK study collected from the literature, a corresponding PBK model simulation was performed by using the predicted physicochemical properties together with study-specific parameters such as the formulation of the vehicle, application area, or dosage. When demographic parameters were not provided in the literature, we assumed healthy adult males as default settings. Skin thickness values were obtained from studies in literature where histological sectioning, Raman spectroscopy, or ultrasound imaging were used to determine stratum corneum, viable epidermis, and dermis thicknesses at different anatomical sites [144–149]. Because dermal administration studies often provided no application-related study conditions, we assumed healthy skin conditions with a skin surface temperature of 30 °C and fully hydrated stratum corneum. Finally, for non-occluded vehicles, we assumed a wind velocity of 16.8 cm/s.

Predictive performance analysis

Predictive performance of PBK models was quantified by comparing the results of PBK model simulations against the observed systemic concentration-time data of the corresponding PK studies. We determined the maximum concentration (Cmax) value of both simulation and observed data, as well as the area under the curve (AUC) of the systemic PK profiles using the linear up-log down method of the PKNCA package version 0.11.0 [138]. Eventually, we calculated the Mean Absolute Log2 Error of Cmax and AUC values of all substances in the dataset to determine the precision of different prediction strategies as

$$\text{Mean Absolute Log2 Error} = \frac{1}{\text{N}} * \sum_{\text{i}=1}^{\text{N}} \left| Log2 \left(\frac{\text{Predicted value}_\text{i}}{\text{Observes value}_\text{i}} \right) \right| \quad \text{(eq. 4)}$$

for all N compounds. Similarly, we calculated the Median Relative Log2 Error of Cmax and AUC values of all substances to evaluate potential biases of different prediction strategies as

$$\text{Median Relative Log2 Error} = \text{Median} \left(Log2 \left(\frac{\text{Predicted value}_\text{i}}{\text{Observes value}_\text{i}} \right) \right) \quad \text{(eq. 5)}$$

For those compounds for which multiple concentration-time profiles had been reported, the median of values across different clinical studies was used to ensure that outliers from individual studies did not disproportionately affect the analysis.

8. Methods for ADME-Based Assessment of Chemical Safety in Chapter 12

Example compound simulations

To give five illustrative examples, identical physiologically based kinetic (PBK) models based on the same generic hepatic clearance process, the PK-Sim Standard partitioning and permeability methods were generated in PK-Sim version 11.1.137 [110,134]. For simplicity, it was assumed that five example compounds have identical molecular weights of 300 g/mol, lipophilicity of 1 and a fraction unbound of 50%. Solubility, intestinal permeability and clearance values of compounds were varied as shown in Table 2. Then, daily oral administration of solutions of different doses ($1x10^{-4}$ to $1x10^{3}$ mg/kg) were simulated for 5 years, and the maximum plasma concentration for each dose was retrieved. The oral bioavailability of the example compounds was predicted by using the standard functionality of PK-Sim, relying on the generated PBK models.

HT-PBK modelling of EPAA NAM Designathon compounds

High-throughput PBK (HT-PBK) simulations of EPAA NAM Designathon compounds were generated based on previous model validations [140,150]. All compound properties required for PBK modelling [39] were generated using the following *in silico* tools: ADMETLab LogD, LogS and fraction unbound [61], ADMET Predictor v12 Fasted State Simulated Intestinal Fluid solubility and intrinsic liver clearance (https://www.simulations-plus.com/software/admetpredictor), OPERA intrinsic liver clearance [121]. Where available, we also collected *in vitro* measured fraction unbound and hepatocyte clearance values of compounds from httk 2.4.0 [49,53] for comparison. For each compound, the default whole-body PBK model implemented in PK-Sim version 11.1.137 was parameterised from R version 4.2.2 [135], and administration of 1 µg/kg bodyweight once daily oral solutions were simulated. From those simulations the total amount of compound in the body after five years was calculated as a readout. For the provision of partitioning coefficients, the method of Berezhkovskiy [151] was applied. Peak Systemic Availability Factor (PSAF) and Mean Systemic Availability Factor (MSAF) values were calculated as shown in equations 9 and 10, respectively.

Manual investigation of EPAA NAM Designathon compounds

We searched for the 150 EPAA NAM Designathon compounds on publicly available data that could be indicative of their toxicokinetic properties, for which we included data from *in vitro*, animal and human studies, as well as potentially relevant regulatory classifications such as persistent organic pollutant (POP), persistent bioaccumulative toxic (PBT), very persistent very bioaccumulative (vPvB), and substance of very high concern (SVHC). Additionally, we gathered molecular properties like molecular weight, predicted water solubility and logP values of these compounds using the above mentioned *in silico* prediction tools.

After having collected potentially relevant data, we classified all endogenous substances, such as amino acids and sugars, as medium PSA concern. All compounds which did possess some TK data but did not exhibit any signs of meaningful accumulation were also classified as medium concern. Substances which did exhibit signs of relevant accumulation over repeated exposures were classified as high PSA concern. Due to the

highly heterogeneous nature of the gathered TK data, which was measured in different species and tissues and using different administration protocols, it was not feasible to define any absolute criteria for what constitutes “meaningful” accumulation. Instead, this had to be determined on a case-by-case-basis by experts who holistically considered all available information. Then, in the absence of additional information, all remaining compounds without TK data and that fell into the “small molecule” space were also classified as medium PSA concern. The two compounds which were out of this domain, both due to large molecular weights above 1500 g/mol, were classified as low PSA concern.

9. Methods for Drug-Induced Liver Injury Prediction in Chapter 13

Extraction of DILI drug data

We conducted a literature search for studies reporting *in vitro* hepatotoxicity assay measurements for the prediction of DILI risks of drugs, and that had provided quantitative toxicity potency values (e.g., IC50 concentrations) rather than binary active/inactive readouts. All retrieved measurements were combined into a single dataset by comparing the various compounds against each other via canonical SMILES identifiers obtained from PubChem [152]. Salt ions were stripped from structures prior to matching which was performed based on the presumed active pharmaceutical ingredient. In total, *in vitro* hepatotoxicity data were compiled from seventeen *in vitro* hepatotoxicity datasets [153–169]. Compounds were further matched against the FDA-approved DILIrank [170] database to obtain clinical DILI risk classifications. While some of the included studies already reported *in vivo* Cmax values of drugs, these did not always align across sources. Therefore, we manually collected information on the highest recommended single oral dose and corresponding *in vivo* Cmax values from clinical studies for each drug. Drugs without available clinical DILI risk information or without a single oral dose Cmax were removed from the dataset. For cases in which multiple *in vitro* potency values were available for the same mechanistic endpoint (cytotoxicity, mitochondrial toxicity), the lowest (strongest potency) value was selected, assuming that assays and cell systems varied in their sensitivity to detect perturbations of the endpoint. For BSEP inhibition, where assays target the same endpoint with comparable sensitivity, multiple IC50 values were integrated by using the median, assuming that differences primarily reflected technical variability.

To specifically investigate drugs associated with cholestatic DILI, we generated a dedicated database, as no resource comparable to DILIrank exists for cholestatic hepatotoxicity. For this purpose, we integrated classifications from the SIDER adverse drug reactions database [171] and additional publications reporting DILI patterns [172–177]. When sources consistently classified a drug as cholestatic, hepatocellular, safe, or agreed on its pattern, those classifications were adopted. In cases where different sources disagreed, the compound was labelled as “contradiction.” Drugs that had failed during clinical development were labelled accordingly, since they would not be represented in the databases used. All remaining compounds for which no classification could be established were labelled as “unknown.”

High-throughput PBK prediction of drug concentrations

High-throughput physiologically based kinetic (HT-PBK) simulations of drugs were performed based on previous model validations [140,150]. All drug properties required as input for PBK simulations [39] were predicted using the following evaluated *in silico* tools: ADMETLab LogD, LogS, fraction unbound, CACO2 permeability [61], ADMET Predictor v12 Fasted State Simulated Intestinal Fluid solubility (https://www.simulations-plus.com/software/admetpredictor), and PKsmart clearance [178]. For each substance, the default whole-body PBK model implemented in PK-Sim [134] version 11.1.137 was parameterised from R version 4.2.2 [135], and oral administration of the highest extracted oral dose was simulated. As partitioning coefficients, the method PK-Sim Standard [179] was used and simulated Cmax values were extracted as model readouts.

In order to account for the fact that the dissolution times of the different drug formulations were unknown for retrospective analysis, while they would be known for prospective evaluations of a given drug formulation, we performed PBK simulations with different Lint80 dissolution times (1-180 min) and eventually selected the simulated Cmax values of drugs closest to their observed values.

Simulating bile acid perturbations

In order to describe the dynamic effects of the inhibition of BSEP by drugs, and the following changes in intrahepatic bile acid levels, we used a previously reported physiologically based bile acid (PBBA) model [180]. In this mode, drug interaction with BSEP is depicted as competitive inhibition in the BSEP-mediated transport rate as

$$V = \frac{Vmax*[S]}{Km,app+[S]} \qquad \text{(eq. 6)}$$

with

$$K_{m,app} = Km*(1+\frac{[I]}{Ki}) \qquad \text{(eq. 7)}$$

Where V is the velocity of the BSEP-mediated transport, [S] is the bile acid substrate concentration, V_{max} is the maximal BSEP transport velocity, [I] is the inhibitor concentration and K_m the Michaelis-Menten constant of the transported bile acid substrates.

To parameterise eq. 7, we converted the reported IC50 values of drugs to Ki values using the equation [181]

$$K_i = \frac{IC50}{1+\frac{[S]}{Km}} \qquad \text{(eq. 8)}$$

where K_i is the inhibition constant of drugs, IC50 is the measured half-maximal inhibition concentration and K_m is the Michaelis-Menten constant of the substrates used as *in vitro* probes. To represent the dynamic changes in drug concentrations, we used the plasma concentration-times profiles of substances that were simulated using the HT-PBK method. Since bile acid model simulations were used for retrospective analysis, we used *in vivo* observed [182] plasma clearances instead of predicted values if available. As readouts of the bile acid model simulations, we extracted the concentration of intrahepatic total conjugated bile acid concentration and evaluated the duration of elevation of this concentration above twofold, relative to the unperturbed state, as well as the duration of elevation to more than 10% above the unperturbed peak concentration.

Part III: Results

10. High-throughput PBK Prediction of Intravenous and Oral Drug Exposure

Abstract

Physiologically based kinetic (PBK) modelling provides a mechanistic framework for predicting the pharmaco-/toxicokinetics of compounds and is central to integrating exposure and toxicity data in non-animal safety assessment. However, conventional PBK models typically depend on animal and human data for parameterisation and validation, which limits their suitability for new approach methodologies (NAMs). High-throughput PBK (HT-PBK) modelling seeks to overcome this limitation by generating models exclusively from *in vitro* and *in silico* sources. Here, we systematically evaluated different strategies for PBK model parameterisation using a combination of publicly available compound-specific input values and a large dataset of human *in vivo* pharmacokinetics. We collected more than 2,000 plasma concentration-time profiles for over 200 compounds after intravenous and oral administration, along with corresponding *in silico*, *in vitro*, and *in vivo* parameter data. Using the open-source PBK modelling software PK-Sim, we assessed the predictive accuracy of all parameterisation strategies against the *in vivo* dataset. Our results show that even simple HT-PBK approaches can achieve good predictive performance, with 87% of maximum plasma concentrations (Cmax) and 84% of areas under the curve (AUC) predicted within a tenfold error. Nonetheless, we also observed substantial variability across parameterisation strategies and between individual compounds. Based on these findings, we propose a workflow for HT-PBK modelling that relies solely on freely available tools. This work provides a systematic evaluation of HT-PBK approaches, offering guidance on their reliability and limitations, and contributes to establishing the confidence necessary for their use in next generation risk assessment.

Chapter 10 is adapted from:

Geci, R. et al. *Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans.* Arch Toxicol 98, 2659–2676; 10.1007/s00204-024-03764-9 (2024).

Author contributions:

Domenico Gadaleta, Marina García de Lomana, Rita Ortega-Vallbona, Erika Colombo, and Eva Serrano-Candelas provided *in silico* predictions of molecule properties. Alicia Paini, Lars Kuepfer and Stephan Schaller provided overall supervision and guidance. René Geci performed all remaining analyses, wrote these chapters and created the figures and tables.

10.1 Introduction

Pharmacokinetics (PK) is the quantitative study of how compounds are absorbed, distributed, metabolised, and excreted (ADME) in the body. By describing the time course of compound concentrations in plasma and tissues, PK provides the essential link between chemical exposure and biological effect and thereby forms the basis for understanding therapeutic efficacy and safety margins. In drug development, PK characterisation is indispensable for selecting promising candidates, designing dosing regimens, and anticipating potential drug-drug interactions. In toxicology, it enables the extrapolation of *in vitro* or animal findings to human exposure scenarios and supports chemical risk assessment.

Among the various administration routes, intravenous (IV) and oral (PO) administration are of particular importance. During IV administration substances are delivered directly into systemic circulation, which allows the investigation of substance distribution and elimination processes in the absence of the confounding influences of absorption. This is why IV data serves as the reference for determining fundamental PK parameters such as clearance (CL), volume of distribution (Vd), and bioavailability (F). Oral administration, by contrast, is the predominant route not only for therapeutic drug use but also for human exposure in chemical risk assessment. It is, however, inherently more variable, as it is influenced by gastrointestinal absorption processes, first-pass metabolism, and solubility limitations. Together, IV and PO PK data capture the key determinants of systemic exposure. This is why they are the most regularly investigated routes of administration and why they are central to both pharmacological and toxicological evaluation.

Traditionally, PK characterisation has relied on experimental measurements in animals and humans. While physiologically based kinetic (PBK) modelling and classical compartmental modelling provide mechanistic and quantitative frameworks, their application is often limited by the need for extensive compound-specific *in vivo* and *in vitro* data, as well as manual model generation and validation. As a result, conventional PK prediction approaches remain relatively low-throughput, costly, and constrained by ethical and regulatory pressures to reduce animal use.

To overcome these limitations, various alternative methods are being explored. High-throughput *in vitro* assays now provide rapid measurements of compound-specific ADME parameters, which can either be incorporated into PBK models and empirical predictions or that can be used directly for compound triaging [183,184]. In parallel, quantitative structure-activity relationship (QSAR) and other *in silico* models allow the estimation of physicochemical and ADME properties directly from chemical structure [60,115]. More recently, artificial intelligence (AI) and machine learning (ML) approaches have been developed to predict substance PK, either by directly estimating summary exposure metrics such as maximum concentration (Cmax) or area under the curve (AUC) [178,185–188], or by predicting mechanistically relevant parameters that can be used as PK model inputs [189–195]. Some of those hybrid approaches also integrate AI with PBK modelling, combining data-driven prediction with mechanistic simulation [18,196].

Despite these advances, important gaps remain. Most importantly the validation of novel PK prediction approaches often remains insufficient. Most validation studies have been limited to evaluating PK predictions using small datasets, were focused only on rodent PK or relied on summary PK parameters [50,72,188,197–199]. Unfortunately, such summary metrics can obscure mechanistic inaccuracies and fail to capture the full temporal dynamics of substance disposition. The reason for this lack of meaningful

validation is, presumably, a lack of high-quality human PK data. Rigorous evaluation of PK prediction requires systematic benchmarking against large human datasets that should ideally include full concentration-time profiles after both IV and PO administration. However, such high-quality PK data do not exist in an easily accessible way and large scale format. The total number of PK studies performed in humans is much smaller than the number of studies performed in animals. And existing data are scattered across in-house databanks of pharmaceutical companies or only exist in non-machine-readable legacy formats. The limited availability of accessible, high-quality human PK data is therefore a key limitation in the development and validation of new PK prediction methods.

The aim of this work is to address this gap by rigorously evaluating already available *in silico* and AI-based approaches for PK prediction against a comprehensive human PK dataset. Specifically, we quantify prediction errors using data that (i) include a sufficient number of structurally diverse compounds, (ii) comprise full time-concentration profiles, and (iii) reflect human PK after IV and PO administration. To this end, we compiled a large dataset of human *in vivo* concentration-time profiles after intravenous (IV) and oral (PO) dosing (210 compounds, 2235 concentration-time profiles). We then systematically assessed the predictive performances of PBK models when parameterised with the different *in silico* tools and compared these results to *in vitro* and *in vivo* benchmark references. This allowed us to identify the *in silico* tools that provide the best input parameters for PBK modelling, as well as to validate the overall accuracy of such high-throughput PBK (HT-PBK) models for predicting human IV and PO PK profiles.

10.2 Results

After retrieving PK data from databases and the literature, we had initially collected 2235 healthy human adult *in vivo* concentration-time profiles, covering 210 unique compounds (Figure 5A). For each compound, we compiled the input data required for PBK model parameterisation, including properties such as lipophilicity and both predicted and measured plasma clearance values. This resulted in a final dataset in which all compounds possessed the required input data comprising 718 IV profiles (143 compounds) and 1402 PO profiles (169 compounds). Some parameters were less widely available, however: *in vitro* measured intrinsic hepatic clearance (CLint) values, retrieved from the httk database, were only available for 97 compounds, and measured solubility values were available for 167 compounds. The final dataset therefore included 182 unique compounds, all small molecules with molecular weights below 900 Da, spanning different ionisation states at physiological pH. The majority of these belonged to ECCS class 2, consistent with metabolism being the dominant clearance mechanism (Figure 5B).

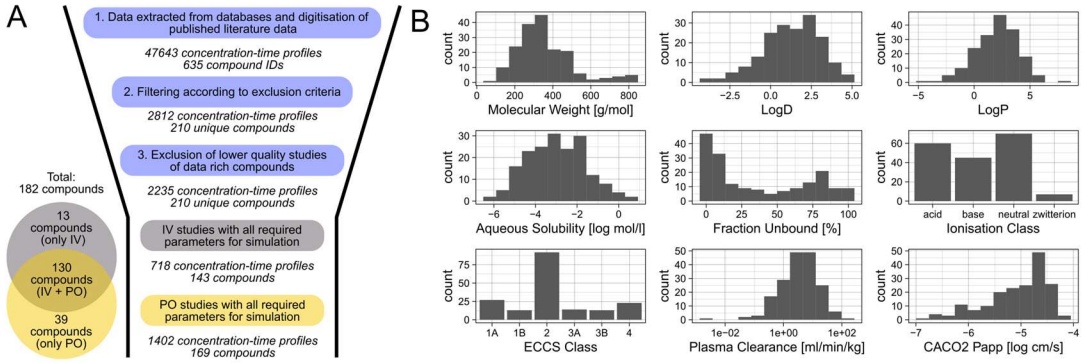


Figure 5: Summary statistics of the collected IV and oral PK dataset and of various compound properties. (A) Overview of the number of concentration-time profiles and compounds during and at the end of the PK data retrieval process. (B) Physicochemical and pharmacokinetic properties of compounds in the retrieved PK dataset. LogD, LogP, aqueous solubility and CACO2 permeability are given as the mean of values predicted *in silico* by the different prediction tools used in this study. Plasma clearance values are the median of *in vivo* values collected from the literature. Ionisation at physiological pH (7.4) was predicted using ChemAxon. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Evaluating the performance of PBK models parameterised with different input data sources against a large *in vivo* PK dataset required extensive simulations and resulted in long computation times. To keep the analyses systematic yet computationally feasible (15+ million simulations), we divided the evaluation of HT-PBK parameterisation strategies into three sequential steps. This design allowed us to assess individual parameterisation decisions one at a time while limiting the number of simulations (Figure 6).

In the first step, we focused on the prediction of passive distribution. For all compounds with available IV data, we generated, simulated, and evaluated every PBK parameterisation strategy, while constraining variability by using only *in vivo* plasma

clearance values and *in vitro* measured fraction unbound (Fu) values as high-quality benchmark references. In the second step, we again used the IV data but restricted the input to the best-performing physicochemical parameter sources identified in step one. On this basis, we systematically tested different Fu and clearance prediction tools to determine which combinations yielded the most accurate PK predictions. In the third step, we extended the evaluation to oral PK data, using the optimal physicochemical, Fu, and clearance sources identified in the previous steps. We then systematically varied solubility and intestinal permeability values to evaluate oral absorption predictions and to assess how well the resulting high-throughput strategies performed overall.

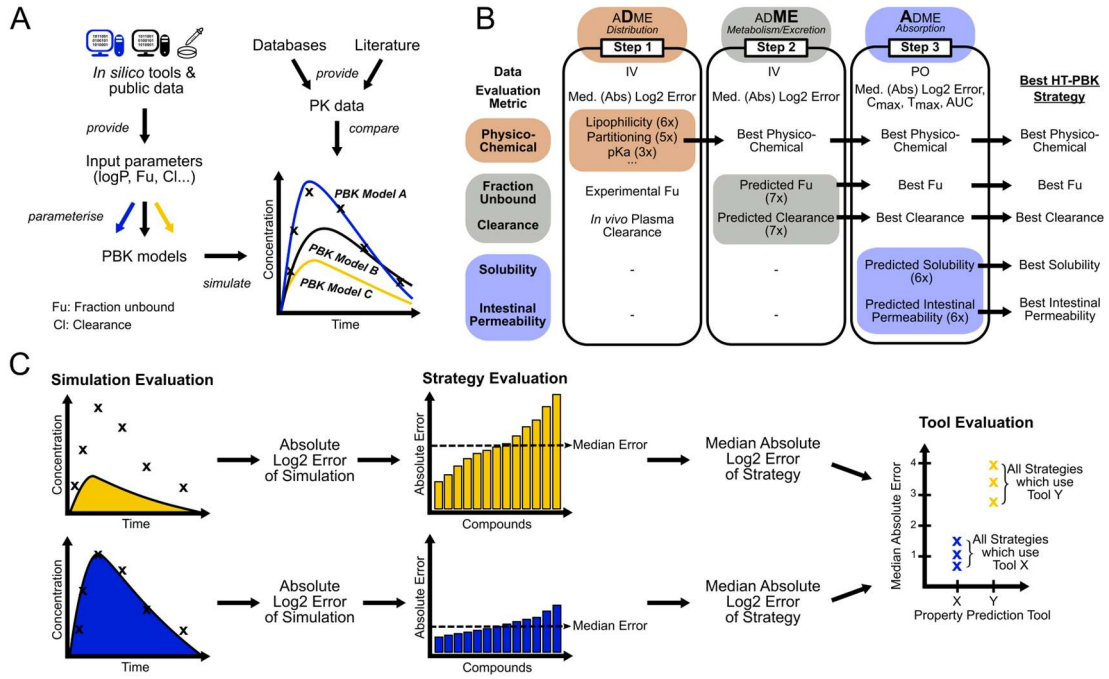


Figure 6: Conceptual overview of the simulation and analysis strategy. (A) For each compound various PBK models were generated using all combinations of available input parameter sources. Their performance was then compared against gathered concentration-time profiles. (B) The simulation and evaluation of PBK models built from the various parameterisation sources was split into three steps, each focusing on the evaluation of a separate level of ADME processes. (C) Model performance was first evaluated at the level of individual simulations as the difference between model simulation and observed data. Then, for every parameterisation strategy, the median error of all compounds was used as a measure of the performance of that parameterisation strategy. Finally, comparing all performances of all strategies using the same input property prediction tools allowed benchmarking of the various tools against each other. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Step 1: Performance of physicochemical property predictions

In step one, we systematically evaluated strategies for setting the physicochemical PBK model parameters that govern passive distribution of compounds within the body. In PK-Sim, these parameters are primarily compound lipophilicity, pKa values, and the choice of partitioning method.

Lipophilicity values were predicted using six *in silico* tools: three LogD tools (SimPlus, ADMETLab, Bayer), two LogP tools (OCHEM, VEGA), and one LogMA tool (Bayer). pKa values were obtained from ChemAxon and SimPlus, and we additionally tested the effect of omitting pKa values altogether, thereby assuming compounds to be neutral. Partitioning was assessed using the five available methods in PK-Sim: PK-Sim [179], Schmitt [200], Rodgers and Rowland [41], Poulin and Theil [201], and Berezhkovskiy [151].

To minimise variability at this stage, *in vivo* plasma clearance and experimentally measured Fu values were fixed as high-quality benchmark inputs, ensuring that remaining simulation inaccuracies reflected only the prediction of passive distribution. We then systematically tested all combinations of physicochemical input sources against the 718 collected IV concentration-time profiles. Model performance was evaluated by calculating median relative log2 error and median absolute log2 error, representing prediction bias (systematic error) and precision (random error), respectively.

Among the tested parameters, the choice of lipophilicity input had the strongest influence on PBK model accuracy (Figure 7). Models based on LogP predictions performed consistently worse than those parameterised with LogD or LogMA values. Even within the same lipophilicity category, performance differed across tools: for example, LogD values from the Bayer predictor outperformed those from ADMETLab or SimPlus (ADMETPredictor). Higher errors with LogP-based predictions were associated with a systematic underprediction of PK data. Compound-level analysis (Figure 7E-J) revealed that in some cases, LogP tools predicted very high lipophilicity values (>5), which in turn led to severe underprediction of plasma concentrations, whereas the same compounds were predicted more accurately when using LogD or LogMA inputs.

The results for partitioning methods were less straightforward. A consistent hierarchy in predictive performance was observed, with the Berezhkovskiy method performing best and the Schmitt method worst under most conditions (Figure 7A). However, this ranking became clear only when lipophilicity values from the poorer-performing LogP tools were used. With better-performing inputs (LogD and LogMA from Bayer), differences in prediction precision across partitioning methods were marginal (Figure 7C). Compound-level analyses (Supplementary Figure 1) showed that Poulin & Theil and Berezhkovskiy were not intrinsically more predictive for most compounds. Instead, their apparent superiority stemmed from greater robustness to the extreme lipophilicity values predicted by LogP tools (Supplementary Figure 2). By contrast, other partitioning methods were more strongly affected by these outliers, which reduced their overall performance (Supplementary Figure 3).

For predicted pKa values, no consistent trend was observed. Although one might expect partitioning methods that incorporate pKa to perform better when these values are provided, this was only reliably observed for the Rodgers & Rowland method, and only when used with LogP lipophilicity inputs (Supplementary Figure 4). For the other methods that use pKa (Schmitt, Poulin & Theil, Berezhkovskiy), predictive performance varied depending on the lipophilicity tool used.

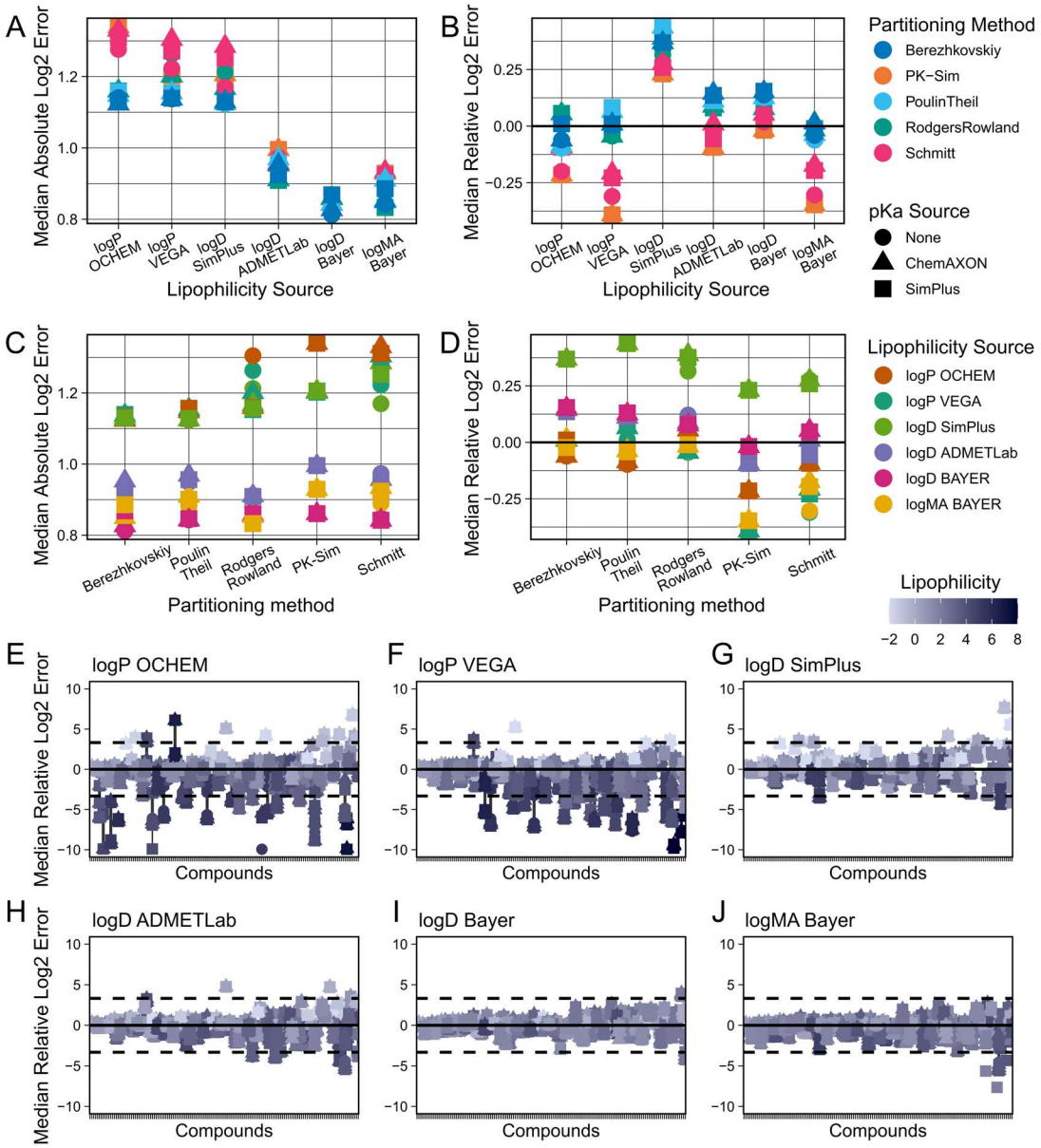


Figure 7: Comparison of predictive performances of different physicochemical parameter sources (step 1). Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* benchmark reference values, respectively. The top row shows Median Absolute Log2 Errors (A) and Median Relative Log2 Errors (B) for different lipophilicity prediction methods. The middle row shows Median Absolute Log2 Errors (C) and Median Relative Log2 Errors (D) for different partitioning methods. The bottom rows (E-J) show the Relative Log2 Errors of predictions for every individual compound for the different lipophilicity prediction tools used. Dashed lines indicate tenfold errors. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Finally, we tested two strategies to further improve lipophilicity inputs for PBK modelling. The first involved generating consensus values across prediction tools by averaging their outputs. For LogP, this was done by taking the mean of all LogP predictions; the same was done for LogD tools. Interestingly, opposite effects were observed (Supplementary Figure 5). Averaging LogP predictions improved PBK model performance compared with any individual LogP tool. In contrast, averaging LogD predictions produced intermediate results: better than the weakest tool (SimPlus) but worse than the stronger ones (Bayer, ADMETLab).

The second strategy was to convert LogP or LogD predictions into membrane affinity values (LogMA) using regression equations. We applied two published equations for converting LogP [117,118] and derived a comparable equation for LogD values based on the dataset of Loidl-Stahlhofen et al. [119]. Of these, only the Yun et al. equation [117] consistently improved PK predictions, regardless of the LogP tool or partitioning method used (Supplementary Figure 6). The other two approaches gave mixed results, and in the case of our LogD-based equation, even reduced predictive accuracy. These findings suggest that the inferior performance of LogP-based models may not reflect inherent inaccuracies of the LogP prediction tools themselves but rather the type of lipophilicity descriptor, which is less suitable for PBK modelling in certain compound classes.

Based on these results, we concluded that there was no universally superior partitioning method and that providing pKa values did not consistently improve predictions. However, lipophilicity values predicted by the Bayer tools (LogD and LogMA) yielded superior results compared with other sources. For this reason, we selected the mean of these Bayer predictions as the optimal lipophilicity input and carried forward all partitioning methods and pKa prediction strategies into the next evaluation step, which focused on clearance and Fu prediction tools (Step 2).

Step 2: Performance of clearance and fraction unbound predictions

After assessing the physicochemical properties relevant for passive distribution, we next evaluated tools for predicting parameters that depend on organism biology, specifically Fu and clearance. For validation, we again used the IV PK dataset but restricted input to the best-performing physicochemical predictions identified in Step 1, while systematically varying Fu and clearance sources.

For Fu, we compiled values from seven *in silico* tools as well as *in vitro* benchmark measurements. As expected, the influence of Fu predictions depended on the clearance input. When models were parameterised with *in vivo* plasma clearance benchmark values, differences among Fu prediction tools were marginal (Supplementary Figure 7). In contrast, when *in vivo* clearance was predicted from *in vitro* measured hepatic CLint values, performance varied more strongly across Fu tools (Figure 8AB). Experimentally measured Fu values provided the most accurate PK predictions, confirming their validity as benchmark references. Nevertheless, overall differences between tools were modest: all Fu prediction sources resulted in median absolute errors within a 2-3-fold range when combined with *in vitro* CLint, and none showed a consistent systematic bias toward under- or overprediction.

For clearance, we evaluated three *in silico* tools directly predicting plasma clearance, our benchmark *in vivo* plasma clearance values, as well as *in vitro* measured hepatic CLint values from httk. In addition, we tested *in silico* predictions of CLint obtained from OPERA, SimPlus, and ADMET-AI.

Before comparing clearance prediction strategies, we first assessed whether activating passive renal excretion improved or worsened PBK model performance. Plasma clearance values already reflect the combined effect of all systemic clearance processes, so adding renal clearance on top of them should theoretically reduce accuracy. By contrast, when *in vivo* clearance is scaled up from hepatocyte-derived CLint values, incorporating renal excretion is expected to improve predictions.

Our results were consistent with these expectations. Adding passive renal clearance to *in vivo* plasma clearance reduced prediction quality and shifted model bias from neutral to underprediction, whereas combining it with hepatocyte-based clearance improved predictions by reducing overprediction (Supplementary Figure 8). Results for *in silico* clearance values were less consistent: for example, CLint predictions from SimPlus and ADMET-AI already underestimated PK, and this bias was further exacerbated when renal clearance was added. Based on theoretical and empirical considerations, we therefore included passive renal clearance only in simulations using hepatocyte-derived CLint, but not when plasma clearance values were applied.

As with Fu, *in vivo* plasma clearance values served as the best benchmark input (Figure 8CD). All other clearance prediction strategies performed markedly worse, with nearly all yielding median absolute errors outside the twofold range. Unlike the relatively small differences observed between Fu tools, clearance tools varied substantially in predictive quality. Among the *in silico* plasma clearance tools, ADMETLab performed best, followed by pkCSM and ScitoVation. While ADMETLab and pkCSM tended to slightly overpredict PK, ScitoVation led to strong underprediction. These findings were confirmed by direct comparison of *in vivo* plasma clearance with *in silico* predictions (Supplementary Figure 9), which showed the closest correlation for ADMETLab.

In vitro hepatocyte CLint values from httk were the second-best clearance source after *in vivo* plasma clearance benchmarks. In contrast, CLint predictions from *in silico* tools performed considerably worse. Among these, OPERA gave the best results, followed by SimPlus and ADMET-AI. However, in some cases OPERA reported values identical to httk measurements (Supplementary Figure 10), indicating that not all outputs represented genuine *in silico* predictions, and suggesting caution in comparing OPERA directly with other tools. Overall, when comparing *in silico* plasma clearance predictors with *in silico* hepatocyte CLint predictors, plasma clearance tools generally produced more accurate PK predictions.

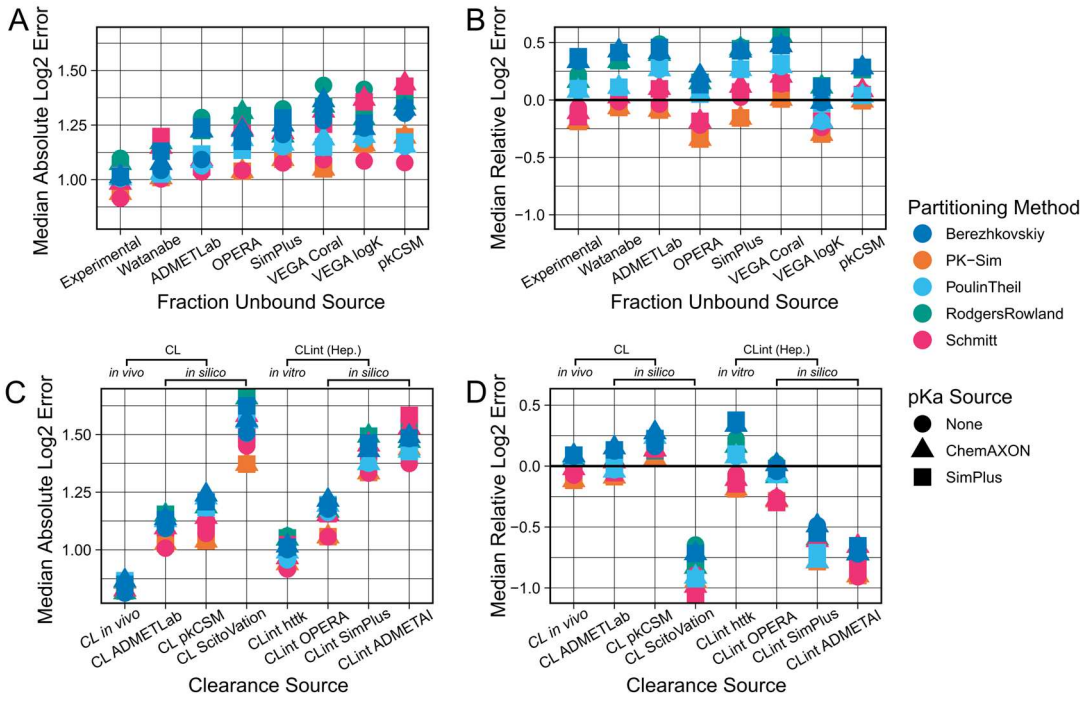


Figure 8: Comparison of predictive performances of different fraction unbound and clearance prediction sources (step 2). Combinations of all available parameterisation sources were evaluated against the collected IV dataset. Results shown were generated using benchmark values for parameterisation of the other parameters, i.e., *in vitro* CLint values from httk when comparing Fu predicting sources, and *in vitro* measured Fu values when comparing clearance predicting sources, as well as the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. The top row shows the Median Absolute Log2 Errors (A) and the Median Relative Log2 Errors (B) for different fraction unbound prediction methods. The bottom row shows the Median Absolute Log2 Errors (C) and the Median Relative Log2 Errors (D) for different clearance prediction methods. CL refers to plasma clearance values (measured or predicted), and CLint signifies hepatocyte intrinsic clearance (measured or predicted). *In vitro* measured CLint values from httk were only available for a subset of compounds (97 out of 143). Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Step 3: Performance of solubility and intestinal permeability predictions

In the third evaluation step, we investigated how to best parameterise PBK models for oral administration. In PK-Sim, the relevant parameters are primarily solubility and intestinal permeability. For solubility, we compiled experimentally measured benchmark values of aqueous solubility as well as predictions from four *in silico* tools (OPERA, ADMETLab, ProtoQSAR, SimPlus). Additional predictions of FaSSIF and FeSSIF solubility were obtained from SimPlus. For intestinal permeability, no benchmark references were available. Instead, we used CACO2 permeability predictions from three *in silico* tools (OPERA, ADMETLab, ProtoQSAR), MDCK permeability predictions from ADMETLab and SimPlus, and predictions generated by PK-Sim's internal equation based on molecular weight and lipophilicity (LogMA Bayer).

For evaluation, we restricted the dataset to PK studies in which the administered formulation indicated the compound was already in solution at dosing (e.g. “solution” or “suspension”), excluding solid formulations such as tablets or capsules that would additionally require knowledge of dissolution kinetics. This yielded 286 concentration-time profiles (94 compounds) from liquid formulations (Figure 9). Across these data, no substantial differences were observed between the parameterisation sources for either solubility or intestinal permeability. For solubility, all *in silico* tools performed comparably to each other and to the experimentally measured benchmark values. Likewise, intestinal permeability predictions from the different tools gave similar results.

We did, however, observe a general trend towards overprediction of absorption velocity, which resulted in consistent underprediction of Tmax values and a slight overprediction of Cmax (Supplementary Figure 11). We hypothesised that this discrepancy arises because the *in silico* tools were trained to predict *in vitro* CACO2 or MDCK permeabilities rather than the PK-Sim-specific intestinal permeability parameter. Under standard procedures, *in vitro* permeability measurements would normally be scaled, e.g. by using reference compounds, before translation into PK-Sim parameters. Only when such reference measurements are unavailable, the *in vitro* permeability values are used directly without scaling, which may explain the observed bias.

To address this, we extracted fitted PK-Sim intestinal permeability values for 56 compounds reported by Willmann et al. [133] and, for each *in silico* tool, derived a scaling factor based on the relationship between its predicted values and these assumed optimal parameters. As expected, CACO2 permeability values tended to be higher than the optimal PK-Sim intestinal permeability values (Supplementary Figure 12). However, applying the scaling factors did not substantially improve overall PK predictions (Supplementary Figure 13), although the adjustment did reduce the bias observed in the underprediction of Tmax.

Finally, we assessed whether the conclusions drawn from the 94 compounds with liquid formulation data also applied to compounds for which only solid formulation data were available. Simulation of solid formulations required specification of an additional parameter for dissolution velocity, which in practice varies across formulations. To approximate suitable average values, we tested different Lint80 dissolution times (10-30 min for capsules, 15-60 min for tablets) and compared prediction errors with those from liquid formulations. Based on this analysis, we selected 25 min for capsules and 40 min for tablets. This extended the oral dataset to 1200 PO concentration-time profiles (161 compounds). Using this larger dataset, all previously established conclusions were confirmed.

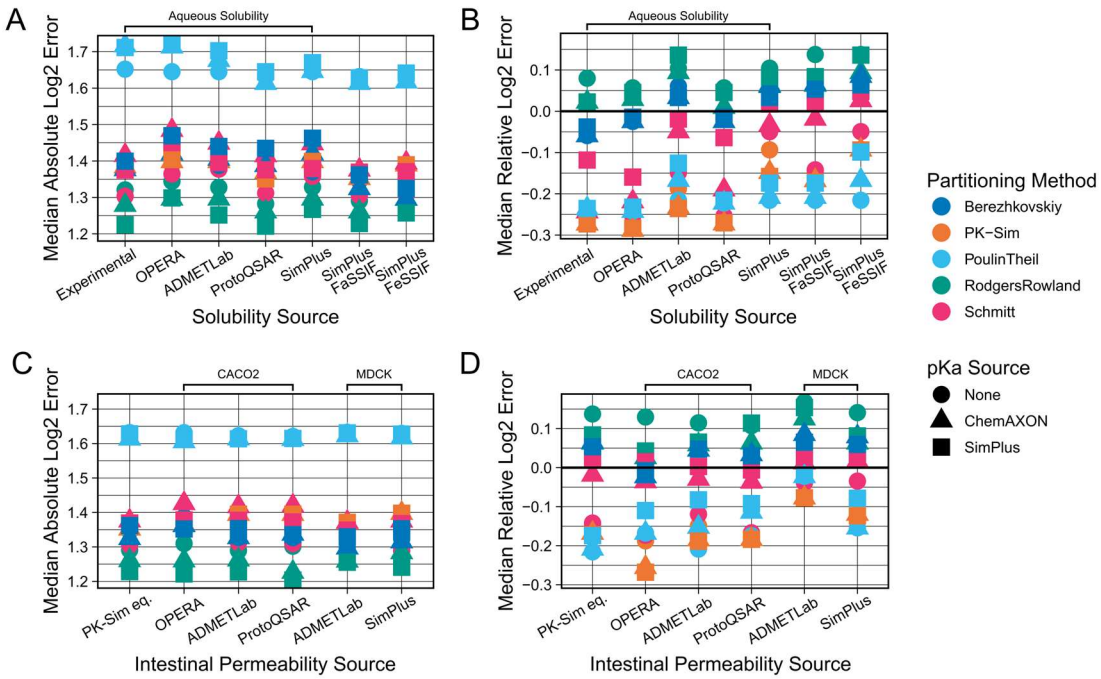


Figure 9: Comparison of predictive performances of different solubility and intestinal permeability prediction sources (step 3). Combinations of all available parameterisation sources were evaluated against the collected PO dataset (dissolved formulations). Results shown were generated using benchmark values for parameterisation of other parameters, i.e., *in vivo* plasma clearances, *in vitro* fraction unbound values and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. The top row shows the Median Absolute Log2 Errors (A) and the Median Relative Log2 Errors (B) for different solubility prediction methods. The bottom row shows the Median Absolute Log2 Errors (C) and the Median Relative Log2 Errors (D) for different intestinal permeability prediction methods. For comparison of solubility predictions, the intestinal permeability values used were the PK-Sim internal equation values (PK-Sim eq.). For comparison of intestinal permeability sources, the solubility values used were SimPlus FaSSIF values. Results for the PK-Sim internal equation were generated with the Bayer LogMA predictions. Intestinal permeability predictions are either CACO2 or MDCK permeability predictions. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Predictivity of full HT-PBK strategies

After stepwise evaluation of the individual compound properties required for HT-PBK modelling, we next assessed how well different overall HT-PBK strategies predicted the collected PK data. We focused on three classes of strategies:

1. Benchmark strategy: using *in vivo* and *in vitro* reference values for plasma clearance and Fu, representing the best possible performance.
2. Best fully *in silico* strategy: based on predicted property values, including those from proprietary tools.
3. Best fully *in silico* and open strategy: restricted to freely available prediction tools.

The specific parameterisation strategies and their performances are summarised in Table 1. As expected, the benchmark strategy using reference values was the most predictive. Nonetheless, fully *in silico* strategies also achieved acceptable accuracy, with 87% of Cmax values predicted within a tenfold error when proprietary tools were used, and 89% when restricted to freely available tools. Importantly, across all strategies, the systematic overestimation of oral absorption velocity meant that Cmax mispredictions outside the tenfold range were predominantly overpredictions rather than underpredictions. Such a bias would lead to conservative and therefore health-protective conclusions in risk assessment. The performance of the best *in silico* HT-PBK strategy is shown in Figure 10.

Table 1: Overview of different HT-PBK modelling strategies and their predictive performances. Table reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

		Best overall (including *in vivo* and *in vitro* benchmark values)	**Best *in silico* - proprietary**	**Best *in silico* - free**
Input parameter sources	**Lipophilicity**	Mean of Bayer LogD and LogMA	Mean of Bayer LogD and LogMA	ADMETLab 2.0: LogD
	Solubility	SimPlus: FaSSIF	SimPlus: FaSSIF	ADMETLab 2.0: Aqueous solubility
	pKa	ChemAxon	ChemAxon	None
	Partitioning method	PK-Sim	PK-Sim	PK-Sim
	Fraction Unbound	*In vitro* measured	ADMETLab 2.0	ADMETLab 2.0
	Clearance	*In vivo* observed Plasma clearance	ADMETLab 2.0: Plasma clearance	ADMETLab 2.0: Plasma clearance
	Intestinal permeability	ADMETLab 2.0: MDCK	ADMETLab 2.0: MDCK	ADMETLab 2.0: MDCK
Performance metrics	**IV: Median Absolute Log2 Error**	0.860	1.065	1.273
	PO: Median Absolute Log2 Error	1.245	1.595	1.780
	PO: Median Absolute Log2 Cmax predicted/observed	0.998	1.127	1.130
	PO: Compounds with Cmax within 10/5/3/2-fold	86% / 78% / 69% / 50%	87% / 78% / 65% / 44%	89% / 71% / 62% / 42%
	PO: Underprediction of compound Cmax of more than 10/5/3/2-fold	1.9% / 4.3% / 8.1% / 16%	0% / 3.1% / 8.1% / 19%	3.1% / 12% / 16% / 25%
	PO: Median Absolute Log2 AUC predicted/observed	0.870	1.285	1.358
	PO: Compounds with AUC within 10/5/3/2-fold	92% / 84% / 75% / 57%	84% / 71% / 58% / 38%	84% / 66% / 54% / 39%

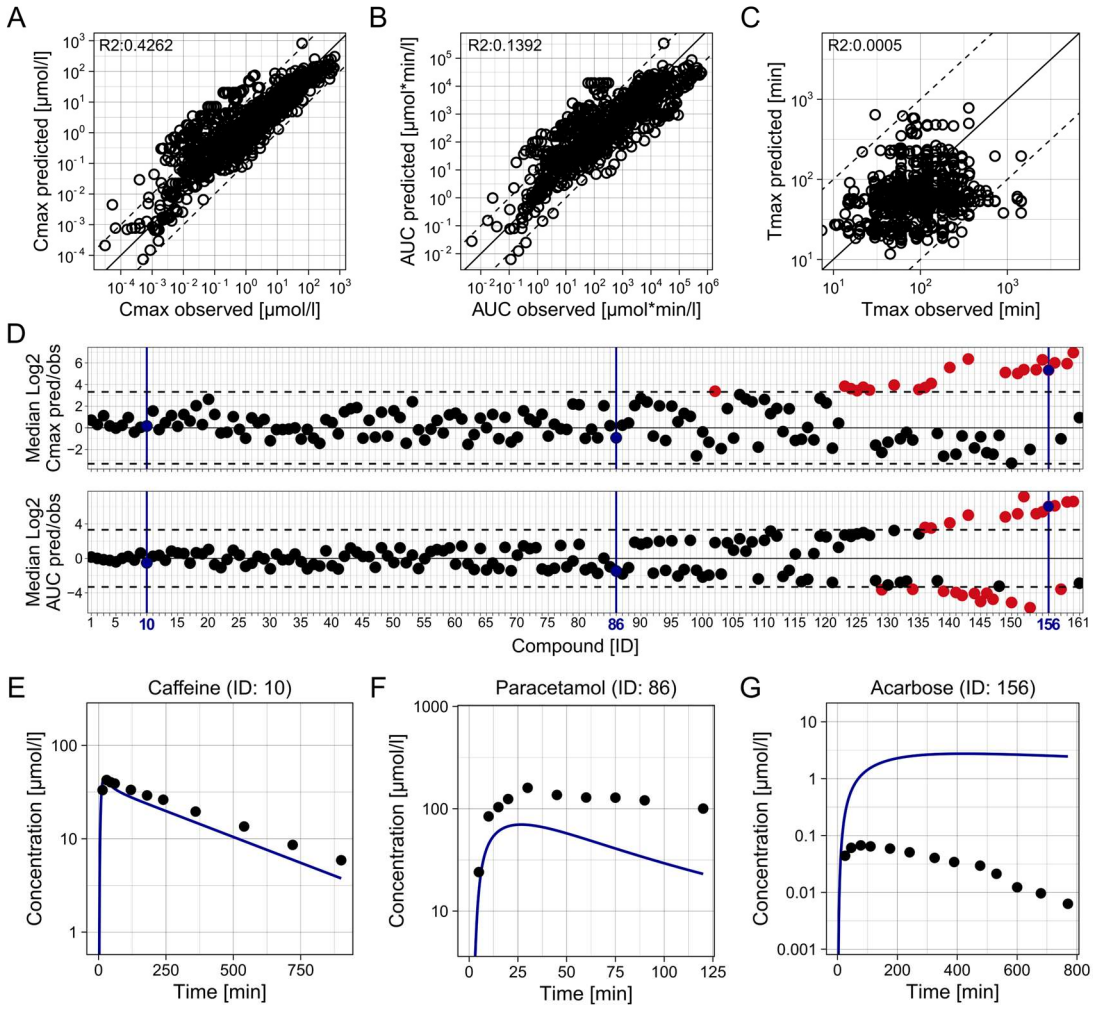


Figure 10: Predictive performance of the best fully *in silico*-based HT-PBK modelling strategy (proprietary). HT-PBK models were generated following the approach outlined in Table 1, then model simulations were compared against the collected PO dataset of 161 compounds (dissolved and solid formulations). (A) Predicted against observed Cmax values of all concentration-time profiles in the PO dataset. (B) Predicted against observed AUC values. (C) Predicted against observed Tmax values. Mind that A-C show individual concentration-time profiles with some compounds only represented by a single and others by multiple PK profiles. (D) The median Log2 Cmax and AUC predicted/observed values of each compound in the PO dataset. Dashed lines indicate tenfold errors. (E-G) Representative concentration-time profile predictions, and corresponding *in vivo* PK data, of three compounds embodying different levels of prediction quality. Blue colour marks the same representative compounds in D. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

10.3 Discussion

We assembled a large dataset of healthy human *in vivo* concentration-time profiles (200+ compounds), together with *in vitro* and *in silico* property predictions from multiple sources required for PBK modelling of the same compounds. Using this resource, we systematically compared all possible HT-PBK parameterisation strategies to determine which prediction tools, and combinations of tools, provided the best performance in PK-Sim. This allowed us to quantify the expected accuracy of HT-PBK predictions for typical pharmaceutical compounds.

For some input parameters, especially lipophilicity and clearance, we observed major differences in PBK performance, while for others predictive accuracy varied little. This likely reflects both the relative differences in predictive power between tools and the varying sensitivity of PBK models to different parameters. These findings suggest that robust HT-PBK predictions may require generating and comparing simulations using multiple prediction tools for the most influential parameters, while less critical parameters could be addressed by averaging predictions across tools or simulating ensembles of models. The latter strategy has the additional advantage of producing a distribution of outcomes, thereby providing confidence intervals around model predictions and enabling probabilistic risk assessment.

A large-scale comparison of commonly used partitioning methods revealed that all methods implemented in PK-Sim performed similarly well when combined with high-quality lipophilicity values. Under conditions of very high predicted lipophilicity, however, the Poulin & Theil and Berezhkovskiy methods performed better than others, likely due to greater robustness against outlier values. More subtle differences between partitioning methods were not detectable, which may reflect the intrinsic variability of our heterogeneous PK dataset compiled from different databases and literature. Even for the same compound, PK profiles differed between studies, implying that our dataset contained variability beyond what even an ideal generic PBK model could capture. This variability likely limited our ability to detect small differences in partitioning method performance when high-quality lipophilicity values were used.

In addition to overall errors, we also examined whether PBK prediction strategies showed systematic biases. Apart from tool-specific tendencies, the most consistent observation was that all HT-PBK strategies overestimated the rate of oral absorption, leading to underprediction of Tmax and slight overprediction of Cmax. The reasons for this are unclear, given the complexity of oral absorption *in vivo*. Some assumptions, such as treating “suspension” formulations as fully dissolved, may have contributed. Our models also represented intestinal absorption as a single passive permeability parameter, ignoring gut efflux transporters that are known to substantially affect absorption. Their omission may have produced faster and more complete absorption in simulations than *in vivo*.

CACO2 and MDCK permeability values should, in principle, account for transporter effects [202] but require scaling to PK-Sim parameters. Our attempt to scale such data using fitted intestinal permeability values from Willmann et al. [133] reduced the bias in Tmax predictions, but did not improve overall prediction errors. One explanation is that underprediction of absorption rates may affect PK predictions more strongly than overprediction. As a result, lower absolute log2 errors were observed when CACO2 values were used directly, despite scaled values being conceptually closer to *in vivo* permeability. For pharmaceutical applications, refining such scaling approaches may be

worthwhile, while for toxicological risk assessment, the conservative bias introduced by faster predicted absorption may even be preferable.

Similar limitations apply to transporter effects beyond the gut, which were not considered here due to the lack of homogeneous input data suitable for high-throughput analyses. Our focus was also limited to parent compounds, excluding metabolite formation, as quantitative bioactivation cannot be reliably predicted. Further work is needed to better understand which compound properties (e.g. physicochemical descriptors, clearance pathways, transporter affinities) correlate with higher or lower prediction accuracy, given the substantial variability we observed between compounds.

Another limitation is that most compounds in our dataset were typical pharmaceuticals, many of which may already be represented in the training data of the *in silico* tools we tested. This could bias our evaluation. The most obvious case was OPERA, which for some compounds returned CLint values identical to httk *in vitro* measurements. Although this reduces the fairness of cross-tool comparisons, from a practical perspective it also highlights that *in silico* tools trained on large or high-quality datasets may provide advantages by effectively recalling high-value data.

Despite these limitations, our study demonstrates that reliable HT-PBK models can be generated for predicting both IV and PO PK in humans. In principle, our approach could be extended to other compound classes, although validation is hindered by the scarcity of comparable concentration-time datasets. The potential applications of HT-PBK are broad, and recent efforts in both pharmacology and toxicology have begun exploring similar approaches. However, many of these studies relied on rodent data for validation [50,51,72,191,192,197–199], or assessed predictions in humans only against summary PK parameters such as Cmax or AUC [71,185,188]. Such metrics, while informative, obscure inaccuracies in model structure and parameterisation that can compensate each other, masking biases and limiting mechanistic insight. By contrast, our evaluation relied on full concentration-time profiles and multiple summary parameters, as is standard in expert PBK model development, thereby allowing a more rigorous and mechanistically meaningful validation.

An alternative strategy increasingly explored for PK prediction is the use of machine learning or AI methods to directly predict summary PK parameters such as Cmax, AUC, or bioavailability [50,185,186,198]. To date, however, such approaches have not matched the performance of using *in silico* predicted properties as inputs for mechanistic PBK models. Our study followed this latter approach, leveraging ML predictions of mechanistic compound properties as inputs to PK-Sim, which incorporates established physiological knowledge. This strategy not only yielded more accurate PK predictions but also preserved interpretability: each PBK model parameter has a physiological meaning, allowing mispredictions to be traced to specific properties. Moreover, PBK modelling provides a mechanistic basis for extrapolation to special populations or novel exposure scenarios, which is not possible with purely data-driven methods.

Overall, HT-PBK modelling shows promise as a tool for both pharmacological and toxicological applications. In drug discovery, HT-PBK models can support rapid compound prioritisation and optimisation. While their accuracy may initially be lower than that of traditional PBK models, they can provide a useful base model that can be refined as drug development progresses. In toxicology, HT-PBK enables estimation of internal tissue concentrations without animal testing, supporting chemical prioritisation, classification, and quantitative *in vitro* to *in vivo* extrapolation. For HT-PBK to play a role in next generation risk assessment, however, rigorous validation is essential to build the

confidence required for regulatory acceptance. Using a large and heterogeneous human PK dataset, we demonstrated that HT-PBK strategies are fit for purpose for such applications.

11. High-throughput PBK Prediction of Systemic PK after Dermal Exposure

Abstract

High-throughput physiologically based kinetic (HT-PBK) modelling offers a mechanistic framework to predict systemic pharmacokinetics (PK) from *in vitro* and *in silico* data, thereby facilitating non-animal chemical safety evaluation within next generation risk assessment (NGRA). In this study, we applied HT-PBK modelling to dermal exposure, an important route of human contact with chemicals. Using the skin permeation model implemented in the Open Systems Pharmacology Suite, we simulated systemic PK profiles for 52 compounds based solely on physicochemical properties predicted by quantitative structure-activity relationship (QSAR) models. Multiple *in silico* tools for lipophilicity, solubility, and related parameters were systematically compared to determine optimal model parameterisation strategies. For all compounds, 75% of predicted Cmax and AUC values fell within tenfold of observed human plasma data, showing a consistent tendency toward overprediction, likely due to the assumption of application occurring on fully hydrated skin. Although prediction variability was greater than for oral exposure, key exposure metrics were reproduced within acceptable limits. These findings demonstrate the feasibility of fully *in silico*, non-animal PBK modelling for dermal absorption and support its integration into regulatory NGRA frameworks for exposure-driven safety assessment.

Chapter 11 is adapted from:

Edizcan, Z., Schaller, S., Kuepfer, L. & Geci, R. *High-throughput PBK modelling for dermal exposure: a pragmatic approach to predict systemic pharmacokinetics.* Arch Toxicol 100, 2421–2432; 10.1007/s00204-026-04357-4 (2026).

Author contributions:

Zeynep Edizcan conducted the literature search and digitisation of dermal PK data, implemented and ran initial simulation scripts, carried out preliminary data analyses, prepared draft figures, and wrote the initial version of the manuscript. Stephan Schaller and Lars Kuepfer provided overall supervision and guidance. René Geci conceived and designed the study, supervised all stages of the project, performed *in silico* prediction of compound properties, refined and executed the final PBK simulations, led the interpretation of the results, wrote these chapters and created final versions of figures and tables.

11.1 Introduction

Next generation risk assessment (NGRA) seeks to evaluate chemical safety without reliance on animal data [203]. It integrates new approach methodologies (NAMs), including *in vitro* and *in silico* methods, within an exposure-driven framework. A key principle of NGRA is that hazard or bioactivity information obtained from NAMs must be interpreted in the context of internal exposure, defined as the concentration of a substance present in the human body under relevant conditions of use [53]. This necessitates approaches capable of predicting systemic pharmacokinetics (PK) in the absence of *in vivo* data [204].

Dermal exposure is a major route of human contact with chemicals across pharmaceutical, occupational, environmental, and consumer product applications [205]. The extent to which a compound penetrates the stratum corneum and reaches dermal blood vessels determines its systemic availability and varies widely among substances [206]. Quantifying systemic uptake is therefore essential to link external dermal application to potential biological effects, whether those are beneficial as in the case of therapeutic agents or adverse as in the case of toxicants.

Traditionally, dermal absorption has been investigated using *in vitro* assays or *in vivo* animal and human studies. However, these approaches are resource-intensive or may raise ethical and consent-related concerns. Consequently, there is a clear demand for predictive, non-animal alternatives. This demand is particularly strong in the cosmetics sector, where European legislation prohibits animal testing for cosmetic ingredients while still requiring systemic safety assessment [207]. Similar trends are emerging in the pharmaceutical and chemical industries, where efforts focus on reducing animal use, accelerating testing, and improving human relevance in safety evaluation.

Physiologically based kinetic (PBK) modelling provides a mechanistic framework to predict the absorption, distribution, metabolism, and excretion (ADME) of substances. Applied to dermal exposure, PBK models integrate skin physiology, compound-specific properties, and systemic circulation to predict internal concentrations. They thus serve as a key tool for linking external exposure with internal dose, enabling integration of hazard and exposure information within NGRA [46,208]. However, conventional PBK model development relies heavily on *in vivo* data for model parameterisation and validation, which constrains its applicability when such data are unavailable.

High-throughput PBK (HT-PBK) modelling overcomes this limitation by parameterising PBK models exclusively from *in vitro* and *in silico* data, including physicochemical properties predicted by machine learning (ML) and quantitative structure-activity relationship (QSAR) models. This approach enables scalable, animal-free predictions of systemic kinetics across large numbers of compounds. While previous work has demonstrated the feasibility of HT-PBK modelling for intravenous and oral exposure [140], its application to dermal administration has not yet been systematically evaluated.

In the present study, we apply HT-PBK modelling to dermal exposure. We compiled an extensive dataset of human *in vivo* plasma concentration-time profiles following dermal administration for 52 compounds and systematically evaluated different strategies for PBK model parameterisation using *in silico* input data. All simulations were conducted using the mechanistic skin permeation model implemented in the Open Systems Pharmacology (OSP) Suite, which represents the stratum corneum, viable epidermis, and dermis as diffusion-partition layers according to the framework established by Kasting and Miller [142] and implemented by Dancik et al. [141]. By benchmarking

against *in vivo* PK profiles and quantifying both systematic and random deviations, we identified the most suitable *in silico* tools for estimating physicochemical parameters and assessed the predictive performance of HT-PBK simulations for systemic plasma concentrations following dermal exposure.

11.2 Results

Retrieval of dermal PK study data

A total of 64 PK studies in healthy humans, covering 52 unique compounds, were collected from the literature (Figure 11). The identified compounds spanned a wide range of physicochemical properties, including molecular weight, lipophilicity, and solubility (Figure 11C-E). Most of the substances were therapeutic drugs, followed by sunscreen agents and cosmetic ingredients. Overall, the majority fell within the drug-like chemical space. Application conditions and experimental protocols, such as the anatomical site of application and formulation type, varied substantially across studies (Figure 11F-H).

For each compound, physicochemical property input data required for dermal PBK modelling were obtained using multiple *in silico* prediction tools. Lipophilicity was predicted from seven different sources, solubility from four, and melting point, boiling point, and vapour pressure from two each. As the various tools apply distinct algorithms and training datasets, all possible combinations of predicted inputs were used to construct a total of 224 PBK models per compound. Across the 64 clinical studies, this resulted in 14,336 individual PBK model simulations (64 studies × 224 model variants). This approach enabled systematic evaluation of which parameterisation strategies yielded the most accurate systemic PK predictions.

For clarity, only results for the parameters with the greatest influence on PK outcomes are presented here, which were lipophilicity and solubility. Results for other compound properties, including melting point, boiling point, and vapour pressure, are provided in Supplementary Figure 14. For these latter parameters, no meaningful differences were observed between prediction sources.

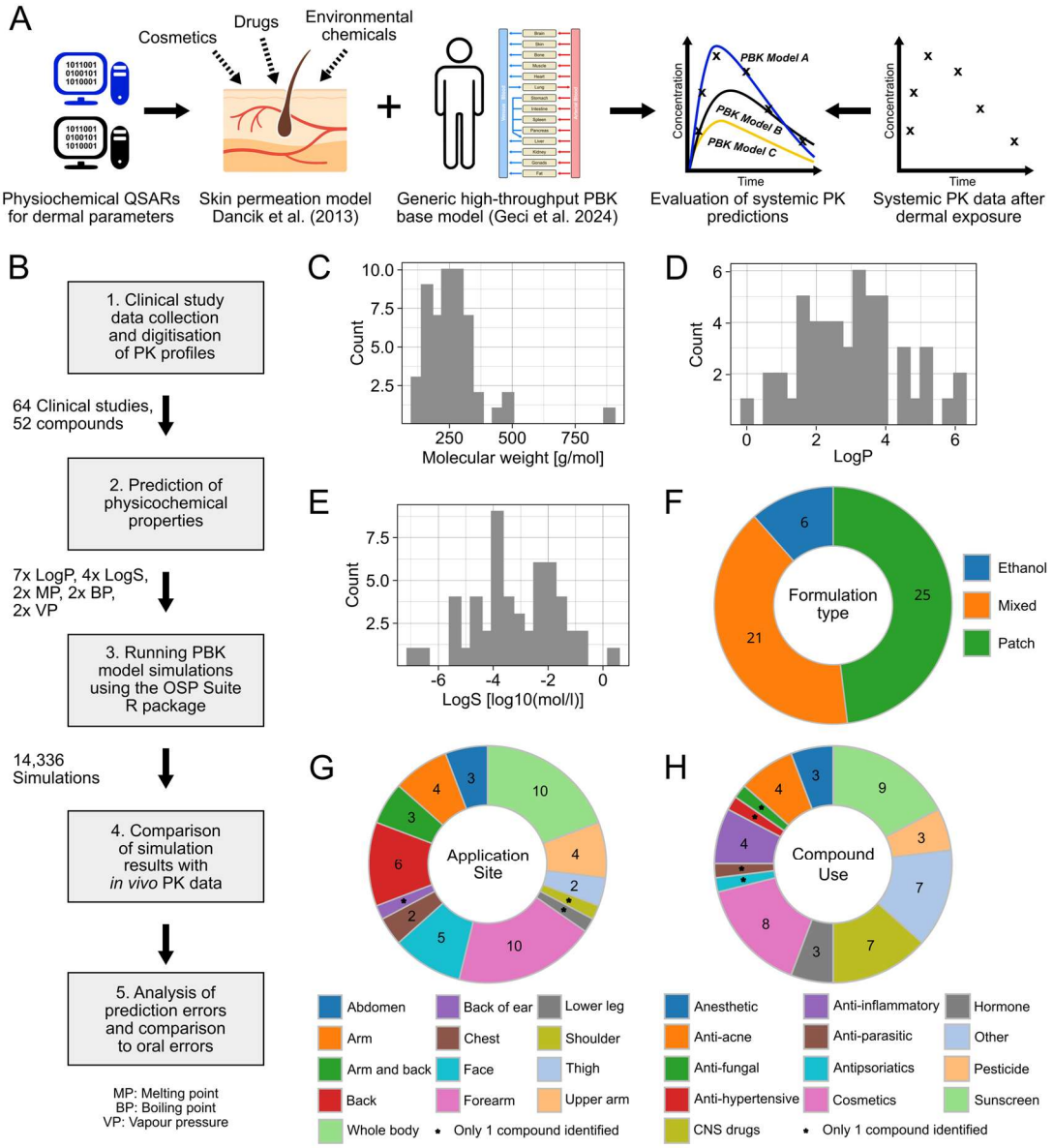


Figure 11: Overview of dermally administered PK data and of compound properties. (A, B) Schematic overview of the HT-PBK modelling and analysis workflow for dermally administered compounds. (C-E) Physicochemical properties of the 52 compounds in the collected PK dataset. LogP and LogS values are given as predicted by ADMETLab 3.0. LogS reported as Log10(mol/l). (F-H) The distribution of number of compounds across different study characteristics, such as the type of compound use, formulation type and application site on the body. Figure reproduced from [209]. No changes made.

Performance of *in silico* property predictions within the dermal PBK model

All dermal absorption simulations were conducted using the mechanistic skin permeation model implemented in the OSP Suite [141]. The model was parameterised exclusively with physicochemical properties predicted *in silico*, enabling systematic evaluation of how different lipophilicity and solubility predictions influenced plasma PK outcomes.

Mean absolute errors for maximum concentration (Cmax) and area under the curve (AUC) were first analysed to identify which lipophilicity prediction tools yielded the most accurate systemic PK simulations following dermal exposure (Figure 12). Simulations based on LogD values generally performed worse than those using LogP, with LogD-based predictions showing a consistent bias toward underprediction of *in vivo* PK data. Performance differences were also observed among tools predicting the same lipophilicity metric. For example, LogP values from SimPlus produced slightly more accurate predictions than those from OPERA.

Model performance also varied markedly with the solubility prediction source. LogS values from SimPlus and ADMETLab achieved the lowest prediction errors, followed by VEGA and OPERA. Overall, discrepancies between solubility prediction tools were greater than those observed for lipophilicity. Heatmaps of Mean Absolute Log2 errors for Cmax and AUC across all combinations of prediction tools are shown in Figure 12CF. The lowest overall errors were obtained when combining SimPlus LogP with LogS values from either SimPlus or ADMETLab.

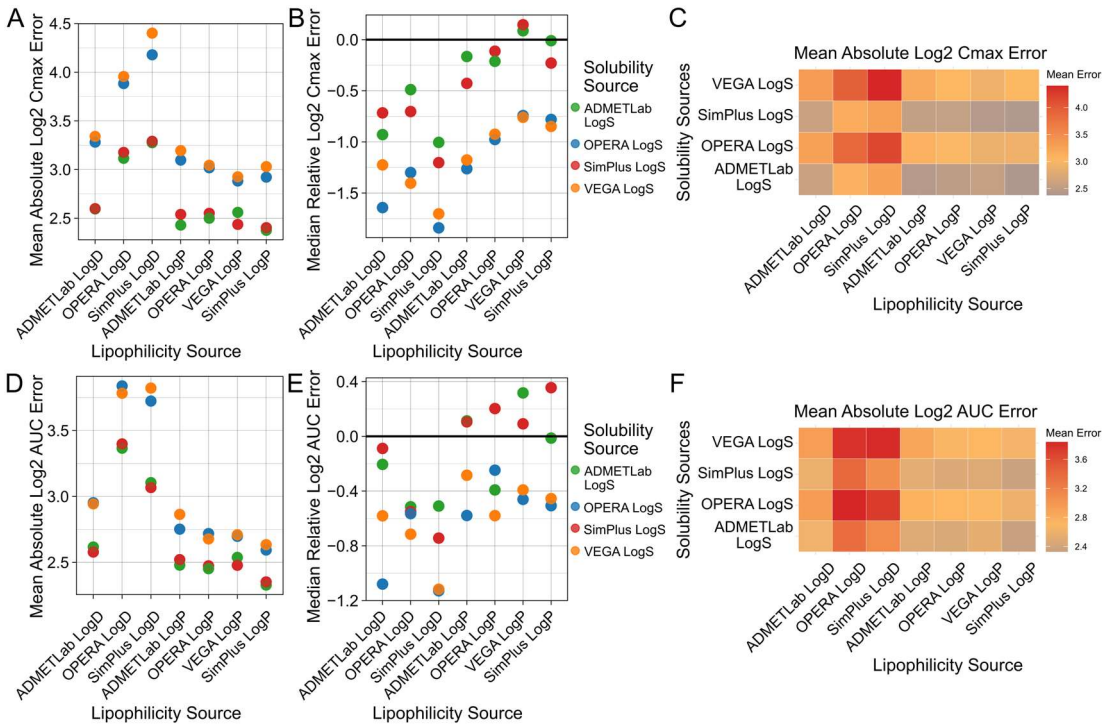


Figure 12: Systemic PK prediction accuracy after dermal exposure across *in silico* input sources. All combinations of physicochemical property prediction tools were used to parameterise the PBK model and evaluated against the collected *in vivo* dermal PK dataset. The first row shows Mean Absolute Log2 Cmax Errors (A), Median Relative Log2 Cmax Errors (B), and a heatmap of Mean Absolute Log2 Cmax Errors (C) for different lipophilicity and solubility prediction sources. The second row shows Mean Absolute Log2 AUC Errors (D), the Median Relative Log2 AUC Errors (E), and a heatmap of Mean Absolute Log2 AUC Errors (F) for different lipophilicity and solubility prediction sources. Figure reproduced from [209]. No changes made.

Simulated plasma concentration-time profiles for five of the best-performing prediction strategies are presented in Figure 13 for all compounds included in the dermal PK dataset. For some substances, the predicted profiles reproduced the qualitative trends of the corresponding *in vivo* data. However, for others, the simulated concentration-time curves deviated substantially from the observed profiles. Both pronounced underpredictions and overpredictions were observed for certain compounds, and the overall temporal dynamics and profile shapes varied considerably across substances.

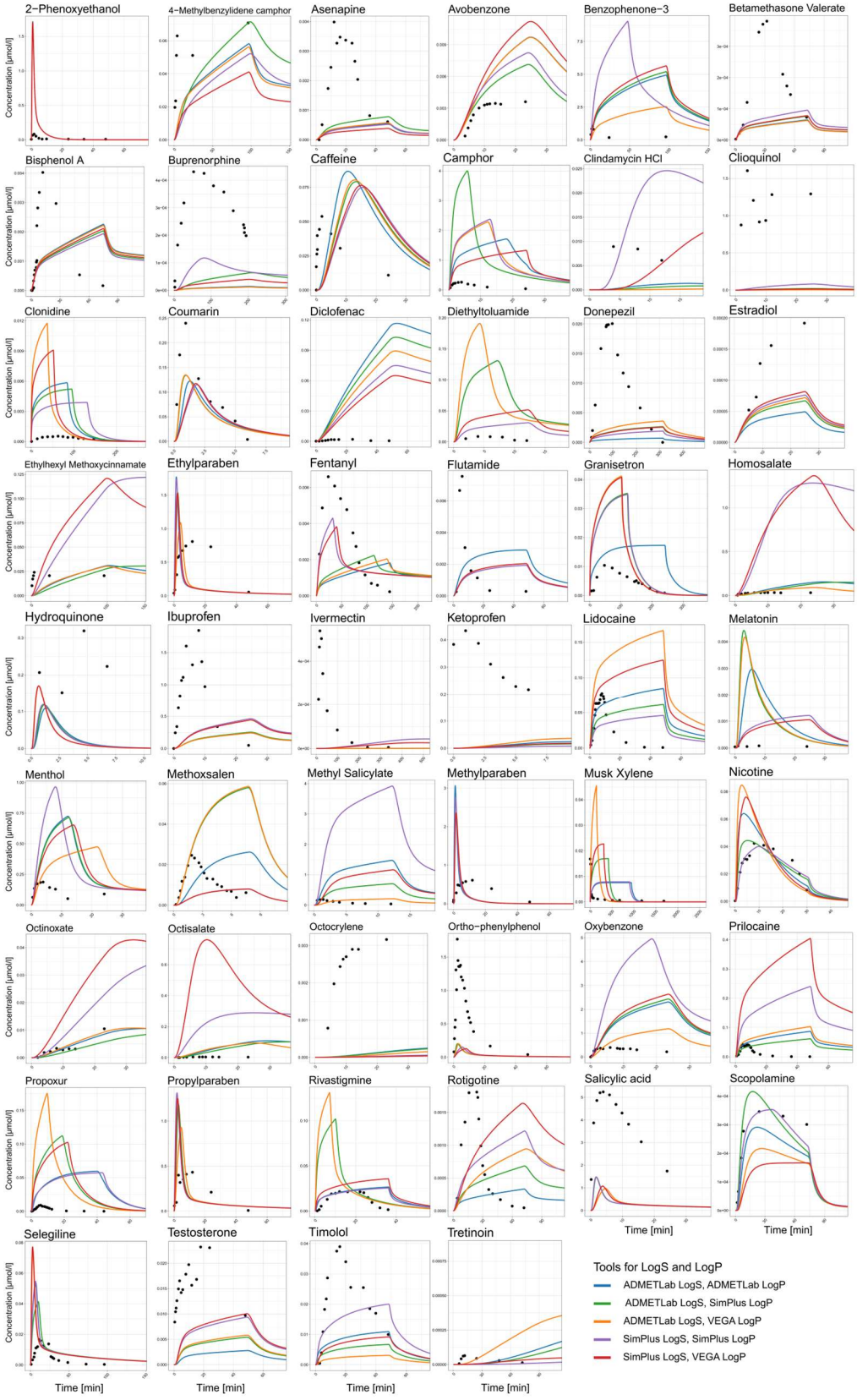


Figure 13: Predicted concentration-time profiles against observed systemic PK data. Shown are the results of five selected best prediction strategies as indicated in the legend for all the compounds in the dataset. Figure reproduced from [209]. No changes made.

Correlation between prediction errors and compound properties

To examine potential systematic relationships between prediction errors and compound properties, we analysed Cmax and AUC errors at the individual compound level. No consistent correlation was observed between prediction errors and molecular weight, predicted LogP, or LogS, despite the wide range of physicochemical characteristics represented (Figure 14). A weak tendency toward underprediction of Cmax and AUC was noted for poorly soluble compounds with low LogS values (Figure 14CF), but this trend was not pronounced due to the high overall variability in prediction errors. Other predicted properties, including density, boiling point, melting point, and vapour pressure, also showed no clear correlation patterns (Supplementary Figure 15). These observations were consistent across all *in silico* tools, suggesting that prediction errors were not specific to individual QSAR models.

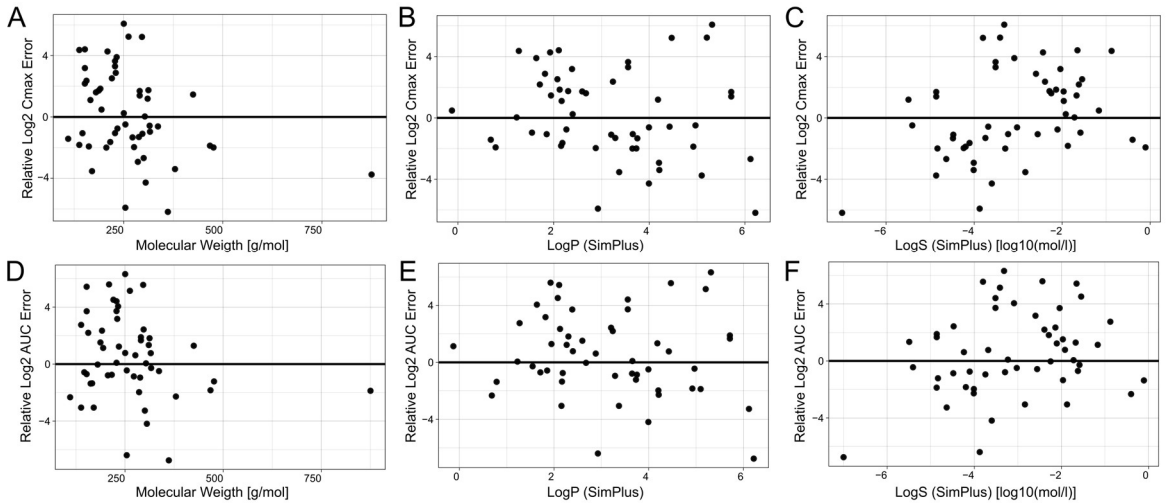


Figure 14: Correlation prediction errors and various compound properties. Results shown are those of the best *in silico* based HT-PBK modelling strategy using SimPlus LogS and LogP. The first row indicates predicted physicochemical parameters including molecular weight, LogP and LogS against Relative Log2 Cmax Errors (A, B, C). The second row shows the same physicochemical parameters against Relative Log2 AUC Errors (D, E, F). Correlations with other compound properties are shown in Supplementary Figure 15. Figure reproduced from [209]. No changes made.

Evaluation of systemic PK prediction reveals overestimation of dermal absorption

After identifying the most suitable physicochemical property prediction tools for HT-PBK model parameterisation, we evaluated the performance of the best *in silico* strategy for predicting systemic PK following dermal absorption (Figure 15). 75% of compounds had Cmax and AUC values predicted within a tenfold range of observed data. However, a systematic tendency toward overprediction of AUC was evident, although three compounds showed more than tenfold underprediction. This systematic bias was not apparent from median error values and was attributed to the model's default assumption of a fully hydrated stratum corneum during exposure. As clinical studies did not consistently report skin hydration conditions, this assumption was applied uniformly to ensure a health-protective approach, given that hydrated skin exhibits higher permeability. Since stratum corneum hydration is a key determinant of dermal

absorption, this simplification likely contributed to overestimation of absorption rates and extents. Conversely, applying the hydration assumption only where explicitly documented would have led to marked underprediction for several compounds (Supplementary Figure 16).

We further compared the prediction errors of dermal HT-PBK simulations with those from previously conducted high-throughput simulations for oral administration [140]. The dermal dataset comprised 52 unique substances, whereas the oral dataset included 161, resulting in limited statistical comparability. Non-normalised Cmax and AUC error distributions are presented in Supplementary Figure 17. Due to the smaller dermal dataset, prediction errors exhibited greater variability and did not follow a clear normal distribution as observed for oral exposure. Nonetheless, dermal predictions showed broader distributions of Cmax and AUC errors, indicating higher variability and uncertainty in systemic PK predictions following dermal exposure (Figure 15EF). For comparison, 87% of Cmax and 84% of AUC values were within a tenfold range for oral administration, suggesting that systemic PK predictions were generally more accurate for the oral than for the dermal route.

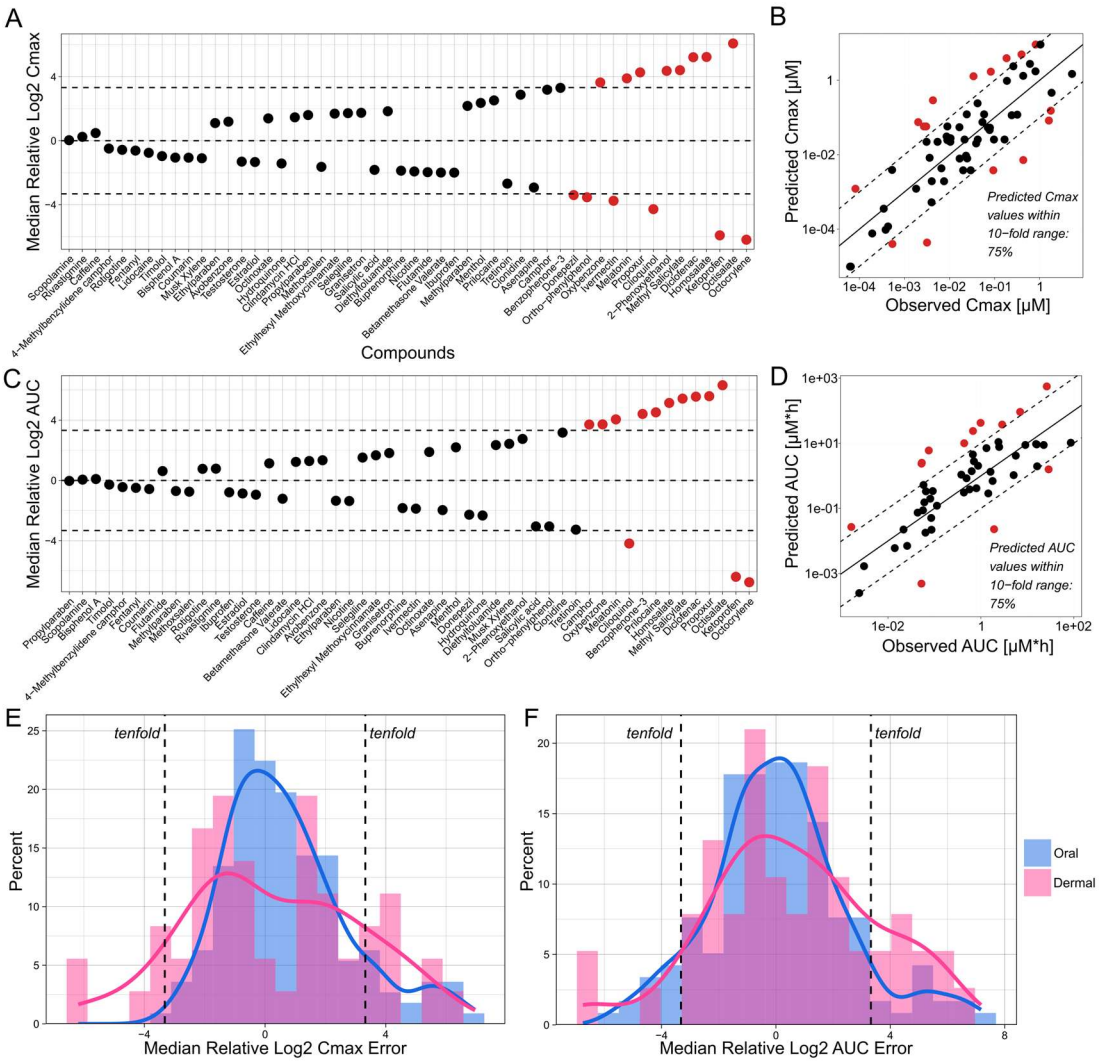


Figure 15: Dataset-wide performance of the best dermal HT-PBK strategy and comparison with oral HT-PBK errors. Model simulations using SimPlus LogS and LogP as the best HT-PBK strategy were evaluated against *in vivo* plasma concentration-time profiles of dermally administered compounds. (A) The Median Log2 predicted/observed Cmax values of each compound in the dataset. (B) Predicted against observed Cmax values of the best strategy for all PK studies. (C) The Median Log2 predicted/observed AUC values of each compound in the dataset. (D) Predicted against observed AUC values of the best strategy for all PK studies. (E) The Median Log2 predicted/observed Cmax values of dermally administered compounds (pink) were compared against errors for prediction of oral administration (blue). (F) The Median Log2 predicted/observed AUC values of dermally administered compounds (pink) were compared against errors for prediction of oral administration (blue). Dashed lines indicate the tenfold error range. Oral PK prediction errors were taken from Geci et al. [140]. Figure reproduced from [209]. No changes made.

11.3 Discussion

In this study, we employed *in silico* predicted physicochemical compound properties in combination with HT-PBK modelling to simulate systemic PK of 52 dermally administered substances. A curated dataset of human *in vivo* plasma concentration-time profiles was compiled to identify which combinations of predicted physicochemical properties yield the most suitable parameters for simulating systemic PK after dermal absorption. All feasible HT-PBK modelling strategies were evaluated, and potential error sources and biases were analysed for compounds relevant to pharmaceutical and cosmetic applications.

Due to the limited availability of dermal PK studies, fewer compounds were retrieved compared to previous analyses of oral administration. Moreover, variability across dermal studies was higher than for intravenous or oral exposure. The investigated compounds were structurally more diverse, ranging from drugs to cosmetic and sunscreen ingredients, and experimental procedures varied considerably. For instance, dermal applications were performed at different body sites with varying skin-layer thickness and perfusion, using diverse vehicles and formulations. Some studies reported occlusion or hydration of the treated skin, while others did not provide such details. Additional factors, such as temperature or airflow, were often unreported. Consequently, our large-scale evaluation of HT-PBK model performance was affected by study heterogeneity and incomplete metadata.

A key objective was to identify the most suitable physicochemical property prediction tools for high-throughput dermal PBK modelling. Substantial differences in simulation accuracy were observed among lipophilicity and solubility prediction sources, whereas properties such as boiling point, melting point, and vapour pressure showed minimal influence. This reflects the importance of lipophilicity and solubility for dermal absorption kinetics, as both strongly affect compound permeability through skin layers [210]. The observation that LogP values resulted in better model performance than LogD was consistent with the model's formal requirement for LogP as the lipophilicity input. LogD was nevertheless included for comparison, as it previously improved predictions in whole-body PBK models. No clear correlation was observed between the magnitude of predicted physicochemical properties and the simulation error in dermal absorption, consistent with earlier reports using experimentally measured data [141].

HT-PBK simulations generally showed a tendency toward overprediction of systemic PK, particularly for AUC values. This bias may partly stem from missing study information. For instance, incomplete release of compounds from formulations could not be accounted for here. Furthermore, all simulations assumed fully hydrated skin, which may not reflect actual study conditions. This assumption, however, ensured conservative (health-protective) estimates that overpredict rather than underpredict internal concentrations. For future prospective applications, however, such protocol-specific formulation and hydration effects could be modelled explicitly for specific substances.

Prediction errors were typically larger for dermal than for oral exposure, although this conclusion remains uncertain due to the smaller dermal dataset. In our previous oral PBK evaluation, 87% of Cmax and 84% of AUC values were predicted within a tenfold range [140], whereas for dermal exposure the corresponding proportion was 75%. The agreement between predicted and observed PK profile shapes was also lower. Previous studies suggest that reservoir formation within the skin, involving slow diffusion or binding processes, may influence release kinetics and systemic uptake [211,212]. Neglecting such processes likely contributes to the higher variability observed for dermal

compared to oral exposure, indicating that high-throughput prediction of dermal PK remains more challenging.

Despite deviations in kinetic profiles, predicted Cmax and AUC values were often within acceptable ranges of observed data. This suggests that while temporal profiles may be imperfectly reproduced, HT-PBK simulations can still provide useful quantitative estimates of key exposure metrics relevant for risk assessment. For example, AUC primarily depends on total absorbed dose and systemic clearance, which may be captured accurately even with simplified absorption kinetics. Overall, HT-PBK modelling appears suitable for predicting systemic PK of dermally applied compounds within an acceptable uncertainty range.

All simulations employed the mechanistic skin permeation model implemented in the OSP Suite [141], which represents diffusion and partitioning across the stratum corneum, viable epidermis, and dermis. Using this established model enabled a focused analysis of parameterisation from *in silico* data rather than structural model development. In earlier work, we applied the same whole-body PBK framework to intravenous and oral exposure [140]. The modular architecture of the OSP framework allows independent validation and optimisation of each exposure route while maintaining consistency across the system model. This modularity supports integrated, route-specific simulations within a unified PBK platform, enabling coherent quantitative exposure assessment for NGRA applications.

Recent studies have introduced mechanistic dermal absorption models based on physicochemical compound properties [213–216], while others have addressed uncertainty quantification in dermal PBK modelling, including Bayesian calibration and tiered safety assessment approaches [217–219]. Partition and diffusion coefficients have also been estimated using ML/QSAR and mechanistic methods [220–223]. However, the integration of QSAR-based physicochemical predictions with PBK modelling to simulate systemic plasma concentrations after dermal exposure has not previously been demonstrated. This represents an important step toward scalable, quantitative *in vitro* to *in vivo* extrapolation [53,224]. It aligns with OECD guidance on PBK model use in regulatory contexts [225] and supports the NGRA vision of systematic, exposure-driven, non-animal safety evaluation [203,226].

By applying QSAR-predicted physicochemical parameters in PK-Sim, our HT-PBK framework enabled prediction of systemic exposure following dermal application for a wide range of substances, including pharmaceuticals, cosmetic ingredients, and pesticides. The approach facilitates quantitative risk assessment without reliance on experimental or animal data, addressing current needs for predictive and scalable tools in regulatory and product development contexts.

12. ADME-based Assessment of Chemical Safety

Abstract

Chemical safety assessment is essential for protecting human health, yet current approaches face two major limitations: they cannot address the vast and growing number of substances requiring evaluation, and they continue to rely heavily on animal testing, which society expects to be phased out. To overcome these challenges, the European Partnership for Alternative Approaches to Animal Testing (EPAA) proposed a novel framework for systemic toxicity classification based on new approach methodologies (NAMs). A central element of this framework is the grouping of chemicals into different levels of Potential Systemic Availability (PSA) concern (low, medium, high). However, both the conceptual basis and practical implementation of PSA classification have so far remained unclear. Here, we develop a health-protective definition of PSA concern classes and propose a quantitative method for their evaluation using high-throughput physiologically based kinetic (HT-PBK) modelling. Applying this approach to 150 benchmark compounds from the EPAA NAM Designathon, we successfully classified PSA concerns for 139 chemicals. Manual annotation enabled comparison of predicted classifications with expert judgement, providing an initial evaluation of plausibility. Our results highlight the conditions under which PSA can be used to prioritise or deprioritise chemicals based on their toxicokinetic properties. However, we also show that PSA alone is insufficient for the majority of compounds, which therefore remain in the medium-concern category to allow further classification once bioactivity information is incorporated. These findings clarify the role of PSA within the NAM Designathon framework and provide a practical step towards its integration with bioactivity data for future systemic toxicity assessment.

Chapter 12 is adapted from:

Geci, R., Paini, A., Worth, A. P., Kuepfer, L. & Schaller, S. *Potential systemic availability classification of chemicals for safety assessment.* Environment international 202, 109636; 10.1016/j.envint.2025.109636 (2025).

Author contributions:

Alicia Paini and Andrew Worth contributed to the manual evaluation of chemical potential systemic availability classification. Lars Kuepfer and Stephan Schaller provided overall supervision and guidance. René Geci performed all remaining analyses, wrote these chapters and created the figures and tables.

12.1 Introduction

The effective assessment and regulation of chemical safety are fundamental to protecting human and environmental health. Yet, despite decades of regulatory oversight in industrialised regions such as the European Union (EU), the majority of chemicals in commercial use have not undergone thorough safety evaluation. The European Environment Agency (EEA) has estimated that only a small fraction of around 0.5% of chemicals on the EU market are robustly characterised with respect to their hazard and exposure profiles, while approximately 70% lack sufficient information to support meaningful risk assessment [227] (Figure 16).

Within the EU, the primary legislation governing industrial chemical risks is the Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) regulation [228]. However, its information requirements scale with the annual production volumes of chemicals. For substances produced in quantities below 1 tonne per year, no safety information is required. And for those in the 1-10 tonnes per year range only a basic set of information is demanded. Even for substances in regular use and high production volume, available data are often limited to basic physicochemical properties or selected test results insufficient for comprehensive risk assessment. This is generally inadequate for comprehensive chemical safety assessment [229]. Consequently, many substances, including some in widespread use, are not subject to systematic safety evaluation.

This lack of information undermines the ability of regulatory authorities to identify and manage hazardous substances, to prioritise regulatory action, and to anticipate long-term or cumulative health risks. This problem is now increasingly recognised as a central limitation of the current chemical safety system, eroding both regulatory effectiveness and public trust in chemical management policies. Addressing this challenge requires both a rethinking of assessment strategies and the development of scalable, resource-efficient methods capable of handling large numbers of substances without compromising scientific rigour.

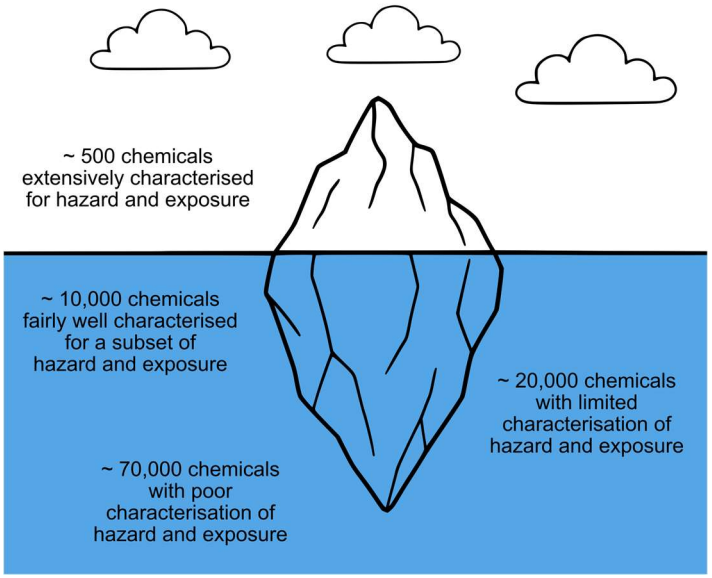


Figure 16: Illustration of the imbalance between the number of chemicals on the EU market and the extent of their hazard and exposure characterisation. Only a small fraction of substances is extensively characterised and regularly monitored, while the majority of chemicals have little or no hazard and exposure information available. The figure is adapted from the European Environment Agency report titled “The European environment – state and outlook 2020” [227].

A central reason for chemical safety data gaps is that current frameworks for chemical regulation were built on the assumption that substances would be assessed on a case-by-case basis, predominantly using animal testing. While this approach established the historical foundation of toxicological evaluation, it is not scalable to the tens of thousands of substances now in circulation across industrial, agricultural, and consumer sectors. Animal studies are resource-intensive, time-consuming, and costly, which makes them impractical for the volume of chemicals requiring assessment. Heavy reliance on them slows regulatory progress, leaving many substances without sufficient toxicological data. Beyond these practical limitations, animal testing is increasingly challenged on ethical grounds. Public demand for stronger animal protection is reflected, for instance, in the fact that six out of only ten successful European Union Citizens’ Initiatives explicitly called for stronger protection of animal rights. Together, these constraints show the inadequacy of animal-based testing as the backbone of modern chemical risk assessment. There is thus a pressing need for alternative methods that are scalable, scientifically robust, and protective of human health while also aligning with societal expectations to move away from animal experimentation.

In light of these challenges, the European Partnership for Alternative Approaches to Animal Testing (EPAA) launched the NAM Designathon to explore how new approach methodologies (NAMs) could support systemic toxicity assessment in a regulatory context that no longer relies on animal testing. At the centre of this initiative is a conceptual framework based on a two-dimensional matrix that combines information on a compound’s toxicodynamic (TD) bioactivity with its toxicokinetic (TK) potential [21] (Figure 17A). By assessing TD and TK properties independently and explicitly decoupling them from external exposure estimates, which are often unavailable or

uncertain, the framework aims to classify chemicals into three levels of regulatory concern. Under this scheme, high-concern substances could be restricted or banned, medium-concern substances would require further application-specific risk evaluation, and low-concern substances could be permitted for general use. The goal is to provide a level of human health protection equivalent to traditional assessment while moving away from reliance on animal studies.

To evaluate the feasibility of this approach, the EPAA compiled a benchmark set of 150 structurally diverse chemicals, equally distributed across the three concern categories and spanning different use types, including industrial chemicals, pesticides, pharmaceuticals, and consumer product substances (Figure 17D). Participants in the Designathon were then tasked with applying NAMs to classify these substances to test the practical value of the proposed framework.

A central component of the Designathon framework is the concept of Potential Systemic Availability (PSA), representing the likelihood that a chemical reaches internal organs at biologically relevant concentrations. While the rationale for including PSA as a dimension in the classification system was clear, no method for its practical evaluation was defined. This omission represents a critical gap, as toxicity depends on the interplay between TK, bioactivity, and exposure. The ability to derive a protective yet practicable PSA classification is essential for the framework's utility.

The aim of this work is therefore to establish the scientific basis for classifying PSA into three levels of concern, as envisioned in the NAM Designathon. We propose explicit definitions for each PSA class and introduce a classification method that is both efficient and health-protective. Finally, we validate the approach by comparing our results with expert judgement-based classifications, thereby assessing its reliability for application in regulatory contexts.

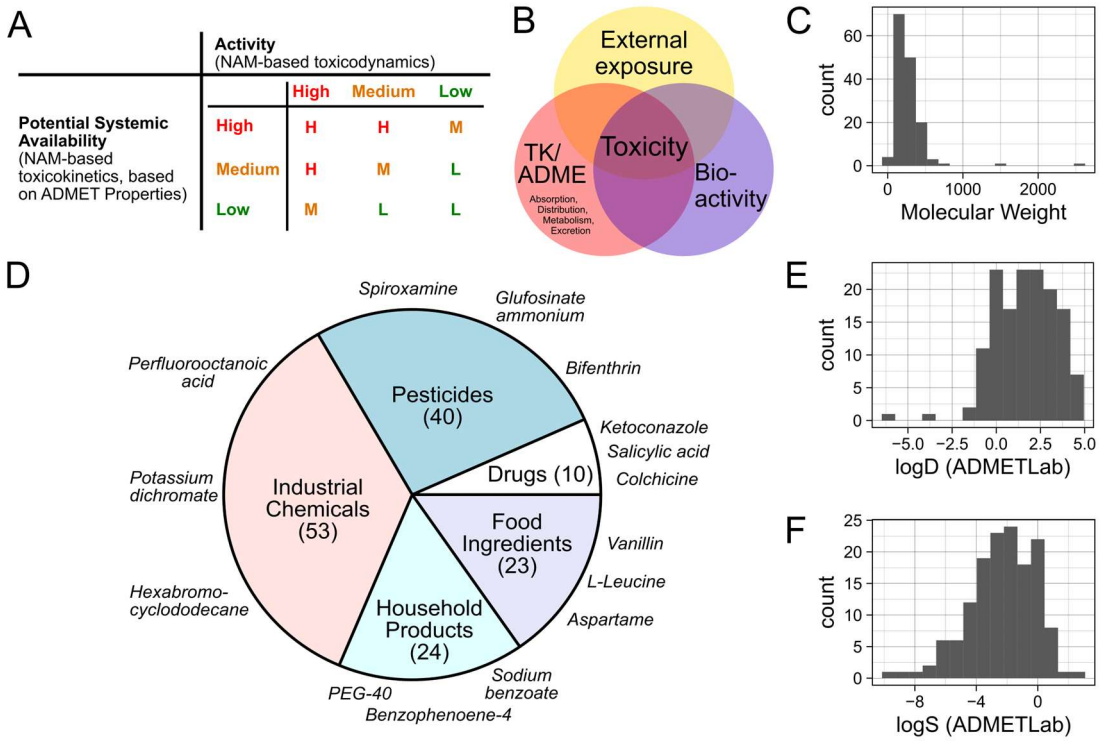


Figure 17: Data overview and general background. (A) The EPAA NAM Designathon classification matrix integrates Activity and Potential Systemic Availability into an overall systemic toxicity concern class (adapted from 21). (B) Scheme illustrating that toxicity is determined by external exposure, bioactivity, and toxicokinetics. (C, E, F) Histograms showing the molecular weight, predicted LogD, and LogS values of NAM Designathon compounds. (D) Distribution of NAM Designathon compounds among different use classes, with three exemplary compounds of each class. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

12.2 Results

Potential Systemic Availability concern class definitions

“Potential Systemic Availability” (PSA) has been proposed as a key element in the chemical safety assessment framework suggested by Berggren and Worth [21]. While the rationale for including PSA is clear, the term itself has not been formally defined. It was broadly stated to encompass toxicokinetics, ADME processes, bioaccumulation, and persistence, but no operational definition or representative examples of different concern categories have been provided. To address this gap, we propose a definition of PSA and contrast it with a possible alternative interpretation.

The definition of PSA concerns we will use in the following is that compounds shall be considered…

… “*Low PSA Concern*” if their *in vivo* concentrations can be assumed to never reach levels required for bioactivity under almost all relevant exposure scenarios, so that such compounds can be assumed to be safe under almost all conditions.

… “*Medium PSA Concern*” if the reaching of critical *in vivo* concentrations depends on the specific exposure scenario and the bioactivity of the compound. For such compounds, we assume that there will be an exposure dependency for their toxicity, with the compound being safe below a certain exposure threshold but unsafe above it.

… “*High PSA Concern*” if already very small exposures can lead to *in vivo* concentrations required for bioactivity, even if the bioactivity of the compound is only modest. Such compounds would pose a risk under almost all relevant exposure scenarios.

This definition represents a “hard” interpretation of PSA, as it sets a stringent standard for classification into the low or high concern categories. By this criterion, most small molecules, including typical drugs and pesticides, fall into the medium concern class, while only extreme cases would be considered low or high concern. Importantly, our definition explicitly refers to whether peak plasma concentrations (Cmax) reach levels required for *in vivo* bioactivity, irrespective of how long such concentrations are sustained. This makes the definition conservative, as it does not depend on average concentrations or overall exposure (e.g., AUC).

A contrasting “softer” interpretation of PSA would equate it with bioavailability, defined in pharmacology as the fraction of an administered dose that reaches systemic circulation. In this view, compounds could be categorised into low, medium, or high PSA concern classes directly based on their bioavailability. Such an approach would lead to a more balanced distribution of typical small molecules across the categories, since drugs and pesticides exhibit widely differing bioavailability values ranging from a few percent to nearly 100%.

To illustrate why we adopt the “hard” definition, we simulated plasma concentrations of five hypothetical compounds differing only in clearance, solubility, and intestinal permeability. Using a generic physiologically based kinetic (PBK) model, we calculated their maximum plasma concentrations following repeated oral exposure over five years, while assuming identical values for other properties (e.g., molecular weight, lipophilicity). Two concentration thresholds were added for context, representing high bioactivity (10 nM) and low bioactivity (100 µM). The results (Figure 18, Table 2) showed that compounds 2-4, despite having markedly different bioavailability, were all capable of reaching both thresholds depending on the applied dose and their intrinsic bioactivity.

Thus, these compounds cannot be considered safe or unsafe *a priori* based solely on PSA defined as bioavailability. This example demonstrates that the “softer” definition does not reliably separate compounds into meaningful concern classes, whereas the “hard” definition better reflects the relationship between exposure, systemic concentrations, and potential biological effects.

Table 2: Properties of five example compounds of different Potential Systemic Availability concerns and different oral bioavailability. Other compound properties are assumed to be identical: molecular weight of 300 g/mol, lipophilicity of 1, fraction unbound of 50%. Table reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Compound No.	**1**	**2**	**3**	**4**	**5**
Potential Systemic Availability	**Low**	**Medium (Low Bioavailability)**	**Medium (Intermediate Bioavailability)**	**Medium (High Bioavailability)**	**High**
Clearance	15 ml/min/kg	15 ml/min/kg	7.5 ml/min/kg	0.5 ml/min/kg	1e-5 ml/min/kg
Solubility (LogS) [log 10 mol/l]	-5	-2	-2	-2	-2
Intestinal Permeability [log 10 cm/s]	-9	-6	-6	-6	-6
Predicted Oral Bioavailability (F)	~0%	~5%	~53%	~97%	~100%

The situation differs for compounds 1 and 5. Compound 1, whose oral absorption is limited by both solubility and permeability, would never be able to reach the 10 nM threshold for “very bioactive” compounds, regardless of daily oral dose. Such a compound can therefore be justifiably classified *a priori* as low PSA concern. Compound 5, in contrast, exhibits bioaccumulation. After a single oral dose, its plasma concentrations would equal those of compound 4 (Supplementary Figure 18). However, with repeated dosing, compound 5 would progressively accumulate, eventually reaching much higher concentrations even at low daily exposures. For example, at 0.1 mg/kg per day, compound 5 would surpass the 100 µM threshold for bioactivity despite having only modest intrinsic potency. Because such compounds carry a heightened risk of attaining concentrations sufficient to cause adverse effects *in vivo*, and because bioaccumulation itself is undesirable, it is warranted to classify them *a priori* as high PSA concern.

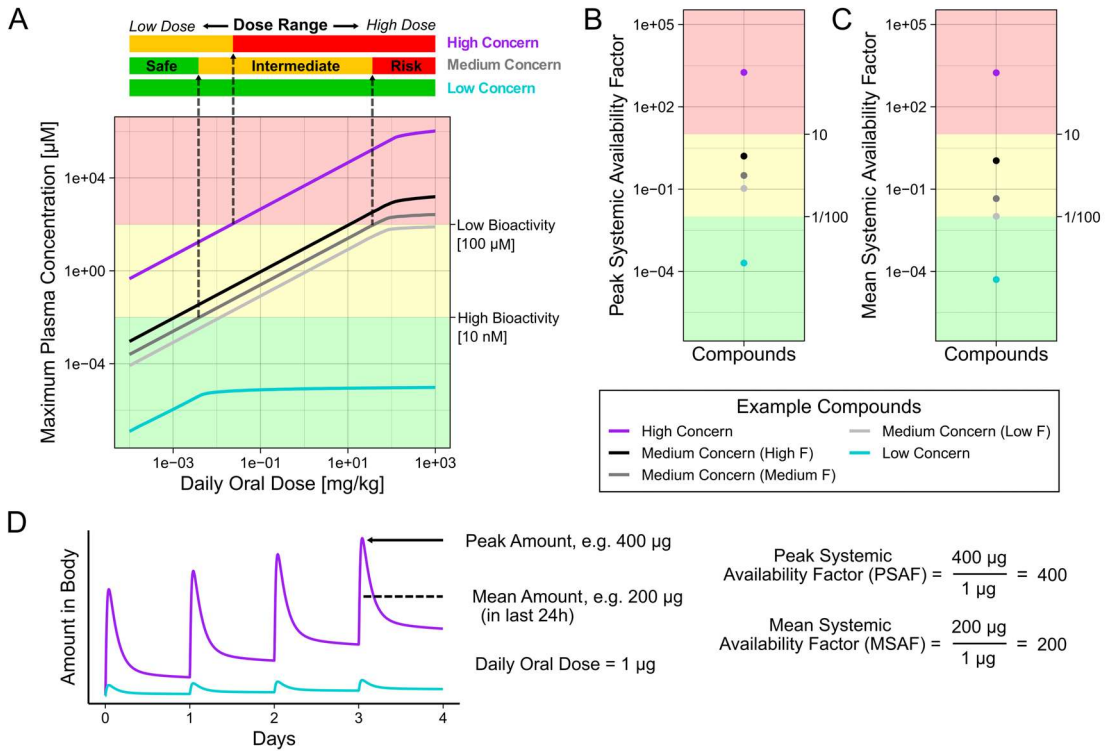


Figure 18: Visual representation for the rationale behind distinguishing different Potential Systemic Availability concern classes. (A) Simulated maximum plasma concentration (Cmax) against oral doses after daily administration for five years of five example compounds as described in Table 2. Marked as safe is the dose range for which the corresponding Cmax value remains below the high bioactivity concentration (10 nM). Risk is the dose range where the Cmax exceeds the low bioactivity threshold (100 µM), and intermediate is the area in between. (B) and (C) show the calculated Peak Systemic Availability and Mean Systemic Availability Factors of the five example compounds, respectively. (D) Example of Peak and Mean Amounts derived from a time-course profile, and calculation of the corresponding PSAF and MSAF values. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

Quantification of Potential Systemic Availability

Having defined the PSA concern classes, a more rigorous conceptual basis for chemical assessment is established. From a practical perspective, however, the question remains how this property can be evaluated efficiently, particularly for large numbers of compounds. Three considerations are especially important:

i) Organ-agnostic assessment. PSA classification should not be based solely on concentrations in a single organ. While Figure 18A illustrates a simplified scenario of a single systemic concentration, the reality is that the human body is composed of multiple compartments in which concentrations can differ substantially. Since the relevant target organs of toxicity are often unknown *a priori*, selecting a single organ for classification is neither trivial nor necessarily appropriate. In fact, it may be misleading. Many bioaccumulating chemicals are highly lipophilic, resulting in higher concentrations in adipose tissue but lower

levels in plasma. If only plasma concentrations were considered, such compounds could erroneously appear less concerning than less lipophilic substances.

ii) Consideration of long-term exposure. To adequately capture differences in the potential of compounds to bioaccumulate, repeated exposures over extended timeframes must be taken into account, rather than focusing only on short-term or single exposures.

iii) Minimisation of arbitrary assumptions. The output should reflect an intrinsic compound property that is as independent as possible from arbitrary assumptions, such as fixed external doses or concentration thresholds. Reliance on such subjective values risks controversy and may undermine broad acceptance of the classification approach.

Based on these considerations, we define the Peak Systemic Availability Factor (PSAF) as

$$\text{PSAF} = \frac{\text{Peak Amount of Compound in the Body during 5 years of Daily Oral Exposure}}{\text{Daily Exposure Dose}} \quad \text{(eq. 9)}$$

This metric can be interpreted as the peak total body burden of a compound relative to its daily exposure. It satisfies the three outlined requirements: (i) it accounts for the total amount of compound distributed in the body, rather than relying on arbitrarily selected organs; (ii) it reflects accumulation under repeated exposure; and (iii) it is unitless and represents an intrinsic compound property. As illustrated in Figure 18B, the metric, here referred to as the Potential Systemic Availability Factor (PSAF), quantitatively distinguishes between the five example compounds: compound 1 has the lowest PSAF (2×10^{-4}), compounds 2-4 range from 0.1 to 1.6, and compound 5 shows the highest value (1789). Figure 18D provides a visual representation of how the PSAF can be derived from a time-course profile.

Although the PSAF itself avoids arbitrary assumptions in its calculation, assigning compounds to the three PSA concern classes ultimately requires threshold values, which are inherently arbitrary. For practical purposes, we defined thresholds at 1/100 and 10. Thus, a compound is classified as high PSA concern if its total body burden exceeds the daily oral exposure by more than tenfold, and as low PSA concern if it never exceeds 1% of the daily exposure. Compounds falling between these thresholds are considered medium PSA concern.

To further explore the concept of PSAF and to contrast it with an alternative evaluation metric, we additionally define the Mean Systemic Availability Factor (MSAF) as

$$\text{MSAF} = \frac{\text{Mean Amount of Compound in the Body in the last 24h after 5 years of Daily Oral Exposure}}{\text{Daily Exposure Dose}} \quad \text{(eq. 10)}$$

The Mean Systemic Availability Factor (MSAF) is conceptually similar to the PSAF but evaluates the time-averaged systemic amount of a compound rather than its peak value. This makes the MSAF more sensitive for distinguishing between chemicals, in a way analogous to bioavailability. For example, compounds with similar PSAF values may differ substantially in their clearance rates, leading to lower MSAF values for those that

are eliminated more rapidly. For consistency, we applied the same classification thresholds (10 and 1/100), although different thresholds could be justified for the two metrics. As shown in Figure 18C, the MSAF also correctly separates the five example compounds into their respective concern classes, although the values for the three medium-concern compounds are positioned closer to the low-concern compound than observed with PSAF.

Evaluation of Potential Systemic Availability classifications

To demonstrate the applicability of the proposed classification metrics, we calculated PSAF and MSAF values for a diverse set of substances, namely the 150 EPAA NAM Designathon compounds. For this purpose, we applied a previously described high-throughput PBK (HT-PBK) modelling approach [140], which enables the rapid generation of generic PBK models and thereby the high-throughput prediction of *in vivo* concentrations. Importantly, our evaluation framework is independent of the specific parameterisation approach and can be applied with any PBK model.

HT-PBK predictions were obtained for 139 of the 150 compounds. The remaining 11 substances, consisting of inorganics, metals, or organometallics, could not be modelled because the required physicochemical property predictions were unavailable from *in silico* tools. Predicted PSAF and MSAF values are shown in Figure 19A and Figure 20A, respectively. Based on PSAF, only 2 compounds were classified as low concern (<1%), 10 compounds as high concern (>10), and the remaining 127 as medium concern. Using MSAF produced broadly similar results, with the same 10 compounds classified as high concern, but a larger number falling into the low concern category (21 instead of 2). As expected, MSAF values showed a broader distribution than PSAF values.

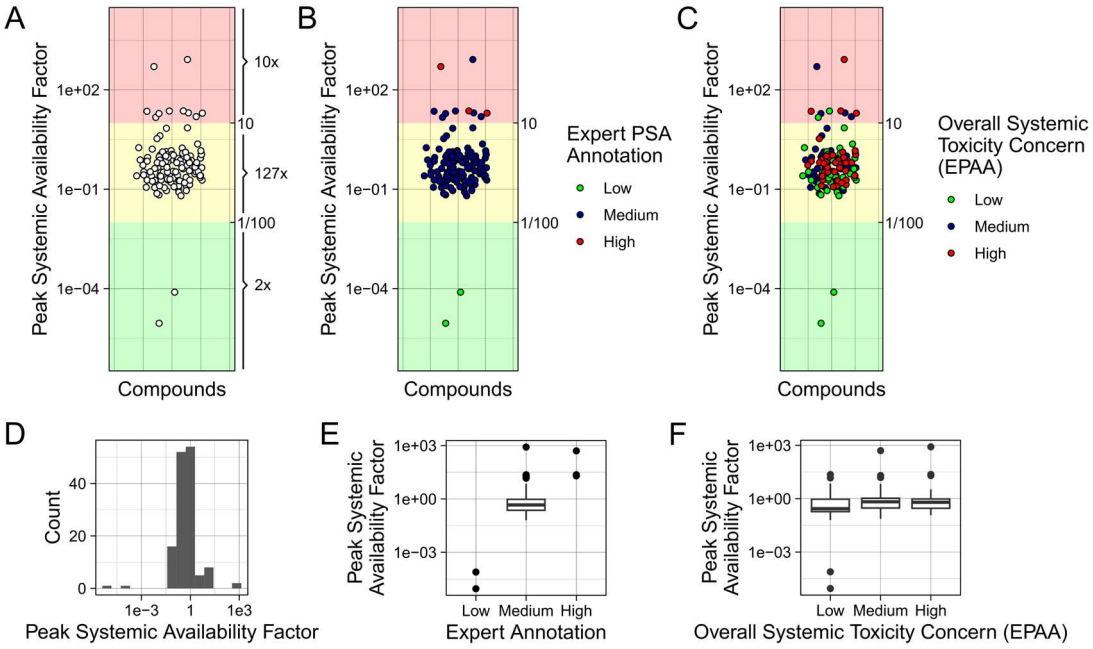


Figure 19: Peak Systemic Availability Factors of the 139 EPAA Designathon Compounds for which HT-PBK predictions could be made. (A,D) Distribution of PSAF values with jittered x-axis. (B) Coloured by manually annotated reference classification. (C) Coloured by overall systemic toxicity concern given by EPAA. (E, F) Comparison of PSAF values of the different manually annotated Potential Systemic Availability or overall concern classes. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

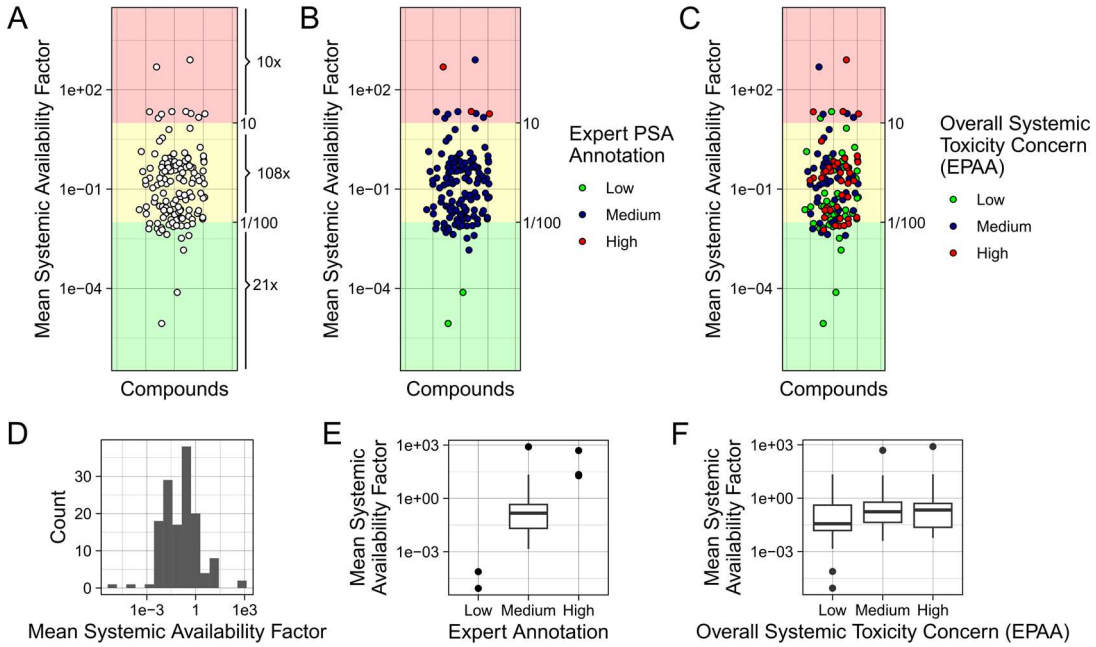


Figure 20: Mean Systemic Availability Factors of the 139 EPAA Designathon Compounds for which HT-PBK predictions could be made. (A,D) Distribution of MSAF values with jittered x-axis. (B) Coloured by manually annotated reference classification. (C) Coloured by overall systemic toxicity compound concern given by EPAA. (E, F) Comparison of MSAF values of the different manually annotated Potential Systemic Availability or overall concern classes. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

To assess the plausibility of our predicted classifications, we manually established a reference classification of the NAM Designathon compounds based on review of publicly available data and expert judgment. Details of this procedure are provided in Methods chapter 8. Establishing such reference categories proved challenging, as the type, quality, and consistency of available data varied considerably. For some compounds, high-quality human *in vivo* data after repeated oral exposure were available, whereas others were represented only by lower-quality animal or *in vitro* data, or lacked toxicokinetic information altogether.

Endogenous compounds (e.g. amino acids, sugars, fatty acids) were classified as medium PSA concern, reflecting the assumption that these substances are efficiently absorbed but not prone to bioaccumulation. One possible exception was ergocalciferol (vitamin D2), which is fat-soluble and can accumulate in the body. However, given its half-life of only a few days, we considered its accumulation potential insufficient to justify a high-concern classification and instead assigned it to medium concern.

All other compounds with available toxicokinetic information were reported to be absorbed orally to some extent, so none were considered low concern. Their reported clearances, half-lives, and tissue accumulation potentials varied substantially. Compounds reported to bioaccumulate strongly or to have half-lives of weeks or longer were examined more closely for potential classification as high concern. On this basis,

three compounds (perfluorooctanoic acid, perfluoroheptanoic acid, and hexabromocyclododecane) were classified as high PSA concern.

Two compounds with very high molecular weights (>1500 g/mol; polyoxyl(40)stearate and 2-hydroxypropyl-β-cyclodextrin) were classified as low concern, under the assumption that they are not absorbed orally to any relevant extent. For the remaining small molecules (<1000 g/mol) lacking specific toxicokinetic data, we defaulted to medium concern. Some compounds carried additional uncertainties due to unusual properties, such as depuration by evaporation (decamethylcyclopentasiloxane) [231] or potential explosiveness (dipicrylamine), but these did not alter the overall classification scheme. In total, the manual reference classification assigned 2 compounds as low PSA concern, 134 as medium, and 3 as high.

Comparing our HT-PBK-derived PSAF classifications with the expert judgment-based reference categories showed a high level of agreement. Out of 139 compounds, 132 were classified identically (Table 3). Seven compounds that were manually assigned to the medium PSA concern category were predicted as high concern by HT-PBK modelling. These represent false positives, which are generally less problematic for safety assessment. Importantly, no compound was predicted to have a lower concern level than indicated by expert judgment. In contrast, MSAF-based classifications yielded some false negatives. Several compounds were predicted as low concern, whereas experts had classified them as medium (Supplementary Table 1). Closer inspection of these cases provided no supporting evidence for a low-concern classification, suggesting that these discrepancies reflect mispredictions.

We further compared our predicted PSA classifications to the overall systemic toxicity concern categories provided by the EPAA. Unlike PSA, which reflects only the toxicokinetic dimension, the EPAA categories integrate both PSA and toxicodynamic bioactivity information and are based on traditional toxicity data. Therefore, exact correspondence was not expected. Nonetheless, since PSA is one of the two factors contributing to overall systemic toxicity concern, broad consistency should be observed. Specifically, no compound of overall high concern should be classified as low PSA, and no compound of overall low concern should be classified as high PSA. Our results were consistent with the first expectation: the two compounds predicted as low PSA were also classified as low overall concern by the EPAA. However, among the ten compounds predicted as high PSA by HT-PBK modelling, two were categorised as low overall concern. Assuming the EPAA classifications are correct, this indicates that these HT-PBK predictions represent false positives.

The two false-positive compounds were decamethylcyclopentasiloxane (CAS 541-02-6) and sodium myristoyl glutamate (CAS 38517-37-2). Both showed high predicted PSAF values because their hepatic clearance was predicted as zero by OPERA [121]. For decamethylcyclopentasiloxane, this prediction was consistent with *in vitro* hepatic clearance data from httk [49] and it has been investigated for potential to bioaccumulate [232,233]. However, studies have shown that in air-breathing organisms it can be eliminated through exhalation [231], a pathway not represented in our generic HT-PBK models. This likely led to underestimation of its clearance. For sodium myristoyl glutamate, no *in vitro* hepatic clearance data were available in httk, but a similar explanation is plausible.

Overall, compounds of low overall concern tended to have lower PSAF and MSAF values, although there was substantial overlap between the concern groups. This trend was more pronounced for MSAF, which in some cases misclassified high overall concern

compounds as low concern. Thus, while MSAF may offer greater sensitivity for identifying lower concern compounds, this comes at the cost of a higher false-negative rate. By contrast, the more conservative, peak-based PSAF metric avoided false negatives, making it the more reliable basis for health-protective classification.

Table 3: Comparison of classifications of 139 EPAA Designathon Compounds based on HT-PBK predictions using the Peak Systemic Availability Factor (PSAF) against expert judged classification of Potential Systemic Availability (PSA). Table reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

	HT-PBK PSAF: Low	HT-PBK PSAF: Medium	HT-PBK PSAF: High	*Sum*
Expert PSA class: Low	2	0	0	2
Expert PSA class: Medium	0	127	7	134
Expert PSA class: High	0	0	3	3
Sum	2	127	10	139

12.3 Discussion

We proposed a definition of PSA concern classes and introduced a quantitative method to measure this property. Using expert judgment, we manually annotated a benchmark dataset of diverse chemicals from the EPAA NAM Designathon and compared these annotations with classifications predicted by HT-PBK modelling. The comparison demonstrated that HT-PBK can reliably predict PSA classifications in a high-throughput manner.

Our conceptual framework builds directly on the central principle of toxicology that chemicals can only cause adverse effects if their *in vivo* concentrations exceed levels required for bioactivity (“the dose makes the poison”). We showed that some compounds are unlikely to ever reach such concentrations and therefore could reasonably be treated more permissively in safety assessment. In contrast, substances with a strong potential to accumulate in the body can reach harmful concentrations even at low daily doses, despite having only modest intrinsic bioactivity. These compounds warrant a more precautionary approach.

Bioaccumulation also carries additional risks that are particularly undesirable from a safety perspective. If the bioactivity of a bioaccumulating chemical is underestimated, its adverse effects may only become evident years after widespread exposure, once steady-state body burdens are reached. At that point, the compound may already be in extensive use, complicating mitigation and regulatory response. This challenge is exemplified by per- and polyfluoroalkyl substances (PFAS), where delayed recognition of their persistence and accumulation has led to widespread environmental and human exposure [234,235].

Our definition of PSA reflects a health-protective interpretation, assuming that even a single peak concentration may be sufficient to trigger adverse effects *in vivo*. However, evidence indicates that for some chemicals, toxicity is more closely associated with cumulative exposure over time (AUC) rather than with peak levels (Cmax) [236,237]. This raises the possibility that an AUC-based metric could be more appropriate for classification. To examine this, we tested whether a time-averaged approach could provide meaningful classifications. Our analysis showed that the simplest option, equating PSA with bioavailability, was not suitable. Instead, we proposed a method that integrates exposures across all organs and accounts for accumulation from repeated dosing.

For the benchmark dataset, classifications based on peak (PSAF) and time-averaged (MSAF) values gave broadly comparable results. Most NAM Designathon compounds clustered in the intermediate range, with only a few falling into the extreme categories. Both metrics therefore appear feasible for chemical classification. Notably, compounds of low overall systemic toxicity concern tended to have lower MSAF values, and the separation between low- and higher-concern groups was slightly clearer with MSAF than with PSAF. This is consistent with the expectation that substances with lower systemic exposure are less likely to cause adverse effects, all else being equal.

We also identified exceptions to the general trend, where compounds classified as high overall concern exhibited low MSAF values. These likely represent cases where toxicity is driven by peak concentrations rather than cumulative exposure. For example, phosphamidon, an organophosphate pesticide, is efficiently cleared [238] and its MSAF-based classification suggests low PSA concern. In the EPAA Designathon, however, it was assigned high systemic toxicity concern. This discrepancy is consistent with the

compound's known mode of action as a cholinesterase inhibitor, where adverse effects occur above defined concentration thresholds [239]. In such cases, toxicity is plausibly driven by Cmax rather than AUC, meaning that only peak-based classification correctly reflects concern. Since it is typically unknown in advance whether toxicity is determined by peak or time-averaged exposure, we conclude that a peak-based approach is preferable for safety assessment. This more conservative metric avoids false-negative classifications, which, although rare, would undermine both regulatory confidence and public trust.

To provide a reference for evaluating our predicted classifications, we manually annotated PSA classes of the Designathon compounds. Consistent with the modelling results, most compounds were best placed into the medium PSA concern category. These cannot *a priori* be considered safe or unsafe, as they may fall into any of the three systemic toxicity concern classes depending on their bioactivity. This outcome appears plausible given that most compounds in the dataset are small molecules such as endogenous metabolites, food ingredients, drugs, and pesticides, substances that are typically bioavailable and often designed or selected for bioactivity. For instance, the food additive tartaric acid, the drug ketoconazole, and the pesticide glufosinate ammonium all had PSAF values of ~80% and were classified as medium PSA concern, despite differing in their overall systemic toxicity classifications by the EPAA (tartaric acid: low, ketoconazole and glufosinate ammonium: high). At first glance, this result may seem counterintuitive, as the Designathon dataset was intentionally balanced across the three systemic toxicity concern classes. One might therefore have expected a more even distribution across PSA categories.

It is important to emphasise that the data available for manual reference classifications were highly heterogeneous, making confident assignments challenging. Our annotations required case-by-case interpretation of disparate information without standardised criteria, and subjectivity was unavoidable. Nevertheless, the overall trends appear robust. In future work, additional data sources such as human biomonitoring could provide more objective evidence for classification, although such analyses were beyond the scope of this study.

Because PSAF and MSAF values were derived using HT-PBK modelling, our results are subject to the general limitations of this approach. Most importantly, prediction accuracy depends on the reliability and applicability of the *in silico* tools that provide input parameters to the PBK models. Errors in input parameters inevitably propagate into the model outputs. Improving these predictions will therefore require continued efforts to validate *in silico* tools, broaden their applicability domains, and enhance their accuracy. For the present application, it is particularly important that tools not only predict average values well but also capture extreme cases, such as compounds with exceptionally low or high clearance, fraction unbound, or intestinal permeability. These extremes are often decisive for assigning substances to low or high PSA concern classes. However, current *in silico* tools frequently perform poorly in this regard, as their training datasets underrepresent such outliers [240,241]. Addressing this limitation is essential for improving the robustness of *in silico* classifications like those presented here.

An alternative would be to parameterise PBK models using *in vitro* data. Yet, *in vitro* methods also have their own biases and may face similar challenges in reliably capturing extreme chemical properties [242,243]. When we compared PBK simulations parameterised with *in silico* predictions against those using *in vitro* measurements of fraction unbound and hepatocyte clearance, which were available for a subset of

compounds in the US EPA's httk package [49,53], no meaningful differences were observed (Supplementary Figure 19).

Limitations of the *in silico* tools used here also explain why certain groups of substances, such as metals or inorganics, could not be assessed with our approach. Nevertheless, coverage was high, with predictions possible for 139 out of 150 compounds of interest, suggesting that this limitation may not pose a major barrier for most safety assessment applications. We also identified several non-standard processes that may influence PSA concern but were not represented in our generic PBK models. These include depuration of volatile compounds via exhalation, chemical instability, gut metabolism, and biliary clearance. Ideally, all relevant processes should be incorporated into PBK models to enable comprehensive and reliable evaluations. However, such detailed modelling was beyond the scope of this work, and predictive tools for many of these processes remain under development. The same applies to transporter activities, which can substantially affect local organ concentrations and, in some cases, be decisive for *in vivo* toxicity [86,244].

Finally, a complete assessment of human exposure should consider multiple routes, including dermal and inhalation uptake. In this study, we focused exclusively on oral absorption, given its high relevance. Nonetheless, the same approach can be applied to other routes. More generally, PBK modelling offers the flexibility to compare and integrate exposures across multiple routes and demographic groups, supporting a holistic evaluation of chemical safety [245].

An alternative to PBK modelling is the use of rule-based classification schemes or regression models based on selected chemical properties. Such approaches are widely applied, for example, in predicting bioaccumulation in aquatic organisms and bioconcentration factors [246]. However, they rely on subjectively defined thresholds and the choice of a limited set of compound properties [247]. In contrast, PBK modelling integrates multiple molecular properties that collectively determine quantitative *in vivo* behaviour. This makes it more suitable for capturing systemic outcomes and the complex interplay between parameters. Accordingly, PBK modelling is increasingly used to evaluate human exposure to chemicals [53,125,248,249], although the optimal method will ultimately depend on context, available data, and accuracy requirements.

For both predicted and manually annotated classifications, most compounds fell into the medium PSA concern class, while their overall systemic toxicity concerns were evenly distributed among low, medium, and high categories. This indicates that systemic toxicity is mainly driven by bioactivity differences rather than toxicokinetics, which raises questions about the utility of TK-based classifications alone. Nonetheless, two considerations are important. First, identifying high PSA concern compounds remains valuable, even if they constitute a minority, as it enables prioritisation of potentially higher-risk chemicals, which is a critical advantage given the large number of untested substances on the market. Second, within the broader chemical universe, many compounds are likely to fall into the low PSA concern category but may not have been captured in the NAM Designathon dataset. Such compounds may be underrepresented because (i) their extreme chemical properties, such as very low permeability, make them difficult to study *in vitro*, and (ii) their low systemic availability means they are unlikely to cause adverse effects severe enough to warrant testing or classification. Our classification framework is consistent with this treatment. For example, polymers would generally be expected to fall into the low PSA concern class under our definition, as they do not readily cross biological membranes and exhibit minimal systemic toxicity. This

aligns with existing regulatory exemptions for polymers [250], although such compounds were not represented in our evaluation dataset.

Finally, the practical application of the presented method within a novel framework for regulatory decision making remains to be explored. There is an ongoing effort to define how chemical safety assessment can be restructured to ultimately phase out animal testing [21,22,226,251,252]. Our results show that considering TK properties in the form of the PSAF can support the *a priori* prioritisation of certain chemicals and the de-prioritisation of others. This could serve as a first step for classification or triaging of substances within a tiered safety assessment framework. Because the HT-PBK modelling approach used here only requires *in silico* predictions of a limited set of compound properties, it allows the rapid classification of large numbers of currently unassessed chemicals. Even when parameterised with *in vitro* data for greater reliability, such analyses could still be conducted efficiently at scale [53,253].

At the same time, many compounds cannot be conclusively prioritised or deprioritised, as most were classified into the medium PSA concern category. From this group, assignment to any of the overall systemic toxicity concern classes remains possible and depends on compound-specific bioactivity. Where bioactivity is also neither clearly low nor clearly high, more detailed evaluation is needed, for instance by comparing predicted internal concentrations to bioactivity thresholds and deriving bioactivity-exposure ratios (BERs) or safe exposure levels [183]. Such in-depth analyses can build directly on the HT-PBK models used for initial PSA classification.

In the absence of bioactivity and internal exposure data, our approach necessarily relied on conservative assumptions. Future integration of bioactivity data into this framework is expected to enable more definitive classification of compounds into low or high systemic toxicity concern [254]. This will ultimately allow systematic comparison of NAM-based classifications with those derived from traditional toxicological studies.

13. High-throughput Prediction of Drug-Induced Liver Injury

Abstract

Drug-induced liver injury (DILI) is a leading cause of drug attrition and a major safety concern in clinical practice. Yet, existing preclinical prediction methods, including heuristic screening rules, *in vitro* assays, machine learning models, and animal studies, remain unreliable. Here, we show that DILI prediction performance can be substantially improved by integrating *in vitro* toxicity data with pharmacokinetic (PK) information. In a retrospective analysis of 241 marketed drugs, we demonstrate that the ratio of maximum plasma concentration (Cmax) to the lowest observed *in vitro* toxicity (cytotoxicity, mitochondrial toxicity, or bile salt export pump (BSEP) inhibition) strongly correlates with clinical DILI outcomes, achieving receiver operating characteristic (ROC) areas under the curve (AUCs) of up to 0.96. Comparable predictivity (ROC AUC up to 91%) was obtained prospectively when Cmax values were predicted *in silico* using high-throughput physiologically based kinetic (HT-PBK) modelling. Furthermore, dynamic simulations of bile acid homeostasis identified compounds likely to cause DILI through BSEP inhibition, providing additional mechanistic insight beyond classification. Together, these results highlight that coupling *in vitro* toxicity assays with PK data, experimentally measured or model-predicted, offers a mechanistically interpretable, animal-free, and scalable approach for DILI risk assessment in early drug development.

Chapter 13 is adapted from:

Geci, R., Sayin, A. Z., Schaller, S. & Kuepfer, L. *Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury.* Arch Toxicol 100, 2029–2046; 10.1007/s00204-026-04305-2 (2026).

Author contributions:

Ahenk Zeynep Sayin performed bile acid metabolism model simulations. Stephan Schaller and Lars Kuepfer provided overall supervision and guidance. René Geci performed all remaining simulations and analyses, wrote these chapters and created the figures and tables unless otherwise specified.

13.1 Introduction

Drug-induced liver injury (DILI) is one of the most severe adverse drug reactions, representing a leading cause of clinical trial termination, post-marketing drug withdrawals, and acute liver failure in patients [88,255]. It has been estimated to cost the pharmaceutical industry billions of dollars annually while also contributing to a significant number of patient deaths worldwide. Despite extensive efforts in toxicology and pharmacology, the ability to reliably predict DILI risk at the preclinical stage remains limited.

In clinical practice, DILI presents with diverse phenotypes, typically classified as hepatocellular, cholestatic, or mixed forms of injury. Moreover, many cases are idiosyncratic, arising without a clear dose dependence or consistent clinical presentation [256]. This heterogeneity makes DILI particularly difficult to detect in conventional preclinical evaluations, and it complicates risk prediction during drug development [23]. Given its impact on both patient safety and drug attrition, improved predictive strategies remain a critical unmet need. Animal studies have historically been the gold standard for assessing drug toxicity, yet their concordance with human liver toxicity outcomes is modest, with reported concordance rates between 40% and 70% [5,107]. This limited predictive value, combined with ethical concerns and increasing regulatory and societal pressure to phase out animal testing, has reinforced the need for alternative, human-relevant approaches to assess liver toxicity risks.

Early preclinical assessments of DILI risk often relied on simple heuristic rules, based on readily accessible drug properties such as lipophilicity, dose, or daily exposure [257]. Although easy to apply, these rules lack a mechanistic basis and are prone to misclassification of compounds. Such rules therefore produce both false positives, leading to unnecessary attrition of viable drug candidates, and false negatives, allowing hepatotoxic compounds to progress into costly clinical trials. In an effort to improve predictivity, a range of *in vitro* assays have been developed to test for toxicity-related endpoints, such as cytotoxicity, mitochondrial dysfunction, and inhibition of the bile salt export pump (BSEP) transporter [24,258]. These assays are now routinely applied in preclinical safety workflows [259,260]. However, they are typically evaluated in isolation and without integration of pharmacokinetic exposure considerations, which limits their predictive accuracy [25,261]. *In vitro* toxicity findings may reflect concentrations that are never reached at relevant drug doses *in vivo*, while conversely, compounds causing subtle *in vitro* effects may still trigger DILI. Without accounting for realistic *in vivo* exposures, *in vitro* assays struggle to reliably distinguish clinically relevant from irrelevant findings.

BSEP inhibition provides a succinct example of these challenges. As a key hepatic transporter in bile acid homeostasis, BSEP dysfunction is mechanistically linked to cholestatic forms of DILI [262–264]. Nevertheless, the predictive performance of *in vitro* BSEP inhibition measurements for clinical outcomes has been inconsistent, in some cases not exceeding that of simple rule-based approaches [265]. This has led some to question the value of BSEP inhibition assays as predictive tools [266]. For mechanisms that are less well understood than BSEP, translation from *in vitro* endpoints to clinical DILI outcomes is even more challenging.

Machine learning (ML) has recently emerged as another strategy for DILI prediction, using models trained on chemical descriptors, *in vitro* assays, animal or clinical toxicity data. While some ML-based models show promising predictive performance [167,267–

269], they typically function as “black boxes,” offering limited mechanistic transparency [270,271]. This lack of interpretability hampers their prospective application, makes extrapolation to novel chemical classes uncertain, and presents major obstacles for regulatory acceptance. Moreover, most published ML models are trained and validated on relatively small or biased datasets, and systematic prospective evaluations remain rare.

In summary, existing DILI assessment strategies, be it animal-based, rule-based, *in vitro*, or ML-driven, fall short of providing the predictive performance, mechanistic interpretability, and regulatory acceptance needed for effective preclinical decision making. There is therefore a pressing need for approaches that combine human relevance, mechanistic insight, and scalability. In this work, we address this gap by demonstrating that accurate DILI risk predictions become possible when *in vitro* toxicity assays are integrated with pharmacokinetic (PK) information. We explore the integration of mechanistic insights from *in vitro* toxicity assays with PK information, thereby linking cellular toxicity data to systemic exposure. We further demonstrate how high-throughput physiologically based kinetic (HT-PBK) modelling can be used to predict PK parameters *in silico*, enabling prospective DILI risk assessment even for novel compounds. This combined approach addresses current limitations in predictivity and transparency, and it provides a framework for more reliable, mechanistic, and human-relevant DILI evaluation.

13.2 Results

Overview of the collated hepatotoxicity dataset

After integrating data from seventeen *in vitro* hepatotoxicity studies, we obtained potency measurements for 487 unique compounds. For 355 of these, a specific toxicity potency value was reported, while the remaining compounds showed no measurable effect even at the highest tested concentration. For the 355 active compounds, we retrieved information on the highest recommended oral dose, the corresponding maximum plasma concentration (Cmax), and their clinical DILI outcomes from DILIrank [170]. This resulted in a final integrated dataset of 241 unique drugs for which oral dose, Cmax, clinical DILI outcome, and at least one *in vitro* toxicity value were available (Figure 21).

Cytotoxicity was the most frequently reported *in vitro* endpoint (205/241 compounds), measured across a range of cell systems. BSEP inhibition (170/241) and mitochondrial toxicity (93/241) were also well represented. Some drugs were tested in multiple studies and had up to 19 distinct potency values available, whereas others had data from only a single assay. The dataset included compounds from all DILIrank concern categories, with the majority classified as Most-DILI concern (34.6%), Less-DILI concern (28.0%), or Clinical Development Failures (11.5%, as previously assigned [153]). The overrepresentation of DILI-positive compounds reflects both the limited testing of DILI-negative compounds *in in vitro* assays and the frequent absence of measurable activity for those that were tested. Nevertheless, 19.8% of the final dataset consisted of No-DILI concern drugs with at least one measurable *in vitro* potency value.

Overall, the 241 compounds represented a structurally diverse set of small molecules with broad variation in molecular properties (Figure 21C-M). Because the hepatotoxicity potential of Ambiguous-DILI concern drugs remains clinically inconclusive, we restricted our analyses to compounds from the conclusive DILIrank categories.

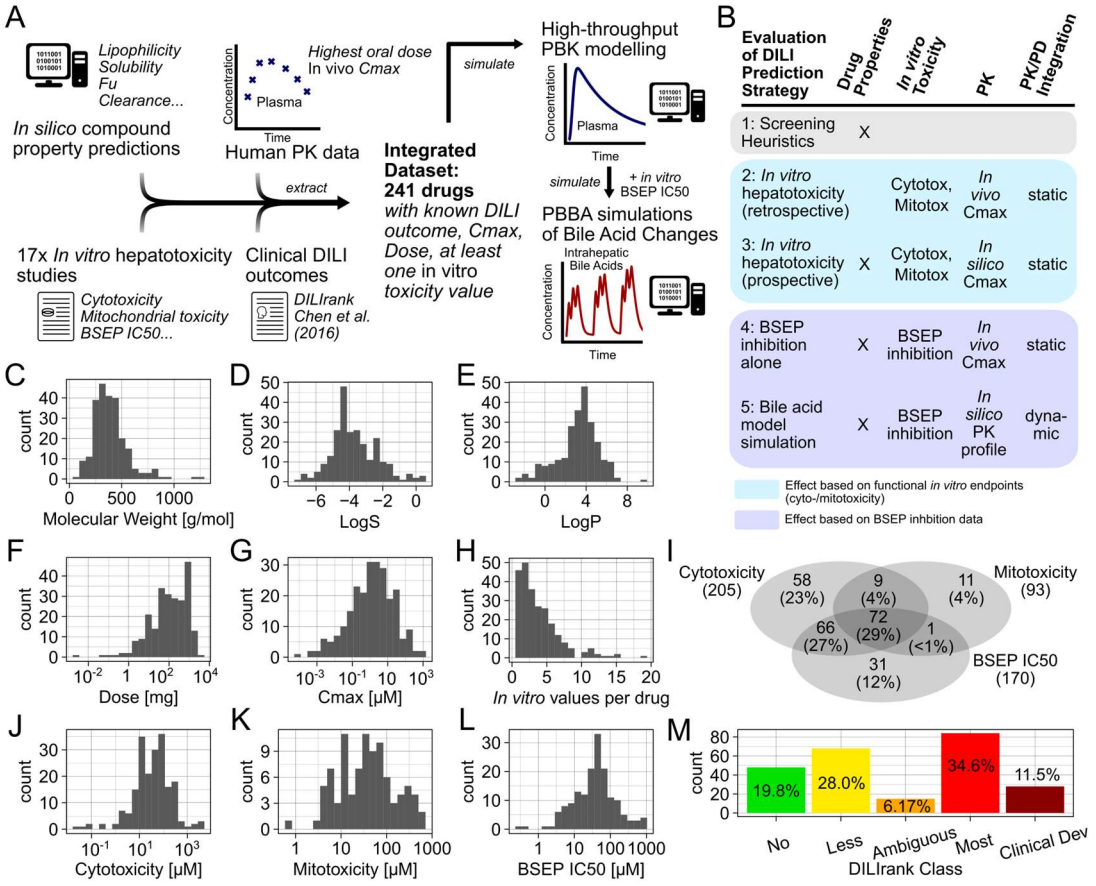


Figure 21: Data collection, analysis workflow and drug properties. Various literature datasets were integrated to obtain a dataset of 241 drugs with complete information on *in vitro* hepatotoxicity, clinical DILI outcome and pharmacokinetics (A). DILI prediction strategies of varying complexity were evaluated using the integrated dataset (B). Statistical plots of various drug properties and parameters of the collected dataset (C-M). Figure reproduced from [272]. No changes made.

Evaluation 1: Simple heuristic screening rules poorly predict DILI

Numerous heuristic screening rules have been proposed for pragmatic prediction of DILI at the preclinical stage, often based on readily available properties such as lipophilicity or clinical dose. Since such heuristics are commonly applied in pharmaceutical research, and will serve as a benchmark for later evaluations, we first examined their predictive performance in our integrated dataset.

One frequently cited heuristic suggests that drugs with higher LogP values are more likely to cause DILI [257]. However, when comparing LogP across DILIrank classes, we found no meaningful separation (Figure 22A). A second heuristic assumes that higher clinical doses are associated with greater DILI risk [273]. For dose, clearer differences were observed: No-DILI concern drugs had the lowest recommended doses, while Most-DILI concern compounds had the highest (Figure 22B).

LogP and dose have also been combined in the so-called “Rule of Two” [257]. Applying this rule improved specificity compared to either property alone, but this came at the expense of sensitivity (Figure 22E; Supplementary Table 2).

Cmax values have been proposed as another heuristic predictor of DILI risk. Consistent with previous reports [154], we found that Cmax outperformed all other simple rules. *In vivo* Cmax discriminated No-DILI from Most-DILI and Clinical Development Failure drugs with a specificity of 89.6% and a sensitivity of 82.1%. This superior performance compared to dose was expected, since dose is only a surrogate, whereas Cmax directly reflects systemic exposure. Nonetheless, the magnitude of this predictive power was unexpected, as it suggests that toxic potency plays a limited role in determining clinical DILI outcomes. From a mechanistic standpoint, both systemic concentration and intrinsic toxic potency should contribute to *in vivo* toxicity and considerable effort has been devoted to developing *in vitro* systems to quantify the latter. Yet, all *in vitro* toxicity values collected here showed limited ability to discriminate between DILIrank classes (Supplementary Figure 20). Even when consolidating *in vitro* values into a single lowest potency per compound, the performance improved compared to individual assays but remained inferior to simple heuristics such as Cmax (Supplementary Table 2).

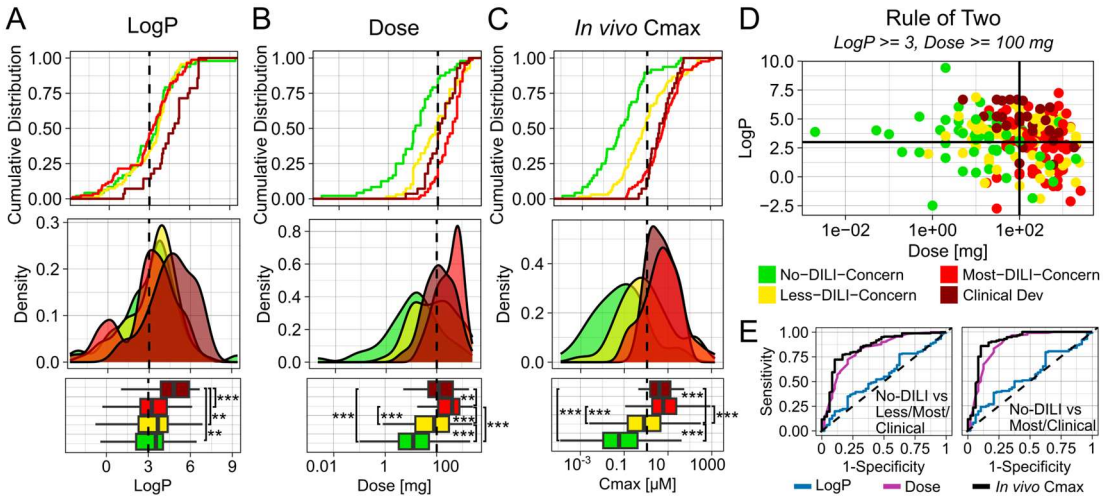


Figure 22: Predictivity of simple DILI screening heuristics (evaluation 1). Cumulative distribution, probability density and boxplots of properties used for heuristic drug candidate screening, like LogP (A), Dose (B) and *in vivo* Cmax (C) for drugs of the different DILIrank classes. LogP values were predicted using ADMETLab. Distribution of drugs across the Rule of Two quadrants (D). ROC curves of the different DILI screening heuristics for distinguishing No-DILI compounds from Less-/Most-DILI/Clinical Development Failures [153], or only from Most-DILI/Clinical Development Failures (E). Boxplots show the median and interquartile range; whiskers extend to 1.5×IQR. Outliers were excluded from the plot for clarity but were included in statistical analyses. Statistical comparisons on log-transformed values were performed using pairwise Wilcoxon rank-sum tests with Holm correction: ** $p < 0.01$, *** $p < 0.001$. Figure reproduced from [272]. No changes made.

Evaluation 2: Cmax to *in vitro* toxicity ratios strongly predict DILI

Although *in vitro* toxicity values alone did not outperform simple heuristics for distinguishing between DILIrank classes, we next assessed whether they would provide additional predictive value when integrated with pharmacokinetic information, specifically *in vivo* Cmax values. Such exposure-activity integrations are increasingly applied in toxicology [183,274,275], but their utility for DILI prediction has remained debated, as previous studies reported mixed results [154,155]. In several cases, predictive accuracy did not clearly exceed that of simple property-based heuristics, raising questions about their practical benefit. For this evaluation, we focused exclusively on functional *in vitro* toxicity endpoints (cytotoxicity and mitochondrial toxicity), initially excluding BSEP inhibition data.

Calculating the ratio of *in vivo* Cmax to the lowest available functional *in vitro* toxicity value for each compound revealed strong correlations with clinical DILI outcomes (Figure 23A). Ratios differed significantly between DILIrank classes (Figure 23C) and increased with severity: No-DILI compounds had the lowest ratios, Less-DILI compounds were intermediate, and Most-DILI compounds showed the highest values. Receiver operating characteristic (ROC) values confirmed this trend, with areas under the curve (AUC) of 96% for distinguishing No- from Most-DILI and Clinical Development Failure drugs, and 90% for separating No- from all positive classes combined (Figure 23D). The optimal threshold determined by Youden's index [276] was 1.1%, yielding a specificity of 82% and sensitivity of 97% for discriminating No- from Most-DILI and Clinical Development Failure compounds.

From a mechanistic perspective, however, an *in vivo* concentration corresponding to only 1.1% of the *in vitro* potency appears implausibly low to cause meaningful perturbations *in vivo*. We therefore applied a more conservative and biologically reasonable threshold of 25% as the minimum ratio required for *in vivo* toxicity. Using this threshold, specificity increased to 100% while sensitivity decreased to 51%. In practice, this means that no false-positive predictions were made, but some false-negatives occurred, corresponding to clinically DILI-positive compounds with Cmax-to-toxicity ratios below 25%.

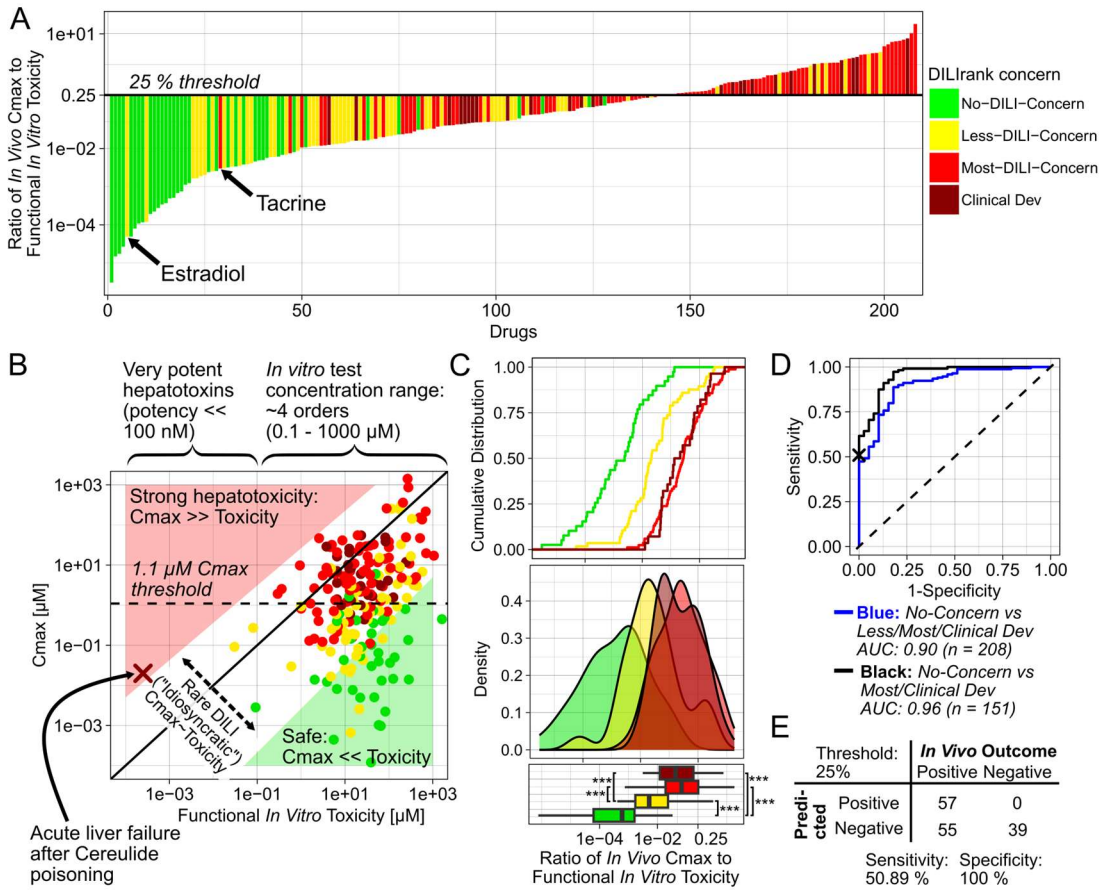


Figure 23: DILI predictivity of integrated functional *in vitro* toxicity and *in vivo* Cmax data (evaluation 2). Ratio of *in vivo* Cmax to lowest functional *in vitro* toxicity of each drug (A). Functional toxicity refers to all *in vitro* toxicity data except BSEP inhibition. *In vivo* Cmax values against lowest functional *in vitro* toxicity values of each drug (B). The red cross indicates a reported case of liver failure after cereulide poisoning. Green and red triangles mark areas where *in vivo* Cmax is much smaller or larger than functional *in vitro* toxicity, respectively. Cumulative distribution, probability distribution and boxplots of the ratio of *in vivo* Cmax to functional *in vitro* toxicity values of different DILIrank classes (C). Receiver operating characteristic curves of *in vivo* Cmax to toxic potency ratios for distinguishing No-DILI compounds from DILI-positive drugs (D). Boxplots show the median and interquartile range; whiskers extend to 1.5×IQR. Outliers were excluded from the plot for clarity but were included in statistical analyses. Statistical comparisons on log-transformed values were performed using pairwise Wilcoxon rank-sum tests with Holm correction: ** $p < 0.01$, *** $p < 0.001$. Predictive performance parameters of ratios using 25% as threshold for DILI positive classification (E). Figure reproduced from [272]. No changes made.

To better understand why some compounds were falsely predicted as DILI-negative, we examined them in more detail. Estradiol, classified as Less-DILI concern, showed the lowest Cmax-to-*in vitro* toxicity ratio among all DILI-positive compounds (4.9×10^{-5}). Estradiol is an endogenous sex hormone used in hormone replacement and gender-affirming therapy. Following oral administration of 2 mg, its *in vivo* Cmax is in the sub-nanomolar range (0.1-0.5 nM), whereas its eight reported *in vitro* toxicity values range

from 13.5 to 103 µM, hence more than 1000-fold higher than the observed Cmax. Given estradiol's endogenous nature, it is unsurprising that cytotoxicity or mitochondrial toxicity are not evident at physiological concentrations. Its observed hepatotoxicity is likely mediated by other mechanisms, such as deregulation of bile acid (BA) synthesis, metabolism, or transporter function [277], which were not captured by the *in vitro* assays considered here. In addition, conjugated metabolites such as estradiol-17β-glucuronide have been implicated in animal studies [278], but such metabolites are unlikely to be adequately generated *in vitro*.

Tacrine, a Most-DILI concern drug, also had a very low Cmax-to-*in vitro* toxicity ratio ($3x10^{-3}$). According to LiverTox [92], its mechanism of liver injury remains unclear, though extensive metabolism and the formation of toxic intermediates are thought to play central roles, with intestinal bacteria possibly contributing [279]. These mechanisms are likewise outside the scope of the functional *in vitro* assays included in our dataset. Based on this manual review, we concluded that the mechanistic threshold of 25% is more appropriate for further evaluations than the lower, statistically optimal threshold of 1.1%.

Direct comparison of *in vivo* Cmax values with *in vitro* toxicity values (Figure 23B) revealed a striking difference in their ranges. Cmax values spanned seven orders of magnitude (10^{-4} to 10^{3} µM), while *in vitro* toxicity values varied across only four (10^{-1} to 10^{3} µM). This narrower spread partly reflects the concentration ranges typically used in *in vitro* testing. As a result, both our integrated dataset and previous analyses based on these data are biased, which helps explain why dose or Cmax alone often emerge as strong predictors of DILI. If toxic potency varies little across compounds, then *in vivo* concentration inevitably becomes the main driver of observed toxicity. Only the inclusion of compounds that are toxic at very low concentrations (high-potency hepatotoxins) or that remain non-toxic even at high concentrations (e.g., endogenous metabolites, natural products) could challenge this pattern. However, such compounds are underrepresented in available *in vitro* datasets, and even if included, their true potencies would likely fall outside the fixed concentration ranges of standard assays. To highlight this, we added a case report of liver failure caused by cereulide poisoning to Figure 23B [280,281], showing that compounds with low Cmax values can still cause severe hepatotoxicity.

Our analysis further offers an explanation for why DILI is often described as "idiosyncratic," with no clear dose dependence. None of the drugs in our dataset exhibited *in vivo* Cmax values that substantially exceeded their corresponding *in vitro* toxicity thresholds, not even those in the Most-DILI concern category. If such exceedance were common, hepatotoxicity would likely be frequent and predictable, and these drugs would not have been approved for clinical use. Instead, because therapeutic Cmax values of DILI-positive drugs typically remain below their *in vitro* toxic potencies, most patients tolerate them safely. Hepatotoxicity then arises only in susceptible subgroups, such as those with genetic predispositions, comorbidities, or interacting comedications. By contrast, in pharmacology it is usually targeted to achieve *in vivo* concentrations several-fold above the *in vitro* potency to ensure predictable therapeutic effects [282]. Our findings show that almost none of the DILI-positive drugs in this dataset reach such levels, which may explain the idiosyncratic nature of clinical DILI.

Evaluation 3: *In silico* predicted Cmax enables prospective DILI evaluation

Retrospective analysis combining functional *in vitro* toxicity data with *in vivo* observed Cmax values revealed that *in vitro* assays provide useful information for the prediction of DILI, and that integrating both data types allows reliable predictions of clinical DILI risks.

However, *in vivo* Cmax values are only available retrospectively, once human clinical studies have already been conducted. A major limitation for prospective DILI prediction of new drug candidates has therefore been the absence of Cmax information prior to clinical testing. Recently, we developed a high-throughput method that integrates machine learning models with mechanistic PBK modelling to predict human pharmacokinetics, thereby enabling the *in silico* prediction of drug Cmax values before clinical data exist.

We applied this method to our integrated dataset and compared *in silico* predicted Cmax values with *in vivo* observations (Figure 24C). Consistent with previous reports, most (90%) drugs had predicted Cmax values within a tenfold range of the observed data. We then substituted the predicted values for the *in vivo* ones to calculate Cmax-to-*in vitro* toxicity ratios (Figure 24A). Unlike the analysis using *in vivo* Cmax values, reliance on *in silico* predictions introduced some false-positive DILI risk classifications. Specifically, the No-DILI concern compounds phenazopyridine and etidronate were predicted to have higher Cmax values than observed, leading to inflated Cmax-to-*in vitro* toxicity ratios. Nevertheless, overall, the predicted ratios still correlated well with DILIrank classifications. ROC AUC values were only slightly lower than when using *in vivo* Cmax (86% instead of 90%, or 91% instead of 96%; Figure 24D). Notably, *in silico* predicted Cmax values alone outperformed *in vitro* toxicity readouts and heuristic screening rules for DILI prediction, second only to observed *in vivo* Cmax (Supplementary Figure 20).

As reported previously [140], *in silico* predictions showed a systematic tendency toward overestimation: 17 drugs were overpredicted by more than tenfold, while only 7 were underpredicted to the same extent. Given the over-representation of DILI-positive compounds in our dataset, the apparent performance of predicted Cmax values may be inflated compared to what would be expected in a balanced dataset. To quantify the additional uncertainty introduced by *in silico* Cmax predictions, we examined how often DILI classifications changed compared with those based on *in vivo* values. In total, 29 out of 208 drugs (13.9%) were assigned different DILI classes (Figure 24B). Although some changes coincidentally improved predictions for individual compounds, all deviations represent technical error introduced by the use of predicted rather than observed Cmax values.

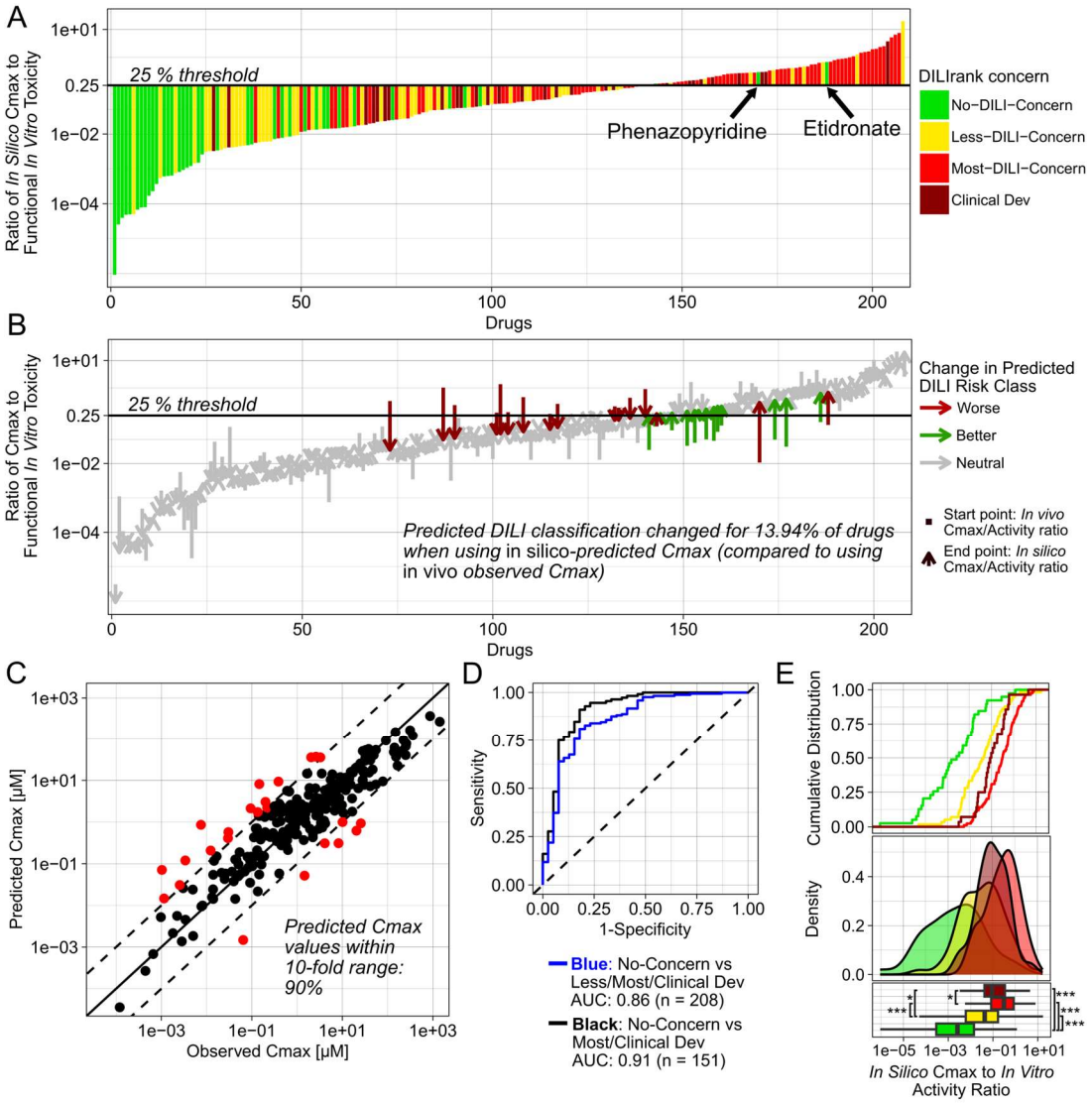


Figure 24: Prospective DILI predictivity of functional *in vitro* toxicity and *in silico* Cmax predictions (evaluation 3). Ratio of *in silico* predicted Cmax to lowest functional *in vitro* toxicity values of each drug (A). Functional toxicity refers to all *in vitro* toxicity values except BSEP inhibition. Change in DILI predictions based on use of *in silico predicted* against *in vivo* Cmax values (B). Green indicates changes to a better DILI prediction using *in silico* Cmax values compared to *in vivo* Cmax, red indicates changes to worse DILI class prediction and grey indicates no change between the two Cmax sources. Start points of arrows indicate the ratio based on *in vivo* Cmax, and end points the ratio based on *in silico* predicted Cmax values. *In silico* predicted values against *in vivo* observed Cmax values of drugs (C). Red indicates Cmax values outside the tenfold range. Receiver operating characteristic curves of ratios for distinguishing No-DILI compounds from DILI-positive drugs (D). Cumulative distribution, probability distribution and boxplots of the ratio of *in vivo* Cmax to functional *in vitro* toxicity values of different DILIrank classes (E). Boxplots show the median and interquartile range; whiskers extend to 1.5×IQR. Statistical comparisons on log-transformed values were performed using pairwise Wilcoxon rank-sum tests with Holm correction: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Figure reproduced from [272]. No changes made.

Evaluation 4: Inclusion of BSEP inhibition improves DILI predictivity

After evaluating the DILI predictivity of functional *in vitro* toxicity assays (cytotoxicity and mitochondrial toxicity), we next analysed the added value of incorporating *in vitro* BSEP inhibition measurements. BSEP inhibition has long been considered a mechanistic driver of cholestatic DILI and is therefore routinely included as a screening assay in pharmaceutical development. However, the utility of *in vitro* BSEP data for predicting clinical DILI outcomes has been debated, with some claiming that “there is no support for *in vitro* BSEP inhibition being DILI predictive” [266].

As for the functional *in vitro* assays, we calculated the ratios of *in vivo* Cmax values to BSEP IC50 values. These ratios correlated well with DILI classifications (ROC AUC up to 91%; Figure 25). No-DILI concern compounds had the lowest ratios, while Most-DILI drugs exhibited the highest. To test whether this signal was independent of other *in vitro* endpoints, we analysed how Cmax-to-toxicity ratios changed when BSEP inhibition data were included alongside other functional endpoints (Figure 25B). This analysis was restricted to drugs with both BSEP IC50 and at least one other *in vitro* toxicity value. Except for mifepristone, inclusion of BSEP data selectively increased the ratios of DILI-positive compounds, with the strongest effects seen in Less-DILI drugs. For compounds lacking functional *in vitro* toxicity data, either because they were not tested or showed no measurable activity, this comparison was not possible. Interestingly, all such compounds were classified as No- or Less-DILI concern (Figure 25C). Further, the distribution analysis showed that including BSEP inhibition particularly shifted Less-DILI compounds toward higher ratio values (Figure 25D-F).

Because BSEP inhibition is expected to drive intrahepatic bile acid accumulation, its effects may preferentially manifest as cholestatic rather than hepatocellular injury. To explore this, we manually collated information on DILI patterns from multiple sources, as no unified database equivalent to DILIrank exists. Using this curated dataset, we observed that inclusion of BSEP inhibition disproportionately increased the ratios of drugs annotated with cholestatic DILI patterns (Supplementary Figure 21).

Unlike the functional endpoints, two drugs appeared to be false positives when analysed with BSEP IC50 data: mifepristone and primidone. Both are labelled as No-DILI concern in DILIrank, yet their *in vivo* Cmax exceeded their BSEP IC50 values. If accurate, this would imply that BSEP inhibition alone may not cause DILI. To investigate, we reviewed these compounds in more detail.

Mifepristone, developed in the 1980s, is primarily used for medical termination of pregnancy at a single 200 mg dose. This short-term use likely explains its apparent clinical safety, since transient BSEP inhibition would not be expected to cause measurable liver injury. However, since 2012, mifepristone has also been approved for the chronic treatment of Cushing’s syndrome, where it is administered daily. We identified three case reports published between 2019 and 2023 [283–285], all after the release of DILIrank in 2016, documenting clear cases of cholestatic DILI associated with chronic mifepristone therapy. We therefore conclude that mifepristone indeed carries a DILI risk and that its DILIrank classification is likely outdated or incorrect.

Similarly, for primidone, we found a case report describing intrahepatic cholestasis in a female patient [286], as well as a veterinary case of liver injury in a dog [287]. These findings suggest that the DILIrank classification of primidone as No-DILI concern may also be inaccurate. Taken together, we interpret our results as evidence that the mechanistic predictions were correct, and that the apparent discrepancies reflect errors in DILIrank annotations rather than shortcomings of the BSEP-based approach.

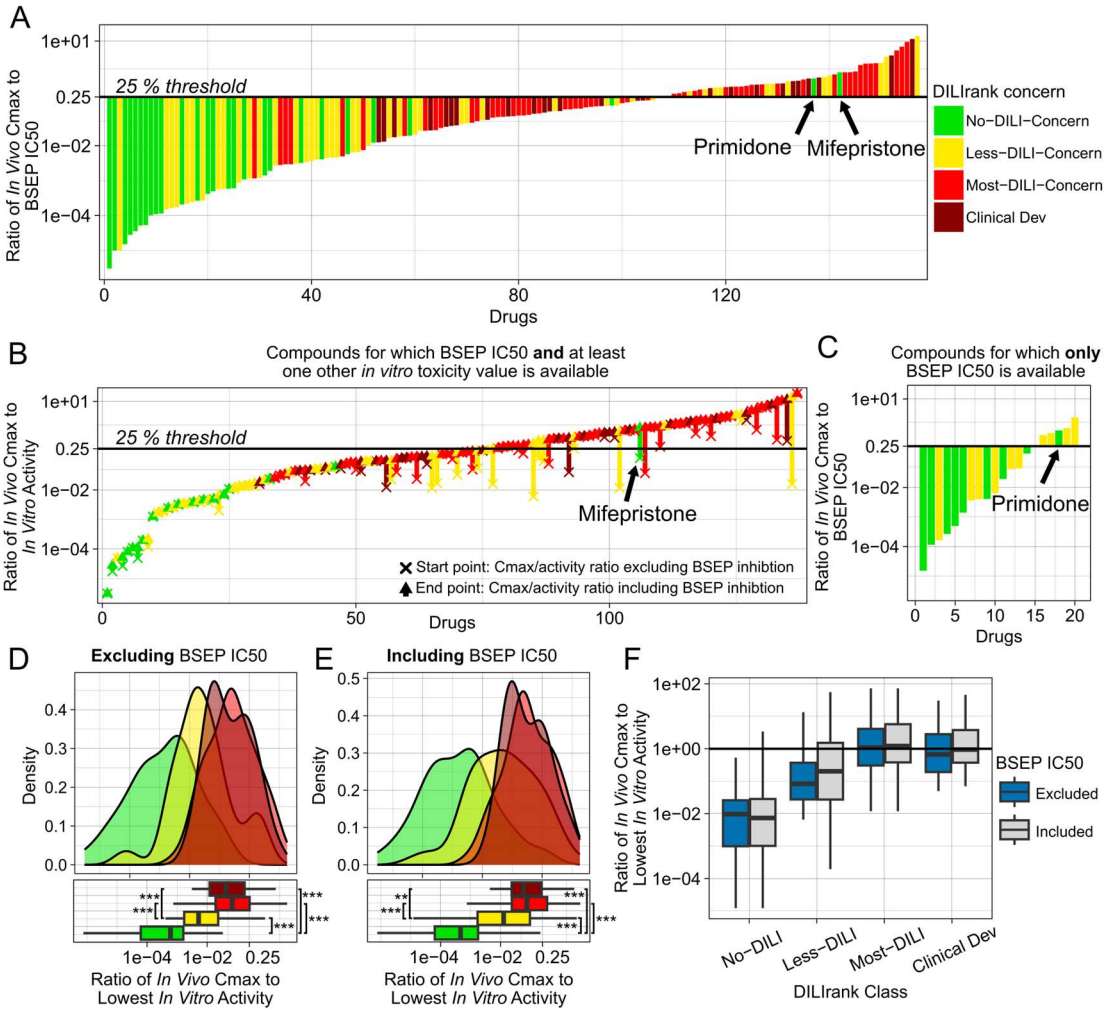


Figure 25: DILI predictivity of *in vitro* BSEP inhibition data (evaluation 4). Ratio of *in vivo* Cmax to BSEP IC50 values of each drug (A). Changes in Cmax to *in vitro* toxicity ratios when including BSEP IC50 values of compounds that have at least one functional *in vitro* toxicity value (B). Ratio of Cmax to BSEP IC50 values of compounds that have no functional *in vitro* toxicity value (C). Probability distribution and boxplot of Cmax to *in vitro* toxicity ratios when BSEP IC50 is excluded (D) or included (E). Boxplot of distributions of Cmax to *in vitro* toxicity ratios of drugs of different DILIrank classes compared with each other when BSEP IC50 is excluded or included (F). Boxplots show the median and interquartile range; whiskers extend to 1.5×IQR. Outliers were excluded from the plot for clarity but were included in statistical analyses. Statistical comparisons on log-transformed values were performed using pairwise Wilcoxon rank-sum tests with Holm correction: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Figure reproduced from [272]. No changes made.

Evaluation 5: Dynamic modelling reveals *in vivo* bile acid perturbations

Following the static pharmacokinetic (PK; i.e., Cmax) and pharmacodynamic (PD; i.e., IC50) analyses, we next evaluated whether integrating PK and PD interactions dynamically would refine predictions of DILI risk. Static ratios do not capture temporal variability, as identical Cmax values may belong to different PK profiles, and DILI often requires sustained perturbations before clinical manifestations become apparent.

To capture such time-dependent effects, we simulated drug concentrations and their inhibitory effects on BSEP within a high-throughput PBK-bile acid (HT-PBK-BA) model. This bile acid (BA) model [180] describes *in vivo* BA kinetics and explicitly represents BSEP as a central transporter. Drug-specific plasma concentration-time profiles predicted by HT-PBK were passed into the BA model to simulate the dynamic consequences of BSEP inhibition on intrahepatic BA concentrations. In the absence of detailed dosing regimens, we assumed once-daily administration for one month.

Dynamic simulations were interpreted by quantifying the duration of BA elevations above two thresholds: (i) twofold increases relative to non-perturbed baseline levels, and (ii) 10% increases above the non-perturbed peak concentration. Results from dynamic simulations generally correlated well with static PK/PD ratios of Cmax to BSEP IC50, identifying similar positive and negative responses (Figure 26CD). However, discrepancies were observed (Figure 26E). For example, troglitazone exhibited Cmax-to-BSEP IC50 ratios greater than 1, yet rapid clearance prevented sustained BA accumulation, suggesting that static ratios overestimated its risk. Conversely, bicalutamide had a low Cmax-to-BSEP IC50 ratio, but its slower clearance and accumulation over repeated dosing caused prolonged BSEP inhibition and meaningful BA perturbations in dynamic simulations.

All compounds identified as causing significant BA elevations in dynamic simulations were classified as DILI-positive in DILIrank, with the exception of mifepristone and primidone, which, as discussed previously, likely represent misclassified positives. Where pattern information was available, these compounds were consistently associated with cholestatic DILI. Together, these findings reinforced the suitability of 25% as a mechanistically plausible threshold for the Cmax-to-potency ratio, since dynamic simulations confirmed that this value effectively discriminated drugs causing clinically relevant BA perturbations from those that did not.

Finally, we combined analyses to identify compounds likely to cause DILI exclusively through BSEP inhibition. We selected drugs predicted to dynamically elevate BA levels, filtered to retain only those with *in vivo* Cmax-to-BSEP IC50 ratios >0.5 and lowest functional toxicity ratios <0.5, making alternative mechanisms unlikely. This yielded eight compounds (Table 4). With the exception of mifepristone, all were classified as DILI-positive in DILIrank, and those with annotated patterns were labelled as cholestatic. Notably, pioglitazone and verlukast exhibited *in vivo* Cmax values substantially exceeding their BSEP IC50. Primidone and dicloxacillin also met this criterion but lacked functional *in vitro* toxicity data for confirmation.

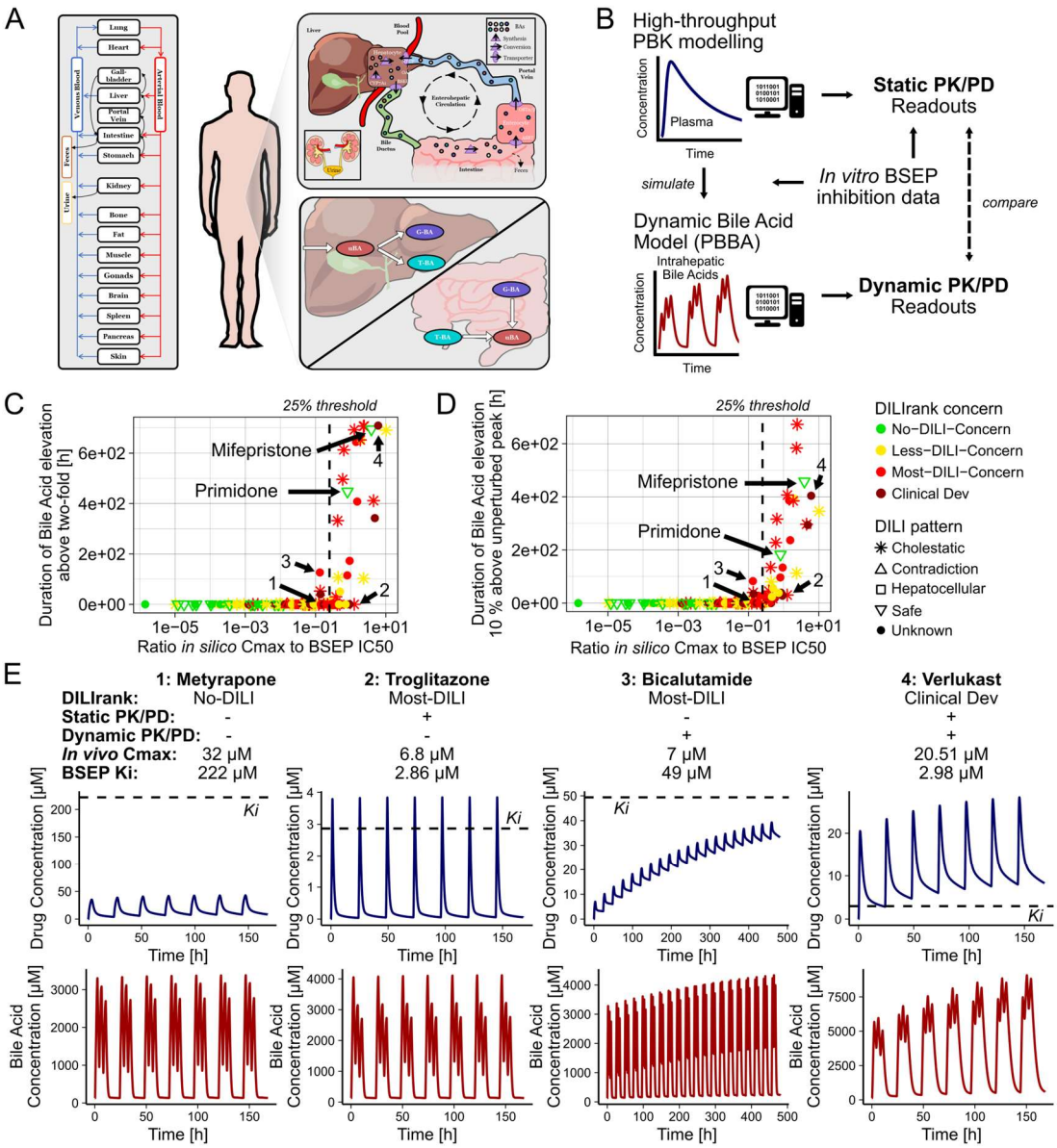


Figure 26: Simulated effects of *in vivo* BSEP inhibition on intrahepatic bile acid levels (evaluation 5). Schematic depiction of processes represented in the physiologically based bile acid (PBBA) model, graphic is taken from Kister et al. [180] (A). Schematic representation of the simulation and analysis strategy (B). Dynamic bile acid model perturbations caused through BSEP inhibition by drugs against their ratios of predicted Cmax to BSEP IC50 values and measured by duration during which Bile Acid levels are elevated more than twofold (C) or duration during which concentrations are more than 10% above physiological peak levels (D). Colour marks DILIrank class and shape the DILI pattern of drugs. Simulated drug and bile acid concentration profiles for four exemplary compounds of which static Cmax to BSEP IC50 ratios and dynamic PBBA model simulation yielded identical or diverging readouts (E). Dashed lines indicate 25% threshold or drug BSEP Ki values. Figure reproduced from [272]. No changes made.

Table 4: Compounds causing meaningful intrahepatic bile acid elevations in dynamic model simulations, having Cmax to BSEP IC50 ratios larger than 0.5 and Cmax to functional *in vitro* toxicity ratios lower than 0.5, or no other *in vitro* toxicity value. *Indicates classifications which may be incorrect, as discussed in text. All available values are provided in brackets. Table reproduced from [272]. No changes made.

Compound	DILI rank	DILI pattern	*In vivo* Cmax	Median BSEP IC50 value(s)	Lowest other, functional *in vitro* toxicity value(s)	Ratio *in vivo* Cmax/BSEP IC50 value	Ratio Cmax/Lowest other, functional *in vitro* toxicity value
Pioglitazone	Less	Cholestatic	4.18 µM	0.3 µM (0.098 µM; 0.3 µM; 0.4 µM)	262.75 µM	13.94x	0.016x
Verlukast	Clinical Dev	Clinical Dev	38.8 µM	3.384 µM	84.04 µM (84.04 µM; 296.75 µM)	11.46x	0.46x
Cyclosporine	Most	Cholestatic	0.67 µM	0.5 µM (0.5 µM; 0.5 µM; 2.393 µM)	2.3 µM (2.3 µM; 10 µM; 10.1 µM; 14.5 µM; 22.49 µM; 30.8 µM; 39.5 µM; 99.18 µM)	1.34x	0.29x
Erythromycin	Most	Cholestatic	10.9 µM	8.55 µM (13.0 µM; 4.1 µM)	293.7 µM (293.7 µM; 357.6 µM; 740.8 µM)	1.27x	0.037x
Mifepristone	No*	Safe*	6.0 µM	4.71 µM (2.02 µM; 7.4 µM)	50 µM	1.27x	0.12x
Fenofibrate	Less	Cholestatic	15.17 µM	25.39 µM (15.3 µM; 35.48 µM)	61.07 µM (61.07 µM; 72 µM; 103.18 µM; 103.67 µM; 295.82 µM)	0.6x	0.25x
Irbesartan	Less	Cholestatic	6.77 µM	7.31 µM	580 µM	0.93x	0.012x
PF-6273340	Clinical Dev	Clinical Dev	1.69 µM	3.0 µM	42 µM (42 µM; 239 µM)	0.56x	0.04x
Dicloxacillin	Less	Cholestatic	148 µM	63.05 µM (56.4 µM; 69.7 µM)	-	2.36x	-
Primidone	No*	Safe*	41.4 µM	48.7 µM	-	0.85x	-

13.3 Discussion

Here, we demonstrated that high predictivity of DILI risk can be achieved by integrating *in vitro* hepatotoxicity measurements with pharmacokinetic (PK) information. Simple heuristic predictors commonly used in the pharmaceutical industry, such as drug lipophilicity and dose, showed only moderate performance, whereas the combination of PK data with *in vitro* toxicity values clearly outperformed them. Importantly, exposure-based predictions not only improved accuracy but also aligned with clinically observed DILI severity classifications, underscoring the biological relevance of this mechanistic approach.

In our analysis, PK information emerged as one of the most influential determinants of DILI risk. This reflected the narrower distribution of *in vitro* toxicity values compared with Cmax, making exposure the dominant driver of clinical outcomes within our dataset. One likely reason for this was the exclusion of compounds with extreme potency values outside the concentration ranges tested in *in vitro* assays. As a result, it remains uncertain whether PK information alone would maintain comparable predictive power in prospective evaluations of less preselected, early-stage drug candidates. On the one hand, developmental compounds may exhibit a broader spectrum of *in vitro* toxicities, making potency more relevant. On the other hand, toxic potencies may generally cluster around intermediate values, since drug candidates are deliberately optimised for bioactivity and highly potent toxins rarely enter development pipelines. Consistent with this hypothesis, toxic potency values in our dataset, including those of No-DILI concern drugs, were approximately normally distributed across the tested concentration range (Figure 21J-L). Depending on the true distribution of toxic activities in novel compounds, PK may thus remain the dominant factor, or represent one of two equally critical determinants of DILI risk, even though DILI is typically described as *idiosyncratic* without obvious dose dependence.

When combined with Cmax data, *in vitro* BSEP inhibition measurements not only improved DILI prediction but also revealed risks that would have been underestimated if only functional toxicity assays had been considered. For instance, BSEP inhibition correctly indicated the DILI potential of mifepristone, a drug classified as No-DILI concern in DILIrank but for which later clinical evidence confirmed cholestatic injury [283–285]. Moreover, we identified a subset of drugs likely to exert hepatotoxicity primarily via BSEP inhibition (Table 4). Although additional mechanisms cannot be ruled out for these compounds, their identification provides an opportunity to advance mechanistic understanding of BSEP-mediated toxicity in humans.

Notably, the compounds whose Cmax-to-toxicity ratios changed most substantially upon inclusion of BSEP inhibition were predominantly classified as Less-DILI concern and exhibited a cholestatic DILI pattern. Likewise, all compounds with BSEP inhibition as their only detected mechanism were classified as No- or Less-DILI concern. These findings suggest that BSEP inhibition alone may manifest with lower clinical severity than cytotoxicity or mitochondrial toxicity, although strong and sustained inhibition, as observed with mifepristone, can nonetheless result in clinically significant liver injury. Consistent with earlier reports [176], our dynamic bile acid (BA) perturbation simulations indicated that most drugs, including those with a cholestatic DILI pattern, are unlikely to cause marked BA level perturbations *in vivo*. This supports the view that BSEP inhibition represents only one of several mechanisms underlying drug-induced cholestasis and DILI, and highlights the importance of considering a broader range of potential toxicity pathways.

In addition, some DILI-positive drugs exhibited Cmax-to-toxicity ratios below the mechanistically expected threshold for meaningful toxicity, despite the overall strong correlation of these ratios with clinical outcomes. Estradiol was a representative example, where the discrepancy likely reflected toxicity mechanisms not captured by the *in vitro* assays included in our analysis. Such cases underscore the potential to further improve DILI prediction accuracy by incorporating additional mechanistic endpoints, including perturbations of gene expression, immune-mediated effects, or metabolite-driven toxicity [23].

It is important to note that our analysis was constrained by sparse and heterogeneous data from previously published studies, with compounds not being systematically tested across all *in vitro* systems. A more comprehensive dataset would likely improve predictive performance. In addition, applying a single fixed threshold value, such as 25%, may not have been optimal given potential differences in assay sensitivities [288]. Our analysis also did not incorporate other relevant factors, including drug dosing schedules, transporter activity, tissue-specific exposures, or potential synergistic effects arising from multiple toxicity mechanisms. Integrating such factors, particularly through dynamic simulations, as exemplified by our modelling of bile acid perturbations, should further enhance the accuracy of clinical toxicity predictions [289]. Nevertheless, the present findings highlight the importance of combining multiple *in vitro* assays that capture distinct toxicity mechanisms. This integration appears to be a key driver of the robust predictive performance achieved here, which in our dataset surpassed that of most previously reported approaches [259] (Supplementary Table 2). Given the near-perfect specificity of our mechanistic framework, increasing its sensitivity through the inclusion of additional influential factors could bring DILI risk predictions close to optimal performance using currently available *in vitro* and *in silico* methods.

Our analysis also provided a mechanistic rationale for why DILI frequently manifests idiosyncratically in clinical practice, without an apparent dose dependence. Even drugs of the Most-DILI concern class exhibited *in vivo* Cmax concentrations below their corresponding *in vitro* toxicity potencies. This observation makes it unsurprising that clinically relevant toxicity only occurs in a subset of susceptible individuals, such as those with genetic predispositions, comorbidities, concomitant medications, or unusually high systemic exposures. We speculate that a clearer dose dependence of idiosyncratic DILI would emerge if drug doses were increased to levels at which *in vivo* concentrations consistently exceeded toxic potencies. However, such testing is not ethically feasible, and it is likely that maximum recommended doses are deliberately selected during drug development to remain below clearly toxic exposure ranges.

In this study, we focused on the prediction of DILI because hepatotoxicity remains a leading cause of drug development failure. Unlike many other forms of toxicity, DILI has been extensively studied, and curated resources such as the DILIrank database facilitate systematic evaluation of predictive methods. Nonetheless, the mechanistic integration of *in vitro* assays with pharmacokinetic information described here is not limited to DILI and can be readily applied to other organ toxicities, provided that assays capturing relevant toxicity mechanisms are available.

Unlike typical machine learning approaches that attempt to predict DILI risk directly in a black-box manner without representing the underlying biological processes, our framework explicitly preserves interpretability. This enables evaluation, continuous refinement, and mechanistic insight into toxicity predictions, thereby increasing trust in its results (Supplementary Figure 22). Such transparency is also critical for regulatory acceptance. Regulatory agencies worldwide are committed to phasing out animal testing

where reliable alternative methods are available. Our findings demonstrate that high levels of reliability, interpretability, and predictive performance can already be achieved today through integrated use of existing *in vitro* and *in silico* methods.

General Conclusions and Outlook

Conclusions

The aim of this thesis was to evaluate the accuracy and robustness of high-throughput physiologically based kinetic (HT-PBK) modelling and to demonstrate how this new method, on its own or when integrated with complementary new approach methodologies (NAMs), can provide a scientifically robust foundation for human-relevant chemical safety assessment. The work thereby addressed a central challenge in contemporary toxicology: the need to predict systemic exposure in the absence of *in vivo* pharmacokinetics (PK) data, so that toxicokinetics can be integrated into new assessment frameworks for chemical safety that do not rely on animal testing.

Across different exposure routes and use cases, this thesis demonstrated that HT-PBK modelling can provide meaningful and, in many cases, quantitatively accurate predictions of human systemic exposure. Using a large reference PK dataset of more than 200 compounds and 2000 plasma concentration-time profiles, we showed that even relatively simple parameterisation strategies based solely on *in vitro* and *in silico* data achieved predictive performance comparable to conventional PBK modelling [290]. This observation was confirmed not only for oral and intravenous administration, but also for dermal exposure, a result that underlines the broad applicability and utility of HT-PBK modelling for safety assessments. At the same time, we identified which *in silico* tools and parameterisation strategies yielded the most reliable PK predictions and outlined a set of freely available prediction strategies with quantified expected error ranges.

Beyond methodological evaluation, the thesis demonstrated the potential utility of HT-PBK modelling as a decision-support tool in industrial and regulatory contexts. In response to the European Partnership for Alternative Approaches to Animal Testing (EPAA) Designathon challenge [291], we developed a systematic approach to classify chemicals into different levels of Potential Systemic Availability (PSA) concern. This quantitative method, applied to 150 benchmark compounds, provided a scientifically sound concept for considering the role of toxicokinetics in new chemical prioritisation frameworks. Finally, the thesis explored how coupling *in vitro* toxicity data with HT-PBK-derived exposure predictions can improve the assessment of complex adverse outcomes. In the case of drug-induced liver injury (DILI), we showed that simple exposure-toxicity ratios achieved high predictive accuracy for clinical outcomes, surpassing the performance of many existing preclinical prediction strategies [259]. Mechanistic simulations of bile acid homeostasis further provided explanatory insight into compound-specific DILI risks. These findings demonstrate that scalable, mechanistically interpretable, and animal-free approaches to human risk assessment are achievable when toxicokinetic modelling is systematically integrated with relevant NAMs. Taken together, this work contributes to building the scientific confidence and practical workflows required for integrating HT-PBK modelling into pharmacological and toxicological assessment practices.

Limitations and challenges

Despite the promising findings presented throughout this thesis, several important limitations must be acknowledged, as they constrain the interpretation and application of these results. These limitations reflect both the current state of computational toxicology

and fundamental challenges inherent to the predictive modelling of complex biological systems. First, the predictive accuracy of HT-PBK models remained highly variable across compounds. It is difficult, for hundreds of substances, to determine the origin of this variability on a large scale, and it may arise from multiple sources. One major contributor is likely the presence of inaccuracies and mispredictions in the predicted input data, i.e., predicted solubility, permeability or metabolic clearance values. Another major contributor may be that some substances are influenced by special processes, such as active transport, that are critical to their PK but inadequately captured by the generic model structure we used here, leading to oversimplified representations of the complex *in vivo* behaviour of these substances. In addition, the PK reference data used for evaluation were heterogeneous in origin and variable in quality, with some datasets generated decades ago using outdated analytical techniques and non-standardised study designs. Such heterogeneity introduces a level of uncertainty that cannot be easily quantified or corrected.

A particularly challenging aspect of this analysis is the lack of transparency surrounding the training data and applicability domains of the machine learning models used for input parameter prediction. It is widely recognised that the reliability of chemical property predictions depends critically on the diversity and quality of the training datasets [292,293]. However, for many commercial and even academic quantitative structure-activity relationship (QSAR) models, this information remains proprietary or incompletely documented. The absence of standardised applicability domain definitions, coupled with the inability to establish whether compounds in our evaluation set overlapped with QSAR training sets, limits the interpretability of our findings. As a consequence, some of the reported predictive performance may be inflated by recall of training compounds rather than reflecting true prospective capability. This uncertainty fundamentally constrains the extrapolation of our results to structurally novel substances.

The Potential Systemic Availability (PSA) classification framework, while conceptually valuable and mechanistically grounded, also proved insufficient as a standalone prioritisation tool. A substantial proportion of compounds clustered in the medium concern category, limiting the framework’s utility for decision making in the absence of complementary bioactivity data. This outcome underscores both the importance of toxicokinetic information and its insufficiency to define systemic hazard potential on its own. Meaningful prioritisation will ultimately require the integration of PSA-derived insights with bioactivity data in order to reach robust and actionable safety conclusions.

The DILI case study further illustrates the opportunities and limitations of the approaches explored in this thesis. While retrospective and prospective analyses demonstrated strong predictive performance, their immediate translation is constrained by several factors. Most importantly, our analysis remained limited in its ability to capture interindividual variability and the influence of confounding factors such as dosing regimens, comorbidities, or polypharmacy. Additionally, certain hepatotoxic compounds were misclassified as safe because their mechanisms of toxicity were not represented in the *in vitro* toxicity assays considered. This limitation highlights a fundamental challenge: predictive frameworks can only account for toxicity pathways that are already known and experimentally accessible. Achieving comprehensive mechanistic coverage remains an ambitious but essential objective, particularly in areas such as hepatotoxicity where the spectrum of underlying mechanisms is broad and not yet fully elucidated.

Outlook

Future progress in HT-PBK modelling will depend not only on refining methodologies but also on addressing the fundamental limitations of available data. While larger and more diverse datasets would certainly enhance the robustness of predictions across chemical classes, the practical reality is that PK data for humans remain scarce. Across the entirety of clinical research, only a few tens of thousands of compounds have ever been evaluated in clinical studies in humans [294,295], and much of this information is inaccessible, being held in proprietary or paper form in the archives of pharmaceutical companies or regulatory agencies. Consequently, the field must adapt to operating in low-data regimes by developing methods that can quantify uncertainty under such constraints and by maximising the value of the limited information that is available.

A key step towards this goal is the deeper integration of the underlying modelling tools, particularly QSAR-based property prediction and PBK simulation. Achieving this requires transparency regarding the training datasets, applicability domains, and potential overlaps between training and evaluation compounds. Only through such systematic alignment can we obtain unbiased estimates of predictive performance and a clear understanding of how these methods generalise to structurally novel substances.

Equally important is the need to move beyond technical demonstrations of PK prediction accuracy and towards applications that test the practical value of HT-PBK in real-world decision making. The DILI case study presented here illustrates how kinetic predictions can be coupled with *in vitro* bioactivity data to yield clinically relevant risk assessments. Extending this principle to other toxicological endpoints will be critical for demonstrating that the approach is not restricted to a specific type of toxicity but reflects a generalisable scientific framework. Ultimately, the decisive test will come from prospective applications in regulatory or industrial contexts, where predictions inform actual safety decisions rather than retrospective academic evaluations. This will require an initial degree of trust in the potential of these methods, a process to which we tried to contribute with our work here.

Final remarks

The work presented in this thesis takes place at a moment of significant transition in toxicology. Advances in experimental and computational methods now allow us to approach the central principle articulated by Paracelsus (“the dose makes the poison” [33]) with a level of mechanistic understanding that was previously unattainable. We are reaching a point where we can determine the true dose of a substance at a specific location within the body over time, without performing any *in vivo* experiments on humans or animals. By integrating *in vitro* data and *in silico* models, we are closer than ever to quantifying human-relevant exposure in ways that can directly inform hazard and risk assessment.

These findings show that the foundations of a new paradigm in chemical safety assessment are already in place. Realising their full potential will require coordinated efforts to improve input data quality, increase transparency of predictive models, and expand mechanistic coverage of toxicity pathways. If these challenges can be met, the field will finally be able to move towards risk assessment approaches that are not only more predictive and scalable but that also align with societal expectations to reduce reliance on animal testing [296]. In this way, the work presented in this thesis contributes to an ongoing transformation in toxicology: moving from descriptive observation in

animals towards predictive, human-relevant science [3]. Significant challenges remain, but the results presented here provide evidence that the tools and concepts needed to achieve this vision are within reach today.

References

1. Karahalil, B. Overview of Systems Biology and Omics Technologies. *Current medicinal chemistry* **23,** 4221–4230; 10.2174/0929867323666160926150617 (2016).
2. Westerhoff, H. V. & Palsson, B. O. The evolution of molecular biology into systems biology. *Nat Biotechnol* **22,** 1249–1252; 10.1038/nbt1020 (2004).
3. National Research Council (US) Committee on Toxicity Testing and Assessment of Environmental Agents. *Toxicity Testing in the 21st Century: A Vision and a Strategy* (National Academies Press, Washington, DC, 2007).
4. Krewski, D. *et al.* Toxicity testing in the 21st century: a vision and a strategy. *Journal of toxicology and environmental health. Part B, Critical reviews* **13,** 51–138; 10.1080/10937404.2010.483176 (2010).
5. Olson, H. *et al.* Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory toxicology and pharmacology : RTP* **32,** 56–67; 10.1006/rtph.2000.1399 (2000).
6. Perel, P. *et al.* Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ (Clinical research ed.)* **334,** 197; 10.1136/bmj.39048.407928.BE (2007).
7. Ingber, D. E. Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nat Rev Genet* **23,** 467–491; 10.1038/s41576-022-00466-9 (2022).
8. Nitsche, K. S., Müller, I., Malcomber, S., Carmichael, P. L. & Bouwmeester, H. Implementing organ-on-chip in a next-generation risk assessment of chemicals: a review. *Arch Toxicol* **96,** 711–741; 10.1007/s00204-022-03234-0 (2022).
9. Sturla, S. J. *et al.* Systems toxicology: from basic research to risk assessment. *Chemical research in toxicology* **27,** 314–329; 10.1021/tx400410s (2014).
10. Hartung, T. *et al.* Systems Toxicology: Real World Applications and Opportunities. *Chemical research in toxicology* **30,** 870–882; 10.1021/acs.chemrestox.7b00003 (2017).
11. Holford, N. H. & Sheiner, L. B. Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. *Clin Pharmacokinet* **6,** 429–453; 10.2165/00003088-198106060-00002 (1981).
12. Rowland, M., Peck, C. & Tucker, G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annual review of pharmacology and toxicology* **51,** 45–73; 10.1146/annurev-pharmtox-010510-100540 (2011).
13. Luzon, E. *et al.* Physiologically based pharmacokinetic modeling in regulatory decision-making at the European Medicines Agency. *Clinical Pharmacology & Therapeutics* **102,** 98–105; 10.1002/cpt.539 (2017).
14. Zhang, X. *et al.* Application of PBPK Modeling and Simulation for Regulatory Decision Making and Its Impact on US Prescribing Information: An Update on the 2018-2019 Submissions to the US FDA's Office of Clinical Pharmacology. *The*

Journal of Clinical Pharmacology **60 Suppl 1,** S160-S178; 10.1002/jcph.1767 (2020).

15. Deepika, D. & Kumar, V. The Role of "Physiologically Based Pharmacokinetic Model (PBPK)" New Approach Methodology (NAM) in Pharmaceuticals and Environmental Chemical Risk Assessment. *International Journal of Environmental Research and Public Health* **20**; 10.3390/ijerph20043473 (2023).
16. Tsamandouras, N., Rostami-Hodjegan, A. & Aarons, L. Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. *British Journal of Clinical Pharmacology* **79,** 48–55; 10.1111/bcp.12234 (2015).
17. Jones, H. M. *et al.* Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. *Clinical Pharmacology & Therapeutics* **97,** 247–262; 10.1002/cpt.37 (2015).
18. Bender, A. & Cortés-Ciriano, I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug discovery today* **26,** 511–524; 10.1016/j.drudis.2020.12.009 (2021).
19. Chou, W.-C. & Lin, Z. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. *Toxicological sciences : an official journal of the Society of Toxicology* **191,** 1–14; 10.1093/toxsci/kfac101 (2023).
20. *Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products* (Publications Office of the European Union, Luxembourg, 2009).
21. Berggren, E. & Worth, A. P. Towards a future regulatory framework for chemicals in the European Union - Chemicals 2.0. *Regulatory toxicology and pharmacology : RTP* **142,** 105431; 10.1016/j.yrtph.2023.105431 (2023).
22. Worth, A. P., Berggren, E. & Prieto, P. Chemicals 2.0 and Why We Need to Bypass the Gold Standard in Regulatory Toxicology. *Alternatives to laboratory animals : ATLA,* 2611929241296328; 10.1177/02611929241296328 (2024).
23. Kullak-Ublick, G. A. *et al.* Drug-induced liver injury: recent advances in diagnosis and risk assessment. *Gut* **66,** 1154–1164; 10.1136/gutjnl-2016-313369 (2017).
24. Weaver, R. J. *et al.* Managing the challenge of drug-induced liver injury: a roadmap for the development and deployment of preclinical predictive models. *Nature reviews. Drug discovery* **19,** 131–148; 10.1038/s41573-019-0048-x (2020).
25. Dirven, H. *et al.* Performance of preclinical models in predicting drug-induced liver injury in humans: a systematic review. *Sci Rep* **11,** 6403; 10.1038/s41598-021-85708-2 (2021).
26. Rowland, M., Tozer, T. N., Derendorf, H. & Schmidt, S. *Rowland' and Tozer's Clinical pharmacokinetics and pharmacodynamics : concepts and applications.* 5th ed. (Wolters Kluwer, Philadelphia, 2020).
27. Geisslinger, G. *et al. Mutschler Arzneimittelwirkungen Pharmakologie - Klinische Pharmakologie - Toxikologie.* 11th ed. (WVG, Wissenschaftliche Verlagsgesellschaft, Stuttgart, 2020).
28. Zhang, D., Luo, G., Ding, X. & Lu, C. Preclinical experimental models of drug metabolism and disposition in drug discovery and development. *Acta Pharmaceutica Sinica B* **2,** 549–561; 10.1016/j.apsb.2012.10.004 (2012).

29. Pardridge, W. M. Drug transport across the blood-brain barrier. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* **32,** 1959–1972; 10.1038/jcbfm.2012.126 (2012).

30. Baillie, T. A. & Rettie, A. E. Role of biotransformation in drug-induced toxicity: influence of intra- and inter-species differences in drug metabolism. *Drug metabolism and pharmacokinetics* **26,** 15–29; 10.2133/dmpk.dmpk-10-rv-089 (2011).

31. *Test No. 417: Toxicokinetics* (OECD, 2010).

32. U. S. Food and Drug Administration. Population Pharmacokinetics: Guidance for Industry.

33. Paracelsus. Die dritte Defension wegen des Schreibens der neuen Rezepte. In *Sämtliche Werke 1965*, Vol. 2, p. 510.

34. Meek, M. E. B. & Lipscomb, J. C. Gaining acceptance for the use of in vitro toxicity assays and QIVIVE in regulatory risk assessment. *Toxicology* **332,** 112–123; 10.1016/j.tox.2015.01.010 (2015).

35. Blaauboer, B. J. The necessity of biokinetic information in the interpretation of in vitro toxicity data. *Alternatives to laboratory animals : ATLA* **30 Suppl 2,** 85–91; 10.1177/026119290203002S13 (2002).

36. Xu, R. N., Fan, L., Rieser, M. J. & El-Shourbagy, T. A. Recent advances in high-throughput quantitative bioanalysis by LC-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis* **44,** 342–355; 10.1016/j.jpba.2007.02.006 (2007).

37. European Medicines Agency & Committee for Medicinal Products for Human Use (CHMP). Guideline on the Investigation of Bioequivalence. European Medicines Agency, January 2010.

38. Aarons, L. Physiologically based pharmacokinetic modelling: a sound mechanistic basis is needed. *British Journal of Clinical Pharmacology* **60,** 581–583; 10.1111/j.1365-2125.2005.02560.x (2005).

39. Kuepfer, L. *et al.* Applied Concepts in PBPK Modeling: How to Build a PBPK/PD Model. *CPT: pharmacometrics & systems pharmacology* **5,** 516–531; 10.1002/psp4.12134 (2016).

40. Rostami-Hodjegan, A. Physiologically based pharmacokinetics joined with in vitro-in vivo extrapolation of ADME: a marriage under the arch of systems pharmacology. *Clinical Pharmacology & Therapeutics* **92,** 50–61; 10.1038/clpt.2012.65 (2012).

41. Rodgers, T. & Rowland, M. Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *Journal of pharmaceutical sciences* **95,** 1238–1257; 10.1002/jps.20502 (2006).

42. Poulin, P. & Theil, F.-P. A Priori Prediction of Tissue:Plasma Partition Coefficients of Drugs to Facilitate the Use of Physiologically-Based Pharmacokinetic Models in Drug Discovery. *Journal of pharmaceutical sciences* **89,** 16–35; 10.1002/(SICI)1520-6017(200001)89:1<16::AID-JPS3>3.0.CO;2-E (2000).

43. Jones, H. & Rowland-Yeo, K. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT: pharmacometrics & systems pharmacology* **2,** e63; 10.1038/psp.2013.41 (2013).

44. Westmoreland, C. Assuring safety without animal testing: Unilever's ongoing research programme to deliver novel ways to assure consumer safety. *ALTEX* **27,** 207–211; 10.14573/altex.2010.3.207 (2010).

45. Marshall, S. *et al.* Model-Informed Drug Discovery and Development: Current Industry Good Practice and Regulatory Expectations and Future Perspectives. *CPT Pharmacom & Syst Pharma* **8,** 87–96; 10.1002/psp4.12372 (2019).

46. Paini, A. *et al.* Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. *Computational toxicology (Amsterdam, Netherlands)* **9,** 61–72; 10.1016/j.comtox.2018.11.002 (2019).

47. Wambaugh, J. F. *et al.* Applying New Approach Methods for Toxicokinetics for Chemical Risk Assessment. *Chemical research in toxicology*; 10.1021/acs.chemrestox.5c00161 (2025).

48. Richard, A. M. *et al.* The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. *Chemical research in toxicology* **34,** 189–216; 10.1021/acs.chemrestox.0c00264 (2021).

49. Pearce, R. G., Setzer, R. W., Strope, C. L., Wambaugh, J. F. & Sipes, N. S. httk: R Package for High-Throughput Toxicokinetics. *Journal of statistical software* **79,** 1–26; 10.18637/jss.v079.i04 (2017).

50. Schneckener, S. *et al.* Prediction of Oral Bioavailability in Rats: Transferring Insights from in Vitro Correlations to (Deep) Machine Learning Models Using in Silico Model Outputs and Chemical Structure Parameters. *Journal of chemical information and modeling* **59,** 4893–4905; 10.1021/acs.jcim.9b00460 (2019).

51. Naga, D., Parrott, N., Ecker, G. F. & Olivares-Morales, A. Evaluation of the Success of High-Throughput Physiologically Based Pharmacokinetic (HT-PBPK) Modeling Predictions to Inform Early Drug Discovery. *Molecular pharmaceutics* **19,** 2203–2216; 10.1021/acs.molpharmaceut.2c00040 (2022).

52. Dorne, J. L. C. M. *et al.* TKPlate 1.0: An Open-access platform for toxicokinetic and toxicodynamic modelling of chemicals to implement new approach methodologies in chemical risk assessment. *EFSA journal. European Food Safety Authority* **21,** e211101; 10.2903/j.efsa.2023.e211101 (2023).

53. Wetmore, B. A. *et al.* Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicological sciences : an official journal of the Society of Toxicology* **125,** 157–174; 10.1093/toxsci/kfr254 (2012).

54. Chang, X. *et al.* IVIVE: Facilitating the Use of In Vitro Toxicity Data in Risk Assessment and Decision Making. *Toxics* **10,** 232; 10.3390/toxics10050232 (2022).

55. Moxon, T. E. *et al.* Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. *Toxicology in vitro : an international journal published in association with BIBRA* **63,** 104746; 10.1016/j.tiv.2019.104746 (2020).

56. Ouedraogo, G. *et al.* Read-across and new approach methodologies applied in a 10-step framework for cosmetics safety assessment - A case study with parabens. *Regulatory toxicology and pharmacology : RTP* **132,** 105161; 10.1016/j.yrtph.2022.105161 (2022).

57. Reale, E. *et al.* Human biomonitoring and toxicokinetics as key building blocks for next generation risk assessment. *Environment international* **184,** 108474; 10.1016/j.envint.2024.108474 (2024).

58. Himstedt, A. *et al.* Beyond CL and VSS: a comprehensive approach to human pharmacokinetic predictions. *Drug discovery today,* 104238; 10.1016/j.drudis.2024.104238 (2024).

59. Walter, M. *et al. In silico PK predictions in Drug Discovery: Benchmarking of Strategies to Integrate Machine Learning with Empiric and Mechanistic PK modelling* (2024).

60. Göller, A. H. *et al.* Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. *Drug discovery today* **25,** 1702–1709; 10.1016/j.drudis.2020.07.001 (2020).

61. Fu, L. *et al.* ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic acids research* **52,** W422-W431; 10.1093/nar/gkae236 (2024).

62. Aleksić, S., Seeliger, D. & Brown, J. B. ADMET Predictability at Boehringer Ingelheim: State-of-the-Art, and Do Bigger Datasets or Algorithms Make a Difference? *Molecular informatics* **41,** e2100113; 10.1002/minf.202100113 (2022).

63. Komura, H., Watanabe, R. & Mizuguchi, K. The Trends and Future Prospective of In Silico Models from the Viewpoint of ADME Evaluation in Drug Discovery. *Pharmaceutics* **15**; 10.3390/pharmaceutics15112619 (2023).

64. Howard, M. L., Hill, J. J., Galluppi, G. R. & McLean, M. A. Plasma protein binding in drug discovery and development. *Combinatorial chemistry & high throughput screening* **13,** 170–187; 10.2174/138620710790596745 (2010).

65. Sodhi, J. K. & Benet, L. Z. Successful and Unsuccessful Prediction of Human Hepatic Clearance for Lead Optimization. *Journal of medicinal chemistry* **64,** 3546–3559; 10.1021/acs.jmedchem.0c01930 (2021).

66. Watanabe, R. *et al.* Predicting Fraction Unbound in Human Plasma from Chemical Structure: Improved Accuracy in the Low Value Ranges. *Molecular pharmaceutics* **15,** 5302–5311; 10.1021/acs.molpharmaceut.8b00785 (2018).

67. Watanabe, R. *et al.* Development of an in silico prediction system of human renal excretion and clearance from chemical structure information incorporating fraction unbound in plasma as a descriptor. *Scientific reports* **9,** 18782; 10.1038/s41598-019-55325-1 (2019).

68. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *Journal of medicinal chemistry* **58,** 4066–4072; 10.1021/acs.jmedchem.5b00104 (2015).

69. Swanson, K. *et al.* ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics* **40**; 10.1093/bioinformatics/btae416 (2024).

70. Pletz, J. Physiologically-Based Kinetics and Mechanistic Models to Assess Exposure to Chemicals. Liverpool John Moores University, 2020/01.

71. Punt, A. *et al.* Predictive performance of next generation human physiologically based kinetic (PBK) models based on in vitro and in silico input data. *ALTEX* **39,** 221–234; 10.14573/altex.2108301 (2022).

72. Punt, A., Louisse, J., Pinckaers, N., Fabian, E. & van Ravenzwaay, B. Predictive Performance of Next Generation Physiologically Based Kinetic (PBK) Model Predictions in Rats Based on In Vitro and In Silico Input Data. *Toxicological sciences : an official journal of the Society of Toxicology* **186,** 18–28; 10.1093/toxsci/kfab150 (2022).

73. Paini, A., Leonard, J. A., Kliment, T., Tan, Y.-M. & Worth, A. Investigating the state of physiologically based kinetic modelling practices and challenges associated with gaining regulatory acceptance of model applications. *Regulatory toxicology and pharmacology : RTP* **90,** 104–115; 10.1016/j.yrtph.2017.08.019 (2017).

74. Stucki, A. O. *et al.* Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Front. Toxicol.* **4**; 10.3389/ftox.2022.964553 (2022).

75. van Norman, G. A. Limitations of Animal Studies for Predicting Toxicity in Clinical Trials: Is it Time to Rethink Our Current Approach? *JACC. Basic to translational science* **4,** 845–854; 10.1016/j.jacbts.2019.10.008 (2019).

76. Hartung, T. The (misleading) role of animal models in drug development. *Front. Drug Discov.* **4,** 1355044; 10.3389/fddsv.2024.1355044 (2024).

77. Shanks, N., Greek, R. & Greek, J. Are animal models predictive for humans? *Philosophy, ethics, and humanities in medicine : PEHM* **4,** 2; 10.1186/1747-5341-4-2 (2009).

78. van der Zalm, A. J. *et al.* A framework for establishing scientific confidence in new approach methodologies. *Arch Toxicol* **96,** 2865–2879; 10.1007/s00204-022-03365-4 (2022).

79. *Detailed Review Paper on In Vitro Test Addressing Immunotoxicity With a Focus on Immunosuppression* (OECD, 2022).

80. Rajagopal, R. *et al.* Beyond AOPs: A Mechanistic Evaluation of NAMs in DART Testing. *Front. Toxicol.* **4**; 10.3389/ftox.2022.838466 (2022).

81. Paini, A. *et al.* Towards a qAOP framework for predictive toxicology - Linking data to decisions. *Computational Toxicology* **21,** 100195; 10.1016/j.comtox.2021.100195 (2022).

82. Daley, M. C., Mende, U., Choi, B.-R., McMullen, P. D. & Coulombe, K. L. K. Beyond pharmaceuticals: Fit-for-purpose new approach methodologies for environmental cardiotoxicity testing. *ALTEX* **40,** 103–116; 10.14573/altex.2109131 (2023).

83. Alépée, N., Tourneix, F., Singh, A., Ade, N. & Grégoire, S. Off to a good start? Review of the predictivity of reactivity methods modelling the molecular initiating event of skin sensitization. *ALTEX* **40,** 606–618; 10.14573/altex.2212201 (2023).

84. Jacobs, M. N. *et al.* Avoiding a reproducibility crisis in regulatory toxicology-on the fundamental role of ring trials. *Arch Toxicol* **98,** 2047–2063; 10.1007/s00204-024-03736-z (2024).

85. Sewell, F. *et al.* New approach methodologies (NAMs): identifying and overcoming hurdles to accelerated adoption. *Toxicology research* **13,** tfae044; 10.1093/toxres/tfae044 (2024).

86. Najjar, A. *et al.* Towards best use and regulatory acceptance of generic physiologically based kinetic (PBK) models for in vitro-to-in vivo extrapolation (IVIVE) in chemical risk assessment. *Archives of toxicology*; 10.1007/s00204-022-03356-5 (2022).

87. Ewart, L. *et al.* Performance assessment and economic analysis of a human Liver-Chip for predictive toxicology. *Commun Med* **2,** 154; 10.1038/s43856-022-00209-1 (2022).

88. Hosack, T., Damry, D. & Biswas, S. Drug-induced liver injury: a comprehensive review. *Therapeutic advances in gastroenterology* **16,** 17562848231163410; 10.1177/17562848231163410 (2023).

89. Andrade, R. J. *et al.* EASL Clinical Practice Guidelines: Drug-induced liver injury. *Journal of Hepatology* **70,** 1222–1261; 10.1016/j.jhep.2019.02.014 (2019).

90. Ahmad, J. & Odin, J. A. Epidemiology and Genetic Risk Factors of Drug Hepatotoxicity. *Clinics in liver disease* **21,** 55–72; 10.1016/j.cld.2016.08.004 (2017).

91. Andrade, R. J. *et al.* Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. *Gastroenterology* **129,** 512–521; 10.1016/j.gastro.2005.05.006 (2005).

92. *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]* (National Institute of Diabetes and Digestive and Kidney Diseases, 2020).

93. Chen, M. *et al.* FDA-approved drug labeling for the study of drug-induced liver injury. *Drug discovery today* **16,** 697–703; 10.1016/j.drudis.2011.05.007 (2011).

94. Lauschke, V. M. & Ingelman-Sundberg, M. The Importance of Patient-Specific Factors for Hepatic Drug Response and Toxicity. *International journal of molecular sciences* **17,** 1714; 10.3390/ijms17101714 (2016).

95. Suzuki, A. & MinjunChen. Epidemiology and Risk Determinants of Drug-Induced Liver Injury: Current Knowledge and Future Research Needs. *Liver international : official journal of the International Association for the Study of the Liver*; 10.1111/liv.16146 (2024).

96. Li, X., Tang, J. & Mao, Y. Incidence and risk factors of drug-induced liver injury. *Liver international : official journal of the International Association for the Study of the Liver* **42,** 1999–2014; 10.1111/liv.15262 (2022).

97. Holt, M. P. & Ju, C. Mechanisms of drug-induced liver injury. *The AAPS journal* **8,** E48-54; 10.1208/aapsj080106 (2006).

98. Tujios, S. & Fontana, R. J. Mechanisms of drug-induced liver injury: from bedside to bench. *Nature reviews. Gastroenterology & hepatology* **8,** 202–211; 10.1038/nrgastro.2011.22 (2011).

99. Mosedale, M. & Watkins, P. B. Drug-induced liver injury: Advances in mechanistic understanding that will inform risk management. *Clinical Pharmacology & Therapeutics* **101,** 469–480; 10.1002/cpt.564 (2017).

100. Antoine, D. J., Williams, D. P. & Park, B. K. Understanding the role of reactive metabolites in drug-induced hepatotoxicity: state of the science. *Expert opinion on drug metabolism & toxicology* **4,** 1415–1427; 10.1517/17425255.4.11.1415 (2008).

101. Malhi, H., Gores, G. J. & Lemasters, J. J. Apoptosis and necrosis in the liver: a tale of two deaths? *Hepatology* **43,** S31-44; 10.1002/hep.21062 (2006).

102. Ramachandran, A., Visschers, R. G. J., Duan, L., Akakpo, J. Y. & Jaeschke, H. Mitochondrial dysfunction as a mechanism of drug-induced hepatotoxicity: current understanding and future perspectives. *Journal of clinical and translational research* **4,** 75–100; 10.18053/jctres.04.201801.005 (2018).

103. Liu, W. *et al.* The Immunological Mechanisms and Immune-Based Biomarkers of Drug-Induced Liver Injury. *Frontiers in pharmacology* **12,** 723940; 10.3389/fphar.2021.723940 (2021).

104. Guo, K. & van den Beucken, T. Advances in drug-induced liver injury research: in vitro models, mechanisms, omics and gene modulation techniques. *Cell & bioscience* **14,** 134; 10.1186/s13578-024-01317-2 (2024).

105. Woodhead, J. L. *et al.* Exploring BSEP inhibition-mediated toxicity with a mechanistic model of drug-induced liver injury. *Frontiers in pharmacology* **5,** 240; 10.3389/fphar.2014.00240 (2014).

106. Vorrink, S. U., Zhou, Y., Ingelman-Sundberg, M. & Lauschke, V. M. Prediction of Drug-Induced Hepatotoxicity Using Long-Term Stable Primary Hepatic 3D Spheroid Cultures in Chemically Defined Conditions. *Toxicological sciences : an official journal of the Society of Toxicology* **163,** 655–665; 10.1093/toxsci/kfy058 (2018).

107. Weitekamp, C. A. *et al.* Quantitative and qualitative concordance between clinical and nonclinical toxicity data. *Toxicol Sci* **206,** 253–272; 10.1093/toxsci/kfaf071 (2025).

108. Grzegorzewski, J. *et al.* PK-DB: pharmacokinetics database for individualized and stratified computational modeling. *Nucleic acids research* **49,** D1358-D1364; 10.1093/nar/gkaa990 (2021).

109. Sayre, R. R., Wambaugh, J. F. & Grulke, C. M. Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. *Scientific data* **7,** 122; 10.1038/s41597-020-0455-1 (2020).

110. Lippert, J. *et al.* Open Systems Pharmacology Community-An Open Access, Open Source, Open Science Approach to Modeling and Simulation in Pharmaceutical Sciences. *CPT: pharmacometrics & systems pharmacology* **8,** 878–882; 10.1002/psp4.12473 (2019).

111. Vinken, M. *et al.* Safer chemicals using less animals: kick-off of the European ONTOX project. *Toxicology* **458,** 152846; 10.1016/j.tox.2021.152846 (2021).

112. Gonzalez Hernandez, F. *et al.* An automated approach to identify scientific publications reporting pharmacokinetic parameters. *Wellcome open research* **6,** 88; 10.12688/wellcomeopenres.16718.1 (2021).

113. Rohatgi, A. *Webplotdigitizer: Version 4.6*. Available at https://automeris.io/WebPlotDigitizer (2022).

114. Ghose, A. K. & Crippen, G. M. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *Journal of Chemical Information and Computer Sciences* **27,** 21–35; 10.1021/ci00053a005 (1987).

115. Benfenati, E., Manganaro, A. & Gini, G. VEGA-QSAR: AI inside a platform for predictive toxicology, 2013.

116. Xiong, G. *et al.* ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic acids research* **49,** W5-W14; 10.1093/nar/gkab255 (2021).

117. Yun, Y. E., Cotton, C. A. & Edginton, A. N. Development of a decision tree to classify the most accurate tissue-specific tissue to plasma partition coefficient algorithm for a given compound. *Journal of pharmacokinetics and pharmacodynamics* **41,** 1–14; 10.1007/s10928-013-9342-0 (2014).

118. Endo, S., Escher, B. I. & Goss, K.-U. Capacities of membrane lipids to accumulate neutral organic chemicals. *Environmental science & technology* **45,** 5912–5921; 10.1021/es200855w (2011).

119. Loidl-Stahlhofen, A., Eckert, A., Hartmann, T. & Schöttner, M. Solid-supported lipid membranes as a tool for determination of membrane affinity: High-throughput screening of a physicochemical parameter. *Journal of pharmaceutical sciences* **90,** 599–606; 10.1002/1520-6017(200105)90:5<599::AID-JPS1016>3.0.CO;2-N (2001).

120. Lee, A. C. & Crippen, G. M. Predicting pKa. *Journal of chemical information and modeling* **49,** 2013–2033; 10.1021/ci900209w (2009).

121. Mansouri, K., Grulke, C. M., Judson, R. S. & Williams, A. J. OPERA models for predicting physicochemical properties and environmental fate endpoints. *Journal of Cheminformatics* **10,** 10; 10.1186/s13321-018-0263-1 (2018).

122. Toma, C. *et al.* QSAR Development for Plasma Protein Binding: Influence of the Ionization State. *Pharm Res* **36,** 28; 10.1007/s11095-018-2561-8 (2018).

123. Hughes, L. D., Palmer, D. S., Nigsch, F. & Mitchell, J. B. O. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P. *Journal of chemical information and modeling* **48,** 220–232; 10.1021/ci700307p (2008).

124. Williams, A. J. *et al.* The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics* **9,** 61; 10.1186/s13321-017-0247-6 (2017).

125. Tonnelier, A., Coecke, S. & Zaldívar, J.-M. Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. *Archives of toxicology* **86,** 393–403; 10.1007/s00204-011-0768-0 (2012).

126. Yamazaki, K. & Kanaoka, M. Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *Journal of pharmaceutical sciences* **93,** 1480–1494; 10.1002/jps.20059 (2004).

127. Lombardo, F., Obach, R. S., Shalaeva, M. Y. & Gao, F. Prediction of Volume of Distribution Values in Humans for Neutral and Basic Drugs Using Physicochemical Measurements and Plasma Protein Binding Data. *J. Med. Chem.* **45,** 2867–2876; 10.1021/jm0200409 (2002).

128. Riley, R. J., McGinnity, D. F. & Austin, R. P. A unified model for predicting human hepatic, metabolic clearance from in vitro intrinsic clearance data in hepatocytes and microsomes. *Drug metabolism and disposition: the biological fate of chemicals* **33,** 1304–1311; 10.1124/dmd.105.004259 (2005).

129. Sohlenius-Sternbeck, A.-K. *et al.* Evaluation of the human prediction of clearance from hepatocyte and microsome intrinsic clearance for 52 drug compounds.

Xenobiotica; the fate of foreign compounds in biological systems **40,** 637–649; 10.3109/00498254.2010.500407 (2010).

130. Votano, J. R. *et al.* QSAR Modeling of Human Serum Protein Binding with Several Modeling Techniques Utilizing Structure−Information Representation. *J. Med. Chem.* **49,** 7169–7181; 10.1021/jm051245v (2006).
131. Zhu, X.-W., Sedykh, A., Zhu, H., Liu, S.-S. & Tropsha, A. The Use of Pseudo-Equilibrium Constant Affords Improved QSAR Models of Human Plasma Protein Binding. *Pharm Res* **30,** 1790–1798; 10.1007/s11095-013-1023-6 (2013).
132. Pearce, R. G., Setzer, R. W., Davis, J. L. & Wambaugh, J. F. Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. *Journal of pharmacokinetics and pharmacodynamics* **44,** 549–565; 10.1007/s10928-017-9548-7 (2017).
133. Willmann, S., Schmitt, W., Keldenich, J., Lippert, J. & Dressman, J. B. A physiological model for the estimation of the fraction dose absorbed in humans. *Journal of medicinal chemistry* **47,** 4022–4031; 10.1021/jm030999b (2004).
134. Willmann, S. *et al.* PK-Sim®: a physiologically based pharmacokinetic 'whole-body' model. *BIOSILICO* **1,** 121–124; 10.1016/S1478-5382(03)02342-4 (2003).
135. R Core Team. *R: A Language and Environment for Statistical Computing*. Available at https://www.R-project.org/ (Vienna, Austria, 2022).
136. Ito, K. & Houston, J. B. Comparison of the use of liver models for predicting drug clearance using in vitro kinetic data from hepatic microsomes and isolated hepatocytes. *Pharm Res* **21,** 785–792; 10.1023/B:PHAM.0000026429.12114.7d (2004).
137. Snyder WS *et al.* Report of the task group on reference man. *Annals of the ICRP* **3,** iii; 10.1016/0146-6453(79)90123-4 (1979).
138. Denney, W. S., Duvvuri, S. & Buckeridge, C. Simple, Automatic Noncompartmental Analysis: The PKNCA R Package. *Journal of pharmacokinetics and pharmacodynamics* **42,** 11-107,S65; 10.1007/s10928-015-9432-2 (2015).
139. Rohatgi, A. *Webplotdigitizer: Version 5.2* (2024).
140. Geci, R. *et al.* Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans. *Arch Toxicol* **98,** 2659–2676; 10.1007/s00204-024-03764-9 (2024).
141. Dancik, Y., Miller, M. A., Jaworska, J. & Kasting, G. B. Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. *Advanced drug delivery reviews* **65,** 221–236; 10.1016/j.addr.2012.01.006 (2013).
142. Kasting, G. B. & Miller, M. A. Kinetics of finite dose absorption through skin 2: volatile compounds. *Journal of pharmaceutical sciences* **95,** 268–280; 10.1002/jps.20497 (2006).
143. R Core Team. *R: a language and environment for statistical computing*. Available at https://www.R-project.org/ (2024).
144. Y., L. & K., H. Skin thickness of Korean adults. *Surgical and Radiologic Anatomy* **24,** 183–189; 10.1007/s00276-002-0034-5 (2002).

145. Sandby-Møller, J., Poulsen, T. & Wulf, H. C. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. *Acta Dermato-Venereologica* **83,** 410–413; 10.1080/00015550310015419 (2003).

146. Egawa, M., Hirao, T. & Takahashi, M. In vivo estimation of stratum corneum thickness from water concentration profiles obtained with Raman spectroscopy. *Acta Dermato-Venereologica* **87,** 4–8; 10.2340/00015555-0183 (2007).

147. van Mulder, T. *et al.* High frequency ultrasound to assess skin thickness in healthy adults. *Vaccine* **35,** 1810–1815; 10.1016/j.vaccine.2016.07.039 (2017).

148. Oltulu, P., Ince, B., Kokbudak, N., Findik, S. & Kilinc, F. Measurement of epidermis, dermis, and total skin thicknesses from six different body regions with a new ethical histometric technique. *Turkish Journal of Plastic Surgery* **26,** 56; 10.4103/tjps.TJPS_2_17 (2018).

149. Jeong, K. M. *et al.* Ultrasonographic analysis of facial skin thickness in relation to age, site, sex, and body mass index. *Skin Research and Technology* **29,** e13426; 10.1111/srt.13426 (2023).

150. Gadaleta, D. *et al.* Comprehensive benchmarking of computational tools for predicting toxicokinetic and physicochemical properties of chemicals. *J Cheminform* **16,** 145; 10.1186/s13321-024-00931-z (2024).

151. Berezhkovskiy, L. M. Volume of distribution at steady state for a linear pharmacokinetic system with peripheral elimination. *Journal of pharmaceutical sciences* **93,** 1628–1640; 10.1002/jps.20073 (2004).

152. Kim, S. *et al.* PubChem 2025 update. *Nucleic Acids Res* **53,** D1516-D1525; 10.1093/nar/gkae1059 (2025).

153. Aleo, M. D. *et al.* Moving beyond Binary Predictions of Human Drug-Induced Liver Injury (DILI) toward Contrasting Relative Risk Potential. *Chemical research in toxicology* **33,** 223–238; 10.1021/acs.chemrestox.9b00262 (2020).

154. Shah, F. *et al.* Setting Clinical Exposure Levels of Concern for Drug-Induced Liver Injury (DILI) Using Mechanistic in vitro Assays. *Toxicological sciences : an official journal of the Society of Toxicology* **147,** 500–514; 10.1093/toxsci/kfv152 (2015).

155. Morgan, R. E. *et al.* A multifactorial approach to hepatobiliary transporter assessment enables improved therapeutic compound development. *Toxicological sciences : an official journal of the Society of Toxicology* **136,** 216–241; 10.1093/toxsci/kft176 (2013).

156. Dawson, S., Stahl, S., Paul, N., Barber, J. & Kenna, J. G. In vitro inhibition of the bile salt export pump correlates with risk of cholestatic drug-induced liver injury in humans. *Drug metabolism and disposition: the biological fate of chemicals* **40,** 130–138; 10.1124/dmd.111.040758 (2012).

157. Warner, D. J. *et al.* Mitigating the inhibition of human bile salt export pump by drugs: opportunities provided by physicochemical property modulation, in silico modeling, and structural modification. *Drug metabolism and disposition: the biological fate of chemicals* **40,** 2332–2341; 10.1124/dmd.112.047068 (2012).

158. Fäs, L. *et al.* Physiological liver microtissue 384-well microplate system for preclinical hepatotoxicity assessment of therapeutic small molecule drugs.

Toxicological sciences : an official journal of the Society of Toxicology; 10.1093/toxsci/kfae123 (2024).

159. Ewald, J. D. *et al.* Cell Painting for cytotoxicity and mode-of-action analysis in primary human hepatocytes. *bioRxiv,* 2025.01.22.634152; 10.1101/2025.01.22.634152 (2025).
160. Albrecht, W. *et al.* Improved identification of human hepatotoxic potential by summary variables of gene expression. *ALTEX*; 10.14573/altex.2403272 (2025).
161. Proctor, W. R. *et al.* Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury. *Arch Toxicol* **91,** 2849–2863; 10.1007/s00204-017-2002-1 (2017).
162. Braak, B. ter *et al.* ToxProfiler: A novel human-based reporter assay for in vitro chemical safety assessment. *Toxicology* **509,** 153970; 10.1016/j.tox.2024.153970 (2024).
163. O'Brien, P. J. *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* **80,** 580–604; 10.1007/s00204-006-0091-3 (2006).
164. Gustafsson, F., Foster, A. J., Sarda, S., Bridgland-Taylor, M. H. & Kenna, J. G. A correlation between the in vitro drug toxicity of drugs to cell lines that express human P450s and their propensity to cause liver injury in humans. *Toxicological sciences : an official journal of the Society of Toxicology* **137,** 189–211; 10.1093/toxsci/kft223 (2014).
165. Persson, M., Løye, A. F., Mow, T. & Hornberg, J. J. A high content screening assay to predict human drug-induced liver injury during drug discovery. *Journal of pharmacological and toxicological methods* **68,** 302–313; 10.1016/j.vascn.2013.08.001 (2013).
166. Schadt, S. *et al.* Minimizing DILI risk in drug discovery - A screening tool for drug candidates. *Toxicology in vitro : an international journal published in association with BIBRA* **30,** 429–437; 10.1016/j.tiv.2015.09.019 (2015).
167. Williams, D. P., Lazic, S. E., Foster, A. J., Semenova, E. & Morgan, P. Predicting Drug-Induced Liver Injury with Bayesian Machine Learning. *Chemical research in toxicology* **33,** 239–248; 10.1021/acs.chemrestox.9b00264 (2020).
168. Porceddu, M. *et al.* Prediction of liver injury induced by chemicals in human with a multiparametric assay on isolated mouse liver mitochondria. *Toxicological sciences : an official journal of the Society of Toxicology* **129,** 332–345; 10.1093/toxsci/kfs197 (2012).
169. U.S. Environmental Protection Agency. *ToxCast retrieved from https://comptox.epa.gov/dashboard/batch-search on Feb 24, 2024* (2025).
170. Chen, M. *et al.* DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug discovery today* **21,** 648–653; 10.1016/j.drudis.2016.02.015 (2016).
171. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res* **44,** D1075-9; 10.1093/nar/gkv1075 (2016).
172. Kotsampasakou, E. & Ecker, G. F. Predicting Drug-Induced Cholestasis with the Help of Hepatic Transporters-An in Silico Modeling Approach. *Journal of chemical information and modeling* **57,** 608–615; 10.1021/acs.jcim.6b00518 (2017).

173. Lima Toccafondo Vieira, M. de & Tagliati, C. A. Hepatobiliary transporters in drug-induced cholestasis: a perspective on the current identifying tools. *Expert opinion on drug metabolism & toxicology* **10,** 581–597; 10.1517/17425255.2014.884069 (2014).

174. Riede, J., Poller, B., Huwyler, J. & Camenisch, G. Assessing the Risk of Drug-Induced Cholestasis Using Unbound Intrahepatic Concentrations. *Drug metabolism and disposition: the biological fate of chemicals* **45,** 523–531; 10.1124/dmd.116.074179 (2017).

175. van Brantegem, P., Chatterjee, S., Bruyn, T. de, Annaert, P. & Deferm, N. Drug-induced cholestasis assay in primary hepatocytes. *MethodsX* **7,** 101080; 10.1016/j.mex.2020.101080 (2020).

176. Bruijn, V. M. P. de & Rietjens, I. M. C. M. From hazard to risk prioritization: a case study to predict drug-induced cholestasis using physiologically based kinetic modeling. *Archives of toxicology*; 10.1007/s00204-024-03775-6 (2024).

177. Drees, A. *et al.* Optimization of the drug-induced cholestasis index based on advanced modeling for predicting liver toxicity. *Toxicology* **514,** 154119; 10.1016/j.tox.2025.154119 (2025).

178. Seal, S. *et al.* PKSmart: an open-source computational model to predict intravenous pharmacokinetics of small molecules. *J Cheminform* **17,** 147; 10.1186/s13321-025-01066-5 (2025).

179. Willmann, S., Lippert, J. & Schmitt, W. From physicochemistry to absorption and distribution: predictive mechanistic modelling and computational tools. *Expert opinion on drug metabolism & toxicology* **1,** 159–168; 10.1517/17425255.1.1.159 (2005).

180. Kister, B., Blank, L. M., Pollmanns, M., Wirtz, T. H. & Kuepfer, L. A physiologically-based model of bile acid metabolism in humans. *bioRxiv,* 2025.07.19.665677; 10.1101/2025.07.19.665677 (2025).

181. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology* **22,** 3099–3108; 10.1016/0006-2952(73)90196-2 (1973).

182. Lombardo, F., Bentzien, J., Berellini, G. & Muegge, I. Prediction of Human Clearance Using In Silico Models with Reduced Bias. *Molecular pharmaceutics* **21,** 1192–1203; 10.1021/acs.molpharmaceut.3c00812 (2024).

183. Wetmore, B. A. *et al.* Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. *Toxicological sciences : an official journal of the Society of Toxicology* **148,** 121–136; 10.1093/toxsci/kfv171 (2015).

184. Breen, M., Ring, C. L., Kreutz, A., Goldsmith, M.-R. & Wambaugh, J. F. High-throughput PBTK models for in vitro to in vivo extrapolation. *Expert opinion on drug metabolism & toxicology* **17,** 903–921; 10.1080/17425255.2021.1935867 (2021).

185. Miljković, F. *et al.* Machine Learning Models for Human In Vivo Pharmacokinetic Parameters with In-House Validation. *Molecular pharmaceutics* **18,** 4520–4530; 10.1021/acs.molpharmaceut.1c00718 (2021).

186. Fagerholm, U., Hellberg, S. & Spjuth, O. Advances in Predictions of Oral Bioavailability of Candidate Drugs in Man with New Machine Learning Methodology. *Molecules (Basel, Switzerland)* **26**; 10.3390/molecules26092572 (2021).

187. Beckers, M., Yonchev, D., Desrayaud, S., Gerebtzoff, G. & Rodríguez-Pérez, R. *DeepCt: Predicting pharmacokinetic concentration-time curves and compartmental models from chemical structure using deep learning* (2024).

188. Li, Y. *et al.* A Combination of Machine Learning and PBPK Modeling Approach for Pharmacokinetics Prediction of Small Molecules in Humans. *Pharmaceutical research* **41,** 1369–1379; 10.1007/s11095-024-03725-y (2024).

189. Danishuddin, Kumar, V., Faheem, M. & Woo Lee, K. A decade of machine learning-based predictive models for human pharmacokinetics: Advances and challenges. *Drug discovery today* **27,** 529–537; 10.1016/j.drudis.2021.09.013 (2022).

190. Pillai, N., Dasgupta, A., Sudsakorn, S., Fretland, J. & Mavroudis, P. D. Machine Learning guided early drug discovery of small molecules. *Drug discovery today* **27,** 2209–2215; 10.1016/j.drudis.2022.03.017 (2022).

191. Mavroudis, P. D., Teutonico, D., Abos, A. & Pillai, N. Application of machine learning in combination with mechanistic modeling to predict plasma exposure of small molecules. *Front. Syst. Biol.* **3**; 10.3389/fsysb.2023.1180948 (2023).

192. Führer, F. *et al.* A deep neural network: mechanistic hybrid model to predict pharmacokinetics in rat. *Journal of computer-aided molecular design* **38,** 7; 10.1007/s10822-023-00547-9 (2024).

193. Walter, M. *et al.* Predicting Pharmacokinetics in Rats Using Machine Learning: A Comparative Study Between Empirical, Compartmental, and PBPK-Based Approaches. *Clinical Translational Sci* **18,** e70150; 10.1111/cts.70150 (2025).

194. Jost, F., Giegerich, C., Grebner, C., Matter, H. & Cordes, H. Machine Learning Meets Pharmacokinetics: A Comparative Analysis of Predictive Models for Plasma Concentration-Time Profiles. *bioRxiv,* 2025.05.05.652167; 10.1101/2025.05.05.652167 (2025).

195. Jia, X. *et al.* Application of Machine Learning and Mechanistic Modeling to Predict Intravenous Pharmacokinetic Profiles in Humans. *Journal of medicinal chemistry*; 10.1021/acs.jmedchem.5c00340 (2025).

196. Gruber, A. *et al.* Prediction of human pharmacokinetics from chemical structure: combining mechanistic modeling with machine learning. *Journal of pharmaceutical sciences*; 10.1016/j.xphs.2023.10.035 (2023).

197. Kamiya, Y. *et al.* In Silico Prediction of Input Parameters for Simplified Physiologically Based Pharmacokinetic Models for Estimating Plasma, Liver, and Kidney Exposures in Rats after Oral Doses of 246 Disparate Chemicals. *Chemical research in toxicology* **34,** 507–513; 10.1021/acs.chemrestox.0c00336 (2021).

198. Obrezanova, O. *et al.* Prediction of In Vivo Pharmacokinetic Parameters and Time-Exposure Curves in Rats Using Machine Learning from the Chemical Structure. *Molecular pharmaceutics* **19,** 1488–1504; 10.1021/acs.molpharmaceut.2c00027 (2022).

199. Handa, K. *et al.* Prediction of Compound Plasma Concentration-Time Profiles in Mice Using Random Forest. *Molecular pharmaceutics* **20,** 3060–3072; 10.1021/acs.molpharmaceut.3c00071 (2023).

200. Schmitt, W. General approach for the calculation of tissue to plasma partition coefficients. *Toxicology in vitro : an international journal published in association with BIBRA* **22,** 457–467; 10.1016/j.tiv.2007.09.010 (2008).

201. Poulin, P. & Theil, F.-P. Prediction of pharmacokinetics prior to in vivo studies. II. Generic physiologically based pharmacokinetic models of drug disposition. *Journal of pharmaceutical sciences* **91,** 1358–1370; 10.1002/jps.10128 (2002).

202. Volpe, D. A. Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. *Future medicinal chemistry* **3,** 2063–2077; 10.4155/fmc.11.149 (2011).

203. Dent, M. P. *et al.* Paving the way for application of next generation risk assessment to safety decision-making for cosmetic ingredients. *Regulatory toxicology and pharmacology : RTP* **125,** 105026; 10.1016/j.yrtph.2021.105026 (2021).

204. Han, P., Li, X., Yang, J., Zhang, Y. & Chen, J. Advancing Toxicity Predictions: A Review on in Vitro to in Vivo Extrapolation in Next-Generation Risk Assessment. *Environment & health (Washington, D.C.)* **2,** 499–513; 10.1021/envhealth.4c00043 (2024).

205. Poet, T. S. & McDougal, J. N. Skin absorption and human risk assessment. *Chemico-biological interactions* **140,** 19–34; 10.1016/s0009-2797(02)00013-3 (2002).

206. Akomeah, F. K., Martin, G. P. & Brown, M. B. Variability in human skin permeability in vitro: comparing penetrants with different physicochemical properties. *Journal of pharmaceutical sciences* **96,** 824–834; 10.1002/jps.20773 (2007).

207. Taylor, K. & Rego Alvarez, L. Regulatory drivers in the last 20 years towards the use of in silico techniques as replacements to animal testing for cosmetic-related substances. *Computational Toxicology* **13,** 100112; 10.1016/j.comtox.2019.100112 (2020).

208. Paini, A., Tan, Y.-M., Sachana, M. & Worth, A. Gaining acceptance in next generation PBK modelling approaches for regulatory assessments - An OECD international effort. *Computational toxicology (Amsterdam, Netherlands)* **18,** 100163; 10.1016/j.comtox.2021.100163 (2021).

209. Edizcan, Z., Schaller, S., Kuepfer, L. & Geci, R. High-throughput PBK modelling for dermal exposure: a pragmatic approach to predict systemic pharmacokinetics. *Arch Toxicol* **100,** 2421–2432; 10.1007/s00204-026-04357-4 (2026).

210. Cross, S. E., Magnusson, B. M., Winckle, G., Anissimov, Y. & Roberts, M. S. Determination of the effect of lipophilicity on the in vitro permeability and tissue reservoir characteristics of topically applied solutes in human skin layers. *The Journal of Investigative Dermatology* **120,** 759–764; 10.1046/j.1523-1747.2003.12131.x (2003).

211. Roberts, M. S., Cross, S. E. & Anissimov, Y. G. Factors affecting the formation of a skin reservoir for topically applied solutes. *Skin Pharmacology and Physiology* **17,** 3–16; 10.1159/000074057 (2004).

212. Anissimov, Y. G. & Roberts, M. S. Diffusion modelling of percutaneous absorption kinetics: 4. Effects of a slow equilibration process within stratum corneum on absorption and desorption kinetics. *Journal of pharmaceutical sciences* **98,** 772–781; 10.1002/jps.21461 (2009).

213. Najjar, A. *et al.* Use of Physiologically-Based Kinetics Modelling to Reliably Predict Internal Concentrations of the UV Filter, Homosalate, After Repeated Oral and Topical Application. *Frontiers in pharmacology* **12,** 802514; 10.3389/fphar.2021.802514 (2021).

214. Clarke, J. F., Thakur, K. & Polak, S. A mechanistic physiologically based model to assess the effect of study design and modified physiology on formulation safe space for virtual bioequivalence of dermatological drug products. *Frontiers in pharmacology* **13,** 1007496; 10.3389/fphar.2022.1007496 (2022).

215. Patel, N. *et al.* Multi-phase multi-layer mechanistic dermal absorption (MPML MechDermA) model to predict local and systemic exposure of drug products applied on skin. *CPT: pharmacometrics & systems pharmacology* **11,** 1060–1084; 10.1002/psp4.12814 (2022).

216. Dancik, Y., Zhang, Y., Telaprolu, K. C. & Polak, S. Physiologically based pharmacokinetic modelling of in vitro skin permeation of sunscreen actives under various experimental conditions. *International Journal of Pharmaceutics* **682,** 125977; 10.1016/j.ijpharm.2025.125977 (2025).

217. Li, H. *et al.* PBK modelling of topical application and characterisation of the uncertainty of Cmax estimate: A case study approach. *Toxicology and Applied Pharmacology* **442,** 115992; 10.1016/j.taap.2022.115992 (2022).

218. Hamadeh, A., Troutman, J., Najjar, A. & Edginton, A. A Mechanistic Bayesian Inferential Workflow for Estimation of In Vivo Skin Permeation from In Vitro Measurements. *Journal of pharmaceutical sciences* **111,** 838–851; 10.1016/j.xphs.2021.11.028 (2022).

219. Hamadeh, A. *et al.* Mechanistic Skin Modeling of Plasma Concentrations of Sunscreen Active Ingredients Following Facial Application. *Journal of pharmaceutical sciences* **113,** 806–825; 10.1016/j.xphs.2023.09.017 (2024).

220. Basak, S. C., Mills, D. & Mumtaz, M. M. A quantitative structure-activity relationship (QSAR) study of dermal absorption using theoretical molecular descriptors. *SAR and QSAR in Environmental Research* **18,** 45–55; 10.1080/10629360601033671 (2007).

221. Arora, S. *et al.* Mechanistic Modeling of In Vitro Skin Permeation and Extrapolation to In Vivo for Topically Applied Metronidazole Drug Products Using a Physiologically Based Pharmacokinetic Model. *Molecular pharmaceutics* **19,** 3139–3152; 10.1021/acs.molpharmaceut.2c00229 (2022).

222. Hamadeh, A. & Edginton, A. Efficient large-scale mechanism-based computation of skin permeability. *Computational Toxicology* **26,** 100263; 10.1016/j.comtox.2023.100263 (2023).

223. Narita, I. *et al.* In silico model to predict dermal absorption of chemicals in finite dose conditions. *The Journal of Toxicological Sciences* **50,** 171–186; 10.2131/jts.50.171 (2025).

224. Punt, A. *et al.* New approach methodologies (NAMs) for human-relevant biokinetics predictions. Meeting the paradigm shift in toxicology towards an animal-free chemical risk assessment. *ALTEX* **37,** 607–622; 10.14573/altex.2003242 (2020).

225. OECD. Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes. OECD Publishing, 2021.

226. Ball, N. *et al.* A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch Toxicol* **96,** 743–766; 10.1007/s00204-021-03215-9 (2022).

227. EEA. The European environment - state and outlook 2020: knowledge for transition to a sustainable Europe. (2019).

228. EU. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). (2006).

229. Botham, P. *et al.* Analysis of health concerns not addressed by REACH for low tonnage chemicals and opportunities for new approach methodology. *Arch Toxicol* **97,** 3075–3083; 10.1007/s00204-023-03601-5 (2023).

230. Geci, R., Paini, A., Worth, A. P., Kuepfer, L. & Schaller, S. Potential systemic availability classification of chemicals for safety assessment. *Environment international* **202,** 109636; 10.1016/j.envint.2025.109636 (2025).

231. Gobas, F. A. P. C., Powell, D. E., Woodburn, K. B., Springer, T. & Huggett, D. B. Bioaccumulation of decamethylpentacyclosiloxane (D5): A review. *Enviro Toxic and Chemistry* **34,** 2703–2714; 10.1002/etc.3242 (2015).

232. Kierkegaard, A., Bignert, A. & McLachlan, M. S. Bioaccumulation of decamethylcyclopentasiloxane in perch in Swedish lakes. *Chemosphere* **93,** 789–793; 10.1016/j.chemosphere.2012.10.050 (2013).

233. Gobas, F. A. P. C. *et al.* Fugacity and activity analysis of the bioaccumulation and environmental risks of decamethylcyclopentasiloxane (D5). *Enviro Toxic and Chemistry* **34,** 2723–2731; 10.1002/etc.2942 (2015).

234. Sunderland, E. M. *et al.* A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J Expo Sci Environ Epidemiol* **29,** 131–147; 10.1038/s41370-018-0094-1 (2019).

235. Brunn, H. *et al.* PFAS: forever chemicals—persistent, bioaccumulative and mobile. Reviewing the status and the need for their phase out and remediation of contaminated sites. *Environ Sci Eur* **35,** 1–50; 10.1186/s12302-023-00721-8 (2023).

236. Tzimas, G., Thiel, R., Chahoud, I. & Nau, H. The area under the concentration-time curve of all-trans-retinoic acid is the most suitable pharmacokinetic correlate to the embryotoxicity of this retinoid in the rat. *Toxicology and Applied Pharmacology* **143,** 436–444; 10.1006/taap.1997.8105 (1997).

237. Groothuis, F. A. *et al.* Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology* **332,** 30–40; 10.1016/j.tox.2013.08.012 (2015).

238. Moerenhout, T. M. J. A., Chen, J., Bouwmeester, H., Rietjens, I. M. C. M. & Kramer, N. I. Development of a Generic Physiologically Based Kinetic Model for the Prediction of Internal Exposure to Organophosphate Pesticides. *Environmental science & technology*; 10.1021/acs.est.4c06534 (2024).

239. Sidhu, G. K. *et al.* Toxicity, monitoring and biodegradation of organophosphate pesticides: A review. *Critical Reviews in Environmental Science and Technology* **49,** 1135–1187; 10.1080/10643389.2019.1565554 (2019).

240. Cherkasov, A. *et al.* QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **57,** 4977–5010; 10.1021/jm4004285 (2014).

241. Kirchmair, J. *et al.* Predicting drug metabolism: experiment and/or computation? *Nature reviews. Drug discovery* **14,** 387–404; 10.1038/nrd4581 (2015).

242. Houston, J. B. & Galetin, A. Progress towards prediction of human pharmacokinetic parameters from in vitro technologies. *Drug metabolism reviews* **35,** 393–415; 10.1081/dmr-120026870 (2003).

243. Proença, S. *et al.* Effective exposure of chemicals in in vitro cell systems: A review of chemical distribution models. *Toxicology in vitro : an international journal published in association with BIBRA* **73,** 105133; 10.1016/j.tiv.2021.105133 (2021).

244. Klaassen, C. D. & Aleksunes, L. M. Xenobiotic, bile acid, and cholesterol transporters: function and regulation. *Pharmacological reviews* **62,** 1–96; 10.1124/pr.109.002014 (2010).

245. Li, Z. & Xiong, J. A dynamic inventory database for assessing age-, gender-, and route-specific chronic internal exposure to chemicals in support of human exposome research. *Journal of environmental management* **339,** 117867; 10.1016/j.jenvman.2023.117867 (2023).

246. Arnot, J. A. & Gobas, F. A. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environ. Rev.* **14,** 257–297; 10.1139/a06-005 (2006).

247. Arnot, J. A. *et al.* Molecular size cutoff criteria for screening bioaccumulation potential: fact or fiction? *Integrated environmental assessment and management* **6,** 210–224; 10.1897/IEAM_2009-051.1 (2010).

248. Coecke, S. *et al.* Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. *Toxicology in vitro : an international journal published in association with BIBRA* **27,** 1570–1577; 10.1016/j.tiv.2012.06.012 (2013).

249. Wambaugh, J. F. *et al.* Toxicokinetic Triage for Environmental Chemicals. *Toxicological sciences : an official journal of the Society of Toxicology* **147,** 55–67; 10.1093/toxsci/kfv118 (2015).

250. U.S. Environmental Protection Agency. *Polymer Exemption Guidance Manual for Premanufacture Notices for New Chemicals under the Toxic Substances Control Act (TSCA): A Technical Manual to Accompany, but not Supersede the "Premanufacture Notification Exemptions; Revisions of Exemptions for Polymers; Final Rule" found at 40 CFR Part 723, (60) FR 16316–16336, Published Wednesday, March 29, 1995* (Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, Washington, DC, 1997).

251. Schmeisser, S. *et al.* New approach methodologies in human regulatory toxicology - Not if, but how and when! *Environment international* **178,** 108082; 10.1016/j.envint.2023.108082 (2023).

252. Worth, A. P. & Berggren, E. A Twin Transition in Regulatory Toxicology-moving towards Chemicals 2.0 and phasing out animal testing. *Toxicological sciences : an official journal of the Society of Toxicology*; 10.1093/toxsci/kfae130 (2024).

253. Rotroff, D. M. *et al.* Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicological sciences : an official journal of the Society of Toxicology* **117,** 348–358; 10.1093/toxsci/kfq220 (2010).

254. Doe, J. E. *et al.* Framework for classifying chemicals for repeat dose toxicity using NAMs. *Arch Toxicol* **99,** 3223–3246; 10.1007/s00204-025-04069-1 (2025).

255. Watkins, P. B. Drug safety sciences and the bottleneck in drug development. *Clinical Pharmacology & Therapeutics* **89,** 788–790; 10.1038/clpt.2011.63 (2011).

256. Hoofnagle, J. H. & Björnsson, E. S. Drug-Induced Liver Injury - Types and Phenotypes. *The New England journal of medicine* **381,** 264–273; 10.1056/NEJMra1816149 (2019).

257. Chen, M., Borlak, J. & Tong, W. High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. *Hepatology (Baltimore, Md.)* **58,** 388–396; 10.1002/hep.26208 (2013).

258. Yucha, R. W. *et al.* In Vitro Drug-Induced Liver Injury Prediction: Criteria Optimization of Efflux Transporter IC50 and Physicochemical Properties. *Toxicological sciences : an official journal of the Society of Toxicology* **157,** 487–499; 10.1093/toxsci/kfx060 (2017).

259. Walker, P. A., Ryder, S., Lavado, A., Dilworth, C. & Riley, R. J. The evolution of strategies to minimise the risk of human drug-induced liver injury (DILI) in drug discovery and development. *Archives of toxicology* **94,** 2559–2585; 10.1007/s00204-020-02763-w (2020).

260. Norman, B. H. Drug Induced Liver Injury (DILI). Mechanisms and Medicinal Chemistry Avoidance/Mitigation Strategies. *Journal of medicinal chemistry* **63,** 11397–11419; 10.1021/acs.jmedchem.0c00524 (2020).

261. Taylor, K. *et al.* Perspective: How complex in vitro models are addressing the challenges of predicting drug-induced liver injury. *Front. Drug Discov.* **5,** 1536756; 10.3389/fddsv.2025.1536756 (2025).

262. Dawson, D. E. *et al.* Designing QSARs for Parameters of High-Throughput Toxicokinetic Models Using Open-Source Descriptors. *Environmental science & technology* **55,** 6505–6517; 10.1021/acs.est.0c06117 (2021).

263. Stieger, B. Role of the bile salt export pump, BSEP, in acquired forms of cholestasis. *Drug metabolism reviews* **42,** 437–445; 10.3109/03602530903492004 (2010).

264. Vinken, M. *et al.* Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Toxicological sciences : an official journal of the Society of Toxicology* **136,** 97–106; 10.1093/toxsci/kft177 (2013).

265. Chan, R. & Benet, L. Z. Evaluation of the Relevance of DILI Predictive Hypotheses in Early Drug Development: Review of In Vitro Methodologies vs BDDCS Classification. *Toxicology research* **7,** 358–370; 10.1039/C8TX00016F (2018).

266. Chan, R. & Benet, L. Z. Measures of BSEP Inhibition In Vitro Are Not Useful Predictors of DILI. *Toxicological sciences : an official journal of the Society of Toxicology* **162,** 499–508; 10.1093/toxsci/kfx284 (2018).

267. Seal, S. *et al.* Improved Detection of Drug-Induced Liver Injury by Integrating Predicted In Vivo and In Vitro Data. *Chemical research in toxicology* **37,** 1290–1305; 10.1021/acs.chemrestox.4c00015 (2024).

268. Garcia de Lomana, M., Gadaleta, D., Raschke, M., Fricke, R. & Montanari, F. Predicting Liver-Related In Vitro Endpoints with Machine Learning to Support Early

Detection of Drug-Induced Liver Injury. *Chemical research in toxicology* **38,** 656–671; 10.1021/acs.chemrestox.4c00453 (2025).

269. Minerali, E., Foil, D. H., Zorn, K. M., Lane, T. R. & Ekins, S. Comparing Machine Learning Algorithms for Predicting Drug-Induced Liver Injury (DILI). *Molecular pharmaceutics* **17,** 2628–2637; 10.1021/acs.molpharmaceut.0c00326 (2020).

270. Jia, X., Wang, T. & Zhu, H. Advancing Computational Toxicology by Interpretable Machine Learning. *Environmental science & technology*; 10.1021/acs.est.3c00653 (2023).

271. Hemmerich, J. & Ecker, G. F. In silico toxicology: From structure-activity relationships towards deep learning and adverse outcome pathways. *Wiley interdisciplinary reviews. Computational molecular science* **10,** e1475; 10.1002/wcms.1475 (2020).

272. Geci, R., Sayin, A. Z., Schaller, S. & Kuepfer, L. Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury. *Arch Toxicol* **100,** 2029–2046; 10.1007/s00204-026-04305-2 (2026).

273. Lammert, C. *et al.* Relationship between daily dose of oral medications and idiosyncratic drug-induced liver injury: search for signals. *Hepatology (Baltimore, Md.)* **47,** 2003–2009; 10.1002/hep.22272 (2008).

274. Middleton, A. M. *et al.* Are Non-animal Systemic Safety Assessments Protective? A Toolbox and Workflow. *Toxicological sciences : an official journal of the Society of Toxicology* **189,** 124–147; 10.1093/toxsci/kfac068 (2022).

275. Cable, S. *et al.* Advancing Systemic Toxicity Risk Assessment: Evaluation of a NAM-based Toolbox Approach. *Toxicological sciences : an official journal of the Society of Toxicology*; 10.1093/toxsci/kfae159 (2024).

276. YOUDEN, W. J. Index for rating diagnostic tests. *Cancer* **3,** 32–35; 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3 (1950).

277. Zu, Y., Yang, J., Zhang, C. & Liu, D. The Pathological Mechanisms of Estrogen-Induced Cholestasis: Current Perspectives. *Frontiers in pharmacology* **12,** 761255; 10.3389/fphar.2021.761255 (2021).

278. Meyers, M., Slikker, W., Pascoe, G. & Vore, M. Characterization of cholestasis induced by estradiol-17 beta-D-glucuronide in the rat. *The Journal of pharmacology and experimental therapeutics* **214,** 87–93 (1980).

279. Yip, L. Y. *et al.* The liver-gut microbiota axis modulates hepatotoxicity of tacrine in the rat. *Hepatology (Baltimore, Md.)* **67,** 282–295; 10.1002/hep.29327 (2018).

280. Schreiber, N. *et al.* Acute Liver Failure after Ingestion of Fried Rice Balls: A Case Series of Bacillus cereus Food Poisonings. *Toxins* **14**; 10.3390/toxins14010012 (2021).

281. Decleer, M. *et al.* Oxygen Consumption Rate Analysis of Mitochondrial Dysfunction Caused by Bacillus cereus Cereulide in Caco-2 and HepG2 Cells. *Toxins* **10**; 10.3390/toxins10070266 (2018).

282. Maurer, T. S., Smith, D., Beaumont, K. & Di, L. Dose Predictions for Drug Design. *J. Med. Chem.* **63,** 6423–6435; 10.1021/acs.jmedchem.9b01365 (2020).

283. Funke, K. & Rockey, D. C. Cholestatic Drug-Induced Liver Injury Caused by Mifepristone. *Hepatology (Baltimore, Md.)* **69,** 2704–2706; 10.1002/hep.30465 (2019).

284. Shah, I., Putnam, T., Daugherty, E., Vyas, N. & Chuang, K.-Y. Mifepristone: An Uncommon Cause of Drug-Induced Liver Injury. *Gastroenterology research* **12,** 181–184; 10.14740/gr1188 (2019).

285. Ault, T. A., Braxton, D. R., Watson, R. A., Marcus, A. O. & Fong, T.-L. Mifepristone induced liver injury in a patient with Cushing syndrome: a case report and review of the literature. *Journal of medical case reports* **17,** 33; 10.1186/s13256-022-03696-x (2023).

286. Rahimi, A. R., Sanders, S. & Warren, A. Primidone and Liver Abnormalities. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY* **62,** S240 (2014).

287. Poffenbarger, E. M. & Hardy, R. M. Hepatic cirrhosis associated with long-term primidone therapy in a dog. *Journal of the American Veterinary Medical Association* **186,** 978–980 (1985).

288. Albrecht, W. *et al.* Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood concentrations. *Archives of toxicology* **93,** 1609–1637; 10.1007/s00204-019-02492-9 (2019).

289. Watkins, P. B. DILIsym: Quantitative systems toxicology impacting drug development. *Current opinion in toxicology* **23-24,** 67–73; 10.1016/j.cotox.2020.06.003 (2020).

290. Shebley, M. *et al.* Physiologically Based Pharmacokinetic Model Qualification and Reporting Procedures for Regulatory Submissions: A Consortium Perspective. *Clinical Pharmacology & Therapeutics* **104,** 88–110; 10.1002/cpt.1013 (2018).

291. Internal Market, Industry, Entrepreneurship and SMEs. EPAA Designathon for Human Systemic Toxicity. Available at https://single-market-economy.ec.europa.eu/calls-expression-interest/epaa-designathon-human-systemic-toxicity_en (2025).

292. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular informatics* **29,** 476–488; 10.1002/minf.201000061 (2010).

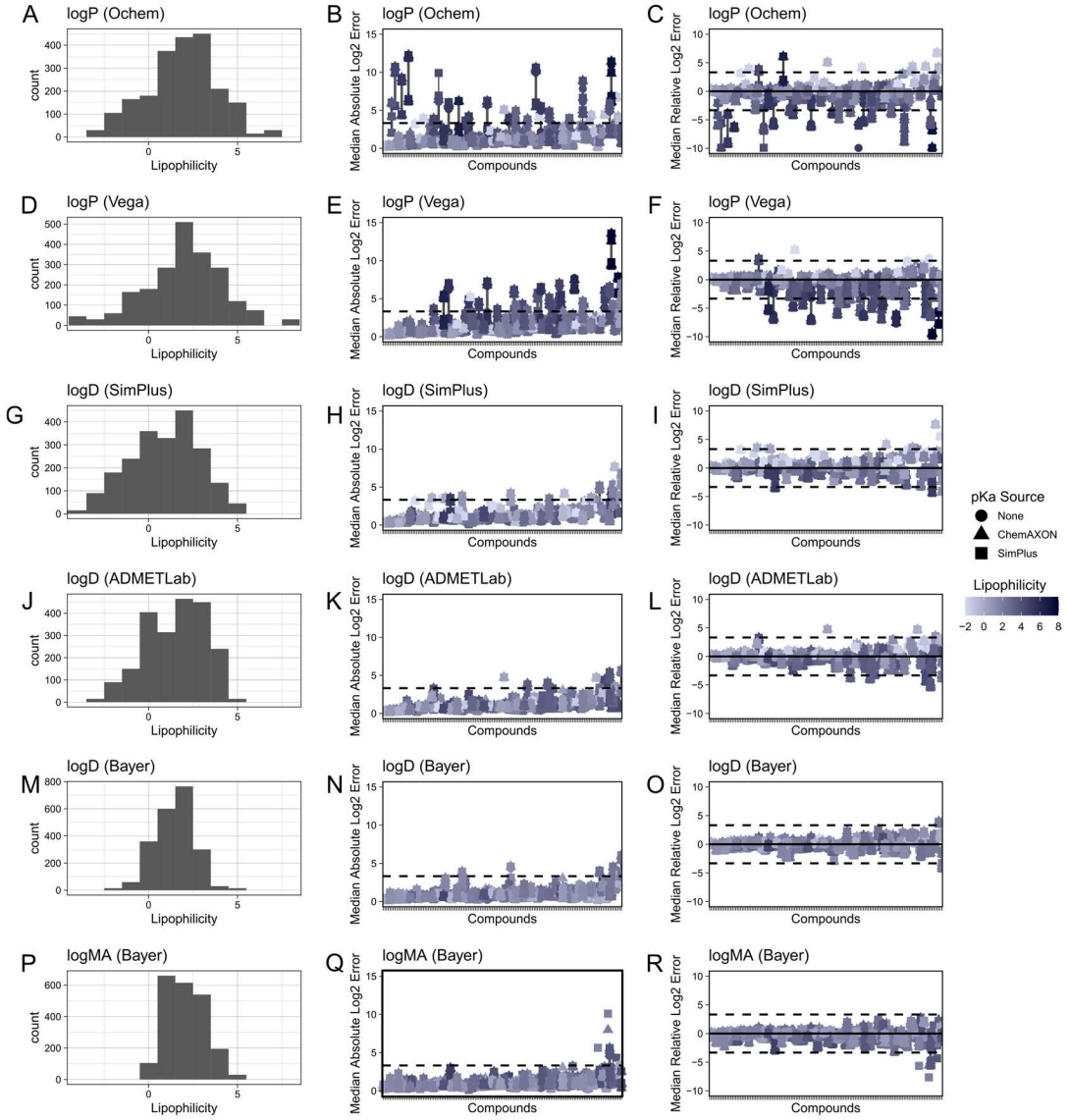
293. Fourches, D., Muratov, E. & Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of chemical information and modeling* **50,** 1189–1204; 10.1021/ci100176x (2010).

294. Lipinski, C. & Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **432,** 855–861; 10.1038/nature03193 (2004).

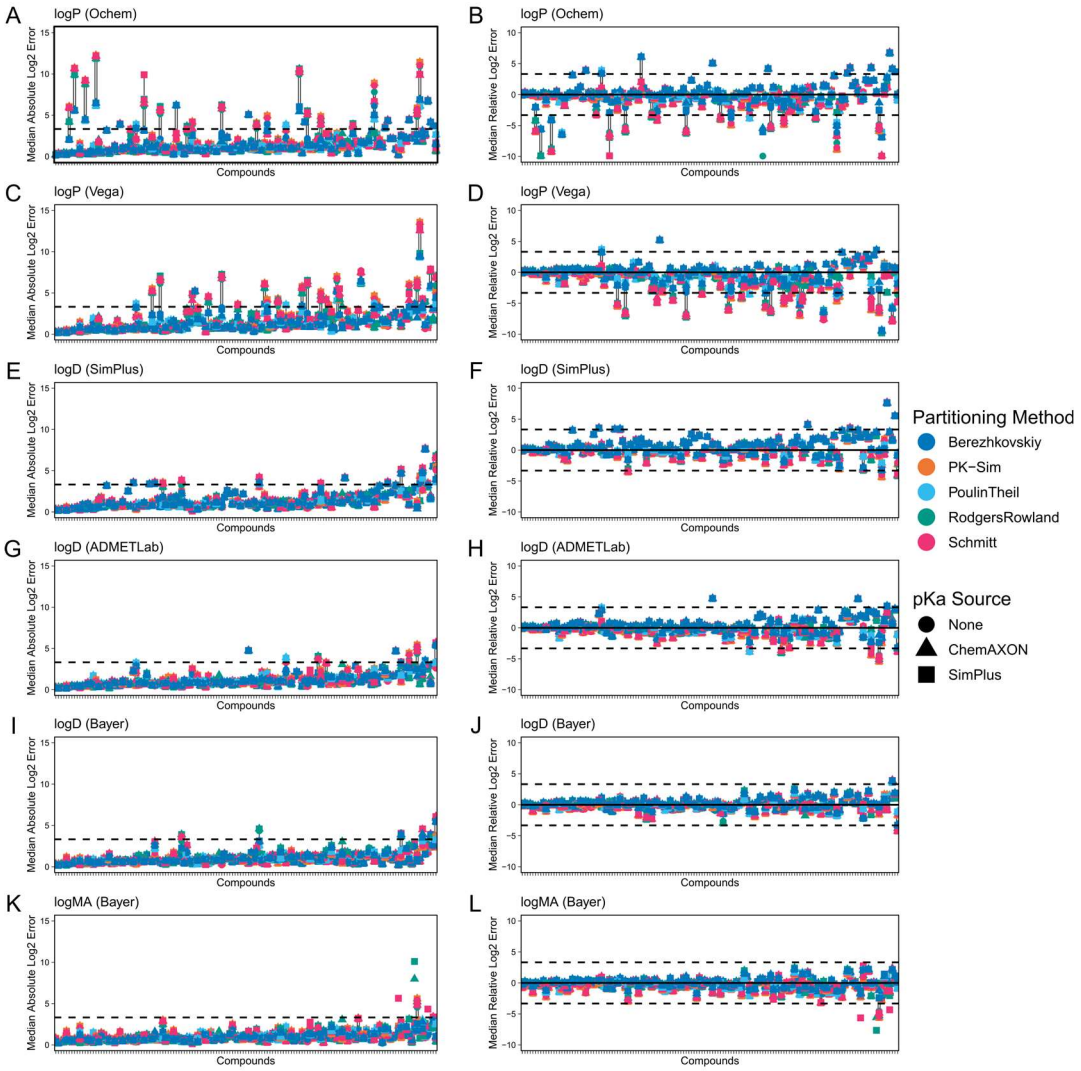
295. Kinch, M. S., Haynesworth, A., Kinch, S. L. & Hoyer, D. An overview of FDA-approved new molecular entities: 1827-2013. *Drug discovery today* **19,** 1033–1039; 10.1016/j.drudis.2014.03.018 (2014).

296. European Commission. Communication from the Commission on the European Citizens' Initiative (ECI) 'Save cruelty-free cosmetics – Commit to a Europe without animal testing'. Publications Office of the European Union, 2023.

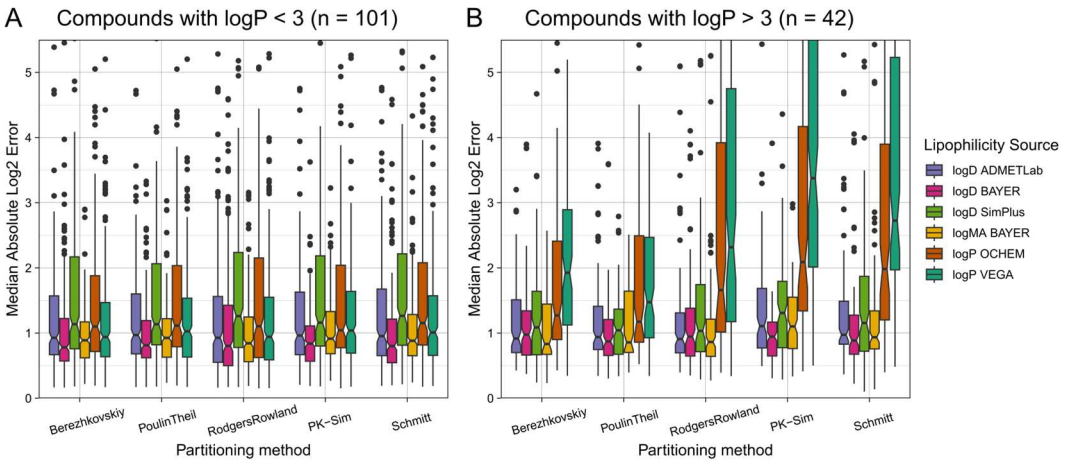
Appendix



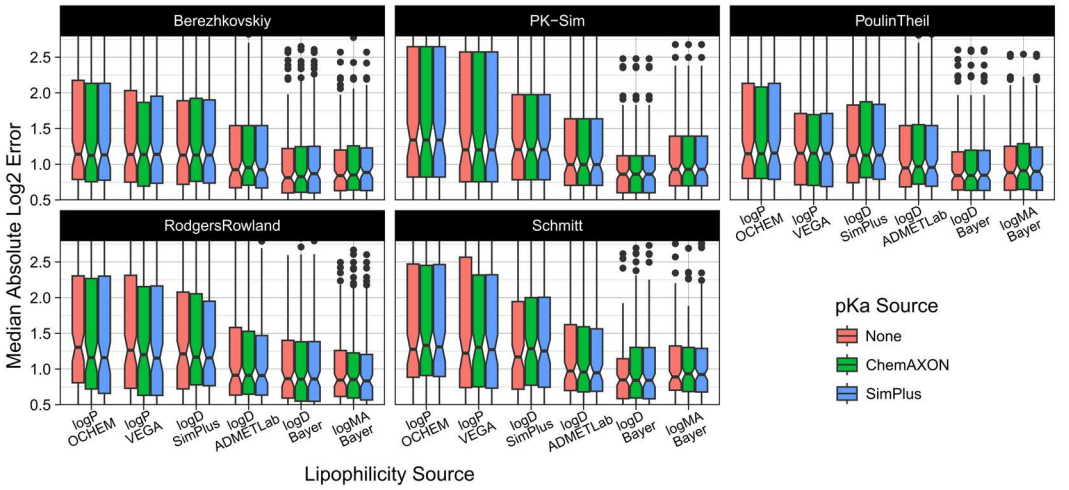
Supplementary Figure 1: Comparison of predictive performances of different lipophilicity prediction tools coloured by lipophilicity value (step 1). Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Left column shows histograms of predicted lipophilicity values by the various prediction tools. Middle column shows Median Absolute Log2 Errors of every compound, colour-coded by lipophilicity value predicted by the tool, and right column shows Median Relative Log2 Errors. Dashed lines indicate tenfold errors. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



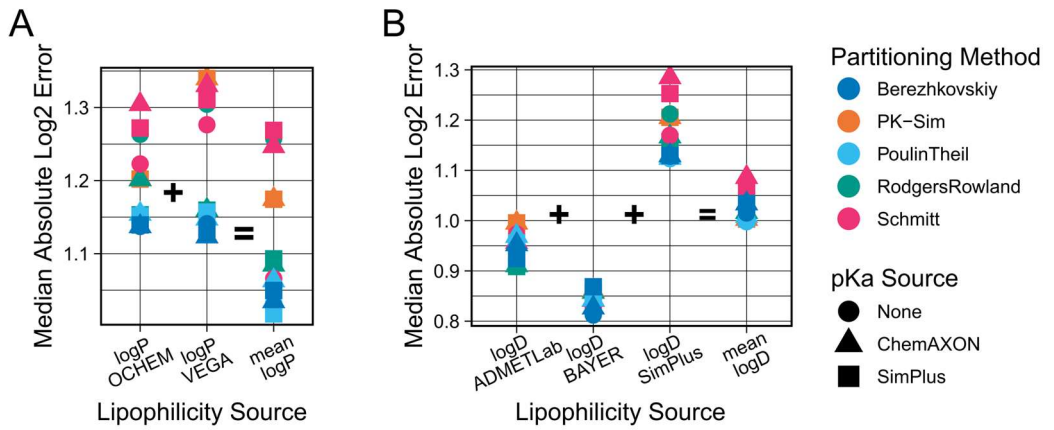
Supplementary Figure 2: Comparison of predictive performances of different lipophilicity prediction tools coloured by partitioning method (step 1). Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Left column Median Absolute Log2 Errors of every compound, colour coded by partitioning method used for simulation, and right column shows Median Relative Log2 Errors. Dashed lines indicate tenfold errors. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



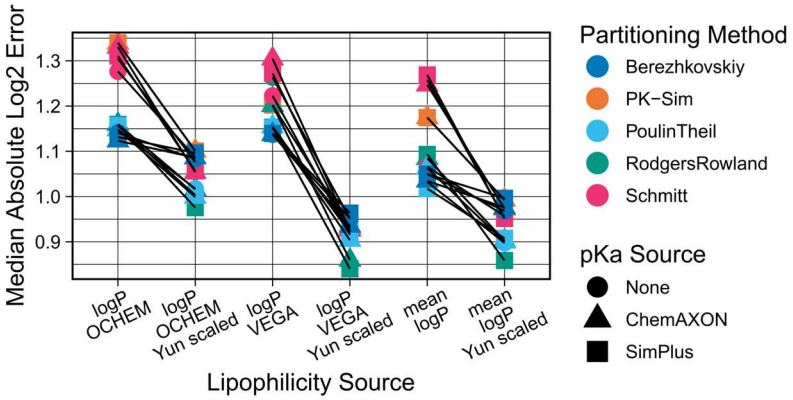
Supplementary Figure 3: Comparison of predictive performances of lipophilicity prediction tools and partitioning methods separated by compound lipophilicity. Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Lipophilicity values were set using the mentioned LogP and LogD *in silico* prediction tools. Plots show Median Absolute Log2 Errors as a measure of prediction precision. A shows results of compounds with LogP < 3 (n = 101), B shows results of compounds with LogP > 3 (n = 42). Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



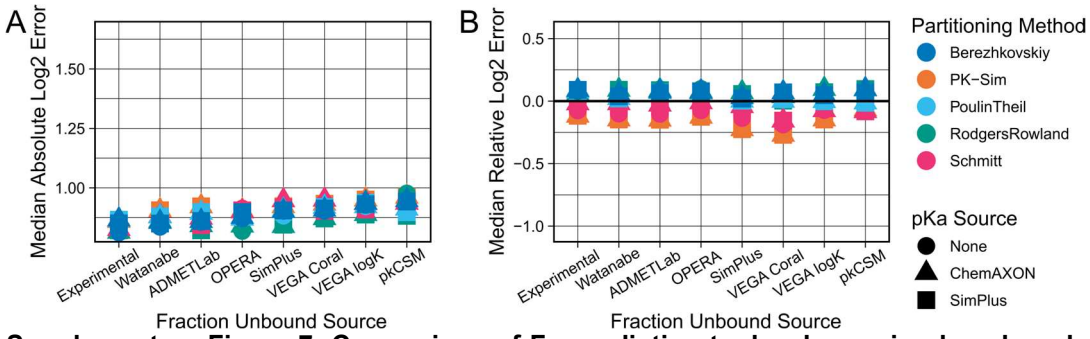
Supplementary Figure 4: Comparison of predictive performances of pKa prediction tools. Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Lipophilicity values were set using the mentioned LogP and LogD *in silico* prediction tools. Plots show Median Absolute Log2 Errors as a measure of prediction precision. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



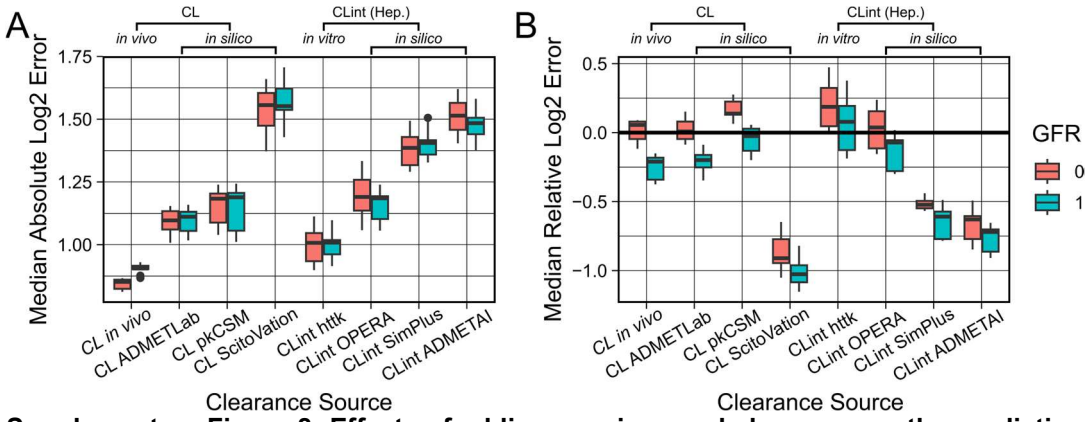
Supplementary Figure 5: Predictive performances of consensus strategies for averaging lipophilicity predictions. Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Lipophilicity values were set using the mentioned LogP and LogD *in silico* prediction tools, as well as the averages of those. Plus signs indicate which predictions of the individual LogP or LogD tools were combined to achieve the mean results marked by equal signs. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



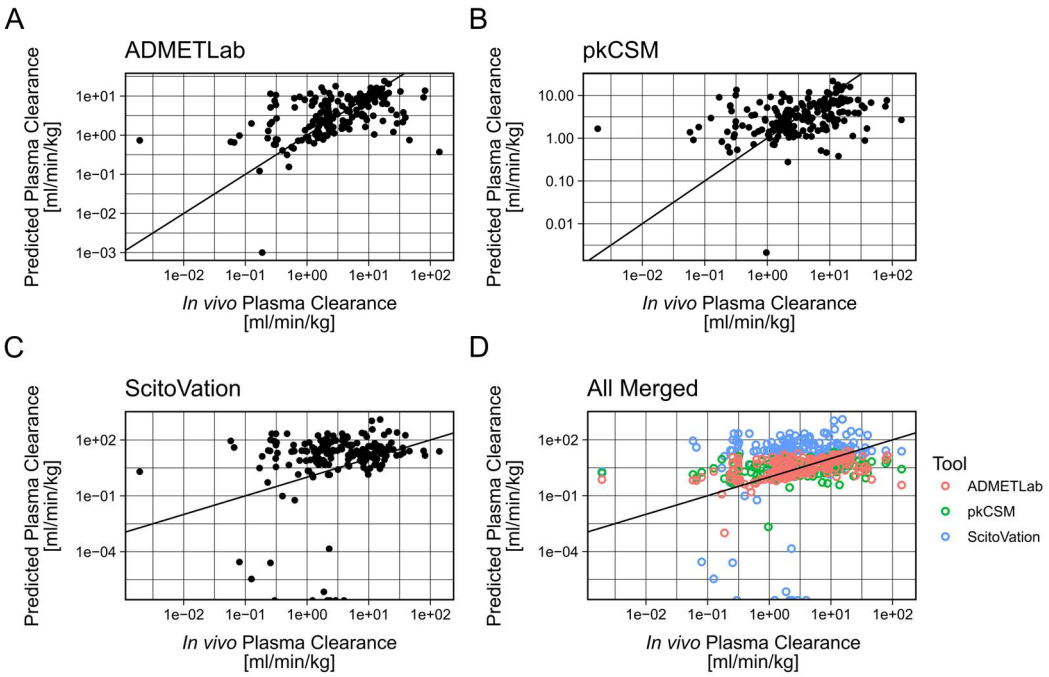
Supplementary Figure 6: Change in predictive performance when using an empirical regression equation to convert LogP lipophilicity to LogMA. Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using *in vivo* and *in vitro* reference benchmark values, respectively. Lipophilicity values were set using the mentioned LogP *in silico* prediction tools VEGA, OCHEM and their average. Additionally, those LogP values were converted to LogMA values by an empirical regression equation established by Yun et al. [117]. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



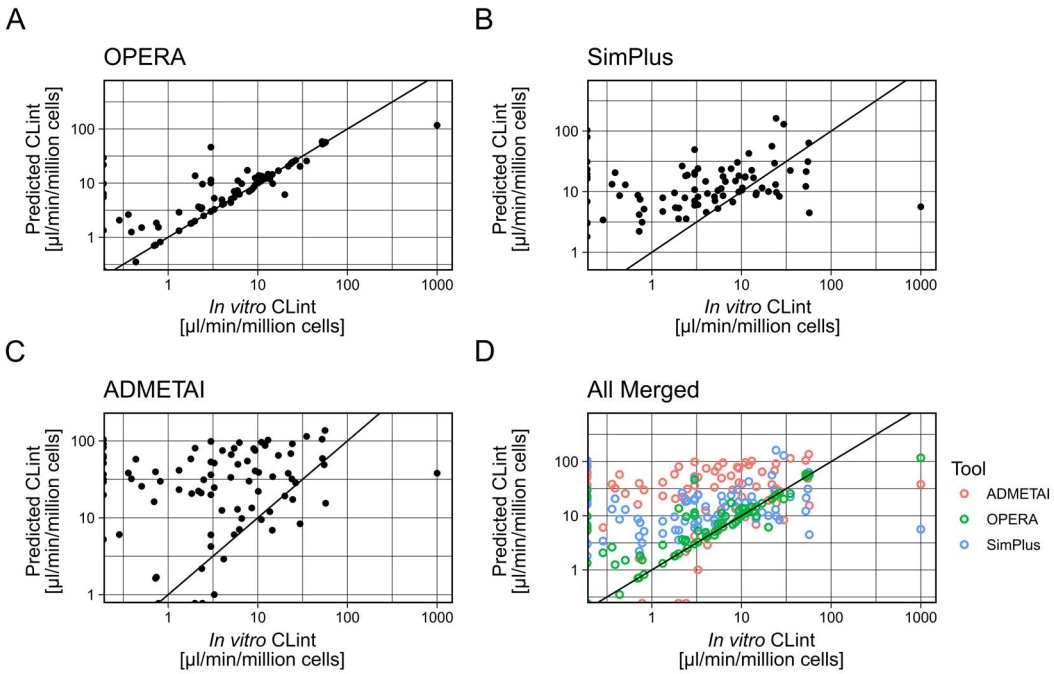
Supplementary Figure 7: Comparison of Fu prediction tools when using benchmark *in vivo* plasma clearance values for PBK model parameterisation. Combinations of all available parameterisation sources were evaluated against the collected IV dataset. Results shown were generated using *in vivo* observed plasma clearance benchmark values, and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. A: Median Absolute Log2 Errors as a measure of prediction precision. B: Median Relative Log2 Errors as a measure of systematic prediction bias. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



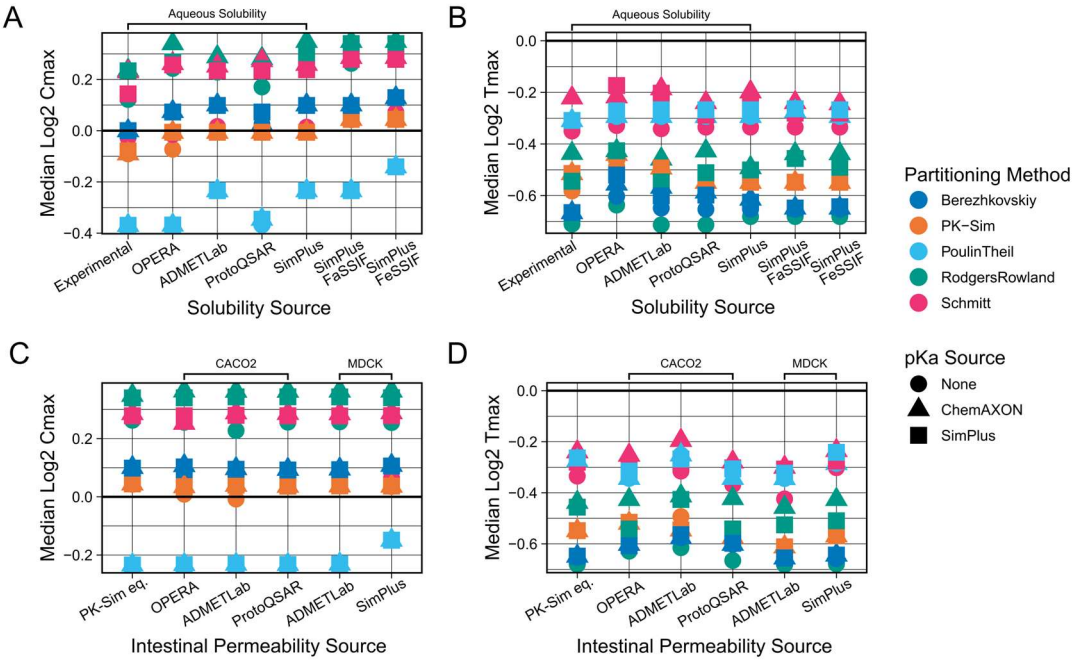
Supplementary Figure 8: Effects of adding passive renal clearance on the predictive performances of HT-PBK predictions. Combinations of all available parameterisation sources were evaluated against the collected IV dataset. Results shown were generated using *in vitro* measured Fu benchmark values, and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. GFR rate of 1 signifies adding passive renal clearance to the PBK model, and a value of 0 indicates leaving it out. CL refers to plasma clearance values (measured or predicted), CLint stands for hepatic intrinsic clearance (measured or predicted). A: Median Absolute Log2 Errors as a measure of prediction precision. B: Median Relative Log2 Errors as a measure of systematic prediction bias. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



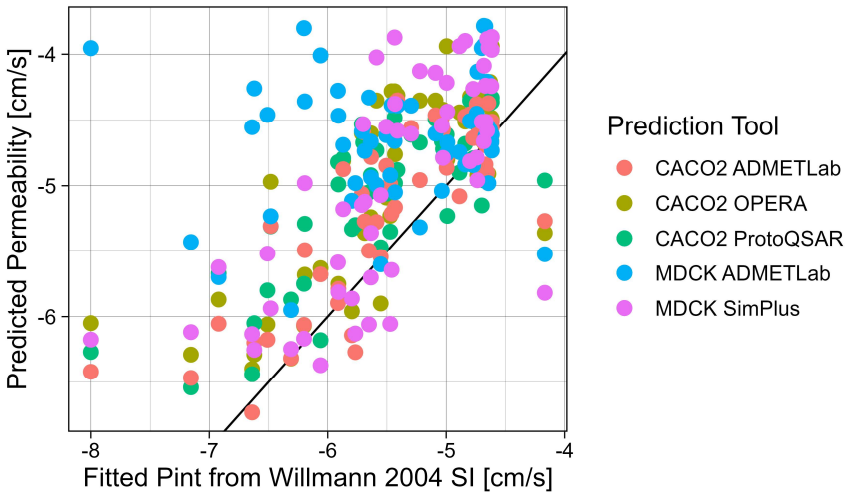
Supplementary Figure 9: *In silico* predicted plasma clearance values from different tools against *in vivo* observed values. *In vivo* values were collected from the literature, *in silico* values were predicted using the corresponding tools. A-C: *In silico* predicted plasma clearance values from ADMETLab, pkCSM, and ScitoVation, respectively. D: All predicted values combined in a single plot. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



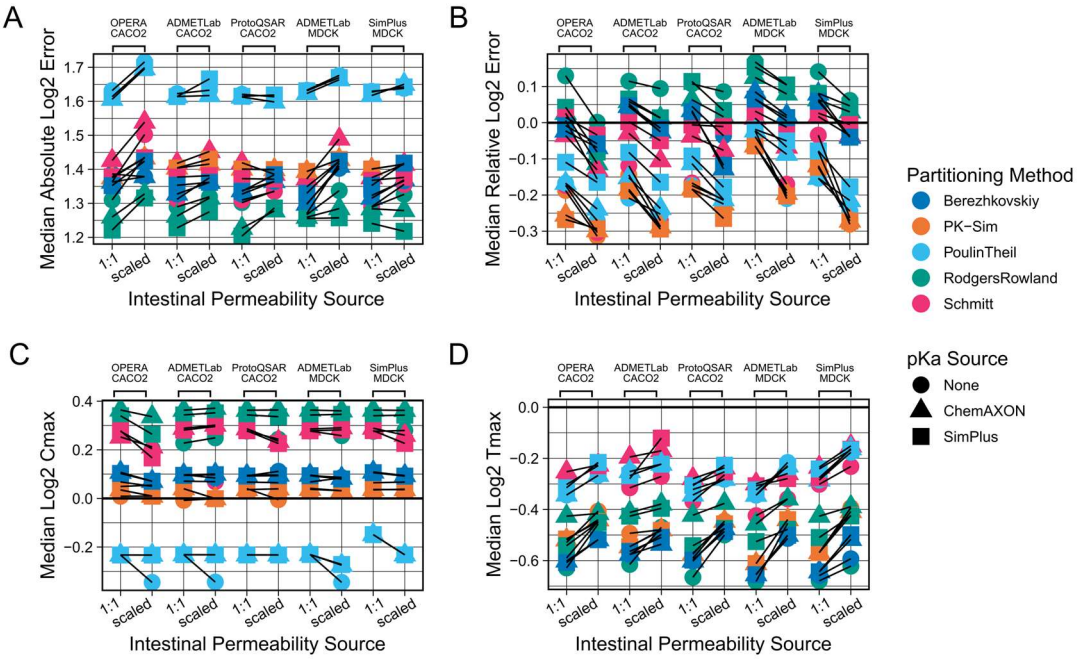
Supplementary Figure 10: *In silico* predicted hepatocyte CLint values from different tools against *in vitro* measured values. *In vitro* CLint values were taken from httk version 2.2.1, *in silico* values were predicted using the corresponding tools. A-C: *In silico* predicted CLint values from OPERA, SimPlus (ADMET Predictor), and ADMET-AI, respectively. D: All predicted values combined in a single plot. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



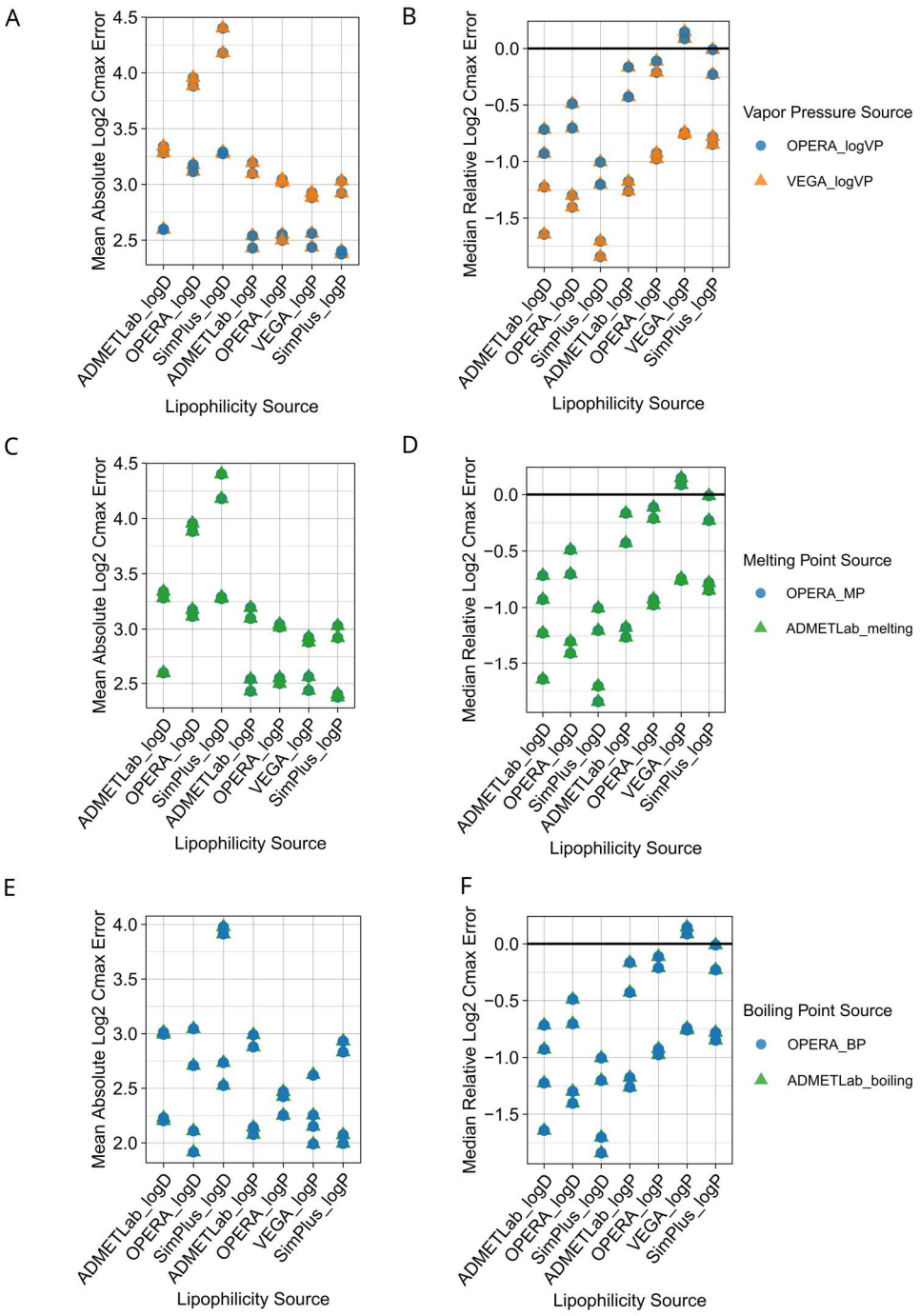
Supplementary Figure 11: Cmax and Tmax prediction biases of different solubility and intestinal permeability prediction sources (step 3). Combinations of all available parameterisation sources were evaluated against the collected PO dataset (dissolved formulations). Results shown were generated using benchmark values for parameterisation of other parameters, i.e., *in vivo* plasma clearances, *in vitro* fraction unbound values and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. The top row shows the Median Log2 Cmax predicted/observed (A) and the Median Log2 Tmax predicted/observed (B) for different solubility prediction methods. The bottom row shows the Median Log2 Cmax predicted/observed (C) and the Median Log2 Tmax predicted/observed (D) for different intestinal permeability prediction methods. For comparison of solubility sources, the intestinal permeability source used was the PK-Sim internal equation (PK-Sim eq.). For comparison of intestinal permeability sources, the solubility values used were the SimPlus FaSSIF values. Results for the PK-Sim internal equation were generated by inputting the Bayer LogMA predictions. Intestinal permeability predictions are either CACO2 or MDCK permeability predictions. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



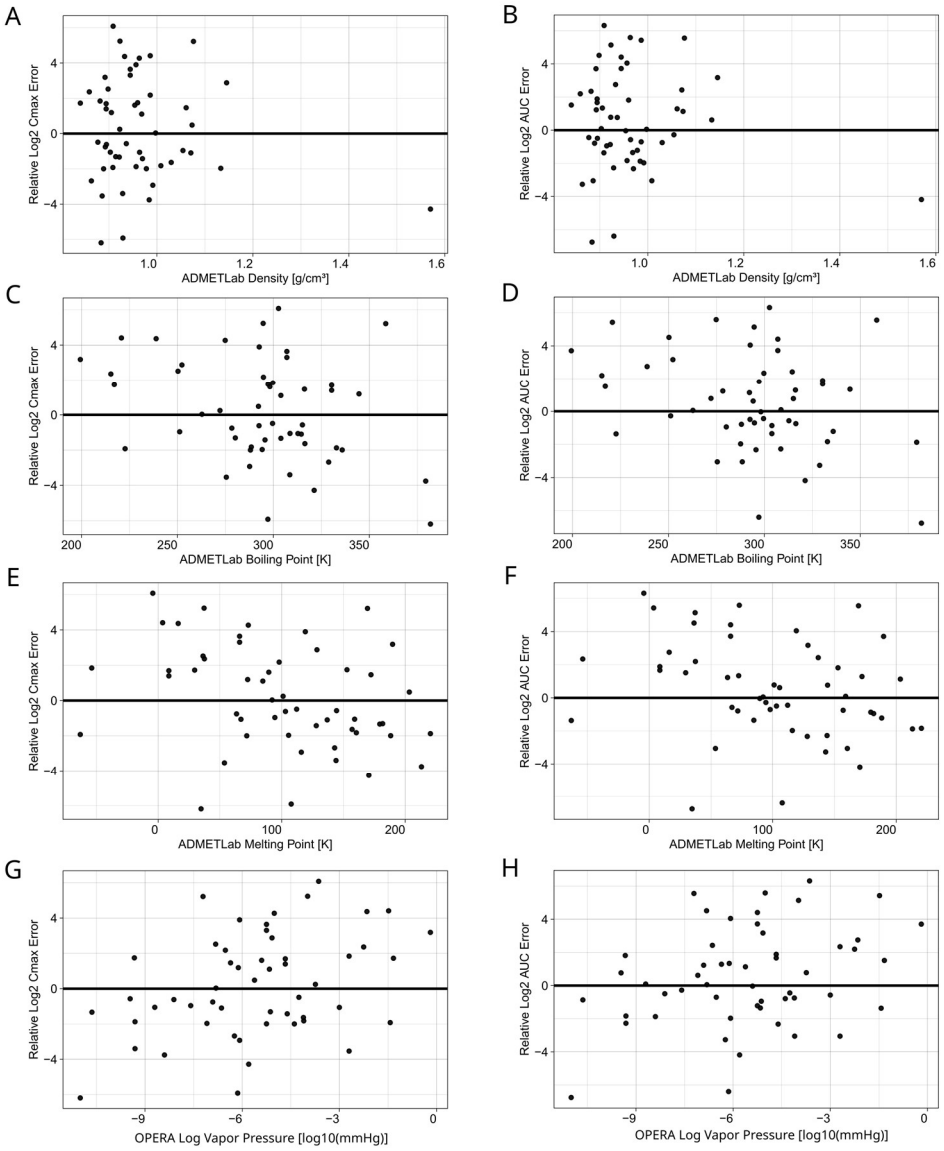
Supplementary Figure 12: *In silico* predicted intestinal permeability values from different tools against optimal fitted Pint values. Optimal fitted Pint values were taken from Willmann et al. [133]. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



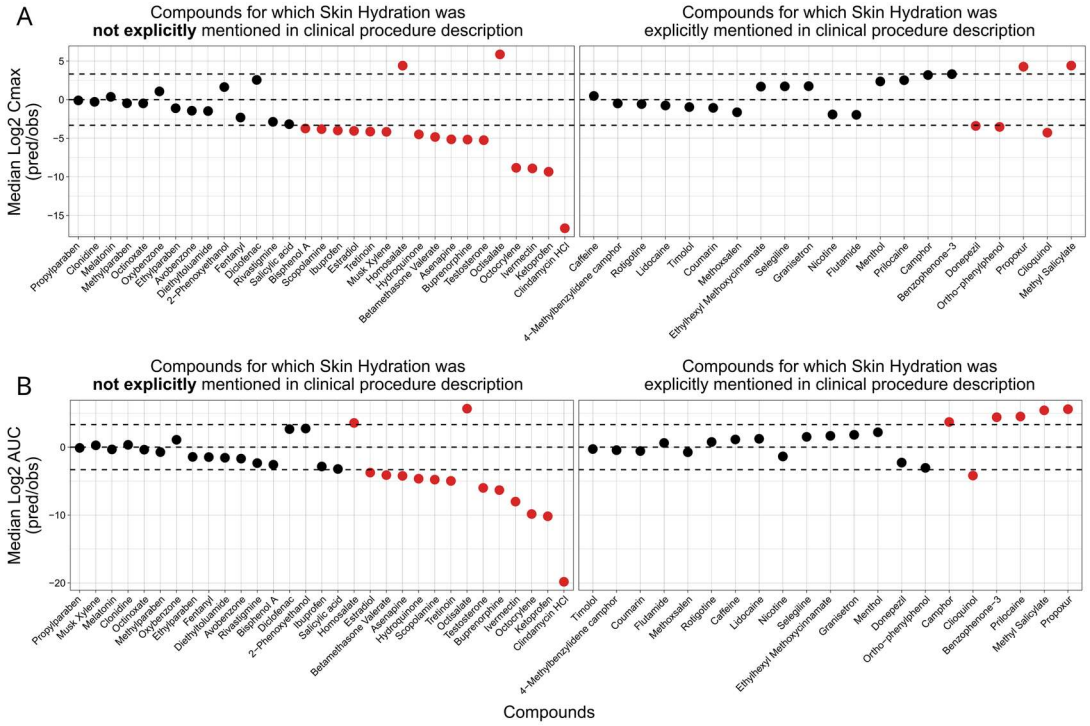
Supplementary Figure 13: Effect of scaling intestinal permeability predictions on different performance metrics. Combinations of all available parameterisation sources were evaluated against the collected PO dataset (dissolved formulations). Results shown were generated using benchmark values for parameterisation of other parameters, i.e., *in vivo* plasma clearances, *in vitro* fraction unbound values and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. SimPlus FaSSIF values were used as solubility values. The top row shows the Median Absolute Log2 Error (A) and the Median Relative Log2 Error (B) for different intestinal permeability prediction approaches. The bottom row shows the Median Log2 Cmax predicted/observed (C) and the Median Log2 Tmax predicted/observed (D). Lines connect the simulation results based on the same *in silico* prediction tool, either used directly (1:1) or after scaling using optimal intestinal permeability values from Willmann et al. [133]. Figure reproduced from [140] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



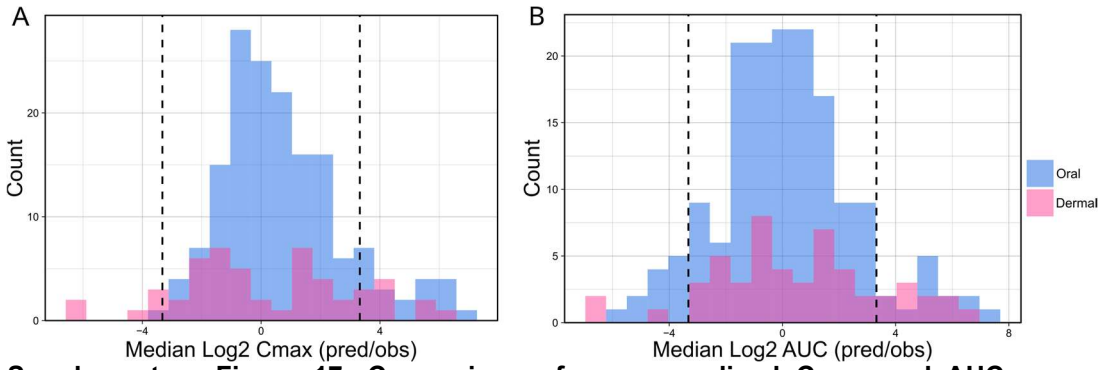
Supplementary Figure 14: Comparison of systemic PK prediction performances for different vapour pressure, melting and boiling point prediction sources. Dermal absorption simulations of all parameterisation sources were examined against the collected *in vivo* dermal PK dataset. Solubility values were taken from SimPlus LogS. The first columns show Mean Absolute Log2 Cmax Errors (A, C, E) for different lipophilicity prediction sources. The second column shows Median Relative Log2 Cmax Errors (B, D, F). VP stands for vapour pressure, MP is melting point, BP is boiling point. Figure reproduced from [209]. No changes made.



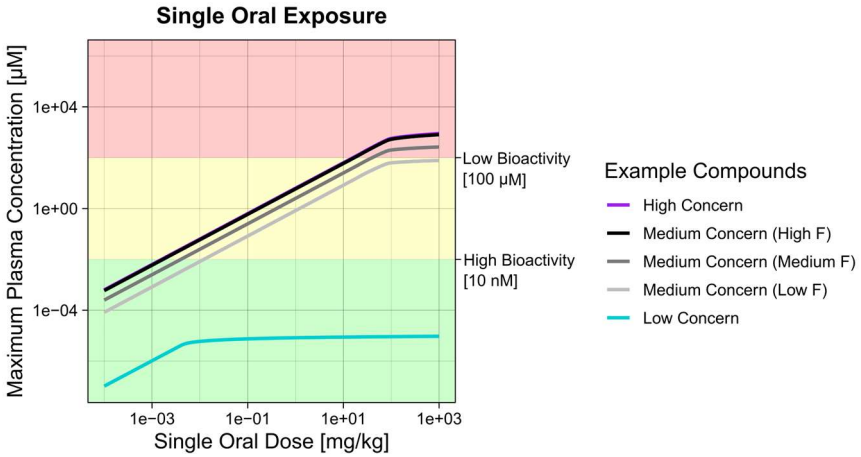
Supplementary Figure 15: Correlation of PK prediction errors of the best dermal HT-PBK strategy with additional compound properties. The predicted compound parameters are compared against the Relative Log2 Cmax and AUC errors. Figure reproduced from [209]. No changes made.



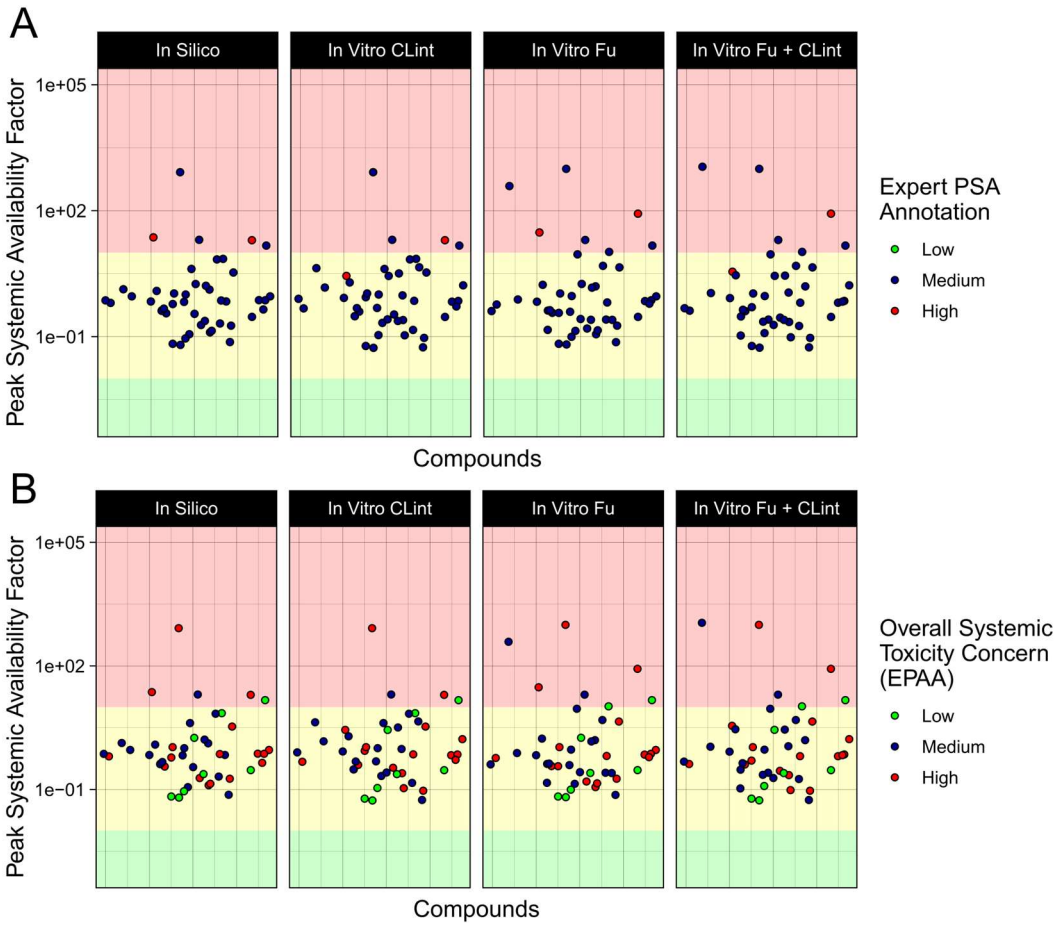
Supplementary Figure 16: The effect of unspecified skin hydration status on systemic PK predictions. Many dermal PK studies did not explicitly mention whether application was performed on hydrated or non-hydrated skin. Shown results are for the assumption that lack of explicit mentioning relates to absence of skin hydration in study procedures (A) The Median Log2 Cmax predicted/observed values of each compound in the dataset. (B) The Median Log2 AUC predicted/observed values of each compound in the dataset. Plots are also split according to which information about the hydration status of the stratum corneum were provided in the clinical study information. Figure reproduced from [209]. No changes made.



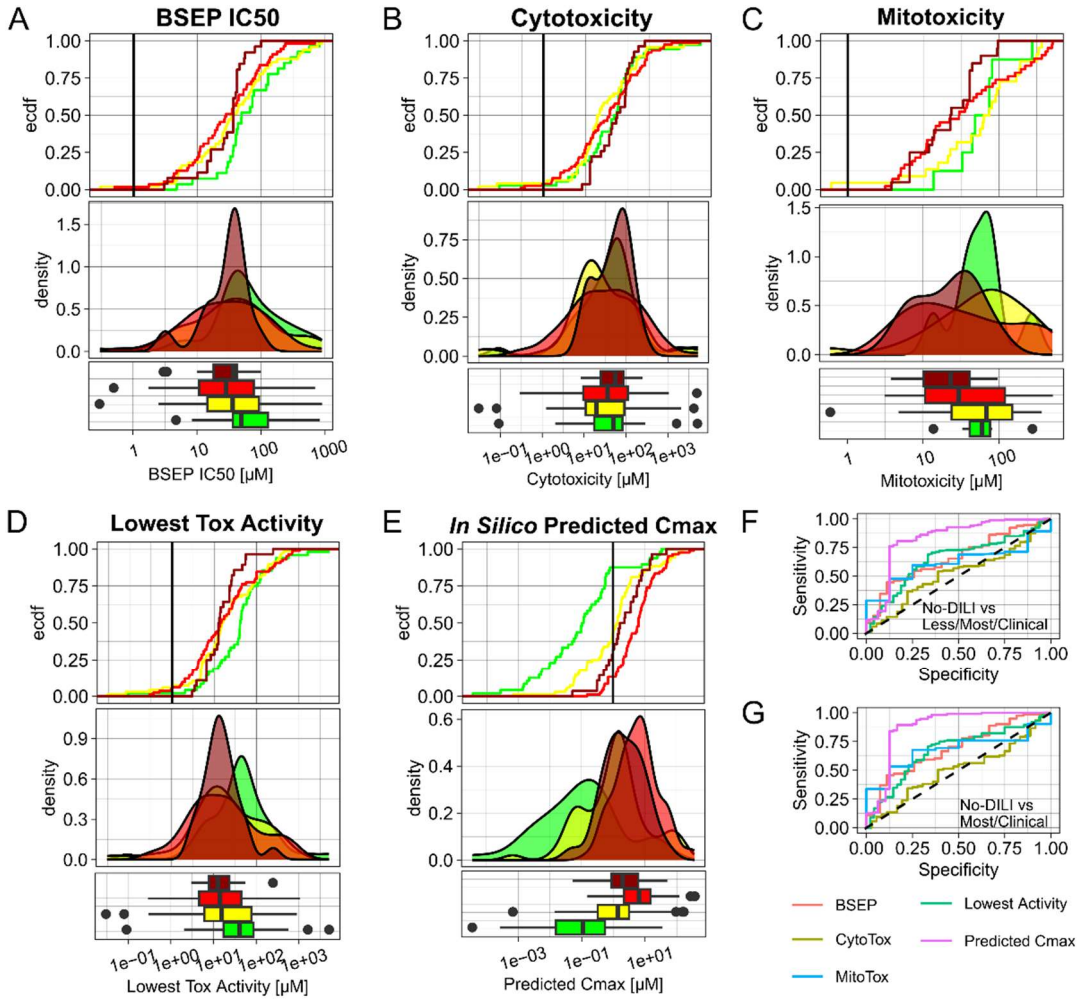
Supplementary Figure 17: Comparison of non-normalised Cmax and AUC error distributions for dermal and oral PK predictions. (A) The Median Log2 predicted/observed Cmax values of dermally administered compounds (pink) were compared against orally administered compounds (blue). (B) The Median Log2 predicted/observed AUC values of dermally administered compounds (pink) were compared against orally administered compounds (blue). Oral HT-PBK prediction errors were taken from Geci et al. 2024 [140]. Figure reproduced from [209]. No changes made.



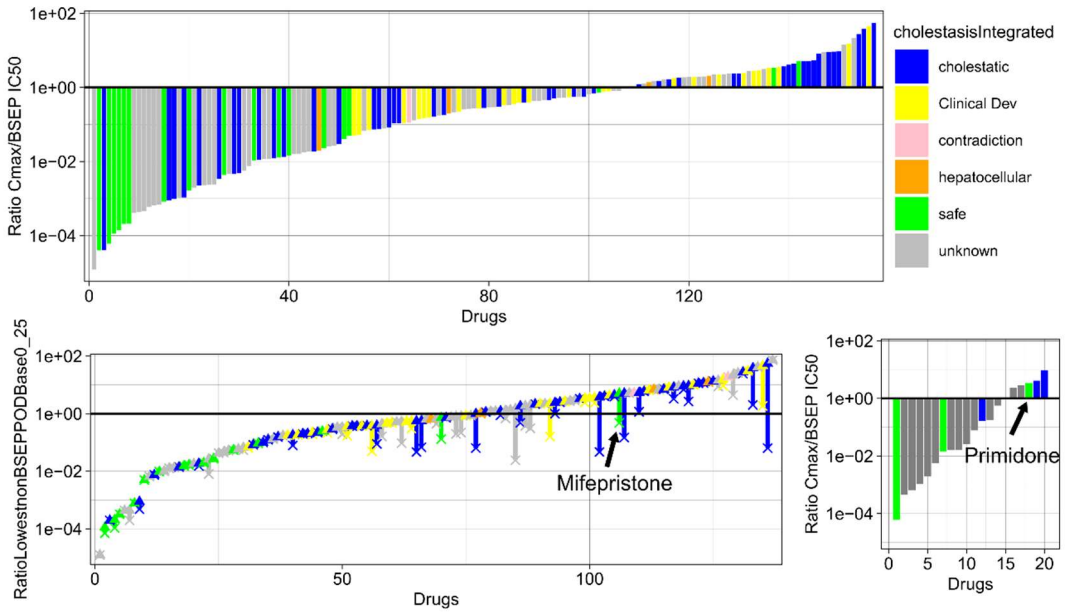
Supplementary Figure 18: (A) Maximum plasma concentration (Cmax) against oral doses after single administration of five example compounds with varying properties as described in Table 2. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



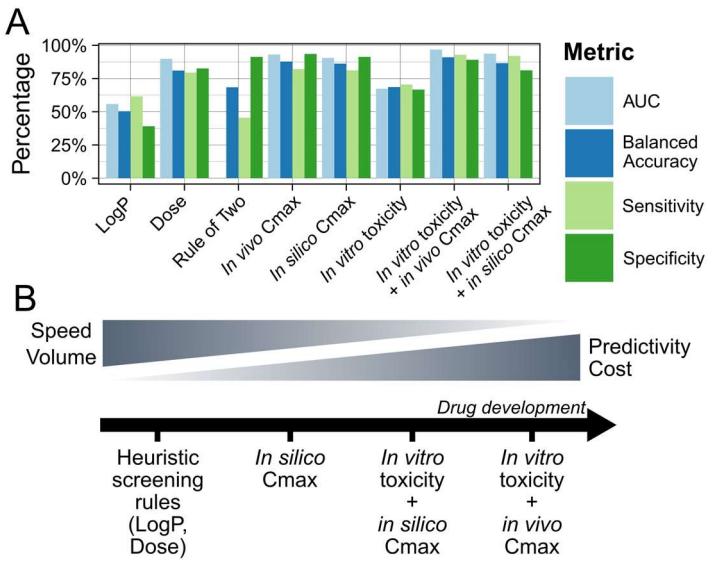
Supplementary Figure 19: Peak Systemic Availability Factors of 45 EPAA Designathon Compounds for which *in vitro* measurements of fraction unbound (Fu) or hepatocyte clearance (CLint) were available from the US EPA's httk package. Results of HT-PBK simulations shown for different parameterisation sources. (A) Coloured by manually annotated reference classification. (B) Coloured by overall systemic toxicity concern given by EPAA. Figure reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.



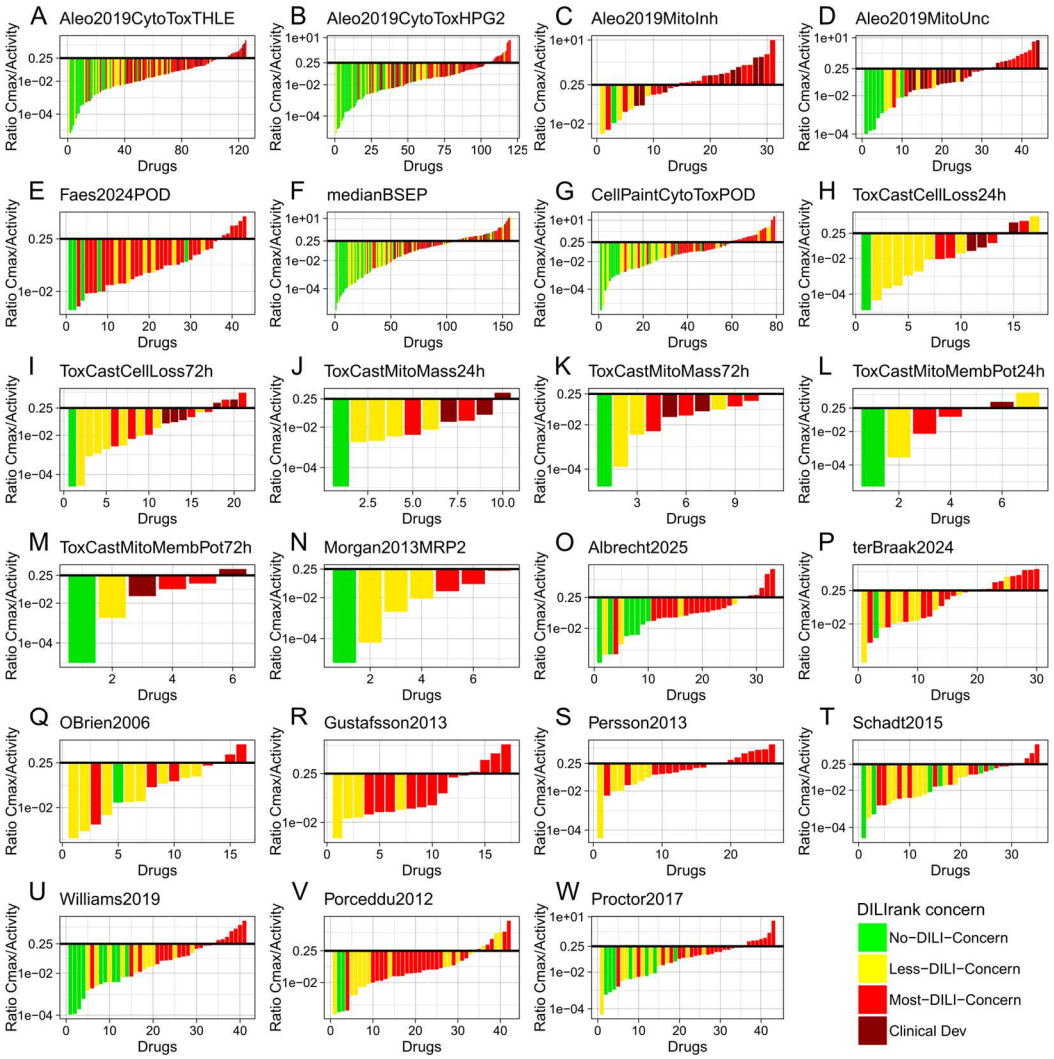
Supplementary Figure 20: DILI predictivity of *in vitro* toxicity and *in silico* predicted Cmax values alone. Cumulative distribution, probability density and boxplots of different properties of drugs of the different DILIrank classes: *In vitro* BSEP IC50 (A), cytotoxicity (B), mitochondrial toxicity (C), the lowest toxicity potency per compound (D) and *in silico* predicted Cmax values (E). ROC curves of the various parameters for distinguishing No-DILI compounds from Less-/Most-DILI/Clinical Development Failures (F) or only from Most-DILI/Clinical Development Failures. (G). Figure reproduced from [272]. No changes made.



Supplementary Figure 21: DILI patterns of drugs with BSEP inhibition data. Ratio of *in vivo* Cmax to BSEP IC50 values of each drug (A). Changes in Cmax to toxicity ratios when including BSEP IC50 values for compounds that have at least one functional *in vitro* toxicity values (B). Ratio of Cmax to BSEP IC50 values of compounds that have no functional *in vitro* toxicity value (C). Figure reproduced from [272]. No changes made.



Supplementary Figure 22: Summary of predictivity evaluations of various DILI prediction strategies. *In vitro* activity refers to the lowest *in vitro* measured toxicity per compound, including cytotoxicity, mitochondrial toxicity and BSEP inhibition (A). Ideal thresholds based on Youden's index are used. Rule of Two does not possess a ROC AUC value, due to absence of drug risk ranking. Schematic representation of the role of different DILI prediction strategies in the drug development timeline (B). Figure reproduced from [272]. No changes made.



Supplementary Figure 23: Individual results of each dataset integrated in this study. Ratio of *in vivo* Cmax to lowest functional *in vitro* toxicity of each drug for each *in vitro* toxicity dataset integrated in this study (A-W). Figure reproduced from [272]. No changes made.

Supplementary Table 1: Comparison of classifications of 139 Designathon Compounds based on HT-PBK predictions using the Mean Systemic Availability Factor (MSAF) expert judged classification of Potential Systemic Availability (PSA). Table reproduced from [230] under CC BY 4.0 (license: https://creativecommons.org/licenses/by/4.0/). No changes made.

	HT-PBK MSAF: Low	HT-PBK MSAF: Medium	HT-PBK MSAF: High	*Sum*
Expert PSA class: Low	2	0	0	2
Expert PSA class: Medium	19	108	7	134
Expert PSA class: High	0	0	3	3
Sum	21	108	10	139

Supplementary Table 2: Predictive performance metrics of various DILI prediction strategies. Calculations were based on distinguishing No- from Most-DILI concern and Clinical Development Failure drugs, and assuming that mifepristone and primidone are DILI-positive, despite being classified No-DILI concern in DILIrank, for the reasons outlined in text. Ideal thresholds are based on Youden's index [276]. Figure reproduced from [272]. No changes made.

DILI Prediction Strategy	**Parameter**	**Classification Threshold Value**	**Sensitivity** (No vs Most/ Clinical Dev)	**Specificity** (No vs Most/ Clinical Dev)	**Balanced Accuracy**	**ROC AUC**
1: Screening heuristics	LogP	3	61.6%	39.1%	50.4%	55.8%
	Dose	100 mg	79.5%	82.6%	81.0%	89.8%
	Rule of Two	LogP >=3, Dose >= 100 mg	45.5%	91.3%	68.4%	-
	In vivo Cmax	1.1 µM	82.1%	93.5%	87.8%	93.1%
	In silico Cmax	1.1 µM	81.2%	91.3%	86.3%	90.5%
	Lowest *In vitro* toxicity (incl. BSEP)	29.44 µM (ideal)	70.5%	66.7%	68.6%	67.34%
2a: *In vitro* hepatotoxicity excl. BSEP (retrospective)	Ratio of *in vivo* Cmax to functional *in vitro* toxicity (excl. BSEP)	0.25 (mechanistic)	50.9%	100%	75.4%	96.5%
		0.011 (ideal)	97.3%	84.2%	90.8%	
2b: *In vitro* hepatotoxicity incl. BSEP (retrospective)	Ratio of *in vivo* Cmax to lowest *in vitro* toxicity (incl. BSEP)	0.25 (mechanistic)	55.4%	100%	77.7%	96.8%
		0.024 (ideal)	92.9%	89.1%	91.0%	
3a: *In vitro* hepatotoxicity excl. BSEP (prospective)	Ratio of *in silico* Cmax to functional *in vitro* toxicity (excl. BSEP)	0.25 (mechanistic)	50.0%	94.7%	72.4%	92.3%
		0.017 (ideal)	91.1%	84.2%	87.6%	
3b: *In vitro* hepatotoxicity incl. BSEP (prospective)	Ratio of *in silico* Cmax to lowest *in vitro* toxicity (incl. BSEP)	0.25 (mechanistic)	55.4%	95.7%	75.5%	93.7%
		0.017 (ideal)	92.0%	81.2%	86.6%	
4: BSEP inhibition alone	Ratio of *in vivo* Cmax to BSEP IC50	0.25 (mechanistic)	42.0%	100%	71.0%	96.4%
		0.012 (ideal)	92.6%	92.0%	92.3%	
5: Bile acid model simulations	Duration of Bile Acid Elevation above twofold	100 min	22.2%	100%	61.1%	61.7%

Curriculum Vitae

Name: René Geci

Date of birth: 16.04.1996

Place of birth: Velbert, Germany

ORCID: 0000-0002-1219-6835

Education

2014	**Abitur: 1.0** Geschwister-Scholl-Gymnasium Velbert, Germany
10/2015 – 09/2018	**Biosciences (B. Sc.): 1.3** University of Heidelberg, Germany
09/2017 – 01/2018	**Exchange year (in BSc Biological Sciences)** University of Leeds, United Kingdom
10/2018 – 10/2021	**Systems Biology (M. Sc.): 1.1** University of Heidelberg, Germany
09/2019 – 06/2020	**Exchange year (in BSc Biochemistry)** Imperial College London, United Kingdom
01/2022 – today	**Doctoral candidate** University Hospital RWTH Aachen, Germany In collaboration with ESQlabs GmbH and within an EU-funded H2020 project (“ONTOX”; grant number 963845)

Professional Experience

10/2018 – 04/2019	**Research Assistant** at Bischofs Lab (BioQuant Heidelberg/MPI Marburg)
07/2020 – 11/2020	**Independent scientific consultant** for Target Malaria
04/2022 – 12/2023	**Research Assistant** at Heidelberg Institute for Theoretical Studies (HITS)

Publications

Geci, R., Willis, K. & Burt, A. *Gene drive designs for efficient and localisable population suppression using Y-linked editors.* PLOS Genetics 18, e1010550; 10.1371/journal.pgen.1010550 (2022).

Geci, R. et al. *Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans.* Arch Toxicol 98, 2659–2676; 10.1007/s00204-024-03764-9 (2024).

Geci, R., Paini, A., Worth, A. P., Kuepfer, L. & Schaller, S. *Potential systemic availability classification of chemicals for safety assessment.* Environment international 202, 109636; 10.1016/j.envint.2025.109636 (2025).

Mathisen, G.H., Svendsen, C., Vist, G.E., ..., Geci, R., et al. *Identification of concepts of importance for the assessment of internal validity of in vitro toxicology studies using a modified Delphi technique.* Evidence-Based Toxicology 3, 2551013; 10.1080/2833373X.2025.2551013 (2025).

Geci, R., Sayin, A. Z., Schaller, S. & Kuepfer, L. *Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury.* Arch Toxicol 100, 2029–2046; 10.1007/s00204-026-04305-2 (2026).

Edizcan, Z., Schaller, S., Kuepfer, L. & Geci, R. *High-throughput PBK modelling for dermal exposure: a pragmatic approach to predict systemic pharmacokinetics.* Arch Toxicol 100, 2421–2432; 10.1007/s00204-026-04357-4 (2026).

Conference Presentations

EUROTOX 2023, Ljubljana, Slovenia: "*Next Generation PBK modeling with PK-Sim for the prediction of chemical distribution in humans*" (Selected Short Talk; 10 min)

ESTIV 2024, Prague, Czech Republic: *"High-throughput PBK modelling providing insights into ADME/TK in systemic toxicity"* (Invited Talk; 20 min)

German Pharm-Tox Summit 2025, Hannover, Germany: *"High-throughput PBK modelling providing insights into IVIVE and quantitative elements relevant within AOPs"* (Invited Talk; 30 min)

PK/PD Expert Meeting 2025, Bad Dürkheim, Germany: *"Evaluating high-throughput PBK modelling strategies for the prediction of human pharmacokinetics"* (Selected Short Talk; 5 min)

OSP Community Conference 2025, Paris, France: "*Integrating In vitro and In silico Approaches Enables Accurate Prediction of Drug-Induced Liver Injury*" (Selected Talk; 30 min)

Awards

5th Aachen Animal Protection Award in the Category "Replacement" (1500€)