Learn from historical animal data – Most frequently observed targets/organs in RDT studies and how can this guide the design of NAM testing batteries?

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Fatemeh Moradiafrapoli

Karshenasi Arshad (Master), Kerman Medical University

2023

Referent:	Prof. Dr. rer. nat. Thomas Scheper Institut für Technische Chemie Gottfried Wilhelm Leibniz Universität Hannover
Korreferent:	Prof. Dr. rer. nat. Bernd Hitzmann Fachgebiet für Prozessanalytik und Getreidewissenschaft Universität Hohenheim
Tag der Promotion:	17.05.2023

A country can't be free unless the women are free

For Women, Life, Freedom

برای زن، زندگی، آزادی

Abstract

Toxicity testing nowadays plays a critical role in decision-making, and in vivo animal studies are still required for hazard assessment. Toxicity assessment of the 21st century, however, aims to replace in vivo studies with new approach methodologies (NAMs) based on human-centric models such as in vitro and in silico approaches. One open question is the scope of the NAM testing battery.

In this study, we like to learn from the existing animal studies on main target organs for the one endpoint of systemic toxicity after repeated dose exposure.

For this purpose, a large inventory of rodent repeated dose toxicity (RDT) studies from high-quality databases was analyzed concerning the most often affected targets/organs in studies with repeated oral exposure.

The project dataset comprised about 7,100 repeated dose toxicity studies with oral and inhalation exposure for approximately 3,000 chemical studies. The data were extracted from several high-quality databases (DBs), namely RepDose, ToxRef (US EPA), and Hess (NEDO). Studies on pharmaceuticals were extracted from eTOX (8000 RDT), which evolved as the largest preclinical toxicity database for drugs and drug candidates, which comprises more than 1,900 different substances.

The analyses distinguish main target organs at the lowest observed adverse effect level (LOEL) from those observed overall.

Detecting the LOEL with high probability is performed with six main TOs, i.e., body weight, liver, clinical chemistry, clinical signs, hematology, and kidney.

In the second part of the analysis, the predictivity of observed target organ/apical findings from short-term was calculated for long-term studies regarding adverse liver outcomes.

The histopathological findings, organ weight changes, and necropsy of the liver in subacute treatment revealed a good predictivity for adverse liver effects in subchronic treatment.

This investigation was a primary step in getting insight into in vivo RDT outcomes.

The coverage and prediction model can be used as a systemic tool to prove and maintain the expert-based validation of human risk assessment.

Key words: NAMs, REACH, 3r principle, in vivo, RDT studies, RepDose, eTOX.

Kurzfassung

Toxizitätstests spielen heutzutage eine entscheidende Rolle bei der Entscheidungsfindung, und In-vivo-Tierversuche sind für die Gefahrenbewertung immer noch erforderlich. Die Toxizitätsbewertung des 21. Jahrhunderts zielt jedoch darauf ab, In-vivo-Studien durch neue Methoden (NAMs) zu ersetzen, die auf humanzentrierten Modellen wie In-vitro- und In-silico-Ansätzen basieren. Eine offene Frage ist der Umfang der NAM-Testbatterie.

In dieser Studie möchten wir aus den vorhandenen Tierstudien zu den wichtigsten Zielorganen für den einen Endpunkt der systemischen Toxizität nach wiederholter Exposition lernen.

Zu diesem Zweck wurde ein großer Bestand an Studien zur Toxizität bei wiederholter Verabreichung an Nagetiere (RDT) aus hochwertigen Datenbanken hinsichtlich der am häufigsten betroffenen Ziele/Organe in Studien mit wiederholter oraler Exposition analysiert.

Der Projektdatensatz umfasste etwa 7.100 Studien zur Toxizität bei wiederholter Verabreichung mit oraler und inhalativer Exposition für etwa 3.000 chemische Studien. Die Daten wurden aus mehreren hochwertigen Datenbanken (DBs) extrahiert, nämlich RepDose, ToxRef (US EPA) und Hess (NEDO). Studien zu Arzneimitteln wurden aus eTOX (8000 RDT) entnommen, der größten präklinischen Toxizitätsdatenbank für Arzneimittel und Arzneimittelkandidaten, die mehr als 1900 verschiedene Substanzen umfasst.

Die Analysen unterscheiden die Hauptzielorgane auf der niedrigsten beobachteten schädlichen Wirkung (LOEL) von den insgesamt beobachteten.

Die Erkennung des LOEL mit hoher Wahrscheinlichkeit wird mit sechs Hauptzielorganen durchgeführt, nämlich Körpergewicht, Leber, klinische Chemie, klinische Zeichen, Hämatologie und Niere.

Im zweiten Teil der Analyse wurde die Vorhersagbarkeit von beobachteten Zielorganen/apikalen Befunden aus Kurzzeitstudien für Langzeitstudien in Bezug auf unerwünschte Leberresultate berechnet.

Die histopathologischen Befunde, die Gewichtsveränderungen der Organe und die Nekropsie der Leber bei subakuter Behandlung ergaben eine gute Vorhersagbarkeit für schädliche Leberwirkungen bei subchronischer Behandlung.

Diese Untersuchung war ein erster Schritt, um einen Einblick in die In-vivo-Ergebnisse der FTE zu erhalten.

Das Erfassungs- und Vorhersagemodell kann als systemisches Instrument zum Nachweis und zur Aufrechterhaltung der Experten basierten Validierung der Risikobewertung beim Menschen verwendet werden.

Schlüsselwörter: NAMs, REACH, 3r-Prinzip, in vivo, RDT-Studien, RepDose, eTOX.

Table of Contents

Abbreviation

ADL	Acceptable Daily Intakes
AF	Assessment Factor
ALT, ALAT	Alanine Transaminase
AOP	Adverse Outcome Pathways
ASP	Aspartate Aminotransferase
AST	Aspartate Aminotransferase
BMDL	Benchmark Dose Level
BWL	Body Weight Loss
ChE	Cholinesterase
CPDB	Carcinogenic Potency Database
CSA	Chemical Safety Assessment
CSR	Chemical Safety Report
DB	Database
DILI	Drug Induced Liver Injury
DNEL	Derived No-Effect Level
eDSD	extended Safety Data Sheet
EPA	United States Environmental Protection Agency
EU CLP	European Union system of Classification, Labelling and Packaging of chemicals
ES	Exposure Scenario
FN	False Negative
FP	False Positive
FPR	False Positive Rate
IATA	Integrated Approaches for Testing and Assessment
IDE	integrated development environment
LOEL	Lowest Observed Effect Level
LR-	Negative Likelihood Ratio
LR+	Positive Likelihood Ratio

MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
NAM	New Approach Methodology
NOEL	No Observed Effect Level
NTP	National Toxicology Program
OECD	Economic Co-operation and Development
PBT	Persistent, Bioaccumulative and Toxic
PCA	Principal Component Analysis
(Q)SAR	(Quantitative) Structure-Activity Relationships
RBC	Red Blood Cell
RDT	Repeated Dose Toxicity
RDW	red cell distribution width
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
ROC	Receiver Operating Characteristic
Sen	Sensitivity
Spe	Specificity
TN	True Negative
TO	Target Organ
TP	True Positive
TPR	True Positive Rate
UF	Uncertainty Factor
vPvB	Very Persistent and Very Bioaccumulative
WBC	White Blood Cell

1. Introduction and Goal

Human risk assessment is moving away from traditional hazard assessment based on in vivo animal data toward mechanistic approaches, which apply human in vitro and in silico models. These alternative methods (NAMs) align with the 3Rs principle, which aims to replace, reduce, and refine animal testing as far as possible [36]. New challenges and uncertainties accompany the testing and integration of NAMs. A central question is the scope of testing and, thus, the design of the test battery, in particular, which types of in vitro and in silico models and which data (e.g., functional readout data, omic data etc.) must be considered.

The information on main TOs in preclinical repeated dose toxicity studies could help to design a test battery by including relevant in vitro and ex vivo models. However, the information from anchoring in vivo data is not always available, indicating that alternative approaches are needed to evaluate the safety of the relatively large number of chemicals in commerce and the environment. This data-poor situation is termed "ab initio" risk assessment, for which comprehensive testing is recommended to generate hypotheses on potential concerns at the beginning of the assessment process.

This thesis addresses the question of the extent to which one can derive the investigational scope of the testing strategy from the most frequently observed adverse findings/target organs in *in vivo* studies. For this purpose, preclinical *in vivo* animal studies with repeated exposure were evaluated because human *in vivo* data are rarely available.

In human risk assessment finding a compound-specific threshold is a critical step in the decision-making process. The lowest observed effect level (LOEL) is used as a point of departure for hazard assessment and defines the lowest dose that causes any adverse toxicological effect. Therefore, effects observed at the LOEL are of higher relevance for the assessment of the hazard Compared to higher dose effects.

High-quality studies of chemicals and drugs from different databases (DBs) were compared to clarify the most frequently observe target organs at the LOEL in repeated dose toxicity (RDT) studies with oral exposure.

RDT studies on chemicals were extracted from RepDose® [40], ToxRef (US EPA), and Hess (NEDO), and drug RDT studies were extracted from eTOX (http://etoxsys.com).

During this investigation, the following key objectives were the prime focus of the project. These objectives can be summarized as follows:

- How to illustrate the complementary structural spaces between drugs and organic chemicals?
- Are all TOs equally crucial in setting the LOEL in the chemical dataset? If not, which target organ is the most sensitive among all others?
- Which T/Os can detect the LOEL with high probability?
- In the interest of reducing de novo animal testing, how reasonable is that to apply the subacute treatment (4 weeks of oral exposure) to predict adverse

effects in subchronic treatment (3 months of oral exposure) in similar hazard properties (similar chemical substance, species, and route of administration)

- Do liver effects in subacute studies predict liver-effect in subchronic studies?
- Which effects are of particular importance for the liver to liver prediction?

2. Theoretical Background

Toxicology dates back to using animal venom and plant extract for warfare and hunting by the earliest human. They were well aware of the toxic effects of many substances, which led to dividing the plant and animals as safe and harmful, venoms of snakes, poisonous plants, and toxic mineral substances, such as arsenic, lead, and antimony [1]. Some of these were used purposely for their toxic effects on committing homicide and suicide. There were always continual efforts to discover and develop preventive and antidotal measures as long as using the toxins. Maimonides (1135–1204) began critically evaluating these measures with his famous *Poisons and Their Antidotes*, published in 1198 [1, 2]. The transition of toxicology began in the sixteenth century and later. Paracelsus, the philosopher of the German Renaissance, stated: "All substances are poisons, and it is the dose (the amount of the exposure) that differentiates a poison and a remedy" [3, 4]. These statements caused the foundation of the concept of the "dose-response relation" and the "therapeutic index" developed later [1].

2.1. Regulatory toxicology

Regulatory Toxicology contains collecting, processing, and evaluating epidemiological and experimental toxicology data to control the production, use, and deposition of hazardous substances to prevent adverse human and environmental health outcomes. Evaluating risks for protecting the populations against probable risks of chemicals, biocides, food additives, cosmetics, pharmaceuticals, medicinal and manufactured products is essential to regulatory toxicology.

Regulatory toxicology requires understanding the relevant regulations and basic principles of toxicology. In addition, it aids the development of standard regulations and new testing strategies to constantly enhance the scientific basis for decision-making procedures. This requires reliable information on the hazardous feature of a chemical substances and their relation to human and environmental exposure. That information is required for a proper risk assessment and decision-making when the regulatory outcomes are evident [5].

Chemical risk assessment contains three steps, hazard identification, evaluation of dose-response relationship, and exposure assessment. In risk assessment, it is necessary to differentiate the reversible and irreversible outcomes that identify a “no observed adverse effect level” (NOAEL) derived from animal experiments using various doses. The precision of NOAEL is related to the count of animals at each dose group and the difference to the LOAEL (Lowest Observed Adverse Effect Level) for the main effects of the substance. In irreversible cases, the solution is to assess the risk at a specific exposure [5-7]. To achieve this objective, regulatory agencies typically use toxicological datasets from industry or experimental findings relevant to evaluating modes of action derived from scientific journals of university institutes [8].

In 1999, the European Union published an investigation of the existing chemical legislation. The result determined several issues and challenges of assigned resources by applying various strategies for new and existing substances. After an intense discussion with all stakeholders, the “REACH” regulation was generated in 2006. The “REACH” is the abbreviation formed from the initial letters of Registration, Evaluation, Authorization, and Restriction of Chemicals [9].

The European Commission attempts to protect human health and the environment from potentially hazardous or toxic chemical substances through the REACH program. The REACH regulation applies to substances, mixtures, and papers and is directly applicable in all member region of the European Union [10].

European Chemicals Agency (ECHA) in Helsinki, Finland, is an institution founded to drive the technological, scientific, and organizational features of REACH at the community level. ECHA supports industries and authorities to complete their commitment and obligations under REACH. The ECHA technical guidance documents were provided by experts from industries, member states, and non-governmental organizations [11].

These documents aim to simplify REACH performance and summarize typically acknowledged good practices. All gathered information should be quality checked. The information must be evaluated for reliability, relevance, and adequacy. Reliability could be defined as the intrinsic quality of the data concerning the standardized method and the experimental design [9].

The essential factors in determining the chemicals toxicity are considered as below:

- The route of exposure (oral, dermal, inhalation),
- The substance dose (amount of the chemical),
- Frequency of exposure (single versus multiple exposures),
- Period of exposure,
- Biological properties (age, gender),
- ADME features (absorption, distribution, metabolism, and excretion/ elimination)
- Structural features.

Animal models have been performed for a long duration for toxicity testing. In the next stage, in-vitro models evolved due to the refinements and high throughput screening.

The reliability category known as Klimisch code system, classifies the data according to quality criteria. The system contains four category of data [9, 12]:

1. Reliable without limitations

2. Reliable with limitations

3. Not reliable

4. Not assignable

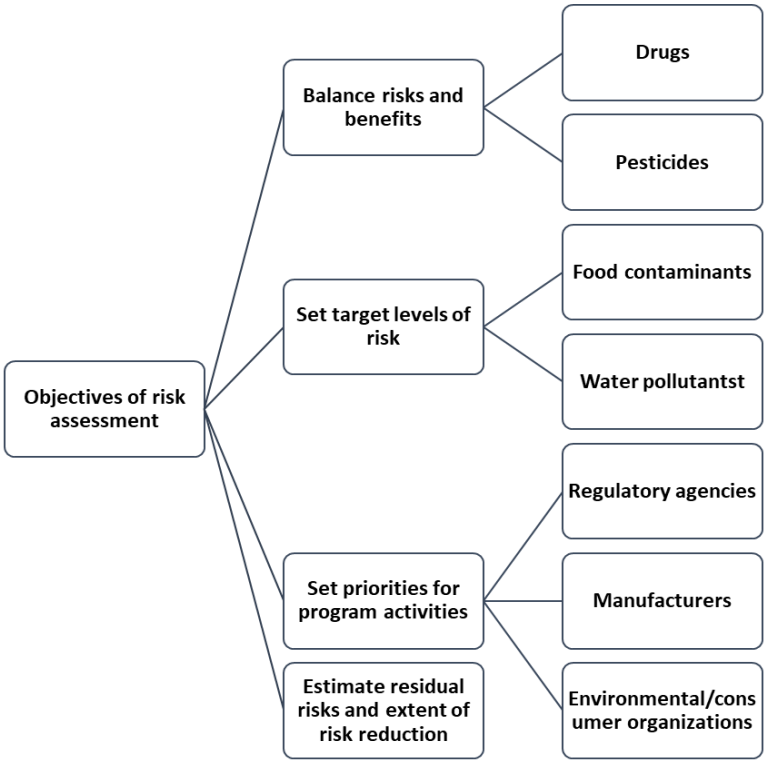

Figure 1 The objectives of risk assessment - adopted from [1].

Generally, the category 1 and 2 are appropriate for the hazard assessment of an endpoint. Such studies meet in general terms the criteria outlined in the OECD guidelines. Klimisch code 3 and 4 studies are considered to provide supporting information but differ significantly from OECD guideline studies with regard to e.g. scope of examination, number of tested animals and other major contsraints, that impact the reliability of the study outcome. However, studies with a Klimisch code 3 or 4 are usually included in the registration dossier of REACH for completeness and clarity [9, 13].

Relevance is the appropriate scope of data and tests for a specific hazard assessment. Adequacy, which combines reliability and relevance, is the utility of the data for hazard and risk evaluation objectives. The most relevant data for hazard and risk assessment is reliable human data, and human data is generally scarce.

REACH has remarkably advanced in the last decade in developing and implementing in vitro methods [9, 14]. Hazard classification in investigating local skin and eye effects without any in vivo tests is a good example. Until 2009 animal-free assays were available only for skin corrosion. After this time, this test method changed by using a reconstructed human epidermis to investigate skin irritation [15].

Physico-chemical properties of chemical substances are typically used to initialize a specific hazard potential. Physico-chemical properties such as water solubility, acidity, alkalinity, hydrophilicity/lipophilicity, and volatility can facilitate understanding whether a substance is bioavailable or be locally toxic. Those data can likewise support a read-across analysis [9].

Chemical Safety Assessment (CSA) classifies the registered substance as dangerous or persistent, bio accumulative or toxic (PBT), or very persistent and very bio-accumulative (vPvB). Chemical safety assessment also characterizes the risk. In chemical safety, risk characterization and exposure assessment must be targeted at the specific hazard harmful to human safety, physicochemical, environmental health, or PBT or vPvB [9, 16].

Chemical safety assessment is necessary for all registered substances under REACH in quantities of 10 t or more per year to confirm that all risks are determined and managed by finding the association between exposure and hazard threshold levels (Figure 3) [17].

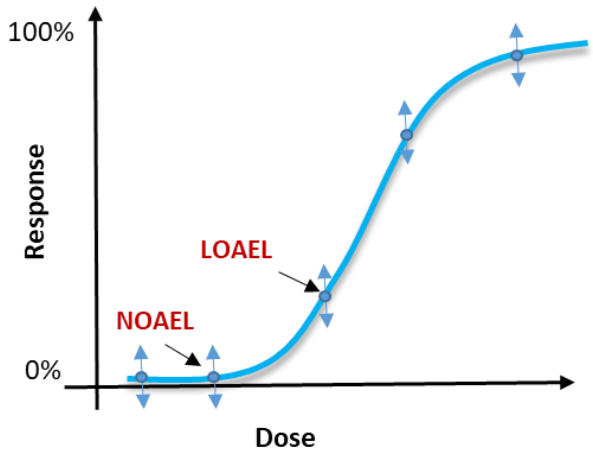

Figure 2 A dose-response curve shows where NOAEL and LOEL occur for a substance.

Exposure levels must be compared to threshold doses with no observed adverse effects, known as derived-no-effect-levels (DNELs). DNEL measurement must consider several factors and conditions. The manufacturing requirement or use must be managed, and depending on the target (workers, consumers), different aspects must be involved. Specific target organ toxicity data after acute and repeated exposure need to be assessed with particular concern for the different features of outcomes (systemic versus local). The initial step in evaluating a DNEL is finding the starting point-of-depart (POD) value. The lowest no observed adverse effect level (NOAEL), which must be reliable, relevant, and adequate, is often used as the POD. At this dose or concentration, there are no adverse treatment-related outcomes. The NOAEL can be used as the initial point for the risk evaluation. It is also required to modify the initial dose descriptors, which can facilitate the specific needs of the risk characterization. Risk characterization is the last step of the chemical safety assessment. Risk characterization compares exposure levels to threshold doses or concentrations with no expected adverse effects, indicating derived-no-effect-levels (DNELs) Figure 3.

The assessment factor (AF), named modifying or uncertainty factor by other regulatory authorities, is applied to the POD. Such factors deal with exposure duration differences, interspecies and intraspecies uncertainty and variability, dose-response relations, and the comprehensive quality of the dataset [9, 11].

Uncertainty factors (UFs) are utilized for evaluating risk from chemical exposure or the acceptable daily intake of chemicals. An uncertainty factor model is the most uncomplicated model for inter-species extrapolation, intra-species extrapolation, or exposure duration extrapolation.

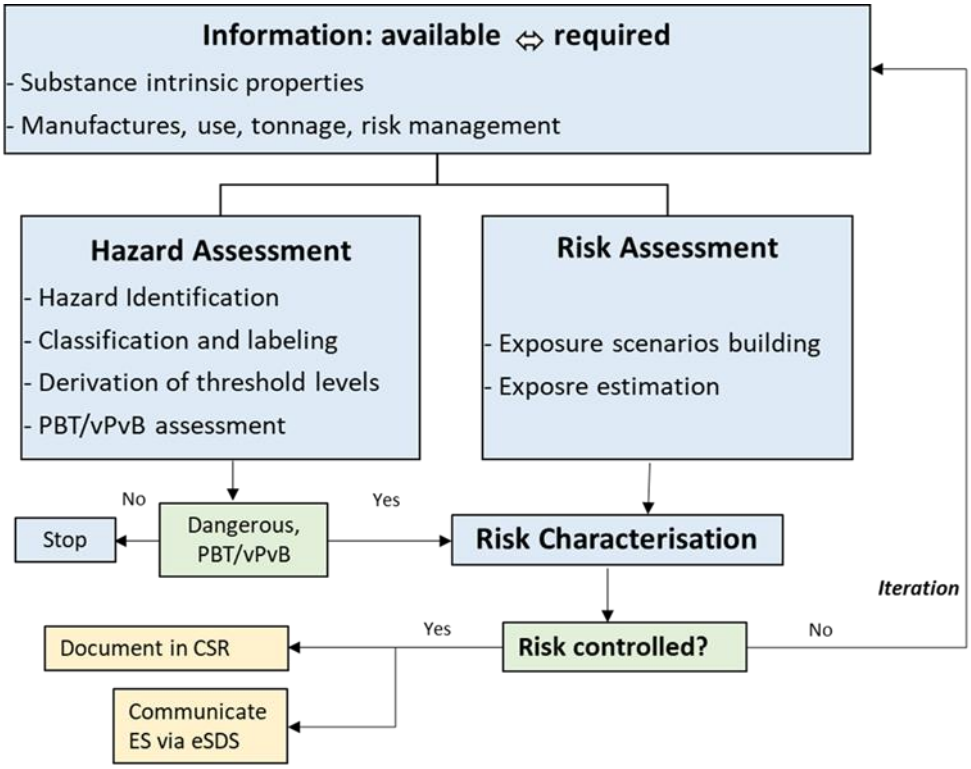

Figure 3 Flow scheme for the chemical safety assessment - adopted from [9, 12].
CSR: chemical safety report, ES: exposure scenario, eSDS: extended safety data sheet

Inter-species extrapolation is performed in converting the animal data to humans. Intra-species extrapolation would be conducted to convert healthy people to a particular population class, such as old people, pregnant women, children, and fetuses. Exposure duration extrapolation is used in short-term exposure to long-duration exposure alteration

In uncertainty factor models, NOAEL and UF are two main factors. NOAEL is the highest dose, not indicating detectable toxicity, and UF is a numerical value to calculate the variation in inter-species, intra-species, exposure period, or contact dose. The dividing of the NOAEL by UF will result in the extrapolation.

Using the NOAEL may lead to some limitations. First, in the description of the NOAEL, the focus is on the absence of the observable risk of toxicity, but it does not mean a

zero-effect threshold. The second point is that the NOAEL value can change by adjusting the study design, such as the number of animals, the count of doses, and the endpoints. For example, the low statistical capability, such as a few counts of tested animals or a few numbers of tested doses, would result in higher NOAEL. With these mentioned deficiencies, it is possible to use the lowest observed adverse effect level (LOEL) or the benchmark dose level (BMDL) instead of the NOAEL.

$$Risk = Exposure > DNEL$$

$$DNEL = \frac{NOAEL}{appropriate\ EFs}$$

$$DNEL = \frac{NOAEL}{inter\ x\ intraspecies}$$

Figure 4 Derived No-Effect Levels (DNELs), the exposure levels above which humans should not be exposed.

In addition to risk assessment, the classification of hazards is an essential step of the approach.

The EU CLP Guidance (European Union system of Classification, Labelling and Packaging of chemical substances and mixtures) [15] explained that hazard classification means identifying all aspects of a hazard, physical, human health, and environmental hazard of a chemical substance or a compound. The hazard classification and identification process will continue by comparing the hazard and its degree with existing criteria to classify an individual substance or a combination [18].

In addition, to hazard identification and classification in chemical safety assessment, the knowledge of a set of safe limit doses, such as derived no-effect levels (DNELs), acceptable daily intakes (ADIs), and acceptable operator exposure levels (AOELs), which estimate the safe dose for human exposure in different strategies based on duration and route of exposure is required [19].

Design and the strategy of repeated-dose toxicity (RDT) animal studies are essential in preclinical regulatory contexts. The most important outcome of RDT studies is defining the lowest observed adverse effect level (LOEL) and the no observed adverse effect level (NOAEL) as toxicity reference values. The NOAL value is applied in calculating the highest safe dose of toxicological and clinical studies. Nevertheless, NOAEL and LOEL modeling have restrictions since they refer to several different endpoints, and the doses used in RDT studies significantly impact the concluded results [7].

The Organization for Economic Co-operation and Development (OECD), which is representative of 34 countries in America, Europe, and the Asia and Pacific region, is responsible for coordinating and harmonizing policies concerning common issues and working together to answer global concerns.

OECD partners have been trying to develop using alternative methods for chemical assessment. The OECD has been evolving guidance documents and tools for alternative methods, such as (Quantitative) Structure-Activity Relationships ((Q)SAR), chemical classifications, and Adverse Outcome Pathways (AOPs), as features of Integrated Approaches for Testing and Assessment (IATA). It is required to investigate the relevance and applicability of these methods/tools for various aspects of regulatory decision-making and use case studies and assessment knowledge through authorities. The OECD follows the purpose of developing a common insight into innovative methodologies and generating guidance and directions from these experiments [20].

An AOP is set of analytical structure representing a consecutive series of causally related events at various groups of biological organizations that result an adverse human or environmental health outcome (Figure 5). AOPs are the prominent factor of a toxicological framework generated to support chemical risk assessment relying on mechanistic basis [21].

IATA or INTEGRATED APPROACHES TO TESTING AND ASSESSMENT is a practical, science-based strategy for chemical hazard description that depends on an integrated analysis of existing information associated with new data using testing processes. IATA tries to answer specific regulatory questions following an iterative approach. Considering the proper level of uncertainty coincides with the decision context. IATA can contain a combination of methods and can be informed by integrating results from one or many methodological approaches [(Q)SAR, read-across, in-chemico, in-vitro, ex-vivo, in-vivo] or omic based technologies (e.g., toxicogenomics).

IATA is mainly the case with non-animal testing approaches and comprehension of the connection between the tested and the predicted apical toxicity endpoint. It can answer why results from innovative approaches are not yet extensively utilized for the regulatory decision-making process. For this aim, an objective and systematic framework is required to describe new methods, biological relationships and toxicological relevance in predicting an adverse effect. The same framework can indicate the potential use of IATA in combination with other tools and methods to profit from an integrated approach. The adverse outcome pathway (AOP) concept is an example of a framework for IATA development (Figure 5).

2.2. In-silico toxicology

Evaluation of existing chemicals is always a desired field in relevant sciences. The existing database is still insufficient for many compounds, and many tests must be repeated or performed from the beginning. This aim will take an extended period, even just for compounds' most basic toxicological testing.

Reducing test repetition and minimizing the study time requires an alternative to prioritize investigation requirements. Using the structure-activity relationships (SARs) or quantitative structure-activity relationships (QSARs) is a scientific alternative applied in the risk assessment of chemical safety [22-26].

In toxicology, the available computer-aided commercial prediction procedures mostly use different (Q)SAR elements, e.g., TOPKAT, CASE/Multi-CASE, DEREK, ONCOLOGIC, etc.

However, for their direct explanation, essential evaluation, and samples of worldwide use, they will refer the reader to the publications [22, 23, 27-31].

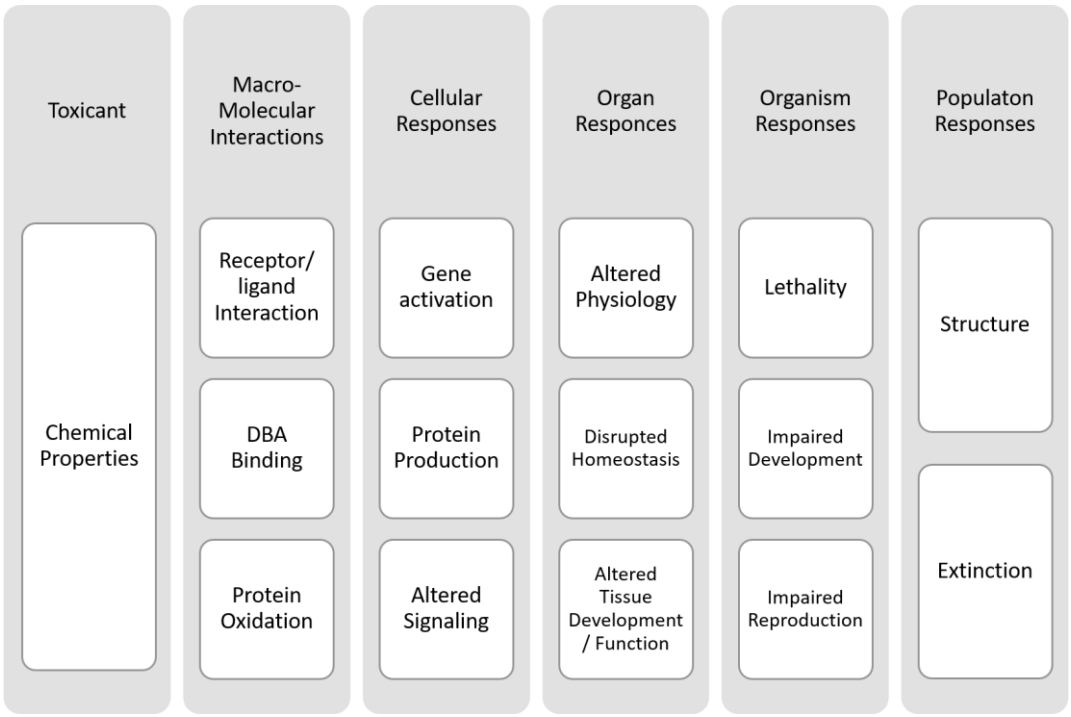

Figure 5 A schematic picture of the AOP described concerning several pathways, adopted from [20].

The (Q)SAR models/ databases mainly concentrate on mutagenicity, carcinogenicity, developmental toxicity, skin sensitization, respiratory sensitization, and skin and eye irritation.

Only limited tools for repeated-dose toxicity studies are developed. In this case, the substance class has been used for repeated dose toxicity assessment of a specific chemical group that is chemically similar [32, 33]. On the other hand, for single cases, certain structural characteristics have been linked to specific target organs [34].

Nevertheless, for complicated cases, the relation of specific target organ toxicity and specific toxic outcomes to certain structural characteristics has not been addressed more broadly and systematically.

The necessity of developing computational models for toxicity assessment became more prominent for using the computational resources in organizing, analyzing, modeling, simulating, visualizing, or predicting the chemicals' toxicity.

The computational toxicology strategy is to complete in vitro and in vivo toxicity tests to reduce the need for animal testing, lower the cost and time of toxicity tests, and

enhance toxicity prediction and safety assessment. On the other hand, in-silico methods can estimate chemicals for toxicity even before they are synthesized [35].

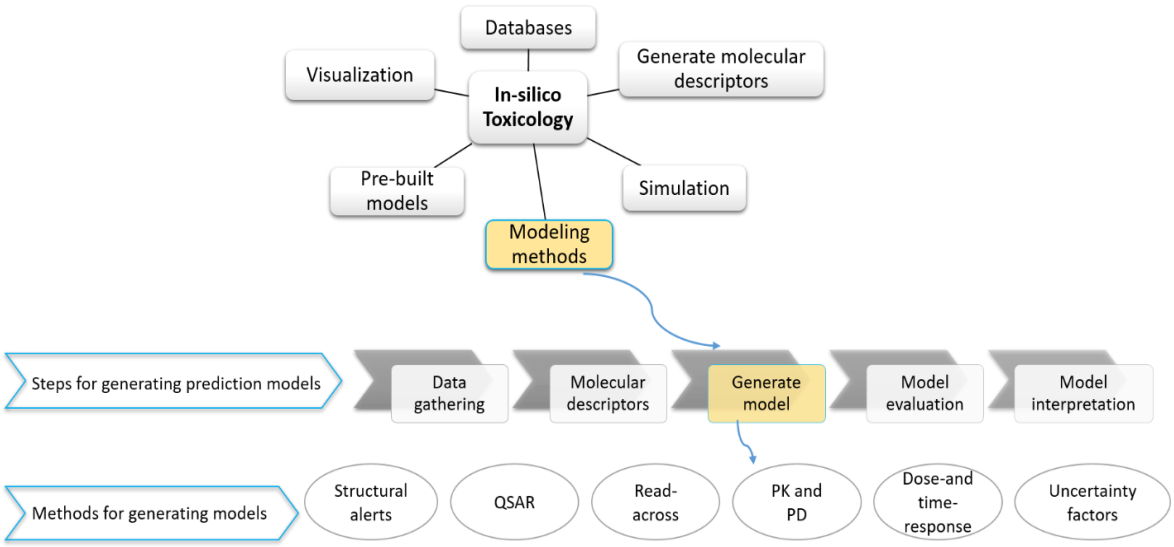

Figure 6 Computational techniques of in-silico toxicology. Gather biological data that contain associations between chemicals and toxicity endpoints, calculate the chemicals' molecular descriptors, generate a prediction model, evaluate the model's accuracy, and interpret the model. Adopted from [36].

In-silico toxicology contains a wide variety of implemented computational techniques and tools:

- Databases comprising data about chemicals properties
- Software for calculating the molecular descriptors
- Simulation tools for molecular dynamics
- Predictive modeling for toxicity data
- Prediction models using statistical packages and software
- Expert systems that include pre-made models in web servers for predicting toxicity
- Visualization tools

Many in silico methods have been designed to predict the toxicity of chemicals. The approaches mentioned above illustrate the historical evolution of in-silico toxicology or define the state-of-the-art procedure for predicting toxicity.

2.3. Systemic toxicity addressed by repeated dose toxicity studies

Repeated dose toxicity testing is performed to ensure safe medicinal product development given repeatedly to the patients. The main goal of repeated-dose toxicity studies is to characterize the toxicological features of the test chemicals following repeated consumption. These studies aim to identify the probable target organs of toxicity, and exposure/response relationships are essential to obtaining this goal. In addition, the reversibility of the toxic effect should be considered.

OECD guidelines are used to standardize repeated dose experiments [37]. These guidelines define the minimal scope of examination as well as study design parameters.

Studies can vary with concerning their study design and reported outcomes. The essential factors in determining the chemicals toxicity are considered as below:

- The route of exposure (oral, dermal, inhalation),
- The substance dose (amount of the chemical),
- Frequency of exposure (single versus multiple exposures),
- Period of exposure (ranging from subacute (28 days) to chronic/life span (720 days)
- Biological properties (tested species, age, gender),
- ADME features (absorption, distribution, metabolism, and excretion/ elimination)
- Structural features

The dose regimen, exposure period, and route of administration should be selected based on the exposure situation (chemicals) or the desired clinical use (drugs) and available pharmacodynamics, pharmacokinetic, and toxicological information.

The other critical points in repeated dose toxicity studies are identifying the substances quality and physicochemical features.

The selection of animal species is based on their similarity to humans regarding pharmacokinetics, biotransformation, and being in contact with the primary human metabolites should be guaranteed. However, also pragmatic considerations play a role, like life span of the tested species as well as body weight. This is the reason for which rodent studies are usually done as first species and dogs are used as mammalian species.

Other quality criteria include, that animals gender should be equal and exact numbers of male and female animals should be utilized.

The size of the exposed group should be adequate to allow a significant scientific arrangement of the data developed. However, moral concerns and practical aspects are also critical, meaning that a high standard of animal husbandry is required. Food intake, general behavior, body weight, hematological parameters, clinical chemistry, urinalysis, and ophthalmology are the points to be observed during the study. For each subject mentioned above, applicable parameters should be selected to identify the toxicity profile.

2.4. Toxicological Database

Evaluation of existing chemicals is always a desired field in relevant sciences. The existing database is still insufficient for many compounds, and many tests must be repeated or performed from the beginning. This aim will take an extended period, even just for compounds' most basic toxicological testing.

Reducing test repetition and minimizing the study time requires an alternative to prioritize. Thus, RepDose database has been developed with a focus on the adverse outcome of the specific target organ, reported in repeated dose studies and on the corresponding LOELs and NOELs. In RepDose, the connection between structural properties and other chemical features, specific target organs, and adverse effects is accessible and can be assessed by applying specific queries. Data on specified commercial organic chemicals with a few functional groups have been used to construct the database. Pharmaceuticals, inorganics, metal compounds, and mixtures with complicated and multifunctional chemical structures have been excluded.

The repeated dose studies have been entered into the RepDose database based on the availability and reliability of the data. Implementing the database for a vast number of chemicals with different structures has been required. It was necessary to pursue peer-reviewed publications for chemical pre-selections to obtain this approach. The imported investigations contain reports on all crucial target organs of repeated-dose toxicity, hematology, clinical chemistry, and histopathology.

Two main study types, oral and inhalation, were entered into the database, and dermal exposure was excluded because of a lack of information. Studies with oral exposure were divided into three subgroups: gavage, feeding, and drinking water. The body dose was estimated for inhalation studies, considering 100% absorption.

Most frequent observations of repeated-dose toxicity for industrial chemicals have been reported in mice or rats. Therefore, the target organs, the LOELs, and mechanisms of toxicity for different chemicals are easily comparable in the database. The investigations with 14-day periods up to the lifetime exposure were selected. In the case of several reliable studies for one compound, all were entered into the database. Therefore, they might report different endpoints, so it is possible to analyze the duration impact on the adverse outcomes and prove any similarity in the results.

The design of our study was addressed in terms of:

- Compound Identity
- Study reliability
- Species
- Route of administration
- Study duration
- Availability of study- and effect-LOEL
- Toxicological endpoints

- Type of adverse outcome

Table 1 indicates the high quality RDT studies with oral and inhalation exposure on chemicals extracted from RepDose® DB, additional oral studies were added from ToxRef DB (US EPA), Hess DB (NEDO). To link external databases such as ToxRef and Hess DB, RepDose has a curated vocabulary and uses standard ontology.

The project database comprises over 7,100 oral (oral food/drinking water and oral gavage) and inhalation RDT studies on about 3,000 chemical compounds. The database includes various study durations, subacute (28±7 days), subchronic (90±10 days) and chronic (>365 days) on mainly rodent species (rat and mouse). Other species contain dog, rabbit and monkey. The toxicological effects per target/organ in RepDose are reported per gender and doses of administration. Reliability scores assigns study quality in reference to general design and scope of examination.

In this experiment, the oral and inhalation treatment in all exposure durations (subacute, subchronic, and chronic) in rat and mouse were extracted from the database for further analysis. In total, 1955 oral studies on rat (424 subacute, 910 subchronic, and 621 chronic studies) and 573 oral studies on mouse (24 subacute, 302 subchronic, and 247 chronic) and 394 inhalation treatment on rat (105 subacute, 199 subchronic, and 90 chronic studies) and 141 studies on mouse (3 subacute, 72 subchronic, and 66 chronic studies) existed in RepDose DB. The selection criteria as mentioned above, contain the reliability of studies and availability of study- and effect-LOEL.

Table 1 Number of studies in RepDose/ToxRef/Hess databases, differentiating by study duration, species, and route of administration.

Examination period						
	Subacute*		Subchronic**		Chronic***	
	oral	inhalation	oral	inhalation	oral	inhalation
Rat	755	280	1229	256	1035	121
Mouse	131	51	417	81	786	87
Dog	45	3	279	7	99	
Rabbit	10	1	1	6		
Monkey	18		3	6	1	
Total	959	335	1929	356	1921	208
<30 days; **>80 days and < 100 days; *>365 days*						

The physicochemical information of each compound, such as molecular weight, solubility in water, physical state, boiling point, dissociation constant, octanol-water partition coefficient (log Kow), and vapor pressure, is enclosed in the database.

The toxicological data has been distinguished between study data and effect data. Effect data include all target organs with all related effects and associated LOELs. Several effects may occur in one target organ at different dose levels. Table 2 shows example of data for a single study, entered into the database. Besides the LOELs for the single effects, the overall lowest observed effect level and the overall no observed effect

level (NOEL) are documented. Study data contain the specification of the animals (strain, sex, number per dose group) and the exposure information (i.e., exposure period, route of administration, post-exposure observation period, dose groups and effect directions).

For a comprehensible computation of chemical features and toxic results, the database entries have to be in an identical form and standardized. Thus glossaries have been designed for all domains, which can be addressed by standard queries. The standardization is available for defined fields: species, strain, sex, route of exposure, organ, and effect.

Table 2 Example for data entry into the database for one study per species, sex. The effects direction is in all observations reported as "increased." The observed effects at LOEL are colored orange.

Target Organ	Effect	LOEL (mg/kg bw/d)
Clinical chemistry	Changed enzyme activity	60
	Bile acids	60
	Urea/nitrogen	60
Clinical symptoms	Mortality	60
Hematology	Hematocrit	60
Kidney	Weight increased	30
Liver	Degeneration	60
	Hypertrophy	60
	Pigmentation	60
	Weight increased	30
Testes	weight increased	60
Overall LOEL		30
Overall NOEL		10

Weight increases in the kidney and liver are observed at LOEL (30 mg/kg bw/d). The studies reported NOEL at 10 mg/kg bw/d.

2.5. eTOX Data

eTOX project (Integrating bioinformatics and chemoinformatics approaches for the development of expert systems allowing the in-silico prediction of toxicities) started on 1st January 2010 and continued for 18 months [38]. The eTOX DB has evolved as the largest preclinical toxicity database for drugs and drug candidates and comprises more than 1900 different chemical structures with 8047 in vivo toxicity studies extracted from over 7,000 legacy reports with repeated exposure, which differ with regard to study duration (from subacute to chronic repeated dose toxicity studies), tested species (rat, mouse, dog, rabbit, monkey, guinea pig, hamster and pig), routes (oral, inhalation, dermal, subcutaneous, subcutaneous and intramuscular) and investigated endpoints (Clinical observation, Clinical chemistry, General toxicity, Hematology, Hemostasis, Histopathology, Necropsy, Organ weight and Toxicokinetic studies).

In this work, the oral treatment in all exposure durations (subacute, subchronic, and chronic) in rat and mouse were extracted from the database for further analysis. In

total, 1366 oral studies on rat (1115 subacute, 225 subchronic, and 26 chronic studies) and 181 oral studies on mouse (55 subacute, 71 subchronic, and three chronic) compared to the RepDose DB. The selection criteria contain the treatment-related status of the study and the availability of LOEL.

Table 3 Number of studies in eTOX differentiating by study duration, species, and route of administration

	Examination period					
	Subacute*		Subchronic**		Chronic***	
	Oral	Inhalation	Oral	Inhalation	Oral	Inhalation
Rat	2345	90	232	18	27	1
Mouse	157	5	80	6	3	1
Dog	1042	60	143	17	14	
Rabbit	19	1				
Monkey	227	8	28	4	5	
Hamster	4	1				
Pig	5					
Total	3799	165	483	45	49	2
<30 days; **>80 days and < 100 days; *>365 days*						

2.6. Primary and secondary adverse outcomes

The liver plays a significant function in metabolism and has several roles in the body. Plasma protein synthesis is one of the liver functions.

When different tissues are damaged, the damaged cells tend to release specific enzymes and biochemicals into the bloodstream, which results in an abnormal level of enzymes and biochemicals. For example, Alanine-Aminotransferase (ALT) and Aspartat-Aminotransferase (AST) are released from damaged hepatocytes in liver disease into the bloodstream. These abnormalities can detect and confirm, diagnose and localize the clinical problems.

Primary toxicity is the toxicity caused by the substance directly. Secondary toxicity would be toxicity resulting from other substances produced as a result of the original toxin in the body or as a result of the primary failure of the organ. An example is elevations in total serum protein levels representative of inflammation in the liver. Primary toxicity would be liver inflammation, and secondary toxicity would be the total protein level changes in the serum. Therefore, in this experiment, the secondary effects such as albumin, globulin, and total protein level changes, classified as chemical chemistry parameters were reallocated to a primary target organ such as liver or kidney.

Albumin and globulin are two types of proteins found in the blood. Albumin is produced by the liver and drives about 60% of the total protein, and globulins fill out the

remaining 40% of proteins in the blood. Globulins have an important role in liver function, blood clotting, and infections. They are various groups of proteins, some constructed by the liver and others initiated by the immune system [39].

Total protein levels may decrease when the production of albumin or globulin proteins fails, such as in undernourishment or extreme liver disorders. When the proteins get denatured or lost, like in kidney failure (nephrotic syndrome), Total protein levels will decrease. Increases in total protein level occur in abnormal high production of proteins, e.g., in inflammation [40].

A low Albumin/Globulin ratio may be due to an overproduction of globulin, underproduction of albumin, or loss of albumin, which may indicate an autoimmune disease, cirrhosis, inflammation, scarring of the liver, multiple myeloma, nephrotic syndrome, and kidney disease [41].

In this study, the observations on total protein level, albumin, and globulin level, and the ratio of albumin to globulin were reallocated to the primary effect/target organ, such as liver and kidney.

Anemia is a typical hematologic disorder and is caused by various mechanisms. Examining blood and blood-formation organs, bone marrow and immunologic disorders are critical steps in diagnosing Anemia [42]. Hematologic diseases can be approached by determining the primary hematologic affected feature: red blood cell (RBCs), white blood cells (WBCs), platelets, or the coagulation system [43, 44]. Anemia is related to a reduction of one or more RBC indices. The erythrocyte indices considered in findings and diagnosing the clinical hematology [45]. The indices provide an estimation of the mean size of circulating erythrocytes (mean corpuscular volume, MCV), the average concentration of hemoglobin per erythrocyte (mean corpuscular hemoglobin concentration, MCHC), and the average count of hemoglobin in erythrocytes (mean corpuscular hemoglobin, MCH) [46]. In addition, the red cell distribution width (RDW) provides a quantitative estimation of the heterogeneity of red cells in the peripheral blood [47].

Different categories of MCV and RDW values of erythrocyte disorders are described. Any of these combinations demonstrated a reasonable differential diagnosis of erythrocyte disorders and consequently will cause Anemia: low MCV/normal RDW, low MCV/high RDW, normal MCV/normal RDW, normal MCV/high RDW, high MCV/normal RDW, high MCV/high RDW [47].

The other critical cause of Anemia is the changes in the count and structure of WBC. Five types of white blood cells with different functions exist, namely Neutrophils, Eosinophils, Basophils (these three are known as granulocytes), Lymphocytes, and Monocytes [48]. In Anemia among the WBC subsets, lymphocyte count is significantly affected [49].

Lymphocytes are group of WBCs in both blood and the lymphatic system. They are classified into three varieties:

- B lymphocytes (B cells) produce antibodies in the body's natural defense and immune responses.
- T lymphocytes (T cells) identify foreign substances and process them for disposal.
- Natural killer cells (NK cells) attack and destroy irregular cells, such as cancer cells or virally infected areas [50].

The high lymphocytic count changes were considered the primary effects of Anemia. According to the mentioned points, the reduction of one or more RBC indices and raised Lymphocyte levels were considered the primary effects of Anemia and reallocated to Anemia in the database.

In this experiment, in the interest of reducing de novo animal testing, the data analysis aims to determine the frequency of affected targets/organs in drugs/chemicals exposure (diet/drinking water and gavage) of the different treatment periods (chronic, subchronic, and subacute) in preclinical studies.

The differences in potency between target organs were the following point of the analysis procedure using the RepDose data, and detecting the LOEL probability by a study with a limited number of targets/organs was conducted in the next step by implementing the RepDose dataset. This expectation was shown by statistical modeling named Coverage Model [51]. In the coverage model, the total count of studies with particular affected organs combination was used to determine the contribution of individual targets/organs and grouped targets/organs in predicting the LOEL in animal studies.

We were eager to realize that it would be reasonable to apply the subacute treatment (4 weeks of oral exposure) to predict adverse effects in subchronic treatment (3 months of oral exposure) in similar hazard properties (similar chemical, species, and route of administration). This prediction model used the RepDose subchronic, oral, rat studies focused on liver effects.

2.7. KNIME Analytics Platform - Open Source

In 2006 the first version of the KNIME Analytics Platform (www.knime.com) was released and was fast accepted by several pharmaceutical organizations. With enough local or cloud-based space and computing power, it is possible to run projects on the KNIME platform with billions of rows. KNIME is developed in Java and based on Eclipse, the open source multi-language software evolution environment including an integrated development environment (IDE) and an extensible plug-in system. KNIME Analytics Platform is released under an Open Source GPLv3 license with an exception that allows others to use the well-defined node API to add proprietary extensions. KNIME analytical platform is used mainly for designing a flexible and user-friendly computational system that allows users to manage the parameters without deep knowledge of their underlying informatics grounds.

Many data analysis frameworks offer possibilities to create and analyze contingency tables. It considers the significance of connections between the classifications of the two variables. The KNIME platform's workflow was designed to measure the association between short- and long-term adverse outcomes. The investigations were carried out by Bayesian analyses based on calculating positive and negative likelihood ratios in the KNIME Analytics platform. The sensitivity and specificity of each test were used for determining the diagnostic power of the tests, and the diagnostic power was used to identify the connection between subacute and subchronic apical findings.

2.8. Available data Repeated Dose Toxicity (RDT) studies

High-quality RDT studies with oral and inhalation exposures were extracted from RepDose [22], and additional oral studies were extracted from ToxRef DB (US EPA) and Hess DB (NEDO) using preexisting ontologies. The project database contains about 7,100 oral (oral food/drinking water and oral gavage) and inhalation RDT studies on approximately 3,000 chemical compounds. The dataset includes studies with various study durations (chronic, subchronic, subacute, and acute) on rodent species (rats and mouse).

Table 4 outlines the count of studies divided by duration, species, and database. In RepDose DB, between all durations, subchronic studies comprise more studies than other durations: 910 oral examinations and 199 inhalation studies on rats and 302 and 72 mouse studies in oral and inhalation treatment, respectively.

Table 4 Data extraction from RepDose and eTOX for further analyses

Study duration	No. of studies					
	RepDose				eTOX	
	Oral		Inhalation		Oral	
	Rat	Mouse	Rat	Mouse	Rat	Mouse
Subacute	424	24	105	3	1115	55
Subchronic	910	302	199	72	225	71
Chronic	621	247	90	66	26	5

The eTOX data, shared with Fraunhofer ITEM, evolved as the largest preclinical toxicity database for drugs and drug candidates and comprised more than 1,900 different substances and more than 8,000 preclinical in vivo toxicity studies, containing 1497 repeated dose oral exposure.

Table 4 lists the number of oral studies in eTOX, differentiating by study duration,1115 subacute, 225 subchronic, and 26 chronic studies on rats and 55 subacute, 71 subchronic, and five chronic studies on mice.

The adverse outcomes in RepDose contain the specification of the animals (strain, sex, number per dose group) and the exposure information (i.e., exposure period, route of administration, post-exposure observation period, and dose groups). Studies are annotated by reliability scores which assign the study quality in reference to the general

design and scope of the examination. In this evaluation, only studies with reliability A (following the OECD regulations or similar quality) and B (some limitations but appropriate for the evaluation) were selected for further analysis. The route of administration contains oral (gavage studies, administration through drinking water or food) and inhalation studies. The species, including rats and mice, and the exposure periods are defined as subacute, subchronic, and chronic. The availability of LOEL was the other criterion applied in the data selection from the RepDose DB.

For chemicals, the dataset on oral RDT studies reports different routes such as continuous exposure via oral diet, drinking water, as well as bolus application via gavage. The frequency of affected TOs as compared by the Chi-square (X^2) test, showed no significant differences across these three different oral routes stratified to the LOEL and overall dose level, as well as both rodent species (data now shown). Therefore, all three oral routes, as mentioned above, were applied without division as an oral administration for all following analyses. For time dependency analysis, the database contains 944 subacute oral studies (gavage, administration through drinking water or food) on 644 chemicals and 434 subchronic studies on 278 compounds in rats.

The selected dataset contains experimental data from a minimum of one subacute and one subchronic rat treatment through oral exposure. The dataset was derived from studies on the same chemical with dose overlap. The finalized dataset comprises 115 compounds with available LOEL. (Figure 7).

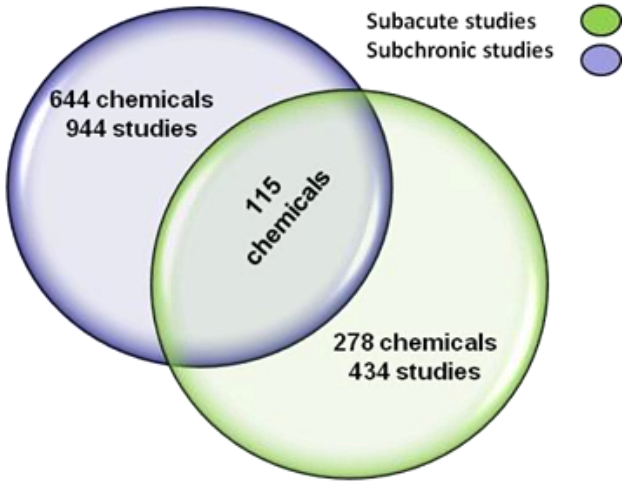

Figure 7 Number of compounds/studies with subacute and subchronic studies (oral exposure, tested species is the rat).

The dataset contains studies with different study designs and scope conducted in various laboratories. Due to these differences, it is required to consider the dose overlapping to increase the prediction power. Therefore, we tested if excluding paired studies that differ regarding study design can enhance the prediction. Also, predicting high-dose effects in 90-day studies is not the most desired aspect. Consequently, we excluded unspecific high-dose weight changes in 90-day studies for prediction improvement.

2.9. Data curation

2.9.1. Differentiation of cancer and chronic studies in RepDose DB

In the RepDose database, the reliability score defines the study quality based on the study scope. Studies are divided into five categories, A, B, C, D, or E to indicate the quality of studies.

Table 5 Reliability Description in five categories

Categories	Quality description
A	Following the OECD regulations or similar quality
B	Some limitations, but appropriate for the evaluation
C	Quality cannot be evaluated (inadequate data)
D	Particular design for a specific target
E	Cancer Study

The distinction of cancer and chronic studies was not fully in place at the beginning of the work. Cancer studies usually report only neoplastic lesions, whereas chronic studies report also non-neoplastic lesions (full scope studies). For a better discrimination, all studies containing different routes of administration, species, regarding the exposure duration (>365 days), with the reliability of A, and B categories, were identified. The next step was to review the papers with the reliability of A and B to determine the assays with true full scope (chronic) and the reports without this element. In the case of cancer studies, based on guidelines and the study scope information, the reliability of studies was allocated to E. Reliability scores were categorized by following the EPA and OECD cancer/chronic toxicity test guidelines (Table 6).

Table 6 Chronic/Cancer Toxicity Test Guidelines

Reliability	EPA Guideline name	Count[1]	OECD Guideline name	Count[2]
A, B	870.4100: Chronic Toxicity	109	452: Chronic Toxicity Studies	6
A, B	870.4300: Combined Chronic /Carcinogenicity	437	453: Combined Chronic/Carcinogenicity	28
E	870.4200: Carcinogenicity	391	451: Carcinogenicity Studies	14

1 Number of studies that follows the mentioned EPA guidelines
2 Number of studies that follows the mentioned OECD guidelines

Studies with high mortality in test and control groups (>30%), low survival, infections, changed dose regimens, and experiments with no control group were excluded from the dataset. Cancer NTP (National Toxicology Program) studies without clinical chemistry parameters, urine analysis, and hematology endpoints were considered equal to the reliability of A and B. The database contains 1970 cancer/chronic studies in various administration routes (oral and inhalation), species (rat and mouse) in more than 365 days of exposure (Table 7).

Table 7 Count of NTP studies with the reliability of E, A, and B compared to Non-NTP cancer/chronic studies at LOEL.

Count of studies				
	NTP		Non NTP	
	E	A, B	E	A, B
Oral	133	245	400	639
Feed	99	117	334	577
Gavage	41	111	29	27
Drinking water	3	17	37	35
Inhalation	9	68	26	72
Total	285	558	826	1350

2.9.2. Reallocation of clinical chemistry parameters in RepDose DB

For the count of target organs that might trigger the LOEL within repeated dose studies, it is essential to reallocate secondary effects to their primary source. Diverse clinical chemistry as well as hematology parameters, such as the number of red and white blood cells, hemoglobin concentration, mean corpuscular volume, coagulation tests, and differential blood count, are reported in the database.

The blood test outcomes can supply the primary assessment of some adverse outcomes, such as anemia and inflammation, which are initial to the histopathology results.

Primary toxicity is the toxicity caused by the substance directly. Secondary toxicity would be toxicity resulting from other substances produced as a result of the original toxin in the body or as a result of the primary failure of the organ. An example is elevations in total serum protein levels representative of inflammation in the liver. Primary toxicity would be liver inflammation, and secondary toxicity would be the total protein level changes in the serum.

Blood sample analysis also comprises clinical chemistry parameters like enzymes, and biochemical analytes in plasma, like transaminases, urea, creatinine, electrolytes levels. The clinical chemistry parameter changes are caused by toxicological effects in tissues and may demonstrate organ toxicity in the liver or the kidneys. Therefore, in this experiment, the secondary outcomes, such as total protein, and creatinine level changes, were reallocated to a primary target organ such as liver or kidney.

The cholinesterase (ChE) level changes remained a clinical chemistry parameter (40 studies) representing a mode of action specific observation for many organophosphate compounds in the database. The analysis is split into two parts, the first part contains the original data from the database, and the second includes the reallocated parameters.1

2

2

Table 9 indicates the count of studies in that clinical chemistry parameters were reallocated to the primary target organs. In 83 studies, the clinical chemistry parameters were changed to the liver; in 86 studies, the parameters were changed to the kidney. Since the clinical chemistry parameters and the primary target/effect were observed at LOEL, in 133 studies the clinical chemistry parameters were excluded from the effect and target lists.

According to this fact, the effects and target organ were allocated or excluded from the dataset:

- If the only effect at the LOEL was the clinical chemistry parameter, we used the next higher dosing to derive the primary target organs.
- If the primary target organ was considered at the next higher dosing, the clinical chemistry parameters were replaced with the primary target organ.
- When the clinical chemistry parameter and the primary target organ were observed at LOEL, the clinical chemistry parameter was excluded from the study because the target organs were already observed at LOEL. This exclusion would prevent any duplications in organ frequency and will help to account for the differences in the reports concerning the granularity.
- While the clinical chemistry was observed at LOEL, and no liver and kidney (as primary target organs) were reported at higher doses, the parameter was excluded.
- If the clinical chemistry was reported at LOEL and no liver and kidney existed at a higher dose, however, other effects which prove the kidney or liver damage were present in the study, the clinical chemistry parameter was replaced with liver or kidney.

Albumin and globulin are two types of proteins found in the blood and essential in liver function, blood clotting, and infections. Due to summarizing the different terms to account for differences in reporting concerning granularity, we combined individual effects into a more general term.

Table 11 demonstrates the count of studies, including the changes in total protein concentration in plasma and related parameters in measuring the total protein concentration (albumin and globulin). In 36 studies, total protein level changes were observed as a single clinical chemistry parameter at LOEL.

Table 8 The frequency of clinical chemistry effects occur in RDT studies at LOEL.

Fraction of studies (%)	
Clinical chemistry parameters	Frequency of effects (N: 334)*
Changed enzyme activity	51
Total protein	19
Cholesterol	15
Urea/nitrogen	11
Albumin	11
Glucose	9
Creatinine	8
Globulin	7
Lipids	7
Phosphorous compounds	7
Calcium	6
Bilirubin	5
Changed hormone status	5
Sodium	5
Bile acids	4
Potassium	3
Chloride	2
lactate	0.6
**N: Total number of studies comprise clinical chemistry parameters: 334*	

Table 9 Count of studies in which clinical chemistry was reallocated to primary organs.

Alternatives for clinical chemistry parameters	No. of studies
Liver	83
Kidney	86
Thyroid	18
Spleen	5
Adrenal gland	2
Urine	1
Body weight	9
Bladder	1
Clinical chemistry	40
Hematology	6
Testes	2
Brain	1
Excluded parameter*	133
** Clinical chemistry parameters and the primary effect/target were observed at LOEL.*	

Table 10 Some examples of the association of clinical chemistry parameters (creatinine level) with adverse outcomes in kidney and urine analysis in similar studies.

Study ID	Target organ [1]	Adverse effect [1]	Target organ [2]	Adverse effect [2]
1	Clinical chemistry	Creatinine	Kidney	Damage Weight increased
2	Clinical chemistry	Creatinine	Kidney	Weight increased
3	Clinical chemistry	Creatinine	Kidney	Weight increased
4	Clinical chemistry	Creatinine	Kidney	Karyomegaly Weight increased
5	Clinical chemistry	Creatinine	Kidney	Weight decreased
6	Clinical chemistry	Creatinine	Kidney	Pigmentation Weight decreased
7	Clinical chemistry	Creatinine	Kidney	Dilatation Hyperplasia Nephropathy Weight increased
8	Clinical chemistry	Creatinine	Kidney	Weight increased
9	Clinical chemistry	Creatinine	Kidney	Mineralization

[1] *The target is clinical chemistry; creatinine is the clinical chemistry parameter related to kidney damage*

[2] *Kidney and urine analyses and related adverse effects (associated with creatinine changes)*

According to primary and secondary toxicity in target organs, these 36 observations of total protein level were reallocated to the primary target organs, such as the liver and kidney. 40 studies reported albumin and 25 studies contain globulin as a single clinical chemistry parameter at LOEL. In these 40 and 25 studies, globulin and albumin were summarized to the term total protein. In 17 studies, total protein was observed with Albumin; in 5 studies, it was reported with globulin level changes; in 4 studies, it was observed with both albumin and globulin. In such studies, when total protein level changes were already reported at LOEL, the albumin and globulin factors were excluded to account for the differences in the reports concerning the granularity.

Table 11 The number of studies containing total protein as a clinical chemistry parameter.

Total protein	Albumin	Globulin	Count of studies
✓			36*
	✓		40*
		✓	25*
✓	✓		17
✓		✓	5
✓	✓	✓	4
	✓	✓	10

**Single parameter occurred at LOEL*

Hematology parameters are the other group of individual effects, which could be summarized as effects representing a general effect, such as anemia. indicates the frequency of individual hematological effects in subchronic oral rat studies. Hematology was reported in 205 subchronic oral rat studies, and the frequency of the effects is defined in percentage. Anemia is a biological process which is characterized by several effects such as a decrease of erythrocytes, hematocrit, MCH, MCHC, MCV, hemoglobin, reticulocytes, red blood cell distribution width, and nucleated erythrocytes. In dataset these effects are sometime reported individually or summarized as one term Anemia (Table 12).

Table 12 Hematology parameters in RepDose DB, the observation are reported in percentage.

Fraction of studies (%)	
Hematology parameters	Frequency*
Hemoglobin	35
Erythrocytes	32
Hematocrit	27
RBC parameters changed	18
Leukocytes	17
Thrombocytes (platelets)	14
MCV	13
Extramedullary hematopoiesis	9
Reticulocytes	8
Anemia	8
Lymphocytes	8
MCH	7
MCHC	7
Clotting time	7
Methemoglobinemia	7
Changes in cellular structures	2
Granulocytes	2
Monocytes	2
Heinz bodies	1
Eosinophils	1
Neutrophils (segmented)	1
Urea/ Nitrogen	1
**N: total count of subchronic oral rat studies: 205*	

Table 13 shows that of the 111 studies comprising the related effects to anemia, 21 contain anemia and related effects together. Therefore 89 studies (111-21= 89) reported individual effects related to anemia. These individual effects were summarized to anemia for the count of effects (Table 13). Those 21 studies containing the related effects to anemia were excluded from the dataset because they simultaneously were reported with anemia at LOEL. The exclusion will help summarize the term to avoid differences in effect pattern only because of differences in reporting and granularity.

Table 14 indicates the individual clinical signs and the combined group of signs used to summarize these effects in 5 new categories: Neurological symptoms, Lethargy, Hyperactivity, Poor general conditions, and Irritation.

Table 13 Count of studies contain anemia and related effects of anemia.

Parameter at LOEL		Count of studies
Anemia[1]	Effects related to Anemia[2]	
✓		16
✓	✓	21
	✓	111

[1] *All studies comprise Anemia at LOEL*

Table 14 Clinical signs whit similarity in target or type of effects combined in one category

Individual clinical signs	Summarized effects
Behavior abnormal	Neurological symptoms
Coma	
Convulsions	
Moving uncoordinated	
Drooping eyelids	
Neurological symptoms	
Tremor	
Ataxia	Lethargy
Hypoactivity	
Lethargy	
Excitation	Hyperactivity
Hyperactivity	
Fur loss	Poor general conditions
Hunched posture	
Piloerection/fur	
Ruffled	
Poor general conditions	
Prostration	
Irritation	Irritation
Encrustation	
Eschar	
Salivation	

2.9.3. Standardization of entries

In addition to the toxicological profile, we calculated the toxicological potency using organ LOEL values. LOEL values show a lognormal distribution. The numeric values were normalized to the log10 scale to realize a normal distribution. The normalized data in our data set were applied to calculate the geometric means, fifth and fiftieth percentiles, quartiles, and confidence interval levels. Percentiles and quartiles are the main ways of measuring the position of values in descriptive statistics of continuous data, and their usage is recommended for reference interval measures [52, 53].

2.10. Simulation of coverage uncertainty

The statistical method developed to calculate the probability of detecting a reference value (LOEL) refers not only to single targets but to combinations of targets. The probability of predicting a reference value can also be identified as the coverage of the target organs included in the model. The following most frequent targets were included in our analysis: Liver, kidney, clinical chemistry, body weight, clinical symptoms, hematology, thyroid gland, spleen, and testes.

The predicted values were calculated through a regression model, and the result of this calculation was used to obtain the empirical and model coverage. We used the regression model because the response variables were considered as count.

The dataset and binary tables, including the 11 most frequently affected TOs, were created to be imported into the regression model. A regression model was provided by Tom Aldenberg in 2013 [51] and fitted for the new dataset in 2020. The models were fitted separately for one to 8 estimated values (organs). For example, in the case of four organs, the Poisson regression equation is as follows:

$$Y = \exp(ß_1\,CCh + ß_2\,Liv + ß_3\,BW + ß_4\,Kid)$$

Which has been performed in R software with this command:

m= glm (fre~ CCh+ Liv+ bw+ Kid, family=poisson,data=dataset)

The table below shows the coefficients related to each organ:

The fit and coefficients related to each organ were estimated and used in the equation:

$$Y = \exp(-0.545CCh - 0.689\,Liv - 1.043\,BW - 1.089\,Kid)$$

By replacing the values in the equation, the model's predicted values were obtained and used for the calculation of the models coverage:

$$Model\ coverage = 1 - (predicted\ fre(0,0,0,0)/Total\ count)$$

$$Model\ coverage = 1 - \left(\frac{212}{910}\right) = 0.77\ or\ 77\%$$

$$Empirical\ coverage = 1 - \left(\frac{194}{910}\right) = 0.79\ or\ 79\%$$

Table 15 Calculated coefficients related to each organ in R software

Organs	Coefficients
CCh	-0.5450
Liv	-.6899
BW	-1.0437
Kid	-1.0898

CCh: clinical chemistry, Liv: Liver, BW: Body weight, Kid: Kidney

2.11. Principal component analysis (PCA)

The SMILES strings for compounds were inputted into the RDKit (v. 2020.03.6; www.rdkit.org) Descriptor Node, accessed through KNIME software (v. 3.4.0; www.knime.com), to obtain the physico-chemical properties for all chemicals. The properties included molecular weight, the number of hydrogen bond donors/acceptors, predicted logarithm of the octanol: water partition coefficient (SlogP), and the topological polar surface area (TPSA) [54]. While it is possible to generate thousands of physicochemical properties, here, only a few calculable properties were selected, representing those most often used to broadly characterize chemicals in terms of size, polarity, and partitioning behavior. Histograms were also generated on KNIME to visualize the different property ranges between the data sets.

Principal component analysis (PCA) was used to visualize and compare the structural properties of drugs and chemicals. eTox does not give the structure information for all tested drugs. Therefore, the curated smiles codes of 1,387 drugs from DrugMapper (kindly provided by the DrugMapper within the Premier project) were compared to the smiles of 1,954 chemicals with oral studies in the project database. MACCs Molecular fingerprints were calculated using the RDKit package in KNIME (version 3.4.0)

2.12. Concordance metrics to analyze time dependencies

Short-term and long-term in-vivo studies concordance relationships were evaluated using two-by-two contingency tables for each duration-observation pair. We treated the short-term observation as a diagnostic test for the long-term observation and used the statistical methods developed to evaluate the efficacy of diagnostic tests in identifying the connection between subacute and subchronic apical findings.

	90-day Observation	*No 90-day Observation*
28-day Observation	True positive TP	False positive FP
No 28-day Observation	False negative FN	True Negative TN

$$Sensitivity = \frac{\Sigma\ True\ positive}{\Sigma\ Condition\ positive}$$

$$Specificity = \frac{\Sigma\ True\ negative}{\Sigma\ Condition\ negative}$$

$$LR+ = \frac{sensitivity}{1 - specificity}$$

$$LR- = \frac{1 - sensitivity}{specificity}$$

Figure 8 2x2 contingency table used for statistical analysis, Cells contain count of studies for the given observation
Condition positive: TP+FN, Condition negative: TN+FP

The values in the two-by-two contingency table, which are counts of the chemicals in each of the four categories for a given RDT observation, were generated as follows and shown in Figure 8 for every unique observation recorded in either short or long periods.

The true Negative (TN) values were measured as observations that occurred neither in the short-term nor long-term. This is an estimation since not all possible observations were measured or attempted to be observed for each duration. In this experiment, we used the positive likelihood ratio, "LR+", computed with the formula TP×(FP+TN)/FP×(TP+FN) from the values in Figure 8 [55].

The likelihood ratio illustrates the chemical risk when an adverse effect is observed in a short-term study. The advantage of the likelihood ratio is that it is independent of the prevalence (Bayesian prior probability) for each observation, so it is more comparable across different adverse events than the conditional probability or positive predictive value [56]. The negative likelihood ratio, "LR-", which defines the decrease in risk if the short-term observation is not made, was also computed [55, 57].

Table 16 Subjective interpretation of positive likelihood ratios [55, 56].

LR+	Interpretation
>10	Large and often conclusive shifts in probability
5 - 10	Moderate shifts in probability
2 - 5	Small, but sometimes important, shifts in probability
1 - 2	Alters probability to a small, and rarely important, degree

As proof of concept, the examination started with measuring the predictive power of the 28-day studies for studies in 90-day containing any effect on the liver, with the same CAS, species, and route of administration in both durations.

The dataset contains studies with different study designs and scopes conducted in various laboratories. Due to these differences, it is required to consider the dose overlapping to increase the prediction power. Therefore, we tested that excluding paired studies that differ regarding study design can effectively enhance the prediction. Also, predicting high-dose effects in 90-day studies is not the most desired aspect. Consequently, we excluded unspecific high-dose weight changes in 90-day studies for prediction improvement.

In the next phase of comparison, the combination of liver effects and clinical liver chemistry parameters, such as liver enzymes alteration (Alanine aminotransferase, Alkaline phosphatase, Gamma-glutamyl transferase, Aspartate aminotransferase, 5'-nucleotidase), and changes in the level of bilirubin, total protein, albumin, and the ratio of albumin to globulin in both durations were added to the calculation. Table 17 outlines the various comparison forms between short and long-term durations.

Table 17 Various comparison forms of short and long-term duration studies.

28-Day study	90-day study	Additional information
Liver	Liver	Organ weight changes assigned to be no-90-day and no-28-day observation. Clinical liver chemistry parameters: liver enzyme , bilirubin, albumin, total protein, ratio of albumin to globulin
Clinical liver chemistry	Liver	
Clinical liver chemistry+ Liver	Liver	
Liver	Clinical liver chemistry+ Liver	
Clinical liver chemistry	Clinical liver chemistry+ Liver	
Clinical liver chemistry+ Liver	Clinical liver chemistry+ Liver	
Organ weight changes	Liver	Predictive power of organ weight changes in 28-day for any liver effect in 90-day
Organ weight changes	Liver+Clinical liver Chemistry	
Liver	Liver	Organ weight is included as a specific effect in both duration
Liver	Liver	Studies with organ weight changes as a single effect were excluded
Clinical liver Chemistry	Liver	
Liver+Clinical liver Chemistry	Liver	
Liver+Clinical liver Chemistry	Liver+Clinical liver Chemistry	

2.13. KNIME Workflow

Many data analysis frameworks offer possibilities to create and analyze contingency tables. It considers the significance of connections between the classifications of the two variables. The KNIME platform's workflow was designed to measure the association between short- and long-term adverse outcomes.

Bayesian analyses were conducted to calculate positive and negative likelihood ratios in the KNIME Analytics Platform (Figure 9).

The investigations were carried out by Bayesian analyses based on calculating positive and negative likelihood ratios in the KNIME Analytics platform. The sensitivity and specificity of each test were used for determining the diagnostic power of the tests, and the diagnostic power was used to identify the connection between subacute and subchronic apical findings.

The analysis was performed using different criteria to find the best prediction model. In the beginning, the liver outcomes in each duration were compared. Then the calculation followed by adding the individual effects of 28-day effects to predict liver outcomes. In the third step, the model was extended by adding the 90-day individual adverse liver outcomes to the comparison. Every effect was iterated in a loop to be compared with all individual effects and the combination of effects in other duration.

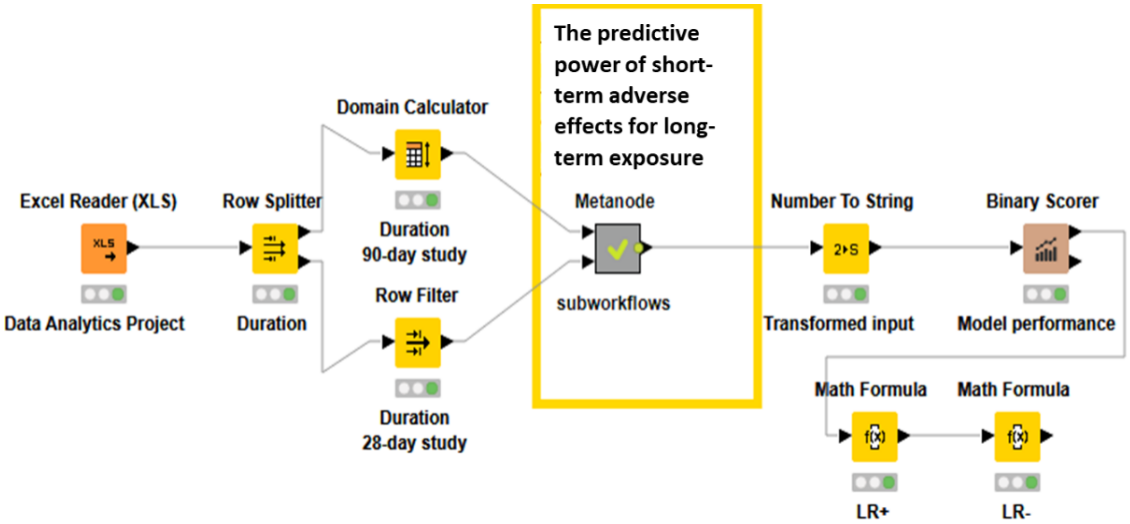

Figure 9 Calculation of positive and negative likelihood ratio in the KNIME Analytics platform The Metanode contains the data resulting from similar chemical substances, the same route of administration, and the same species.

Table 18 outlines the different conditions of comparing short and long-term TOs and effects in the prediction model.

Table 18 Examples of compared criteria in the contingency table

Comparison									
28-DAY	**90-DAY**	**TP[1]**	**FN[2]**	**FP[3]**	**TN[4]**	**Sen[5]**	**Spe[6]**	**LR+[7]**	**LR-[8]**
Liver	Liver	92	21	41	23	0.8	0.66	2.63	0.3
Clin Chem*	Liver	46	16	46	69	0.4	0.74	1.53	0.82
Liver+Clin Chem	Liver	97	28	34	18	0.84	0.55	1.87	0.29
Liver	Liver+Clin Chem	99	14	30	34	0.74	0.68	2.33	0.38
Clin Chem	Liver+Clin Chem	55	7	37	78	0.41	0.84	2.6	0.7
Liver+Clin Chem	Liver+Clin Chem	109	16	28	24	0.82	0.6	2.54	0.29

[1] TP: True Positive, [2] FN: False Negative, [3] FP: False positive, [4] TN: True Negative, [5] Sen: Sensitivity, [6]Spe: Specificity, [7] LR+: Positive Likelihood, [8] LR-: Negative Likelihood
**Clinical Liver Chemistry parameters*

2.14. ROC Space

The ROC space is used for a better and worse classification of the prediction model results.

The contingency table is the potential to provide several evaluation metrics. Still, the ROC space graph needs the true positive rate (TPR) and false positive rate (FPR) as parts of some classifier parameters. The TPR illustrates the count of correct positive results occurring among all positive observations available during the test. On the other hand, FPR defines how many incorrect positive results appear among all negative measures available during the examination.

FPR and TPR define a ROC space as x and y axes, respectively, illustrating relative trade-offs between true positive and false positive.

Since the true positive rate is identical to sensitivity and the false positive rate is equivalent to 1 − specificity, the ROC diagram is known as the sensitivity vs. 1 − specificity.

The prediction models which result in, or instance of, a contingency matrix could be illustrated at one point in the ROC space.

The most useful and potential prediction method would induce a point in the top left of the plot or coordinate (0,1) of the ROC space, describing 100% sensitivity (zero false negatives) and 100% specificity (zero false positives). This (0,1) point is known as a perfect classification [58].

3. Results and Discussion

A comparison of structural and physicochemical properties between drugs and chemical datasets through two-dimensional principal component analysis (PCA) shows that the chemical domain between both differs (Figure 10, red dots drugs, chemicals purple dots). Drugs and organic chemicals cover complementary structural spaces. It is also noted, that chemicals have, on average, a lower molecular weight compared to drugs (data not shown). As this analysis was not possible with the data from eTOX compounds because of confidentiality reasons, curated smiles data from well-known drugs from the DrugMapper inventory are shown. As drugs and chemicals do not cover the same structural space, the following analysis will distinguish these two compound classes within the analysis of main T/O for chemicals and drugs.

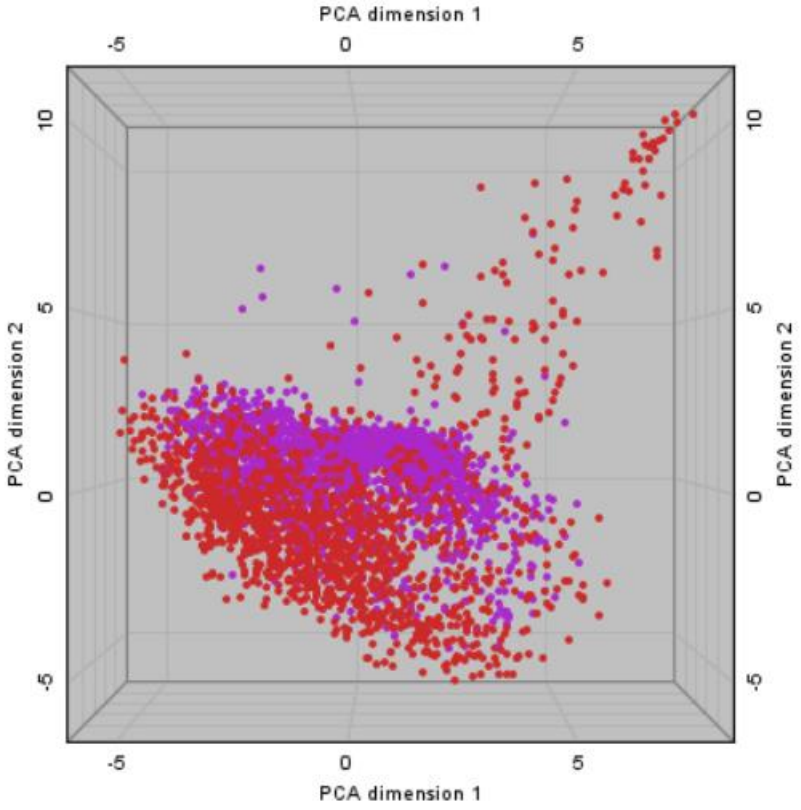

Figure 10 Visualization of the chemical domain of organic chemicals used as drugs (red) and chemicals (purple).

3.1. Number and Frequency of Target Organs (TOs)

3.1.1. Main TOs observed for drugs (eTOX DB)

The frequency of affected TOs induced by drugs in studies with repeated oral exposure stratified to species (rat and mouse) and study duration (subacute and subchronic) is shown in the form of percentage for the entire study (overall) and at LOEL (Table 19). The liver is the main target organ in RDT studies with oral exposure, as it appears to be the TO with the highest frequency in both exposure periods, both species, at LOEL and overall.

In subacute rat/mouse, the liver is affected in 39/44 percent of studies, respectively, and this number changes only slightly to 48/35 percent in subchronic duration. Overall, liver effects are observed in 91/78 percent in rat/moues studies with a subacute exposure period, and a similar increase is observed for longer-term subchronic studies. In descending order, the following most frequently observed TOs at LOEL and overall for the here analyzed conditions are clinical chemistry, clinical observation, and kidney.

Table 19 Main TOs observed for drugs in RDT studies at study level (overall) and at LOEL. Percentage of most frequently affected TOs in subacute and subchronic oral rodent studies are shown (most frequent in grey); main TO are indicated in gray.

Drug TO	Fractions (% of studies) Oral*							
	LOEL				Overall			
	Subacute		Subchronic		Subacute		Subchronic	
	Rat	Mouse	Rat	Mouse	Rat	Mouse	Rat	Mouse
Liver	39	44	48	35	91	78	89	90
Clinical chemistry	34	33	40	31	72	67	65	65
Clinical observation	28	27	25	23	58	55	52	56
Kidney	20	29	30	21	85	73	83	83
Adrenal gland	19	9	16	13	82	60	82	76
Spleen	19	27	17	31	81	73	80	80
Body weight	18	9	19	11	81	85	52	76
Thymus	17	24	11	8	77	62	69	58
Lymph node	12	13	8	14	59	49	60	52
Stomach	11	4	8	8	59	58	59	69
Ovary	11	5	8	10	67	56	66	63
Lung	9	5	16	11	75	58	76	72
Uterus	9	7	4	7	63	45	57	55
Testes	7	15	10	8	67	64	69	65
Hematology	4	24	9	18	38	45	39	51

**Total number of studies: subacute Rat: 1115, Mouse: 55; subchronic Rat: 225, Mouse: 71. The targets/organs with more than 10% of observations are colored grey.*

At the overall dose, the most frequently observed organs in subacute rat studies are the liver, kidney, adrenal gland, spleen, thymus, and lung. Moreover, clinical chemistry is observed in 58% of studies located after the lung.

At the overall dose in subchronic rat studies, the liver, kidney, adrenal gland, spleen, lung, and testes are the most frequently affected TOs.

In mouse studies in a subacute period of exposure, body weight was the most frequently observed organ in 85% of studies, and liver, kidney, spleen, and clinical chemistry were reported in 78%, 73%, 73%, and 67% of the studies, respectively. In subchronic treatment on the mouse, the most frequently affected organs are the liver, kidney, spleen, body weight, and adrenal gland in 90%, 83%, 80%, and 76% of the studies.

Table 19 demonstrates that at a reference dose, such as LOEL, liver, clinical chemistry, clinical symptoms (signs), and kidney are the most sensitive organs that show the adverse effect earlier than other target organs. When the standard threshold (LOEL) is not considered, all target organs are frequently observed in repeated dose toxicological studies.

3.1.2. Main TOs observed for chemicals (RepDose DB)

RepDose comprises studies with oral and inhaltion route of exposure. Main target organs differ between these two routes [59] are therefore distinguished in the following analyses. Due to the limited number of mouse studies, we combined rat and mouse studies as rodent to compare different exposure durations with reliable count of studies.

After oral exposure, all study durations at LOEL listed the liver as the most sensitive organ with 41%/42% in subacute rat/rodent, 33%/38% in subchronic rat/rodents, and 32%/35% in chronic rat/rodents studies. Clinical chemistry as the second most frequently affected TO was reported in 40% rat/ rodent subacute studies, and 37%/ 32% of subchronic rat/rodent treatment and in 24%/20% of chronic rat/rodent experiment.

As observed before, affected organs are evident in a higher count of studies overall compared to LOEL. In subacute studies both in rat and rodent, liver is still on the top (72%), in the subchronic studies it is in the same range as body-weight changes (increased or decreased) (73%), and in chronic exposure, body-weight is observed about 15% more than liver (62%/ 61% in rodent/rat). More than 10% of subacute studies at LOEL report spleen and urine analysis besides the six main targets/organs (in chemicals and 52% in drugs).

Table 20 outlines that at a reference dose, such as LOEL, liver, clinical chemistry, kidney, clinical symptoms, and hematology are the most sensitive organs that indicate the adverse effect earlier, in a lower dose of exposure than other target organs. When the standard threshold (LOEL) is not considered, all target organs are frequently observed in repeated dose toxicological studies. Body weight change is shown as the most frequently affected TOs at the overall doses. Body weight is a sensitive straightforward sign of toxic effects. Body weight changes could be associated with decreased food consumption which has been influenced by the potential toxicity of chemicals or, on the other hand, due to local irritation in the gastrointestinal tract in oral administration.

The results of more than 10% of affected organs in chemical exposure were compared to the results of drugs under the same condition. The significant differences between these two groups were determined by the Chi^2 statistical test when $p = 0.05$ or less. This test was used for testing the independence or determining the relationship between two categorical variables. The comparison of the most frequently affected target organs in preclinical subchronic oral studies between drug and chemicals shows that the most frequently observed targets/organs are similar in both databases at LOEL, and comprise namely liver, clinical chemistry, clinical symptoms, hematology,

and kidney. Body weight changes belong to the most affected and sensetive targets but are less often observed in eTOX dataset. Body weight is the most frequently observed target in chemical exposures at the overall dose. In drug studies, it is located in the mid of the frequency table, 25% less than the chemical table (77% in chemicals and 52% in drugs).

Table 20 Main TOs observed for chemicals in RDT studies at study level (overall) and at LOEL. Percentage of most frequently affected TOs in subacute and subchronic, and chronic oral rodent studies are shown (most frequent in grey); main TO are indicated in gray.

Chemical TOs	Fractions (%of studies) Oral**											
	LOEL						Overall					
	Subacute		Subchronic		Chronic		Subacute		Subchronic		Chronic	
	Rat	Ro-	Rat	Ro-	Rat	Ro-	Rat	Ro-	Rat	Ro-	Rat	Ro-
Body	18	18	26	26	33	34	56	57	77	75	79	76
Liver	41	42	33	38	32	35	72	72	73	73	62	61
Kidney	26	27	25	23	21	19	54	55	60	56	45	40
Clinical	20	19	13	13	14	15	50	50	45	45	40	39
Clinical	40	40	37	32	24	20	70	69	63	56	46	37
Hematol-	20	21	23	21	12	11	46	46	53	49	36	30
Thyroid	6	6	5	4	8	8	12	11	13	11	22	19
Testes	4	4	5	5	7	6	16	16	29	26	21	18
Spleen	10	10	8	8	7	7	22	21	27	26	20	18
Lung	2	2	2	2	3	0	6	6	11	12	16	17
Adrenal	5	5	5	5	7	6	19	19	18	17	19	16
Urine anal-	13	13	8	6	6	5	29	29	25	20	20	15
Brain	2	2	4	3	3	3	9	9	17	16	14	11
Heart	3	3	4	4	3	0	9	10	18	18	13	11
Thymus	3	2	3	2	0	0	14	14	16	15	3	3

**Rodent: rat and mice*

***Total number of studies: subacute Rat: 424, Rodent: 448- subchronic Rat: 910, Rodent: 1212- chronic Rat: 624, Rodent: 941*

The targets/organs with more than 10% of observations are colored in gray.

The same analysis was done for inhalation studies. In inhalation studies, the TO of the respiratory tract (nose, respiratory tract, and lung) was combined to summarize different terms with similar outcomes to help to account for the differences in the reports.

Table 21 indicates the frequency of affected target organs in inhalation rat and rodent studies under chemicals exposure at LOEL and the overall dose level. The most frequently affected targets/organs are marked in gray. Six primary reported targets/organs in oral exposure (liver, kidney, clinical chemistry parameters, clinical signs, hematology, and body weight) were reported in more than 10% of subchronic and chronic inhalation studies. In addition, nose and respiratory tracts were observed at the highest rate in rats and rodents. At LOEL nose was affected in 47%/ 48%, 67%/ 60%, 58%/ 54%, and the respiratory tract was observed in 36%/ 35%, 23%/ 44%, and 22%/ 29% of subacute, subchronic, and chronic duration, respectively in rats/rodents.

Table 21 Main TOs observed for chemicals in RDT studies at study level (overall) and at LOEL. Percentage of most frequently affected TOs in subacute and subchronic, and chronic inhalation rodent studies are shown (most frequent in grey); main TO are indicated in gray.

Chemical TOs	Fractions (%of studies) Inhalation											
	LOEL						Overall					
	Subacute		Subchro-		Chronic		Subacute		Subchro-		Chronic	
	Ra	Ro-	Ra	Ro-	Ra	Ro-	Ra	Ro-	Ra	Ro-	Ra	Ro-
Nose	47	48	67	60	58	54	88	88	47	48	67	62
Clinical che-	30	31	28	22	12	7	46	45	51	42	22	15
Respiratory	36	35	23	44	22	29	52	51	45	44	36	43
Body weight	24	23	21	21	15	18	55	54	62	63	64	65
Hematology	21	0	23	21	15	11	39	39	46	45	28	20
Clinical signs	20	19	20	16	16	19	48	46	43	45	42	44
Liver	19	20	21	26	19	26	44	44	54	60	33	42
Lung	18	18	15	11	20	23	34	33	32	31	32	37
Spleen	12	13	4	5	2	4	24	25	15	17	9	9
Kidney	9	8	16	17	21	18	15	16	41	41	40	32
Testes	3	3	3	3	16	12	8	7	16	16	20	15
Adrenal gland	2	2	2	2	6	4	13	13	10	10	16	12
Thyroid gland	2	2	2	2	7	5	4	4	4	4	15	11
Brain	2	2	3	3	2	1	5	5	12	11	12	10
Urine analysis	2	2	6	4	5	4	13	13	7	10	11	8
Heart	1	1	2	2	2	2	6	6	12	15	9	6
Thymus	1	1	3	4	1	1	10	9	14	15	1	1

**Rodent: rat and mice*

***Respiratory tract: pharynx, larynx, trachea, bronchi and lungs*

N (count of studies): subacute Rat:105, Rodent: 108- subchronic Rat: 193, Rodent: 256- chronic Rat: 87, Rodent: 153

The targets/organs with more than 10% of observations are colored in gray

At the overall dose level, the nose was reported in 88%/ 88%, 47%/ 48%, 67%/ 62% of subacute, subchronic, and chronic studies in rats and rodents. The count of studies with an effect on respiratory tract increased at overall dose level to 52%/51%, 45%/44%, and 36%/43% in rat/rodent in subacute, subchronic, and chronic periods. The adverse nose effects, besides the toxicity of the substance, could occur because of the local injuries on the exposure site. The combination of the respiratory tract organs (pharynx, larynx, trachea, bronchi, and lungs) increased the frequency of the respiratory tract at a LOEL and overall dose levels. Accordingly, it is concluded that the main systemic target organs in inhalation exposure follow the six main TOs in oral studies. Moreover, the respiratory tract was observed in addition to inhalation exposure.

3.2. Exclusion of eTOX data from the following analysis

Data curation in eTOX was required to ensure the entered data held in the correct columns, terms, and units were standardized to identify the missing information. Since

the observed effects in treatment-related findings contain phrases that do not indicate the adverse outcomes (no abnormality detected, normal pathology, normal immune response, etc.), in some cases, judgments on the treatment-relatedness of findings were hardly made plausible. Finding the most frequently affected target/organs and analyzing the principal component (PCA) were performed on eTOX DB, and further investigations continued with RepDose data.

3.3. Number of targets/organs in RepDose DB

Table 22 shows the number of affected target/organs in different study durations (subacute, subchronic and chronic) conducted in different species, rat or mouse, at LOEL and overall dose level in oral and inhalation exposures. The average number of targets/organs at LOEL are more than two target/organs per study type in oral and inhalation treatments. The average number of target/organs in subacute studies increased from 2.9 at LOEL to 9.25 at the overall dose level in oral rat treatments. In inhalation, this number changed from 2.7 at LOEL to 6.5, due to increases in the dose level in subacute period. The average number of target/organs in subchronic rat studies changed with a range of 8.5 targets/organs for oral and 6.1 targets/organs for an inhalation study. In chronic rat treatments, the average count of observed targets/organs increased with a range of 9 in oral and 6.5 in inhalation. In mouse models, the highest average count of affected targets/organs at LOEL was reported in a chronic exposure period with 3.3 targets/organs for an oral study, which increased to 6.8 at the overall dose level. In chronic inhalation studies, 4.3 targets/organs were observed at LOEL and 9.5 at the overall dose level. The difference between average organ counts at different dose levels in subchronic mouse studies also confirmed the increase in number of organs with dose increases. At the overall dose level, 5.3 targets/organs more than at LOEL are reported in oral exposure, and this number reached 6 in inhalation models. Subacute oral treatments were more effective than subacute inhalation exposures in increasing the average count of targets/organs from LOEL to overall dose levels (an increase of 7 organs in oral to 0.8 organs in inhalation).

Table 22 The average number of affected target/organs per study for different study types

Study duration	species	N		No. of targets [a]			
				RepDose DB			
		Oral		LOEL		Overall	
				Oral	Inhalation	Oral	Inhalation
Subacute	Rat	44	31	2.9(2.3)	2.7(2.2)	9.52(2.5)	6.5(2.4)
Subchronic	Rat	55	35	2.5(2.3)	2.9(2.2)	11(2.5)	9(2.5)
Chronic	Rat	60	37	3.6(2.4)	4.8(3.5)	12.6(8.7)	11.3(8.4)
Subacute	Mouse	14	6	2.2(2.4)	3.5(1.2)	9.2(2.1)	4.3(2)
Subchronic	Mouse	45	24	2.2(2.1)	2(2.1)	7.5(2.4)	8(2.4)
Chronic	Mouse	45	29	3.3(2.4)	4.3(3.2)	6.8(5)	9.5(7)

a: Geometric means and geometric standard deviation in parentheses
n: number of organs per study type in different species

3.4. Descriptive statistics of organ-LOEL

The data describe the total number of observed target organs at the end of the descriptive statistical analysis table. In total, 5210 observations were reported for those 28 target organs. The mean value in total was calculated at 2.79 mg/kg bw/d with a -0.35 geometric mean in which 5th% of the reported target organs showed an adverse outcome below -1.99.

Table 23 outlines the descriptive statistics of the lowest dose level, which indicates the target/organ disorders (organ-LOEL), in subchronic, oral studies on rats. Twenty-eight target/organs with a minimum of 20 observations were selected for the analysis. N indicates the observation of each organ with body-weight on the top (686 reports) and intestine, lymph node, and prostate with the lowest count of observations (22 times).

The organ-LOEL values were normalized by converting to the logarithmic scale and applied for calculations of the geometric means, fifth percentiles, the quartiles, and the 95% confidence interval levels of the mean organ-LOEL value.

As observed before in Table 20, the most frequently affected target organ in subchronic oral rat studies is body weight changes with a mean LOEL value of 2,83 mg/kg bw/d. The adverse body weight changes were observed in 686 studies with the geometric mean value of -0.26, while in 5% of the studies, the body weight changed at the dose below -1.81, and in 50% of the studies, the body weight changes were influenced below -0.18.

The liver, clinical symptoms, clinical chemistry parameters, and kidney are located in the table after that, with average LOEL of 1.78, 2.92, 2.14, and 2.15 mg/kg bw/d.

The highest mean values belong to sperm parameters, thymus, and testes.

The sperm parameter, as a less sensitive organ, was affected at the mean LOEL value of 9.2 mg/kg bw/d and the geometric mean of 0.27. In contrast, 5% of observations were affected below -1.46, and the median (50^{th}%)of the sperm parameter LOEL was calculated at 0.35.

The data describe the total number of observed target organs at the end of the descriptive statistical analysis table. In total, 5210 observations were reported for those 28 target organs. The mean value in total was calculated at 2.79 mg/kg bw/d with a -0.35 geometric mean in which 5^{th}% of the reported target organs showed an adverse outcome below -1.99.

Table 23 Descriptive statistics of organ-LOEL in subchronic, oral rat studies. The green rows indicate the most and less sensitive TOs, clinical chemistry, and sperm parameters.

	N[a]	%N	Mean	G[b]	5P[c]	Q1[d]	50P	Q3	CI[e]	
Body weight	686	13.2	2.83	-0.26	-1.81	-0.74	-0.18	0.37	-0.32	-0.19
Liver	610	11.7	1.78	-0.47	-2.03	-0.99	-0.36	0.14	-0.54	-0.39
Clinical signs	526	10.1	2.92	-0.27	-1.82	-0.87	-0.18	0.42	-0.35	-0.19
Clinical chemistry	523	10.0	2.14	-0.68	-2.62	-1.24	-0.54	0.05	-0.77	-0.59
Kidney	485	9.3	2.25	-0.32	-1.81	-0.78	-0.27	0.27	-0.39	-0.24
Hematology	445	8.5	2.02	-0.51	-2.17	-1.03	-0.43	0.13	-0.60	-0.42
Testes	217	4.2	3.86	-0.08	-1.36	-0.60	0.00	0.49	-0.19	0.03
Spleen	205	3.9	2.34	-0.34	-2.02	-0.87	-0.24	0.29	-0.46	-0.21
Urine analysis	200	3.8	2.2	-0.3	-2	-0.90	-0.12	0.41	-0.43	-0.16
Heart	152	2.9	2.62	-0.26	-1.8	-0.73	-0.19	0.25	-0.41	-0.12
Adrenal gland	147	2.8	2.25	-0.25	-1.61	-0.77	-0.20	0.42	-0.39	-0.11
Brain	144	2.8	3.38	-0.15	-1.47	-0.63	-0.16	0.39	-0.29	-0.02
Thymus	123	2.4	3.91	-0.14	-1.55	-0.67	-0.01	0.42	-0.28	0.01
Thyroid gland	89	1.7	1.16	-0.63	-2.32	-1.08	-0.51	-0.01	-0.82	-0.43
Lung	88	1.7	2.16	-0.21	-1.53	-0.69	-0.12	0.30	-0.37	-0.06
Ovary	61	1.2	1.96	-0.24	-1.72	-0.76	-0.20	0.31	-0.45	-0.03
Bone marrow	59	1.1	1.77	-0.29	-1.93	-0.76	-0.28	0.30	-0.49	-0.08
Stomach	49	0.9	2.02	-0.29	-1.85	-0.86	-0.25	0.28	-0.53	-0.05
Epididymis	45	0.9	6.67	0	-1.73	-0.33	0.18	0.55	-0.29	0.30
Eye	41	0.8	1.19	-0.69	-2.92	-1.00	-0.34	-0.11	-1.06	-0.33
Sperm parameters	36	0.7	9.2	0.27	-1.46	-0.08	0.35	0.80	-0.02	0.57
Pituitary gland	35	0.7	1.99	-0.38	-1.88	-0.97	-0.25	0.13	-0.66	-0.10
Uterus	32	0.6	0.95	-0.53	-2.08	-1.03	-0.65	0.22	-0.83	-0.24
FOB	26	0.5	2.79	-0.4	-1.93	-1.04	-0.49	0.42	-0.79	-0.01
Respiratory tract	24	0.5	3.19	-0.29	-2.34	-0.91	-0.37	0.65	-0.75	0.16
Prostate	22	0.4	1.24	-0.49	-1.33	-0.87	-0.43	0.01	-0.88	-0.09
Lymph node	22	0.4	2.2	-0.08	-1.22	-0.61	0.08	0.34	-0.39	0.23
Intestine	22	0.4	1.95	-0.12	-1.05	-0.39	-0.18	0.24	-0.40	0.17
All	5210	100	2.51	-0.35	-1.99	-0.87	-0.26	0.29	-0.38	-0.33

a: N: Count of studies with affected organs
b: G: Geometric mean, the mean on logarithmic scale.
c: P: Percentile
d: Q: Quantile
e: 95% Confidence interval

The normalized organ-LOEL values were applied also for determining the differences in potency between targets/organs by the One-Way ANOVA test. The result revealed significant differences between the minimum LOEL values of 28 targets/organs, with sperm parameters being less sensitive and clinical chemistry parameters being more sensitive than the majority of all other organs. Liver and kidney appeared after clinical chemistry (Figure 11).

Quantiles and percentiles are used as statistical tools for visualizing the distribution of organ-LOELs in subchronic oral rat studies. They are considered a sample estimate of a population parameter; thus, it is required to be presented with a confidence interval [60]. Overlapping the confidence interval levels (C.I.95%) for the 5th and 50th percentiles, respectively, of organ-LOEL, confirm the result of the One-Way ANOVA test.

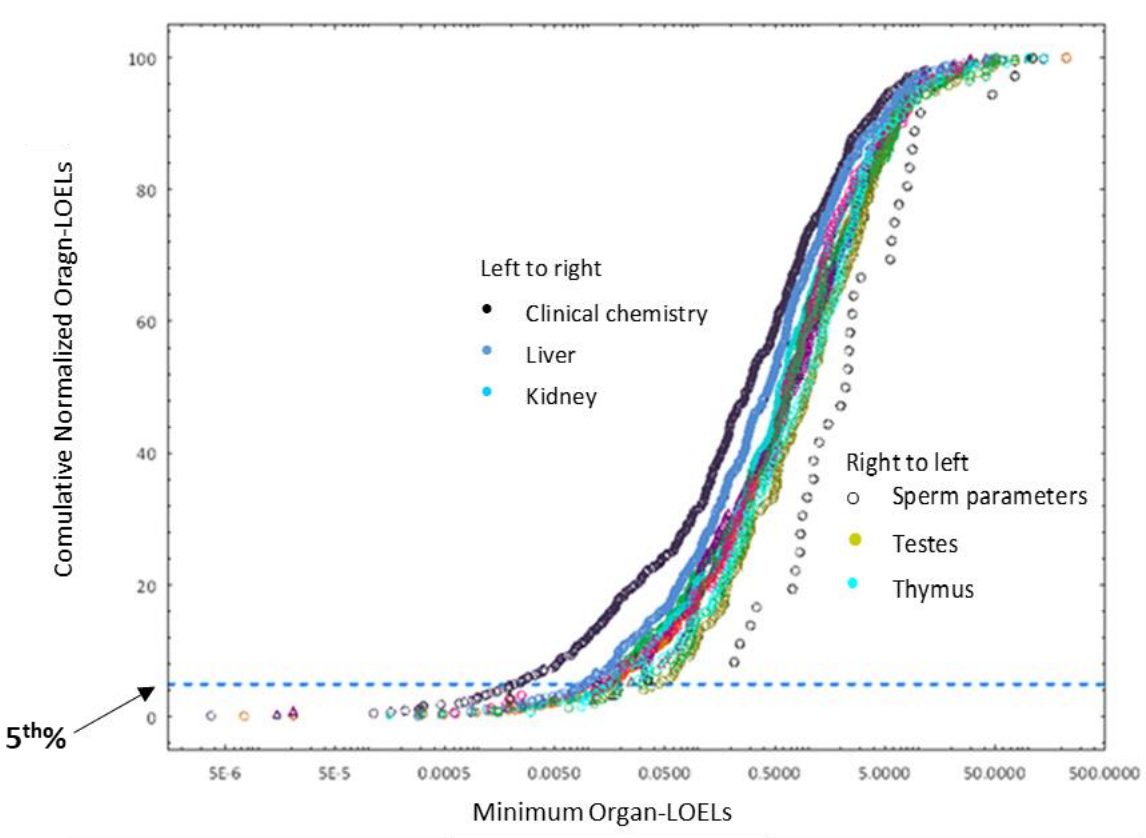

Figure 11 Cumulative distribution of normalized organ-LOELs, Blue line indicates the 5th percentile of the LOELs.

3.5. Coverage model

Coverage model detects the probability of LOEL in a study with a limited number of targets/organs [61]. The coverage model developed in close cooperation with Tom Aldenberg (RIVM).

Total counts of studies with particular affected organ combinations was used to determine the contribution of individual targets/organs and combination of targets/organs in predicting the LOEL in animal studies. The analysis started with the four most frequently affected TOs, clinical chemistry, liver, body weight, and kidney. 716 studies of 910 subchronic oral rat studies showed this combination.

The main four frequent TOs predicted the LOEL with 78.7% probability and 5.3% uncertainty. In the following step, hematology was added as the fifth TO. the count of studies reported the combination of mentioned TOs increased to 774 studies, with an 85% prediction power and 4.6% uncertainty.

With the six most frequently observed TOs (clinical chemistry, liver, body weight, kidney, hematology, and clinical symptoms), the LOEL prediction power reached 89%. Moreover, adding more TO did not remarkably increase the prediction of LOEL; therefore, we decided to remain the analysis with the six TO model.

The six main target organs which are most frequently affected are the most predictive for LOEL in the whole dataset. This combination of targets/organs showed the coverage with 89% probability and 4% uncertainty.

Table 24 Target combinations and the coverage at the LOEL per number of combined TO (N). TO combinations with the highest probability are indicated in orange.

N	Targets												Coverage	Uncertainty		Range
	Che	Liv	Bod	Kid	Hem	Sym	Uri	Spl	Tes	Adr	Thy	n	Empirical	2.5%	97.5%	
4	X	X	X	X								716	**78.7%**	*76.0%*	*81.3%*	5.3%
5	X	X	X	X	X							774	**85.1%**	*82.7%*	*87.3%*	4.6%
	X	X	X	X		X						752	**82.6%**	*80.1%*	*85.0%*	4.9%
	X	X	X	X			X					737	**81.0%**	*78.4%*	*83.5%*	5.1%
	X	X	X	X				X				739	**81.2%**	*78.6%*	*83.7%*	5.1%
	X	X	X	X					X			728	**80.0%**	*77.4%*	*82.5%*	5.2%
	X	X	X	X						X		731	**80.3%**	*77.7%*	*82.8%*	5.2%
	X	X	X	X							X	726	**79.8%**	*77.1%*	*82.3%*	5.2%
6	X	X	X	X	X	X						810	**89.0%**	*86.9%*	*91.0%*	4.1%
	X	X	X	X	X		X					791	**86.9%**	*84.7%*	*89.0%*	4.4%
	X	X	X	X	X			X				787	**86.5%**	*84.2%*	*88.6%*	4.4%
	X	X	X	X	X				X			784	**86.2%**	*83.8%*	*88.3%*	4.5%
	X	X	X	X	X					X		786	**86.4%**	*84.1%*	*88.5%*	4.5%
	X	X	X	X	X						X	783	**86.0%**	*83.7%*	*88.2%*	4.5%
7	X	X	X	X	X	X	X					824	**90.5%**	*88.6%*	*92.4%*	3.8%
	X	X	X	X	X	X		X				823	**90.4%**	*88.5%*	*92.3%*	3.8%
	X	X	X	X	X	X			X			818	**89.9%**	*87.8%*	*91.8%*	3.9%
	X	X	X	X	X	X				X		821	**90.2%**	*88.2%*	*92.1%*	3.9%
	X	X	X	X	X	X					X	818	**89.9%**	*87.8%*	*91.8%*	3.9%
8	X	X	X	X	X	X	X	X				837	**92.0%**	*90.1%*	*93.7%*	3.5%
	X	X	X	X	X	X	X		X			832	**91.4%**	*89.5%*	*93.2%*	3.6%
	X	X	X	X	X	X	X			X		835	**91.8%**	*89.9%*	*93.5%*	3.6%
	X	X	X	X	X	X	X				X	832	**91.4%**	*89.5%*	*93.2%*	3.6%

Che: Cinical chemistry, Liv: Liver, Bod: body weight, Kid: Kidney, Hem: Hematology, Sym: Clinical signs, Uri: Urine analysis, Spl: Spleen, Tes: Testes, Adr: Adrenal gland, Thy: Thyroid gland
n: count of studies

3.6. Reallocation of clinical chemistry parameters and elimination of in vivo specified effects

One open question in human risk assessment is the selection of in vitro and in silico models within integrated approaches for testing and assessment (IATA), which aim to replace preclinical in vivo animal studies. Although the use of these NAMs does not intend to replace the in vivo animal study organ by organ, it would help to a better comprehension of the most frequently observed TOs in preclinical studies and cover their biology by appropriate NAM models.

Some of the above most frequently observed in vivo parameters cannot be directly tested by in vitro models, particularly general signs of toxicity such as body-weight changes or symptomatic observations covered in clinical symptoms. Other parameters, such as changes in enzyme activities frequently observed in clinical chemistry, are secondary to tissue toxicological effects.

Few specific effects remain in clinical chemistry, especially the inhibition of cholesterin esterase, a primary effect of neurotoxic compounds such as organophosphates and carbamates. These observations were reallocated to the primary effect/target organ. With the resultant dataset, the LOEL was predicted for four remaining organs (liver, kidney, hematology, and clinical chemistry) with a probability of 72.1% with about a 6% uncertainty range. This increasing trend continued when the thyroid gland or testes were added as a fifth organ to the list, with 74.3%/74.2% probability, respectively, and a 5.8% uncertainty range. The uncertainty range and the coverage indicate no significant differences when the 6th, 7th, and 8th organ was appended to the target/organ combination (data not shown).

Before reallocating the secondary effect to the primary TO, the most frequently affected target organs (clinical chemistry, liver, kidney, body weight, hematology, and clinical symptoms) predict the LOEL with a probability of 89% and 4% uncertainty.

In comparison, after the reallocation, the prediction power with six main target organs (liver, kidney, hematology, clinical chemistry, and thyroid gland/ testes) changed to 74% and 5.8% uncertainty. The prediction power showed a 15% decrease but is still above 70%, which confirms that the LOEL prediction is probable with a different combination of TOs.

Table 25 Target combinations and the coverage at the LOEL per number of combined TO (N). TO combinations with the highest probability are indicated in orange.

Targets									Coverage	Uncertainty		Range
Liv	Kid	Hem	Chem	Thy	Tes	Adr	Spl	n	Empirical	*2.5%*	*97.5%*	
x	x	x						607	**69.1%**	*66.0%*	*72.1%*	6.1%
x	x	x	x					634	**72.1%**	*69.1%*	*75.0%*	5.9%
x	x	x		x				626	**71.2%**	*68.2%*	*74.2%*	6.0%
x	x	x			x			625	**71.1%**	*68.1%*	*74.1%*	6.0%
x	x	x				x		622	**70.8%**	*67.7%*	*73.7%*	6.0%
x	x	x					x	620	**70.5%**	*67.5%*	*73.5%*	6.0%
x	x	x	x	x				653	**74.3%**	*71.4%*	*77.1%*	5.8%
x	x	x	x		x			652	**74.2%**	*71.2%*	*77.0%*	5.8%
x	x	x	x			x		647	**73.6%**	*70.6%*	*76.5%*	5.8%
x	x	x	x				x	647	**73.6%**	*70.6%*	*76.5%*	5.8%
x	x	x		x	x			643	**73.2%**	*70.2%*	*76.0%*	5.8%
x	x	x		x		x		641	**72.9%**	*69.9%*	*75.8%*	5.9%
x	x	x			x	x		639	**72.7%**	*69.7%*	*75.6%*	5.9%
x	x	x		x			x	638	**72.6%**	*69.6%*	*75.5%*	5.9%
x	x	x			x		x	637	**72.5%**	*69.5%*	*75.4%*	5.9%
x	x	x				x	x	635	**72.2%**	*69.2%*	*75.2%*	5.9%

Che: Clinical chemistry, Kid: Kidney, Hem: Hematology, Liv: Liver, Bod: body weight, Spl: Spleen, Tes: Testes, Adr: Adrenal gland, Thy: Thyroid gland
n: count of studies

Clinical chemistry parameters were reallocated to primary effects/organs like liver and kidney. Clinical chemistry parameters here represent the changes in cholinesterase level due to many organophosphate compound in the DB.

3.7. Prediction of LOEL probability by non-main targets

The reallocated dataset (879 subchronic oral rat studies) was used for LOEL probability estimation using the non-main TOs.

Table 26 outlines the number of studies in which the spleen, testes, thyroid gland, and adrenal gland were reported as a single target organ in studies (1, 2, 2, and 1 time, respectively). In contrast, column B demonstrates the number of studies reporting that the spleen (9 studies), testes (6 studies), thyroid gland (6 studies), and adrenal gland (5 studies) were reported with the four most frequent TOs (liver, kidney, hematology, and clinical chemistry).

The comparison between single effects and column B concludes that the non-main TOs are mainly observed with other organs, and their adverse effects result from the primary TOs' failure.

Column C indicates the number of studies that the four main target organs observed at LOEL, and the non-main TOs are reported at the next higher dose (LOEL+1). The spleen was reported in 17 studies, and the testes, thyroid gland, and adrenal gland

were observed in 19, 8, and 11 studies. This outcome confirms that the non-main TOs are mainly the consequence of the primary targets' adverse outcomes.

This result also proves that non-main TOs did not significantly impact the LOEL probability prediction; when the 7th or 8th organs were appended to the prediction model, the coverage and uncertainty range did not significantly increase.

Table 26 Percentage of studies indicating the coverage of the target in subchronic rat oral studies, Most frequent TOs are colored in blue.

Targets	Fraction (% of studies)			
	LOEL		LOEL+1	
	A	single target	B	C
Liver	43	17		
Kidney	34	12		
Hematology	23	7		
Clinical chemistry	5	3		
Spleen	9	1	5	17
Testes	6	2	1	19
Thyroid gland	6	2	2	8
Adrenal gland	5	1	3	11

A: Percentage of studies
B: Target organs are observed with 4 most frequent targets (liver, kidney, hematology, and clinical chemistry) in the same study
C: Four main targets were reported at LOEL and targets colored in gray were observed at the next higher dose
Total count of studies: 879

3.8. Prediction of adverse effects in preclinical subchronic studies by analysis of adverse effects from shorter-term studies

For time dependency analysis, the database contains 944 subchronic oral rat studies (gavage, administration through drinking water or food) on 644 chemicals and 434 subacute oral rat studies on 278 compounds.

The selected dataset contains experimental data from a minimum of one subacute and one subchronic rat treatment through oral exposure. The dataset was derived from studies on the same chemical with dose overlap, including adverse liver outcomes. The finalized dataset comprises 115 compounds with available LOEL.

As proof of concept, the examination started with measuring the predictive power of the 28-day studies for studies in 90-day containing any effect on the liver, with the same CAS, species, and route of administration in both durations. According to the level of sensitivity and specificity resulting from compared studies in two durations, sensitivity of 75% and specificity of 65% (Table 27, ID 1*), the liver in 28-day can be predictive for 90-day liver effects. In this comparison, the occurrence of different specific effect was not considered; only the same target organs were included in the calculation.

For hazard identification high dose effects such as weight changes in different tissues are of low relevance. Therefore, in the next steps we excluded unspecific high-dose weight changes in 90-day studies for prediction improvement.

Also paired studies with very different dose sections, are unlikely to provide a good relation between short and long-term studies. Therefore, we tested whether excluding paired studies that differ regarding study design can effectively enhance the prediction

Table 27 The comparison between all 28-day liver effects to all 90-day liver effects

ID	Bac[1]	Acc[2]	Sen[3]	Spe[4]	PPV[5]	NPV[6]	N[7]	TP[8]	FP[9]	TN[10]	FN[11]	LR+[12]	LR-[13]
*1**	0.7	0.72	0.75	0.65	0.84	0.52	177	95	18	33	31	2.14	0.38
*2**	0.73	0.75	0.80	0.66	0.81	0.64	177	92	21	41	23	2.63	0.30
*3**	0.78	0.78	0.78	0.78	0.73	0.82	177	60	22	78	17	3.54	0.28
*4**	0.79	0.79	0.79	0.80	0.77	0.82	165	60	18	71	16	3.90	

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] positive predictive values, [6] negative predictive values, [7] Total study number, [8] True positive, [9] False positive, [10] True negative, [11] False negative, [12] Positive likelihood ratio, [13] Negative likelihood ratio.

1-Dose overlap between two durations is considered*

2- Dose overlap between two durations is considered, organ weight changes at the highest dose in 90-day duration were excluded.*

3- Organ weight changes at the overall dose were assigned to be negative (effect was filtered out in both durations, but the studies remained in calculation)*

4- studies that report organ weight changes as single adverse effect were excluded.*

After considering the dose overlap between long and short-term duration, and the exclusion of organ weight changes at the highest dose in 90-day experiments, the count of False Negatives was reduced to 23, with 80% sensitivity and 66% specificity (Table 27, ID 2*). In this analysis, TP included 92 comparisons on the same chemical substances, which comprise both 28-day and 90-day liver effects. The third calculation was performed by filtering the organ weight changes out in both durations. In this step, the weight changes were assigned negative, but the studies containing these high dose effects remained in the calculation. Then the FN showed even more reduction to 17 (Table 27, ID 3*) with acceptable specificity and sensitivity (both 78%).

Often weight changes alone are not considered to be adverse if observed in isolation in human hazard assessment. In the next step, studies that reported organ weight changes as a single adverse effect were excluded from both durations. Implementing this criterion caused a reduction in the number of False Negatives from 31 to 16, and the total number of studies decreased from 177 to 165 (Table 27, ID 4*).

According to Table 27, we understood that the liver effect in the 28-day study could predict the liver effects in 90-day exposure with a sensitivity (Sen) of ≥75% and specificity (Spe) of ≥65%. Implementing the criteria from steps one to four increased the specificity and sensitivity. In the highest range (Sen=79%, Spe=80), the organ weight changes as a single adverse effect were excluded from the dataset.

3.9. The predictive power of individual and grouped 28-day effects for liver in 90-day studies

So far only target organ prediction was investigated. The next analysis adress the question if individual effects in short-term studies can predict specific effects in 90-day study. Again liver is used as an example. Every 25 effects that occurred under the same substance exposure in the subacute period were compared to subchronic liver studies. The comparisons with significant TP and the positive likelihood higher than two are colored in red

Table 28). The 28-day liver effects, with the highest TP (59 comparisons), was weight increase. This means that in 59 studies, when the organ weight increased in 28-day studies, the liver injuries are possible to happen in 90-day studies.

Hypertrophy occurred with 31 TP and 10.06 LR+. Vacuolization with 17 TP and 7.36 LR+ also indicated a good prediction. Discoloration, fatty degeneration, and necrosis with 9, 6, and 6 TP and 11.69, 7.79, and 7.79 LR+ located in the significant range of LR+ (>2). These effects showed small to significant and often conclusive shifts in probability regarding the acceptable positive likelihood ratio domain in Table 16.

Table 28 revealed that comparing 28-day individual effects to 90-day liver outcomes did not show a good predictivity . Therefore, we combined the related effects into grouped categories. Table 30 contains some examples of these comparisons. Seventy-seven comparisons showed weight increased in both durations with 74% and 77% sensitivity and specificity, and a 3.21 positive likelihood ratio. The highest sensitivity and specificity (greater than 90%) belong to these two combinations:

Hypertrophy and vacuolization in 28-day studies compared to changes in organ structure and pigmentation in 90-day treatment.

Hypertrophy, vacuolization, and weight increased in 28-day studies were compared to changes in organ structure and pigmentation in 90-day exposure (Table 30).

Group of vacuolization, hypertrophy, and fatty degeneration were associated with 40 true positive and 6.49 positive likelihood ratios and 52% sensitivity and 92% specificity. The following prediction group (weight increased and decreased) indicated 79% sensitivity and 63% specificity. Combining vacuolization, hypertrophy, and fatty degeneration illustrated low sensitivity (9%) and high specificity (96%).

So far, the subacute combined adverse liver effects compared to the target organ (liver) adverse outcome indicated an acceptable range of LR+ (>2); it is still required

to enhance the prediction therefore we implemented a new condition. To improve the prediction, the model was extended by adding the 90-day individual and grouped adverse liver outcomes to the comparison. Each effect was iterated in a loop to be compared with each effect and the combination of effects in other duration.

Table 28 Predictive power of individual 28-day liver effects for 90-day liver disorders.

Liver Effects (28-day)	Bac[1]	Acc[2]	Sen[3]	Spe[4]	N[5]	TP[6]	FP[7]	TN[8]	FN[9]	LR+[10]	LR-[11]
Apoptosis	0.52	0.58	0.04	1	177	3	0	100	74		0.96
Changes in cellular structures	0.54	0.60	0.09	0.99	177	7	1	99	70	9.09	0.92
Changes in organ structure	0.55	0.60	0.10	0.99	177	8	1	99	69	10.39	0.91
Congestion	0.50	0.56	0.01	0.99	177	1	1	99	76	1.30	1.00
Damage	0.51	0.57	0.01	1	177	1	0	100	76		0.99
Deposits	0.51	0.57	0.01	1	177	1	0	100	76		0.99
Discoloration	0.55	0.61	0.12	0.99	177	9	1	99	68	11.69	0.89
Enlargement	0.53	0.59	0.05	1	177	4	0	100	73		0.95
Eosinophilic structures	0.53	0.59	0.05	1	177	4	0	100	73		0.95
Extramedullary-haematopoiesis	0.51	0.58	0.03	1	177	2	0	100	75		0.97
Fatty degeneration	0.53	0.59	0.08	0.99	177	6	1	99	71	7.79	0.93
Functional disorders	0.51	0.57	0.01	1	177	1	0	100	76		0.99
Glycogen	0.50	0.56	0.03	0.97	177	2	3	97	75	0.87	1.00
Hyperaemia	0.52	0.58	0.04	1	177	3	0	100	74		0.96
Hypertrophy	0.68	0.72	0.40	0.96	177	31	4	96	46	10.06	0.62
Infiltration	0.50	0.56	0.03	0.98	177	2	2	98	75	1.30	0.99
Inflammation	0.49	0.55	0.01	0.96	177	1	4	96	76	0.32	1.03
Karyomegaly	0.51	0.57	0.01	1	177	1	0	100	76		0.99
Necrosis	0.53	0.59	0.08	0.99	177	6	1	99	71	7.79	0.93
Nodules	0.52	0.58	0.04	1	177	3	0	100	74		0.96
Peroxisome-proliferation	0.51	0.57	0.08	0.95	177	6	5	95	71	1.56	0.97
Protein	0.51	0.58	0.03	1	177	2	0	100	75		0.97
Vacuolization	0.60	0.64	0.22	0.97	177	17	3	97	60	7.36	0.80
Weight decreased	0.50	0.55	0.06	0.93	177	5	7	93	72	0.93	1.01
Weight increased	0.73	0.72	0.77	0.69	177	59	31	69	18	2.47	0.34

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] Total study number, [6] True positive, [7] False positive, [8] True negative, [9] False negative, [10] Positive likelihood ratio, [11] Negative likelihood ratio.

Table 30 contains some examples of these comparisons. Seventy-seven comparisons showed weight increased in both durations with 74% and 77% sensitivity and specificity, and a 3.21 positive likelihood ratio. The highest sensitivity and specificity (greater than 90%) belong to these two combinations:

- Hypertrophy and vacuolization in 28-day studies compared to changes in organ structure and pigmentation in 90-day treatment.
- Hypertrophy, vacuolization, and weight increased in 28-day studies were compared to changes in organ structure and pigmentation in 90-day exposure (Table 30)

Table 29 Predictive power of class of 28-day liver effects for 90-day liver disorders.

Combination of Associated 28-day Effects	Bac[1]	Acc[2]	Sen[3]	Spe[4]	N[5]	TP[6]	FP[7]	TN[8]	FN[9]	LR+[10]	LR-[11]
Inflammation+Fibrosis+Necrosis+Infiltration	0.53	0.58	0.09	0.96	177	7	4	96	70	2.27	0.95
Vacuolization+Hypertrophy+Fatty degeneration	0.72	0.75	0.52	0.92	177	40	8	92	37	6.49	0.52
weight increased+Weight decreased	0.71	0.70	0.79	0.63	177	61	37	63	16	2.14	0.33

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] Total study number, [6] True positive, [7] False positive, [8] True negative, [9] False negative, [10] Positive likelihood ratio, [11] Negative likelihood ratio.

Table 30 Results from effects combination, containing the iteration of adverse liver outcome in 28-day and 90-day duration. The best prediction belongs to the colored rows in blue.

Liver effect 28-day	Liver effect 90-day	Bac[1]	Acc[2]	Sen[3]	Spe[4]	N[5]	TP[6]	FP[7]	TN[8]	FN[9]	LR+[10]
Weight increased	Weight increased	0.75	0.75	0.74	0.77	169	77	15	50	27	3.21
Hypertrophy	Hypertrophy	0.77	0.80	0.69	0.84	169	27	21	109	12	4.29
Hypertrophy	Hypertrophy & weight	0.78	0.81	0.72	0.83	169	26	22	111	10	4.37
Hypertrophy & weight in-	Hypertrophy & weight	0.77	0.85	0.64	0.91	169	23	12	121	13	7.08
Vacuolization	Vacuolization	0.85	0.93	0.75	0.95	169	12	7	146	4	16.39
Hypertrophy	Necrosis	0.66	0.73	0.57	0.76	169	12	36	112	9	2.35
Vacuolization	Pigmentation	0.87	0.93	0.79	0.95	169	11	8	147	3	15.22
Hypertrophy & weight in-	Changes in organ struc-	0.77	0.83	0.69	0.84	169	11	24	129	5	4.38
Hypertrophy	Pigmentation	0.77	0.76	0.79	0.76	169	11	37	118	3	3.29
Hypertrophy	Necrosis & weight in-	0.75	0.76	0.73	0.76	169	11	37	117	4	3.05
Hypertrophy	Vacuolization	0.72	0.75	0.69	0.76	169	11	37	116	5	2.84
Hypertrophy & vacuolization	Changes in organ structure & pigmentation	0.94	0.98	0.90	0.98	169	9	3	156	1	47.70
Hypertrophy & vacuolization & weight increased	Changes in organ structure & pigmentation	0.94	0.98	0.90	0.98	169	9	3	156	1	47.70

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] Total study number, [6] True positive, [7] False positive, [8] True negative, [9] False negative, [10] Positive likelihood ratio, [11] Negative likelihood ra-

In the next phase of comparison, the combination of liver effects and clinical liver chemistry parameters, such as liver enzymes alteration (Alanine aminotransferase, Alkaline phosphatase, Gamma-glutamyl transferase, Aspartate aminotransferase, 5'-nucleotidase), and changes in the level of bilirubin, total protein, albumin, and the ratio of albumin to globulin in both durations were added to the loop of calculation.

Table 31 outlines the various comparison forms between short- and long-term durations effects that iterated in a loop. The clinical liver chemistry parameters in 28-day studies predict adverse liver outcomes with 49% Sen and 0.73% Spe and 37 TP. To find the contribution of liver enzyme changes in this prediction, the individual enzyme level

changes in 28-day studies compared to TO adverse effect (liver) (more details in Table 32).

A combination of adverse liver effects and clinical liver chemistry parameters in 28-day and 90-day studies showed an acceptable range of Sen (76%) and Spe (71%) and 2.58 LR+ (Table 31, the row is colored yellow). Following these results, we finalized the comparison with a combination of clinical liver chemistry parameters and liver effects individually and as a combined group for both durations (details in Table 33).

Table 31 Clinical liver chemistry parameters were added to liver effect in both durations.

ID	Bac[1]	Acc[2]	Sen[3]	Spe[4]	PPV[5]	NPV[6]	N[7]	TP[8]	FP[9]	TN[10]	FN[11]	LR+[12]	LR-[13]
*1**	0.61	0.62	0.49	0.73	0.61	0.63	165	37	24	65	39	1.81	0.7
*2**	0.72	0.71	0.83	0.61	0.64	0.81	165	63	35	54	13	1.81	0.7
*3**	0.73	0.74	0.76	0.71	0.83	0.61	165	81	17	41	26	2.58	0.34

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] positive predictive values, [6] negative predictive values, [7] Total study number, [8] True positive, [9] False positive, [10] True negative, [11] False negative, [12] Positive likelihood ratio, [13] Negative likelihood ratio.
Studies that reported only organ weight changes as adverse effect were excluded in both durations.
1 Clinical chemistry parameters in 28-day studies compared to liver studies in 90-day.*
2 Liver effects and clinical liver chemistry parameters in 28-day studies were compared to liver 90-day studies.*
3 Liver effects and clinical liver chemistry parameters in 28-day studies were compared to liver effects and clinical liver chemistry parameters in 90-day studies.*

Table 32 outlines the individual clinical liver chemistry parameters in 28-day studies, which were imported to the comparison to the target organ (liver) in the 90-day duration of exposure. The enzyme level changes were measured individually and in a group of all enzymes.

Alanine aminotransferase demonstrated the highest TP, 19 but low sensitivity, 25%, and 84% specificity between all other enzymes. The best sensitivity and specificity belong to a grouped enzyme level changes, 54%, and 66%, and the LR+ of 1.60, which is lower than our reference range of LR+ (>2). Individual liver enzyme level changes did not provide a perfect prediction result. Therefore, the enzyme group was examined, indicating a better Spe and Sen than the individual enzyme.

In the following analysis to measure the prediction power of short-term liver effects, the associated effects were grouped in one category and iterated in a loop to be compared with the other combination of effects in a longer duration. The true positive (TP) count of long-term and short-term exposure durations greater than six and LR+ ≥ 10 indicated the most significant relationships [55].

Table 32 Individual clinical liver parameters in 28 day were compared to liver studies in 90-day exposure.

28-day Clinical chemistry parameters	Bac[1]	Acc[2]	Sen[3]	Spe[4]	N[5]	TP[6]	FP[7]	TN[8]	FN[9]	LR+[10]
Alanine aminotransferase	0.55	0.57	0.25	0.84	165	19	14	75	57	1.59
Alkaline phosphatase	0.51	0.54	0.14	0.88	165	11	11	78	65	1.17
Aspartate aminotransferase	0.52	0.56	0.08	0.97	165	6	3	86	70	2.34
Gamma glutamyl transferase	0.52	0.55	0.07	0.97	165	5	3	86	71	1.95
Albumin	0.52	0.55	0.13	0.90	165	10	9	80	66	1.30
Albumin/ Globulin	0.52	0.56	0.04	1.00	165	3	0	89	73	
Bilirubin	0.52	0.55	0.12	0.92	165	9	7	82	67	1.51
Globulin	0.52	0.55	0.09	0.94	165	7	5	84	69	1.64
Total protein	0.52	0.55	0.14	0.90	165	11	9	80	65	1.43
Changed enzyme activity*	0.60	0.61	0.54	0.66	165	41	30	59	35	1.60

[1] Balanced accuracy, [2] Accuracy, [3] Sensitivity, [4] Specificity, [5] Total study number, [6] True positive, [7] False positive, [8] True negative, [9] False negative, [10] Positive likelihood ratio, [11] Negative likelihood ratio.

**All enzyme parameters in one category.*

Table 33 contains some examples of prediction parameters of grouped liver effects. The highest TP (N= 86) belongs to the weight changes (category 1) with an LR+ at 2.5.

The reported prediction parameters in Table 33 are visualized in a ROC space plot (Figure 12). Each label indicates a comparison of grouped liver effects. The best prediction power is illustrated with the dots above the diagonal line, which divides the graph into two parts. When Spe and Sen=1, the dots will show the perfect classification. In this plot, all dots locate over the diagonal line, and numbers 16, 17, 18, 20, 22, 23, 29, 30, and 31 determine the perfect classification in the graph.

As the numbers are clarified in Table 33, row 16 includes group 8 (the granulation, fatty degeneration, and vacuolization) and group 1 (weight decreased, and weight increased) in subacute studies that were compared to group 9 (functional disorders, changes in organ structure) and group 7 (hemorrhage, discoloration, pigmentation) in subchronic studies. This comparison indicates 100% Sen and 90% Spe, 12 TP, and 11.2 LR+.

Row 31 indicates groups 7, 4, and 1 (hemorrhage, discoloration, pigmentation+ dilatation, enlargement, karyomegaly, hypertrophy + weight decreased and weight increased) in 28-day studies that were compared to groups 7 and 9 (hemorrhage, discoloration, pigmentation+ functional disorders, changes in organ structure) in the 90-day studies. The result showed 100% Sen and 90% Spe, 11 TP, and 48 LR+.

Table 33 Some examples of prediction power of grouped liver effects in 28 and 90-day studies

*Labels**	28d*	90d**	Accuracy	Sensitivity	Specificity	TP	FP	TN	FN	LR+
1	1	1	0.7	0.7	0.7	86	16	38	29	2.5
2	4	4	0.8	0.8	0.8	30	23	106	10	4.2
3	4	4, 1	0.8	0.7	0.8	29	24	106	10	4.0
4	4	13, 4	0.8	0.8	0.8	18	35	110	6	3.1
5	4	13, 4, 1	0.8	0.8	0.8	18	35	110	6	3.1
6	13, 1	13, 4	0.8	0.7	0.8	17	25	120	7	4.1
7	13, 1	13, 4, 1	0.8	0.7	0.8	17	25	120	7	4.1
8	4	4, 2	0.8	0.8	0.8	17	36	111	5	3.2
9	4	4, 1, 2	0.8	0.8	0.8	17	36	111	5	3.2
10	4	7	0.7	0.7	0.8	17	36	109	7	2.9
11	13, 1	4, 2	0.8	0.7	0.8	16	26	121	6	4.1
12	13, 1	4, 1, 2	0.8	0.7	0.8	16	26	121	6	4.1
13	4	13, 4, 2	0.7	0.8	0.7	13	40	112	4	2.9
14	4	13, 4, 1, 2	0.7	0.8	0.7	13	40	112	4	2.9
15	4	3, 13	0.7	0.7	0.7	13	40	111	5	2.7
16	8, 1	7, 9	0.9	1.0	0.9	12	14	143	0	11.2
17	8, 1	13, 8	0.9	0.8	0.9	12	14	140	3	8.8
18	8	7, 9	0.9	1.0	0.9	12	18	139	0	8.7
19	8, 1	9	0.9	0.7	0.9	12	14	138	5	7.7
20	8	13, 8	0.9	0.8	0.9	12	18	136	3	6.8
21	6	8	0.9	0.6	1.0	12	6	143	8	14.9
22	8, 1	7, 9	0.9	1.0	0.9	12	14	143	0	11.2
23	7, 8, 1	7, 9	1.0	0.9	1.0	11	1	156	1	143.9
24	7, 8, 1	9	1.0	0.6	1.0	11	1	151	6	98.4
25	4, 8	8	0.9	0.6	1.0	11	1	148	9	82.0
26	4, 8, 1	8	0.9	0.6	1.0	11	1	148	9	82.0
27	7, 8, 1	7, 1	0.9	0.5	1.0	11	1	146	11	73.5
28	7, 8, 1	7	0.9	0.5	1.0	11	1	144	13	66.5
29	7, 4	7, 9	1.0	0.9	1.0	11	3	154	1	48.0
30	7, 8	7, 9	1.0	0.9	1.0	11	3	154	1	48.0
31	7, 4, 1	7, 9	1.0	0.9	1.0	11	3	154	1	48.0
32	13, 8	4, 2	0.9	0.5	1.0	11	2	145	11	36.8
33	13, 8, 1	4, 2	0.9	0.5	1.0	11	2	145	11	36.8
34	13, 8	4, 1, 2	0.9	0.5	1.0	11	2	145	11	36.8
35	13, 8, 1	4, 1, 2	0.9	0.5	1.0	11	2	145	11	36.8
36	7, 4	9	0.9	0.6	1.0	11	3	149	6	32.8
37	7, 8	9	0.9	0.6	1.0	11	3	149	6	32.8
38	7, 4, 1	9	0.9	0.6	1.0	11	3	149	6	32.8

**Labels: indication code for any comparison in Figure 12*
***28-day liver effects*
****90-day liver effects*
1 Weight decreased and weight increased
2 Albumin, globulin, albumin/globulin, protein, total protein
4 Dilatation, enlargement, karyomegaly, hypertrophy
6 Infiltration, inflammation, hyperemia, foci, macrophages, nodules, eosinophilic structures, congestion
7 Hemorrhage, discoloration, pigmentation
8 Granulation, fatty degeneration, vacuolization
9 Functional disorders, changes in organ structure
13 Changed enzyme activity

The prediction parameters shown in Table 33 are visualized in a ROC space plot (Figure 12). Each comparison of grouped liver effects is outlined with a number. The prediction model is successful when the dots are located above the diagonal line. The Spe and Sen=1 indicate the perfect classification of the graph. In the figure 12 plot, all predictions are located over the diagonal line, and numbers 16, 17, 18, 20, 22, 23, 29, 30, and 31 can define the perfect prediction classification. The numbers are described in Table 33.

Figure 13 visualizes the distribution of LR+ in the dataset with TP> 6. The blue horizontal line indicates the LR+= 10. The graph shows that the LR+ values with the minimum level of 0.152, median of 23.31, maximum value of 168, first quartile of 7.18 and the third quartile of 61.87 are mainly located over the line of LR+= 10.

Table 16 Subjective interpretation of positive likelihood ratios [55, 56]. interprets the LR+ value. The LR+>10 means large and often conclusive shifts in probability. Therefore, this prediction model regarding other parameters, such as Sen and Spe, showed acceptable values.

The predictions possessing LR+≥ 10 and TP≥ 6 were imported to the ROC plot (Figure 14) to illustrate the range and acceptability of the predictions. Figure 14 indicates that all 2207 comparison possessing LR+≥10 and TP≥ 6 located over the diagonal line of the diagram. these 2207 comparisons are the perfect result of this work regarding the level of sensitivity, specificity, LR+ and count of TP.

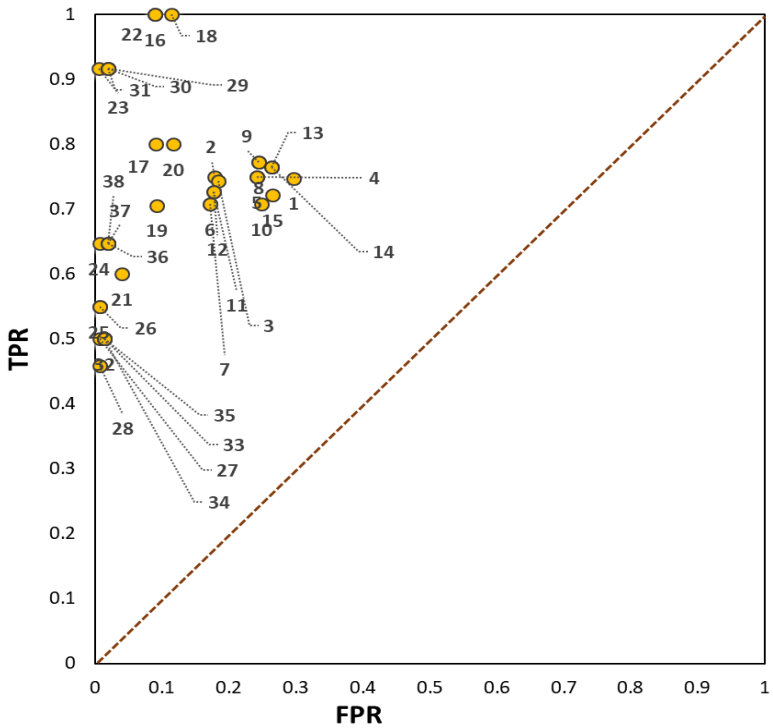

Figure 12, 38 best prediction result of grouped liver effect iteration visualized by ROC space graph,
TPR: true positive rate or sensitivity, FPR: false positive rate or 1-specificity

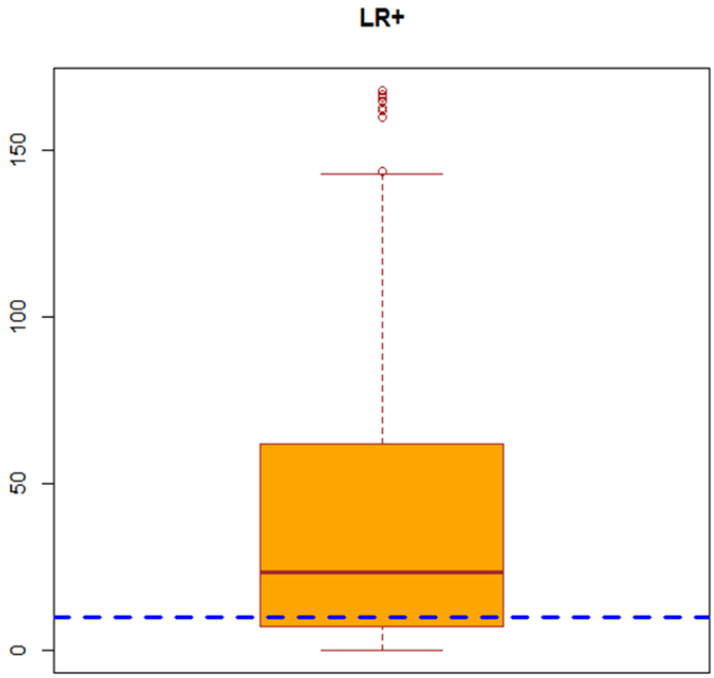

Figure 13 Box plot visualizing the positive likelihood ratio for the predictions possessing TP>6 The blue line outlines the LR+ =10

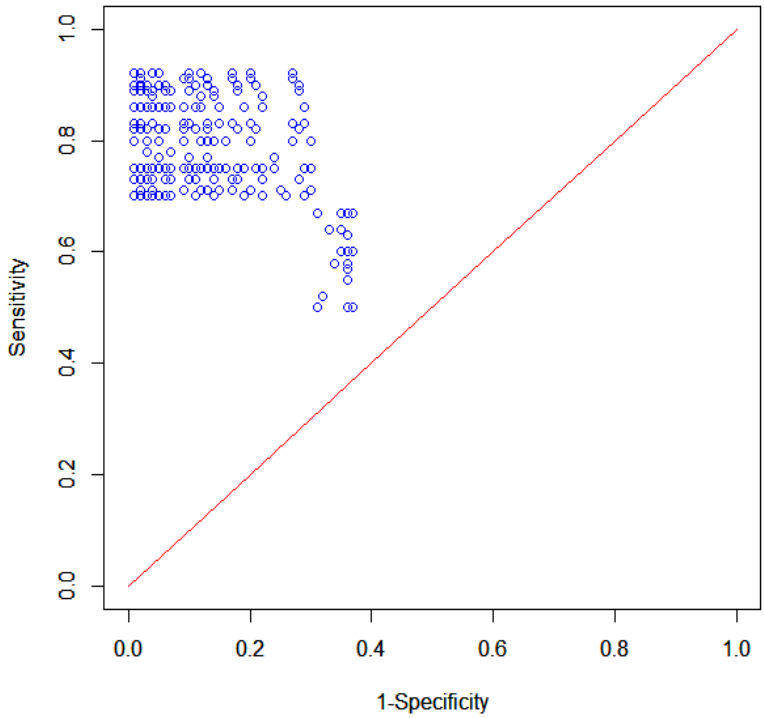

Figure 14 ROC space graph, true positive rate or sensitivity (TPR), false positive rate or 1-specificity (FPR)
2207 data with TP>6 were imported to the ROC Space graph

This experiment confirmed that the histopathological findings, organ weight changes, and necropsy of the liver in subacute treatment revealed a good predictivity for adverse liver effects in subchronic treatment.

The exclusion of studies with organ weight changes as a single effect of the study in both duration improves the prediction LR+ changed to 3.54 after filtering the effect (organ weight changes) out and 3.90 after exclusion of the studies with only organ weight changes as single effect (Table 27, ID 3 and 4).

Clinical liver chemistry parameters from subacute treatment alone could not predict the liver effects observed in subchronic treatment.

In short-term studies, individual and combined liver effects plus clinical liver parameters compared to long-term effects were predictable for the same criteria of the 90-day studies. This combination of liver toxicity data and clinical liver chemistry parameters in subacute and subchronic treatments shows the highest count of true positive and sensitivity values among investigated conditions.

ROC space plot as a visualizing tool enabled the presentation of the prediction model in an intelligible, straightforward, and graphical form. The ROC space graph confirms that the condition of TP>6 and LR+>10 for selecting the proper prediction is correct. Those selections are all located over the diagonal line based on the ROC graph characteristic.

4. Summary and Outlook

The main goal of this investigation was to develop a strategy to define preclinical studies' capabilities in guiding the testing scope of NAM-based testing. To achieve this goal, we analyzed the target organs that mainly triggered the LOEL in legacy animal data for different routes of exposure. The second topic was to analyze the predictivity of observed target organ/apical findings from short-term to long-term studies.

A comparison of structural and Physico-chemical properties shows that drugs and organic chemicals cover complementary structural spaces. It is also noted that chemicals have, on average, a lower molecular weight compared to drugs.

The target organs that often trigger the LOEL in legacy animal data with oral and inhalation exposure in RepDose and eTOX DB were similar (body weight, liver, clinical chemistry, clinical signs, hematology, and kidney). In RepDose DB, body weight change is shown as the most frequently affected TOs at the overall doses. While in drug studies, it is located in the mid of the frequency table, it is observed to be 25% less than the chemical (77% in chemicals and 52% in drugs). Body weight is a sensitive parameter in health care and drug development [62]; this was proved in our analysis of repeated-dose oral and inhalation toxicity studies based on RepDose DB. The difference in observing the body weight changes in RepDose and eTOX DB could be due to the sensitivity of body weight change as a clear sign of toxic effects.

In RepDose, the frequency of affected targets/organs did not differ significantly between all oral routes of administration (diet/drinking water or gavage). Atcha, Rourke, et al. proved that the alternatives for gavage administration provide similar results in the treatment. Simultaneously reduce the distress and morbidity related to standard gavage [63].

As expected, inhalation studies show that in addition to these main TO, the target organs of first contact are the respiratory tract. These findings confirm previous investigations with a smaller dataset [51].

Some recent studies with a much smaller amount of repeated dose studies came to similar results as in our studies. S. Horner et al. determined that the most frequently affected organs in toxicity studies for 77 AstraZeneca candidate drugs in the rodent was the liver, followed by adrenal glands, kidney, spleen, bone marrow, and thymus [64]. Although in our investigation, the adrenal gland, spleen, and thymus were observed in more than 10% of drug studies, the different sequences of these observations could be due to different sizes of databases. Therefore, the more minor observations may vary in some details.

Emma Gustafson and colleagues defined a complete screening of the oral repeated dose toxicity data in safety assessment reports of cosmetic ingredients. The data were extracted from the oral repeated dose toxicity experiments of 90-day studies for 79 cosmetic compounds provided by the Scientific Committee on Consumer Safety between 2009 and 2019. They identified that the potentially most frequently affected

organs in the oral administration of cosmetic ingredients to animals were the liver and the hematological system. The liver was reported as the most frequently affected organ in 20 RDT studies of 79 available studies on cosmetic ingredients [65].

The kidney was reported as the third most affected organ, with lower occurrence than the liver and the hematological system. The histopathological changes were the observed effects in most cases [65].

Martin, Dix, et al., in 2010, profiled in vivo toxicities across 310 chemicals as a model application of ToxRef DB, meeting the need for detailed anchoring endpoints for developing ToxCast predictive signatures. Using query and structured data-mining approaches, they generated toxicity profiles from ToxRef DB based on long-term rodent bioassays. These chronic/cancer data were analyzed for suitability as anchoring endpoints based on incidence, target organ, severity, potency, and significance. Under conditions of the bioassays, the pathology observation for 273 of 310 chemicals, with > 90% occurring in the liver, kidney, thyroid, lung, testis, and spleen [66].

We determined the differences in potency between targets/organs by applying the minimum organ-LOEL values of each compound, the lowest exposure dose in which the specific target organ was most frequently affected. The significant differences between the minimum LOEL and cumulative frequency of target organs illustrated the more and less sensitive target organs. Clinical chemistry parameters showed the highest sensitivity and sperm parameters was reported as the less sensitive organ.

The distribution of organ-LOELs in subchronic oral rat studies was visualized using quantiles and percentiles. The quantiles and percentiles are considered sample estimates of a population parameter; thus, they must be presented with a confidence interval [60]. Overlapping the confidence interval levels (C.I.95%) for the 5th and 50th percentiles of organ-LOEL confirm that clinical chemistry parameters are the most sensitive target organ among all other tested TOs.

The sensitivity of clinical chemistry parameters means that clinical chemistry endpoints in blood samples and biochemical analyses of plasma can detect organ toxicity and could be used as an early indicator of onset and progression of liver and kidney toxicological outcomes.

Although the testis is sensitive to numerous toxicants, including pharmaceuticals and industrial chemicals, our experiment shows less sensitivity for sperm parameters than other target organs. This result does not disregard the importance of infertility and adverse health outcomes but demonstrates that the adverse effects will occur in higher doses.

We further defined a testing strategy by developing the coverage model to assess the high probability of LOEL with low uncertainty. The coverage concentrated on the most frequently affected targets/organs at LOEL. The most frequently affected target organs at the LOEL in 910 repeated-dose oral subchronic toxicity studies based on RepDose DB were applied to predict the LOEL reliability. Examination of the six main

target organs (liver, kidney, body weight, clinical chemistry, hematology, and clinical signs) illustrated the most predictive combination for LOEL in the whole dataset. This combination of targets/organs showed coverage with 89% probability and 4% uncertainty. This finding is in line with the examination in 2013 by Batke, Aldenberg, et al. based on a smaller dataset from the RepDose DB on subchronic oral exposure studies. The same six main target organs were identified, and a probability of 86% was derived to detect the LOEL [51].

In our investigation, the spleen was the next most frequent organ observed after the liver, kidney, clinical signs, and body weight changes and was added to the list of relevant target organs. It was observed in 17% of studies at LOEL and was reported as a single organ in 1% of studies at LOEL.

Batke, Aldenberg et al. reported that the spleen was not frequently affected and did not increase the detection fraction significantly compared to the six relevant organs; it occurred as a single organ in 2.3% of the studies in their investigation [51]. In our experiment, in 16% of studies, the adrenal gland was observed at LOEL, which was reported in only 1% of studies as a single target organ, and 2% of studies reported the thyroid gland as a single affected organ at LOEL.

Histopathology of the testes is a sensitive endpoint in toxicological studies [67]. In limited analyses of Batke, Aldenberg, et al., testes did not show up as a single target organ, which shows that toxicity to the testes is most often associated with other toxicological effects. In our calculation, testes were observed in 10% of the studies, and 2% showed testes as a single affected organ at LOEL. These data showed that when the sixth, seventh, and eighth organs were appended to the target/organ combination, the uncertainty range and the coverage indicated no significant differences.

Calculation of LOEL probability in the coverage model indicated that if the examination concentrates on the most frequently affected target organs (liver, kidney, hematology, and clinical chemistry), the LOEL prediction power is higher, with the lowest uncertainty value.

One open question in human risk assessment is the selection of in vitro and in silico models within integrated approaches for testing and assessment (IATA). Although the use of these NAMs does not intend to replace the in vivo animal study organ by organ, it would help to a better comprehension of the most frequently observed TOs in preclinical studies. The clinical chemistry data are an essential component that, together with the gross pathology, organ weights, and histopathology data analysis, allow the detection of specific organ toxicity with a high degree of certainty [68]. Association between clinical chemistry parameters and the primary effect/target organ made us reallocate these observations to the primary effect/target organ, in most cases, the liver or kidney. Besides the reallocation of clinical chemistry parameters, some of the most frequently observed in vivo parameters, which cannot be directly tested by in vitro models, were not considered in LOEL prediction. The general signs of toxicity,

such as body weight changes or symptomatic observations covered in clinical symptoms, were considered the in vivo specific outcomes.

After reallocating clinical chemistry, and hematological parameters, the coverage model and the LOEL probability calculation were performed with targets/organs for which in vitro models can be generated.

The LOEL was predicted with 72% coverage for four main targets, liver, kidney, hematology, and clinical chemistry, with 5.9% uncertainty. This is already a remarkably high coverage taking into account that only four targets out of 8 examined targets contribute to this probability, and these findings may serve as an indication that in vitro models that cover the biological processes of these four targets should be part of a testing battery addressing systemic toxicity after repeated exposure. The remaining uncertainty of 28% will have to be covered by other models, e.g., indicating general toxicity as a surrogate for other often observed effects such as changes in body weight.

The investigation resumed by comparing the short- and long-term rat oral exposure studies with the primary aim of recent European legislation to reduce and minimize animal testing.

In general, long-term animal studies will be used in risk assessment to evaluate the risk of a given chemical substance for long-term human exposure. Since long-term animal studies are expensive and a significant number of animals are needed, only short-term studies with subacute to subchronic exposure periods are accessible for many substances.

In the interest of reducing de novo animal testing, we explore the relationship between short-term effects (subacute treatment) and adverse effects in longer-term studies (subchronic treatment).

We conducted a statistical analysis using the RepDose database. The dataset comprises 944 subchronic RDT studies on 644 chemical substances and 434 subacute studies on 343 chemical substances. The analysis was followed by the dataset possessing experimental data from a minimum of one subacute and one subchronic rat treatment through oral exposure.

In our experiment, the true positive (TP) count of long-term and short-term exposure durations greater than six and LR+ ≥ 10 indicated the most significant relationships.

In this work, the specificity calculated for liver combined effects for long and short-term comparison is greater than 75%, which is generally lower than reported specificity in the Olson study. For example, Olson reported 90% sensitivity for liver disorders for the animal to human prediction [69].

In the prediction model provided by Mathew Clark in 2018, the concordance of the toxicity of pharmaceuticals in animals and humans mentioned a low sensitivity of the liver, which is about 16%. The main reason for differences between these values can

come from different datasets. Clark used PharmaPendium, covering marketed drugs that passed strict clinical safety rules [55].

Table 27 indicates that when the studies possessing organ weight changes as a single effect were excluded from the calculation, the sensitivity changed from 75% to 79%, and specificity showed a 15% increase up to 80%. This may prove that weight changes are a sensitive parameter in healthcare. Weight changes in longer duration and at higher doses are more probable than shorter duration, and most subchronic studies will reveal organ weight changes. The exclusion of organ weight changes will reduce the count of false negatives, and consequently, it will increase the true positive rate.

In the inclusion of all sub-categories of adverse liver effects in 28-day studies (

Table *28*), weight increase and hypertrophy showed better prediction parameters than the majority of effects. Therefore, some effects were combined into one category to increase the prediction probability. In addition, the association of liver effects in one study would have been considered for the following analysis.

In another study, Clark and colleagues [70] calculated the likelihood ratio of individual major terms of liver-related clinical observation given by nonclinical observation. The sensitivity and specificity of each effect were comparable to the liver effect prediction in our examination.

Blood bilirubin increase, aspartate aminotransferase, alanine aminotransferase increase, blood alkaline phosphatase increase, hepatic necrosis, hepatocellular damage, and liver disorders were considered in Clark's experiment. The similar effects in our investigation did not show good predictivity. Low sensitivity and high specificity were due to the low number of TP. The weight increased with higher TP; 59 studies out of 177 indicated a Sen and Spe of 70%. A high count of FN and TN reduces the sensitivity; therefore, the LR+ presents a low prediction power.

Unlike most liver observations, liver enzyme level changes, including the markers alanine aminotransferase (ALAT), and aspartate aminotransferase (ASP), showed a narrow range of LR+. The reason could be that a slight enzyme level elevation indicates liver injury in a shorter duration, revealing a prominent stage of damage after a longer period of exposure. Thus, a high count of FN will reduce the sensitivity and, consequently, lower LR+ because of a lack of enzyme changes at LOEL and more liver damage in 90-day studies.

As Chen Zhang mentioned in their in silico prediction of drug-induced liver toxicity model, the specificity values of the hepatotoxicity prediction were lower than Sen values, which will cause a higher prediction accuracy for hepatic toxicants. The reason might be that the dataset contains more DILI (drug-induced liver injury)-positives compounds than DILI-negative.

The other reason could be that hepatotoxicity is one of the most complicated toxicity endpoints, and there is a deficiency of specific criteria for classifying DILI-positive and negative compounds [71].

In our work, the sensitivity was lower than the specificty value in some cases, possibly because of a higher amount of FN than TP. It means that adverse liver effects occurred in 90-day but not in 28-day. It is clear that in longer duration, many effects will appear at LOEL while similar effects at 28-day may happen at LOEL+1 or higher doses.

Combination and iteration of the class of liver effects in Table 33 revealed good predictivity, a high TP, and LR+. It showed that the prediction methodology functioned perfectly with high Sen and Spe. Besides high TP for the comparisons like rows 1, 2, 3, and 25, that the similar effects occurred in both durations, finding the relation between non-similar liver effects can prove the prediction model. As an example, row 21 can clarify this point. In row 21, category 8 comprises the combination of granulation, fatty degeneration, and vacuolization in a 90-day study and infiltration, inflammation, hyperemia, foci, macrophages, nodules, eosinophilic structures, and congestion (group 6) in 28-day studies. K. Yasuhara and colleagues, in their investigation, confirmed the association between karyomegaly and inflammation of hepatic cells. They explained that karyomegalic epithelial cells were often observed when the inflammatory changes were prominent in the liver as a tissue response to a significant injury by a toxic agent [72].

Due to complex liver function, uncovering the connection between different toxic outcomes require expert knowledge of pathology and histology. Although the prediction may not find a relevant connection between all groups of effects, we are persuaded that the methodology of the prediction model is working accurately and could be implemented for further toxicological data, such as pharmaceutical compounds in different target organs.

These short- to long-term aspects are not yet explored for drugs and still need to be examined. The perspective of these analyses is to improve the new approach methodology, such as developing alternative in vitro methods by targets and outcomes identified by the coverage and prediction models.

Our analysis was a primary step in getting insight into in vivo RDT outcomes. The coverage can be used as a systemic tool to prove and maintain the expert-based validation of human risk assessment.

We also highlighted the concordances with the highest numbers of TP and the positive likelihood of liver effects. Our results answer the questions posed in the introduction; do liver effects in subacute studies predict liver-effect in subchronic studies?

This experiment confirmed that the histopathological findings, organ weight changes, and necropsy of the liver in subacute treatment revealed a great predictive power for adverse liver effects in subchronic treatment.

This study has created a metric for event-duration pair that can be used to predict other durations and preclinical- clinical comparison to improve human risk assessment methodologies.

5. List of Figures

6. List of Tables

7. References

1. Klaassen, C.D., *Casarett and Doull's toxicology: the basic science of poisons*. Vol. 1236. 2013: McGraw-Hill New York.
2. Lee, B.-M., S. Kacew, and H.S. Kim, *Lu's basic toxicology: fundamentals, target organs, and risk assessment*. 2017: CRC press.
3. Gantenbein, U.L., *Poison and its dose: Paracelsus on toxicology*, in *Toxicology in the Middle Ages and Renaissance*. 2017, Elsevier. p. 1-10.
4. Langley, G., et al., *Lessons from toxicology: developing a 21st-century paradigm for medical research.* Environmental health perspectives, 2015. **123**(11): p. A268-A272.
5. Greim, H., *Aims and mission of regulatory toxicology.* Regulatory Toxicology, 2021: p. 3-22.
6. Schmid, O. and F.R. Cassee, *On the pivotal role of dose for particle toxicology and risk assessment: exposure is a poor surrogate for delivered dose*. 2017, BioMed Central. p. 1-5.
7. Pizzo, F. and E. Benfenati, *In silico models for repeated-dose toxicity (RDT): prediction of the no observed adverse effect level (NOAEL) and lowest observed adverse effect level (LOAEL) for drugs*, in *In Silico Methods for Predicting Drug Toxicity*. 2016, Springer. p. 163-176.
8. Henkler, F., et al., *Risk assessment of nanomaterials in cosmetics: a European union perspective.* Archives of toxicology, 2012. **86**(11): p. 1641-1646.
9. Aulmann, W., et al., *REACH and CLP. Its Role in Regulatory Toxicology.* Regulatory Toxicology, 2020: p. 1-19.
10. Schaafsma, G., et al., *REACH, non-testing approaches and the urgent need for a change in mind set.* Regulatory Toxicology and Pharmacology, 2009. **53**(1): p. 70-80.
11. Spielmann, H., U.G. Sauer, and O. Mekenyan, *A critical evaluation of the 2011 ECHA reports on compliance with the REACH and CLP regulations and on the use of alternatives to testing on animals for compliance with the REACH regulation.* Alternatives to laboratory animals, 2011. **39**(5): p. 481-493.
12. Klimisch, H.-J., M. Andreae, and U. Tillmann, *A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data.* Regulatory toxicology and pharmacology, 1997. **25**(1): p. 1-5.
13. Haighton, L., et al., *Systematic review and evaluation of aspartame carcinogenicity bioassays using quality criteria.* Regulatory Toxicology and Pharmacology, 2019. **103**: p. 332-344.
14. Combes, R., M. Barratt, and M. Balls, *An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system.* Alternatives to Laboratory Animals, 2003. **31**(1): p. 7-19.
15. ECHA, *Guidance on the Application of the CLP Criteria; Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Version 5.0, July 2017. Reference: ECHA-17-G-21-EN; ISBN: 978-92-9020-050-5.* 2017.
16. Foth, H. and A. Hayes, *Concept of REACH and impact on evaluation of chemicals.* Human & experimental toxicology, 2008. **27**(1): p. 5-21.
17. Jagadeeswaran, I. and H. Sriram, *EU 1907/2006–Registration, Evaluation, Authorisation and Restriction of Chemicals*, in *Medical Device Guidelines and Regulations Handbook*. 2022, Springer. p. 237-260.
18. Ball, N., et al., *A framework for chemical safety assessment incorporating new approach methodologies within REACH.* Archives of Toxicology, 2022. **96**(3): p. 743-766.
19. Pastoor, T.P., et al., *A 21st century roadmap for human health risk assessment.* Critical reviews in toxicology, 2014. **44**(sup3): p. 1-5.

20. OECD, *Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment*. 2017.
21. Webster, F., I.B. Lambert, and C.L. Yauk, *Adverse Outcome Pathway on Cyp2E1 activation leading to liver cancer.* 2021.
22. Bitsch, A., et al., *REPDOSE: A database on repeated dose toxicity studies of commercial chemicals--A multifunctional tool.* Regul Toxicol Pharmacol, 2006. **46**(3): p. 202-10.
23. Bristol, D.W., J.T. Wachsman, and A. Greenwell, *The NIEHS Predictive-Toxicology Evaluation Project.* Environmental Health Perspectives, 1996. **104**(suppl 5): p. 1001-1010.
24. Eriksson, L., et al., *Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs.* Environmental health perspectives, 2003. **111**(10): p. 1361-1375.
25. Handbook, R.C., *Science Policy Council US Environmental Protection Agency, Washing ton, DC 20460*. 2000, EPA 100-B-00-002, December.
26. Hansen, C.D., *Danish EPA Use of Models for Assessment of Pesticides Mobility*, in *Modelling of Environmental Chemical Exposure and Risk*. 2001, Springer. p. 183-192.
27. Cronin, M.T., et al., *Use of QSARs in international decision-making frameworks to predict health effects of chemical substances.* Environmental health perspectives, 2003. **111**(10): p. 1391-1401.
28. Lewis, D., C. Ioannides, and D. Parke, *COMPACT and molecular structure in toxicity assessment: a prospective evaluation of 30 chemicals currently being tested for rodent carcinogenicity by the NCI/NTP.* Environmental Health Perspectives, 1996. **104**(suppl 5): p. 1011-1016.
29. Mack, C.M., et al., *Burst and principal components analyses of MEA data for 16 chemicals describe at least three effects classes.* Neurotoxicology, 2014. **40**: p. 75-85.
30. Pölloth, C. and I. Mangelsdof, *Commentary on the application of (Q) SAR to the toxicological evaluation of existing chemicals.* Chemosphere, 1997. **35**(11): p. 2525-2542.
31. Richard, A.M., *Commercial toxicology prediction systems: a regulatory perspective.* Toxicology letters, 1998. **102**: p. 611-616.
32. Greim, H., et al., *Assessment of structurally related chemicals: toxicity and ecotoxicity of acrylic acid and acrylic acid alkyl esters (acrylates), methacrylic acid and methacrylic acid alkyl esters (methacrylates).* Chemosphere, 1995. **31**(2): p. 2637-2659.
33. Greim, H., et al., *Toxicity and ecotoxicity of sulfonic acids: structure-activity relationship.* Chemosphere, 1994. **28**(12): p. 2203-2236.
34. Group, E.W., *The toxicology of glycol ethers and its relevance to man.* 1995.
35. Raies, A.B. and V.B. Bajic, *In silico toxicology: computational methods for the prediction of chemical toxicity.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2016. **6**(2): p. 147-172.
36. Thomas, R.S., et al., *The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency.* Toxicol Sci, 2019. **169**(2): p. 317-332.
37. OECD, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies*. 2002.
38. Schwab, C.H., et al., *eTOXsys–An Integrated Platform to Profile And Data Mine Across Multiple Databases and Prediction Models.*
39. Hall, R.L., *Clinical pathology of laboratory animals.* Animal models in toxicology, 2007. **2**: p. 787-830.
40. Ike, C., et al., *Liver Enzymes and Total Protein Levels as Index of Hepatotoxicity of Naphthalene.* OSR Journal of Pharmacy and Biological Sciences, 2016. **11**(2): p. 28-33.
41. Beamer, N., et al., *Fibrinogen and the albumin-globulin ratio in recurrent stroke.* Stroke, 1993. **24**(8): p. 1133-1139.

42. Littlewood, T. and F. Mandelli. *The effects of anemia in hematologic malignancies: more than a symptom*. in *Seminars in oncology*. 2002. Elsevier.
43. Weingand, K., et al., *Harmonization of animal clinical pathology testing in toxicity and safety studies.* Fundamental and Applied Toxicology, 1996. **29**(2): p. 198-201.
44. Birare, S.D., N. Jain, and R.S. Abhange, *A study of hematological indices and bone marrow morphology in patients presenting with anemia.* Indian Journal of Pathology and Oncology, 2019. **6**(2): p. 266-274.
45. Wintrobe, M., *A simple and accurate hematocrit.* The Journal of Laboratory and Clinical Medicine, 1929. **15**(3): p. 287-289.
46. Christensen, R., et al., *The erythrocyte indices of neonates, defined using data from over 12 000 patients in a multihospital health care system.* Journal of Perinatology, 2008. **28**(1): p. 24-28.
47. Monzon, C.M., B.D. Beaver, and T.D. Dillon, *Evaluation of Erythrocyte Disorders With Mean Corpuscular Volume (MCV) and Red Cell Distribution Width (RDW).* Clinical Pediatrics, 1987. **26**(12): p. 632-638.
48. Smith, G., *Applied clinical pathology in preclinical toxicology testing.* Handbook of toxicologic pathology, 2002: p. 123-156.
49. Aly, S.S., et al., *Assessment of peripheral blood lymphocyte subsets in children with iron deficiency anemia.* BMC pediatrics, 2018. **18**(1): p. 1-6.
50. Freitas, A.A. and B. Rocha, *Population biology of lymphocytes: the flight for survival.* Annual review of immunology, 2000. **18**(1): p. 83-111.
51. Batke, M., et al., *Relevance of non-guideline studies for risk assessment: The coverage model based on most frequent targets in repeated dose toxicity studies.* Toxicology Letters, 2013. **218**(3): p. 293-298.
52. Nick, T.G., *Descriptive statistics.* Topics in biostatistics, 2007: p. 33-52.
53. Schoonjans, F., D. De Bacquer, and P. Schmid, *Estimation of population percentiles.* Epidemiology (Cambridge, Mass.), 2011. **22**(5): p. 750.
54. Thompson, C.V., et al., *A systematic review of published physiologically-based kinetic models and an assessment of their chemical space coverage.* Alternatives to Laboratory Animals, 2021. **49**(5): p. 197-208.
55. Clark, M. and T. Steger-Hartmann, *A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans.* Regulatory Toxicology and Pharmacology, 2018. **96**: p. 94-105.
56. Chien, P.F. and K.S. Khan, *Evaluation of a clinical test. II: Assessment of validity.* BJOG: An International Journal of Obstetrics & Gynaecology, 2001. **108**(6): p. 568-572.
57. Machin, D., et al., *Statistics with confidence: confidence intervals and statistical guidelines*. 2013: John Wiley & Sons.
58. Flach, P.A. *The geometry of ROC space: understanding machine learning metrics through ROC isometrics*. in *Proceedings of the 20th international conference on machine learning (ICML-03)*. 2003.
59. Escher, S., et al., *Evaluation of inhalation TTC values with the database RepDose.* Regulatory Toxicology and Pharmacology, 2010. **58**(2): p. 259-274.
60. Ialongo, C., *Confidence interval for quantiles and percentiles.* Biochemia medica, 2019. **29**(1): p. 5-17.
61. Batke, M., et al., *Relevance of non-guideline studies for risk assessment: the coverage model based on most frequent targets in repeated dose toxicity studies.* Toxicol Lett, 2013. **218**(3): p. 293-8.
62. Holemans, K., L. Aerts, and F.A. Van Assche, *Fetal growth restriction and consequences for the offspring in animal models.* Journal of the Society for Gynecologic Investigation, 2003. **10**(7): p. 392-399.

63. Atcha, Z., et al., *Alternative method of oral dosing for rats.* Journal of the American Association for Laboratory Animal Science, 2010. **49**(3): p. 335-343.
64. Horner, S., et al., *Target organ toxicities in studies conducted to support first time in man dosing: an analysis across species and therapy areas.* Regulatory Toxicology and Pharmacology, 2013. **65**(3): p. 334-343.
65. Gustafson, E., et al., *Screening of repeated dose toxicity data in safety evaluation reports of cosmetic ingredients issued by the Scientific Committee on Consumer Safety between 2009 and 2019.* Archives of Toxicology, 2020. **94**(11): p. 3723-3735.
66. Martin, M.T., et al., *Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program.* Chemical research in toxicology, 2010. **23**(3): p. 578-590.
67. Mangelsdorf, I., J. Buschmann, and B. Orthen, *Some aspects relating to the evaluation of the effects of chemicals on male fertility.* Regulatory Toxicology and Pharmacology, 2003. **37**(3): p. 356-369.
68. Kaufmann, W. and M.C. Jacobsen, *Examination of Organ Toxicity.* Regulatory Toxicology, 2021: p. 117-127.
69. Olson, H., et al., *Concordance of the toxicity of pharmaceuticals in humans and in animals.* Regulatory Toxicology and Pharmacology, 2000. **32**(1): p. 56-67.
70. Clark, M., *Prediction of clinical risks by analysis of preclinical and clinical adverse events.* Journal of Biomedical Informatics, 2015. **54**: p. 167-173.
71. Zhang, C., et al., *In silico prediction of drug induced liver toxicity using substructure pattern recognition method.* Molecular informatics, 2016. **35**(3-4): p. 136-144.
72. Yasuhara, K., et al., *Ultrastructure and cell proliferative activities of karyomegalic alveolar epithelial cells in early pulmonary inflammatory lesions of Syrian golden hamsters induced by N-methyl-N-nitrosourethane.* J Vet Med Sci, 1996. **58**(9): p. 825-31.

8. List of Publications

List of Publication in the context of the PhD

Batke, M., **Afrapoli, F. M.**, Kellner, R., Rathman, J. F., Yang, C., Cronin, M. T., & Escher, S. E. (2021). Threshold of Toxicological Concern—An Update for Non-Genotoxic Carcinogens. Frontiers in Toxicology. https://doi.org/10.3389/ftox.2021.688321

Poster Presentations

Moradiafrapoli F., Wehr M., Bitsch A. and Escher S.E., Prediction of adverse effects in preclinical subchronic studies by analysis of adverse effects from shorter-term studies using e.g. the RepDose® database, 55th Eurotox Congress (2019) Helsinki, Finland

Moradiafrapoli F., Bitsch A., Escher S.E., Frequency and sensitivity of affected organs in preclinical studies from the RepDoseand eTOXdatabases, 85th DGPT Congress (2019) Stuttgart, Germany

Other Publications

Moradiafrapoli F, Mandegary A, Heydari MR, Shaki F, Hepatoprotective effect of Polygonum hyrcanicum methanolic extract on the carbon tetrachloride induced hepatotoxicity in mice. Iranian Journal of Physiology and Pharmacology, 2015; 1: 43-48

Khaksari M, Rajizadeh MA, Bejeshk MA, Soltani Z, Motamedi S, **Moradiafrapoli F**, Islami M, Shafa Sh, Khosravi S. Does inhibition of angiotensin function cause neuroprotection in diffuse traumatic brain injury, Iran J Basic Med Sci 2018; 21:615-620

Ahmadipour, A., Sharififar, F., Pournamdari, M., **Moradiafrapoli F**, et al.Hepatoprotective effect of Zataria multiflora Boiss against malathion-induced oxidative stress in male rats. Oriental Pharmacy and Experimental Medicine 2016; 16: 287-293

9. Resume

Personal data

Name	Fatemeh Moradiafrapoli
Date of Birth	07.01.1988
Place of Birth	Babol, Iran
Nationality	German-Iranian

Education

PhD candidate Oct 2017 – May 2023	Gottfried Wilhelm Leibniz University Hannover, Institute for Technical Chemistry, Hannover, Germany
Karshenasi Arshad (M.Sc.) Sep 2012 - Mar 2015	Toxicology, Kerman University of Medical Sciences, Iran
B.Sc. Sep 2007 - Jun 2012	Plant Protection and Integrated Pest Management, University of Tehran, Iran

Professional Experience

Research Assistant Oct 2017 – Dec 2020	Chemical risk assessment and Toxicology, Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Hannover, Germany
Senior Expert Product Stewardship Hazard and Risk Management Jan 2023 – Present	Röhm GmbH, Darmstadt, Germany

10. Acknowledgment

I would like to thank Dr. Sylvia Escher for giving me the opportunity to join her research group and for her support, advice, and encouragement during my entire Ph.D. period.

My sincere appreciation goes to Prof. Dr. Thomas Scheper for his great support and for accepting to supervise my Ph.D. thesis.

I would like to express my gratitude to Prof. Dr. Bernd Hitzmann, who accepted to be the co-referee of my Ph.D. examination.

My warm thanks to Tom Aldenberg from RIVM National Institute for Public Health and the Environment, for his cooperation with the simulation of The selection of animal Coverage Uncertainty model.

A big thanks to my colleagues at Fraunhofer ITEM, Matthias Wehr and Walter Zobl, for their assistance in the project's progression. And all my friends and former colleagues at the department of chemical risk assessment and in silico toxicology at Fraunhofer ITEM for their friendship, constant support, and great atmosphere, which made me always feel at home.

Finally, I give very special thanks to my family, my husband, Ebi, and my dear son, Diako, for their love, support, and encouragement.