



الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة وهران للعلوم والتكنولوجيا "محمد بوضياف"
كلية علوم الطبيعة والحياة

People's Democratic Republic Of Algeria
Ministry of Higher Education and Scientific Research
University of Science and Technology of Oran Mohamed BOUDIAF



Faculty of Natural and Life Sciences

COURSE MANUSCRIPT

Experimental Toxicology

*A Course Manual for Master 1 Students
Fundamental and Applied Toxicology*

(Aligned with the Official USTO-MB Curriculum)

Author

Dr. Sofiane Mohamed El Amine BENGHALI
Department of Life and Environment - USTO-MB
Laboratory Network for Environmental
Monitoring (LRSE)
ORCID: 0009-0001-7488-0608

Academic Year 2025–2026

Foreword

This course manuscript is intended for first-year Master's students (M1) enrolled in the Fundamental and Applied Toxicology programme at the Department of Living and Environment, Faculty of Natural and Life Sciences, Université des Sciences et de la Technologie d'Oran — Mohamed Boudiaf (USTO-MB).

This second edition has been restructured to align with the official USTO-MB Master 1 curriculum in Experimental Toxicology of Reproduction and General Experimental Toxicology. Where the first edition organised content around ten thematic chapters covering conceptual foundations, mechanisms, testing, ecotoxicology, and analytical methods, this version follows the six-part official programme — Animal Experimentation, In Vivo Toxicity Testing, In Vitro Toxicity Evaluation, Stochastic Effects, Good Practices in Animal Experimentation, and Alternative Methods — supplemented by structured Tutorial (TD) and Practical (TP) programmes.

The scientific content of the original edition has been preserved, reorganised, and enriched. Chapters on ecotoxicology and major chemical classes, which fall outside the formal programme scope, have been condensed into worked examples integrated throughout the text (lead, cadmium, benzo[a]pyrene, aflatoxin B1, BPA, glyphosate). The toxicokinetics chapter is distilled into a focused ADME section within Part 2; dose-response content is distributed between Part 2 and Part 4. The Algerian and Mediterranean context — including Loi 12-06 (2012) on animal use in research, the Oran industrial basin, and West Algerian coastal fish biomonitoring — is highlighted throughout.

Each Part includes explicit learning objectives, a structured introduction, body sections with tables and figures, key take-home messages, and further reading with recent 2020–2025 references. The Tutorial programme provides a structured framework for critical analysis of scientific articles; the Practical programme offers six laboratory protocols covering the full range of skills expected of an M1 graduate in experimental toxicology.

All chemistry and biology facts, OECD guideline numbers, IARC classifications, and regulatory references have been verified against primary sources.

Sofiane Mohamed El Amine BENGHALI
Oran, Algeria — 2026

Table of Contents

Foreword.....	i
Table of Contents	ii
List of Figures	viii
List of Tables	ix
Part 1 — Animal Experimentation.....	1
1.1 Learning Objectives	1
1.2 Introduction.....	1
1.3 Ethical Principles and Directives	1
1.3.1 Russell and Burch (1959): The 3R Principle.....	1
1.3.2 Declaration of Helsinki and Animal Research	2
1.3.3 EU Directive 2010/63/EU.....	3
1.3.4 ARRIVE Guidelines	4
1.4 Organisation of Animal Facilities and Governance.....	4
1.4.1 The Animal Facility (Animalerie).....	4
1.4.2 Institutional Animal Care and Use Committee (IACUC) / Ethics Committee.....	5
1.4.3 Roles and Responsibilities	6
1.5 Regulation.....	6
1.5.1 European Union: Directive 2010/63/EU.....	6
1.5.2 United States: Animal Welfare Act and PHS Policy.....	6
1.5.3 Algerian Framework: Loi 12-06 and Institutional Context.....	6
1.5.4 OECD Principles and Mutual Acceptance of Data.....	8
1.6 Housing, Husbandry and Experimentation Conditions	8
1.6.1 Cage Standards and Space Requirements	8
1.6.2 Environmental Parameters	9
1.6.3 Feeding, Water, and Acclimation	9
1.6.4 Environmental Enrichment.....	9
1.7 Standards	10

1.7.1 ISO 10993: Medical Devices	10
1.7.2 Good Laboratory Practice (GLP)	10
1.7.3 AAALAC International	10
1.7.4 FELASA Recommendations	10
1.8 Advantages and Disadvantages of Animal Experimentation	11
1.9 Case Study: The ARRIVE 2.0 Lifecycle	12
Part 2 — In Vivo Toxicity Testing.....	15
2.1 Learning Objectives	15
2.2 Introduction.....	15
2.3 Acute Toxicity Tests	15
2.3.1 Historical Context and the Demise of TG 401	15
2.3.2 OECD TG 420: Fixed Dose Procedure.....	15
2.3.3 OECD TG 423: Acute Toxic Class Method	16
2.3.4 OECD TG 425: Up-and-Down Procedure (UDP).....	16
2.3.5 Dermal and Inhalation Acute Toxicity	17
2.3.6 GHS Classification Based on Acute Oral Toxicity.....	17
2.4 Short-Term Toxicity Tests	18
2.4.1 28-Day Repeated Dose Study: OECD TG 407	18
2.4.2 90-Day Repeated Dose: OECD TG 408 (Rodent) and TG 409 (Non-Rodent)	19
2.5 Long-Term Toxicity Tests	19
2.5.1 Chronic Toxicity: OECD TG 452	19
2.5.2 Carcinogenicity and Combined Chronic/Carcinogenicity: TG 451, TG 453	19
2.5.3 Reproductive Toxicity: TG 415, 416, 443 EOGRTS.....	20
2.5.4 Developmental Toxicity: OECD TG 414	21
2.6 ADME Considerations in In Vivo Testing Design.....	22
2.6.1 Absorption.....	23
2.6.2 Distribution.....	23
2.6.3 Metabolism and Bioactivation.....	24

2.6.4 Excretion and Half-Life.....	24
2.7 Dose Selection, Statistical Design and Sample Size	24
2.7.1 Principles of Dose Selection	25
2.7.2 Sample Size Justification.....	25
2.7.3 NOAEL and LOAEL Determination.....	25
2.7.4 Benchmark Dose (BMD) Modelling.....	25
Part 3 — In Vitro Toxicity Evaluation	27
3.1 Learning Objectives	27
3.2 Introduction.....	27
3.3 Cell Models	28
3.3.1 Primary Cells	29
3.3.2 Immortalised Cell Lines	29
3.3.3 iPSC-Derived Models.....	29
3.4 Three-Dimensional Models.....	29
3.4.1 Spheroids	31
3.4.2 Organoids.....	31
3.4.3 Liver-on-Chip and Microphysiological Systems	31
3.5 Cytotoxicity Assays.....	31
3.6 Mechanistic In Vitro Endpoints	32
3.6.1 Oxidative Stress.....	33
3.6.2 Apoptosis	35
3.6.3 Mitochondrial Function	35
3.7 Genotoxicity In Vitro	35
3.8 Endocrine Disruption In Vitro.....	37
3.11 Case Study: Integrated In Vitro Strategy for Hepatotoxicity	37
3.9 High-Throughput Screening and Tox21	39
3.10 Strengths and Limitations.....	39
Part 4 — Special Case of Stochastic Effects	41

4.1 Learning Objectives	41
4.2 Introduction: Deterministic vs. Stochastic Effects.....	41
4.3 Theoretical Framework	41
4.3.1 The Linear Non-Threshold (LNT) Model	41
4.3.2 The Threshold Debate	43
4.3.3 Hormesis as Counterpoint	43
4.4 Carcinogenesis as the Paradigmatic Stochastic Effect	43
4.4.1 Initiation, Promotion, Progression.....	43
4.4.2 IARC Carcinogen Classification	45
4.5 Heritable Genetic Effects	45
4.6 Quantification: Slope Factor, Unit Risk, Excess Lifetime Cancer Risk.....	46
4.7 Risk Assessment for Stochastic Effects.....	47
4.8 Examples of Stochastic Carcinogens.....	48
4.8.1 Benzo[a]pyrene	50
4.8.2 Aflatoxin B1.....	50
4.8.3 Vinyl Chloride.....	50
4.8.4 Formaldehyde	50
4.9 Regulatory Implications	50
Part 5 — Good Practices in Animal Experimentation	53
5.1 Learning Objectives	53
5.2 Introduction.....	53
5.3 Good Laboratory Practice (GLP): OECD Principles	53
5.3.1 Scope of GLP	55
5.4 GLP Organisation: Key Roles	56
5.4.1 The Study Director	57
5.4.2 Quality Assurance Unit.....	57
5.4.3 Archives and SOPs.....	57
5.5 ARRIVE 2.0 Guidelines	57

5.6 Experimental Design Best Practices	59
5.6.1 Randomisation and Blinding	60
5.6.2 Sample Size Justification.....	60
5.6.3 Pre-Registration.....	60
5.7 Humane Endpoints and Severity Classification	61
5.8 Refinement Techniques	62
5.9 Reduction Strategies	63
5.10 Data Integrity and the Reproducibility Crisis	63
5.11 Algerian Context: Implementation at USTO-MB	64
Part 6 — Alternative Methods to Animal Experimentation.....	67
6.1 Learning Objectives	67
6.2 Introduction: The 3R Drive Toward NAMs.....	67
6.3 Validation of Alternative Methods	67
6.3.1 ECVAM and EURL ECVAM.....	69
6.3.2 ICCVAM and JaCVAM.....	69
6.3.3 OECD Validation Process.....	69
6.4 In Silico Modelling and Analysis.....	69
6.4.1 QSAR — Quantitative Structure-Activity Relationships.....	69
6.4.2 Read-Across and Chemical Grouping	70
6.4.3 PBPK Modelling.....	71
6.4.4 Machine Learning and Deep Learning in Toxicology.....	73
6.4.5 Software Tools	74
6.5 Tools of Predictive Toxicology	75
6.5.1 Adverse Outcome Pathways (AOPs) and AOP-Wiki.....	75
6.5.2 Tox21 and ToxCast.....	77
6.5.3 Integrated Approaches to Testing and Assessment (IATA)	77
6.5.4 Defined Approaches for Skin Sensitisation and Eye Irritation	77
6.5.5 Next-Generation Risk Assessment (NGRA)	77

6.6 Examples of Fully Validated Alternative Methods.....	79
6.7 Limitations and Future Perspectives	80
Tutorial Programme (Travaux Dirigés) Critical Analysis of Scientific Articles	82
T.1 Pedagogical Objectives.....	82
T.2 Selection Criteria for Articles	82
T.3 Critical Reading Grid	84
T.4 Six Suggested Articles with Rationale.....	88
T.7 Sample Worked Example: Full Critical Analysis.....	89
T.5 Assessment Grid	92
T.6 Suggested Calendar	92
Practical Programme (Travaux Pratiques) Setting Up a Toxicology Study	94
P.1 Learning Objectives	94
P.2 General Framework: The Toxicology Study Lifecycle.....	94
P.3 Practical 1: Designing a Dose-Response Study (Acute Oral Toxicity, OECD TG 425)	94
P.3.1 Objectives.....	96
P.3.2 Background.....	96
P.3.3 Protocol	97
P.3.4 Discussion Questions	97
P.3.5 Safety Considerations	97
P.4 Practical 2: In Vitro Cytotoxicity Assay (MTT on HepG2 with CdCl ₂).....	97
P.4.1 Objectives.....	99
P.4.2 Background.....	99
P.4.3 Materials.....	100
P.4.4 Step-by-Step Protocol.....	101
P.4.5 Safety Considerations	101
P.4.6 Discussion Questions	101
P.5 Practical 3: Histopathological Observation of Liver and Kidney Sections	101
P.5.1 Objectives.....	101

P.5.2 Background.....	101
P.5.3 Lesion Scoring System.....	102
P.5.4 Discussion Questions	103
P.6 Practical 4: Heavy Metal Quantification by AAS in Fish Tissue	103
P.6.1 Objectives.....	103
P.6.2 Background — Mediterranean Fish Example.....	103
P.6.3 Protocol	104
P.6.4 Method Validation Parameters	105
P.6.5 Discussion Questions	105
P.7 Practical 5: QSAR Prediction Using OECD QSAR Toolbox	105
P.7.1 Objectives.....	105
P.7.2 Background.....	105
P.7.3 Protocol	106
P.7.4 Discussion Questions	106
P.8 Practical 6: Critical Statistical Analysis (NOAEL/LOAEL Determination, BMD Calculation)	106
P.8.1 Objectives.....	106
P.8.2 Provided Dataset	106
P.8.3 Statistical Analysis Protocol.....	108
P.8.4 Discussion Questions	108
P.8.5 Safety Considerations	108
P.9 Practical Report Template and Marking Grid.....	109
Glossary of Key Terms and Abbreviations	110
Additional Terms (Expanded)	113
Bibliography.....	118

List of Figures

Figure 1.1. The 3R Principle in Experimental Toxicology.	2
---	---

Figure 1.E1. Functional organigram of a typical animal research facility, showing governance structure from the Institutional Animal Welfare Body (AWB/IACUC) through Named Officers to study personnel and procedural functions (SOPs).	5
Figure 2.E1. OECD TG 425 Up-and-Down Procedure worked example.	16
Figure 2.E2. OECD TG 443 Extended One-Generation Reproductive Toxicity Study (EOGRTS) cohort design.	20
Figure 3.1. The Three Experimental Paradigms in Toxicology.	27
Figure 4.E1. Multistage chemical carcinogenesis — initiation, promotion, and progression.	42
Figure 4.1. The Adverse Outcome Pathway (AOP) Framework, illustrated with an aromatase-inhibition example.	50
Figure 5.E1. Organisational structure of a GLP test facility according to OECD Principles of GLP (ENV/MC/CHEM(98)17).	53
Figure 6.1. The 3R Principle — the ethical and regulatory backbone driving the development of alternative methods.	65
Figure 6.E1. In Vitro to In Vivo Extrapolation (IVIVE) workflow using PBPK modelling.	70
Figure 6.E2. Adverse Outcome Pathway (AOP 25): Estrogen Receptor α agonism leading to reproductive impairment in oviparous organisms (AOP-Wiki, OECD 2022).	74
Figure 6.E3. Next-Generation Risk Assessment (NGRA) framework.	77
Figure T.1. Critical Reading Flowchart for Toxicology Articles.	81
Figure T.2. Evidence Pyramid for Toxicological Risk Assessment.	82
Figure T.3. CONSORT-Style Flow Diagram for an In Vivo Toxicology Study.	85
Figure T.4. IMRaD Structure of a Scientific Article in Toxicology.	85
Figure P1.1. OECD TG 425 Up-and-Down Procedure — Dosing Decision Algorithm.	93
Figure P2.1. 96-Well Plate Layout for the MTT Cytotoxicity Assay (CdCl ₂ model toxicant on HepG2 cells).	97
Figure P2.2. Example dose-response sigmoid curve: HepG2 cell viability (%) vs.	97
Figure P3.1. Hepatic Lobule — Annotated Schematic with Hallmark Toxicological Lesions.	99
Figure P3.2. Nephron Segments — Key Sites and Types of Nephrotoxic Lesions.	99
Figure P4.1. Atomic Absorption Spectrometry (AAS) — Instrument Block Diagram.	100
Figure P4.2. AAS Calibration Curve for Cadmium (Cd) at 228.	101
Figure P5.1. QSAR Prediction Workflow: from SMILES input through descriptor calculation, model application, and applicability domain check, to final report.	103
Figure P6.1. NOAEL, LOAEL, BMD, and BMDL on an Annotated Dose-Response Curve.	104
Figure P6.2. Statistical Decision Tree for In Vivo Toxicity Data Analysis.	104

List of Tables

Table 1.2. Housing and husbandry standards for common laboratory species (FELASA / EU Directive 2010/63/EU Annex II).	9
Table 1.3. Advantages and disadvantages of animal experimentation — a balanced assessment.	13

Table 2.1. Comparison of modern acute oral toxicity test guidelines (TG 420, 423, 425) with the deleted TG 401.....	16
Table 2.2. Overview of subacute and subchronic repeat-dose OECD test guidelines.	18
Table 2.3. OECD long-term and reproductive/developmental toxicity test guidelines.	21
Table 2.4. Summary of key toxicokinetic (ADME) parameters and their relevance to in vivo study design.	23
Table 3.1. Commonly used cell lines in toxicology with metabolic competence and principal uses.	28
Table 3.2. Comparison of cytotoxicity assays commonly used in in vitro toxicology.	30
Table 3.4. OECD in vitro genotoxicity battery: test guidelines and endpoints.....	35
Table 4.1. IARC carcinogen classifications with illustrative examples.	43
Table 4.2. Stochastic effect examples: slope factors, unit risk, and key mechanistic basis...45	
Table 5.1. OECD GLP Principles — summary of key elements (OECD No.	54
Table 5.2. ARRIVE 2.	57
Table 5.3. Humane endpoint severity classifications (EU Directive 2010/63/EU).	60
Table 6.2. Key QSAR and in silico software tools in predictive toxicology.	73
Table 6.1. Fully validated alternative test methods for regulatory use (OECD Test Guidelines).....	78
Table T.1. Critical reading grid — 10-question rubric for article analysis (TD sessions).	84
Table T.2. Six suggested articles for TD sessions with rationale and focus points.	89
Table T.3. Assessment rubric for student article presentations (TD sessions).....	90
Table T.4. Suggested TD calendar (6 × 2-hour sessions across the semester).....	90

Part 1 — Animal Experimentation

1.1 Learning Objectives

- Define the ethical principles governing animal experimentation and explain the 3R framework (Replacement, Reduction, Refinement) in the context of modern regulatory toxicology.
- Describe the organisation of a certified animal facility, including governance structures, roles of the Institutional Animal Care and Use Committee (IACUC), study director, and veterinarian.
- Compare the regulatory frameworks governing animal experimentation in the EU (Directive 2010/63/EU), USA (Animal Welfare Act, PHS Policy), and Algeria (Loi 12-06, 2012).
- Specify housing, husbandry, and environmental conditions recommended by FELASA and AAALAC for the principal laboratory species.
- Explain GLP principles and ISO 10993 requirements as they apply to animal studies.
- Provide a balanced critical appraisal of the advantages and limitations of animal models in toxicology.

1.2 Introduction

Animal experimentation has been central to biomedical and toxicological research for well over a century. The systematic observation of toxic effects in whole organisms provides an integrated biological readout that no current in vitro or computational system can fully replicate: the interaction of a xenobiotic with the totality of an organism's physiology — its immune system, endocrine axes, metabolic organs, and nervous system — unfolds in real time, generating the complex toxicological phenotypes that constitute the basis of regulatory hazard assessment.

At the same time, animal experimentation raises profound ethical questions. The suffering of sentient beings in the pursuit of scientific knowledge demands rigorous justification, and modern toxicology has responded with a set of principles and governance structures designed to minimise harm while maximising scientific validity. Part 1 of this manual surveys those principles and structures — the ethical frameworks, the institutional architecture, the applicable regulations, and the practical standards that govern animal experimentation in a research laboratory operating at international quality level.

1.3 Ethical Principles and Directives

1.3.1 Russell and Burch (1959): The 3R Principle

Historical context: The 3Rs originated in a landmark monograph commissioned by the Universities Federation for Animal Welfare (UFAW). William Moy Stratton Russell (1925–2006), a biologist, and Rex Leonard Burch (1926–1996), a microbiologist, spent three years reviewing the scientific literature and conducting site visits to laboratories across the United Kingdom. Their report, "The Principles of Humane Experimental Technique," was published by Methuen in 1959 and received little immediate attention. It was only in the 1980s and 1990s, as the animal rights movement gained traction and as regulatory agencies began to demand explicit justification for animal use, that the 3Rs framework was formally embedded in legislation and guidelines. The European Centre for the Validation of Alternative Methods (ECVAM, now EURL ECVAM) was established in 1991 explicitly to implement the Russell–Burch vision. The 3Rs are now statutory requirements under EU Directive 2010/63/EU (Articles 4, 13), the US Health Research Extension Act (1985), and a growing number of national frameworks worldwide, including Algeria's Loi 12-06 (2012) which references the protection of animals used for scientific purposes.

The 3Rs framework is not merely an ethical aspiration — it has direct scientific benefits. Refinement reduces stress-induced variability in experimental animals, thereby improving data quality and reproducibility. Reduction demands rigorous statistical power calculations, which guard against both underpowered (false negative) and overpowered (wasteful) designs. Replacement accelerates high-throughput screening of thousands of chemicals that would be impractical to test in whole animals. The OECD has incorporated 3Rs principles throughout its Test Guideline (TG) system: many modern TGs (e.g., TG 420, 423, 425 for acute toxicity) were designed explicitly to reduce animal numbers compared to the classical LD₅₀ test (TG 401, deleted in 2002). Today, the 3Rs are a central pillar of the OECD Guidance Document 19 on the recognition, assessment, and use of clinical signs as humane endpoints in acute toxicity and safety evaluation studies (OECD, 2000/updated 2022).

The intellectual foundation of ethical animal experimentation was laid by William Russell and Rex Burch in their 1959 monograph *The Principles of Humane Experimental Technique* (Russell & Burch, 1959). Their central thesis was that animal welfare and scientific quality are not in opposition — that unnecessary pain and distress in experimental animals introduces biological variability that degrades data quality. The 3Rs they defined were:

- Replacement: substituting animal models with non-animal methods (in vitro, in silico, ex vivo) wherever scientifically valid alternatives exist.
- Reduction: using the minimum number of animals consistent with adequate statistical power, by means of better experimental design and data sharing.
- Refinement: modifying procedures to minimise pain, distress, and suffering, and to improve animal welfare throughout the experimental process.

Although articulated in 1959, the 3Rs gained regulatory force from the 1990s onward, driven by ECVAM's establishment (1991), growing societal concern, and the recognition that classical testing paradigms are inadequate for the modern chemical regulatory landscape.

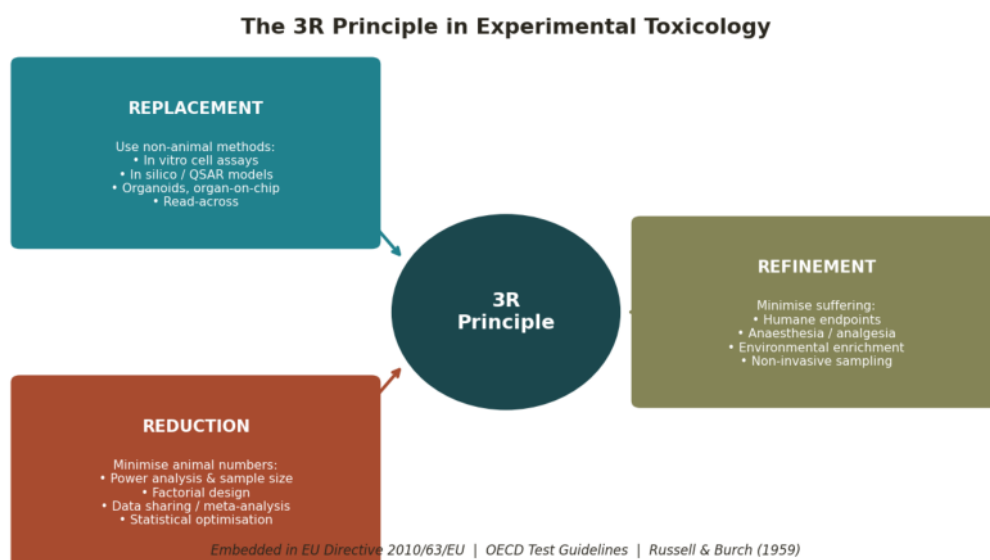


Figure 1.1. The 3R Principle in Experimental Toxicology. The three components — Replacement, Reduction, and Refinement — are legally mandated in EU Directive 2010/63/EU and form the ethical backbone of modern experimental toxicology. (Original figure.)

1.3.2 Declaration of Helsinki and Animal Research

Although the Declaration of Helsinki (World Medical Association, 2013) formally governs human research, its ethical principles — that any research on living beings must be preceded by adequate laboratory testing, that potential benefits must justify risks, and that results must be honestly reported

— have informed the development of ethical standards for animal research. The ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines, discussed further in Part 5, represent the translational extension of Helsinki principles to animal experimentation reporting.

1.3.3 EU Directive 2010/63/EU

Detailed comparison of regulatory frameworks governing animal experimentation:

EU Directive 2010/63/EU was transposed into national law across all EU Member States by January 2013. It applies to all vertebrate animals (including cyclostomes) and cephalopod molluscs. The Directive establishes: (1) a prospective authorisation system — any project involving animals must receive prior approval from a competent authority, typically within 40 days; (2) a severity classification of procedures into four categories (non-recovery, mild, moderate, severe), with restrictions on the use of severe procedures; (3) a mandatory non-technical project summary for public transparency; (4) a retrospective assessment for all severe studies and long-term studies; (5) a National Committee system for advice on implementation; and (6) a 3Rs requirement embedded in Articles 4 and 13.

The US Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals (2015 revision) applies to all institutions receiving PHS funding and requires: (1) adherence to the Guide for the Care and Use of Laboratory Animals (ILAR, 2011, 8th edition); (2) maintenance of an Institutional Animal Care and Use Committee (IACUC) that reviews all protocols; (3) a veterinarian with training and experience in laboratory animal medicine; (4) a programme of veterinary care including preventive medicine; and (5) semi-annual programme review by the IACUC. Severity classification under the US framework follows the USDA Animal Welfare Act categories B–F (pain/distress classification), but lacks the formalised retrospective assessment required in the EU.

Algeria's Loi 12-06 (Loi n° 12-06 du 12 janvier 2012 relative aux associations) does not address animal experimentation directly; the broader legislative basis in Algeria comes from Décret exécutif n° 94-399 (1994) on the protection of animals and from the ministerial circulars issued by the Ministry of Higher Education and Scientific Research. The MESRS (Ministère de l'Enseignement Supérieur et de la Recherche Scientifique) issued guidance in 2018 requiring university ethics committees to review protocols involving vertebrate animals. USTO-MB maintains an institutional ethics committee (Comité d'Éthique Institutionnel) under the Faculty of Natural and Life Sciences (FSNV). However, a formal national regulatory framework equivalent to EU 2010/63/EU remains under development, and the CONASA (Conseil National de l'Animal de la Science et de l'Agriculture) initiative represents the most advanced effort to formalise these standards at the national level.

Key comparison table (EU vs US vs Algeria):

- Scope: EU covers all vertebrates + cephalopods; US covers warm-blooded animals (plus birds, rats, mice explicitly added 2002); Algeria covers vertebrates under administrative guidance.
- Ethics review: EU requires competent authority authorisation; US requires IACUC approval; Algeria requires institutional ethics committee review.
- Personnel training: EU requires documented competence assessment; US requires training programme overseen by IACUC; Algeria requires compliance with MESRS guidelines.
- Severity classification: EU has 4 formal categories; US uses AWA pain/distress categories; Algeria follows EU model by guidance.
- Retrospective assessment: Mandatory in EU for severe/long-term studies; voluntary in US; not formally required in Algeria.
- Public reporting: EU requires annual statistical reporting to EC; US reports to USDA annually; Algeria has no centralised reporting system currently.

Directive 2010/63/EU on the protection of animals used for scientific purposes (European Parliament & Council, 2010) replaced the earlier Directive 86/609/EEC and substantially strengthened animal welfare requirements across all EU member states. Key provisions include:

- Mandatory application of the 3R principle to all experimental protocols.
- Coverage of all vertebrate animals (including foetal stages in the last third of development) and certain invertebrates (decapod crustaceans, cephalopods).
- Establishment of national competent authorities and ethics committees that evaluate all animal use projects before approval.
- Severity classification of procedures (mild, moderate, severe, non-recovery) to facilitate ethical review.
- Requirements for retrospective assessment of projects involving severe procedures.
- Recognition of validated alternative methods as mandatory replacements where they exist.

1.3.4 ARRIVE Guidelines

ARRIVE 2.0 (2020) represents a major revision of the original ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments, Kilkenny et al., 2010, PLOS Biology). The revision was driven by evidence that fewer than 50% of published animal studies reported sufficient information to allow replication (Percie du Sert et al., 2020, PLOS Biology). The 2020 guidelines restructured the 20 items into two tiers: the Essential 10 (items that must be reported in all manuscripts) and the Recommended Set (items that should be reported when relevant). The Essential 10 covers: study design, sample size, inclusion/exclusion criteria, randomisation, blinding, outcome measures, statistical methods, experimental animals, experimental procedures, and results. The Recommended Set adds: abstract, background, objectives, ethical statement, housing, animal characteristics, protocol registration, statistical analysis details, baseline data, numbers analysed, and adverse events.

A critical innovation in ARRIVE 2.0 is the Experimental Design section, which requires authors to state explicitly whether a priori power calculations were performed and to report the resulting justification for the chosen animal numbers. This responds to the widespread problem of underpowered studies (typically $n=6-8$ per group, often insufficient to detect biologically relevant effect sizes with 80% power). The guidelines also emphasise the distinction between exploratory and confirmatory studies: confirmatory studies require pre-registered hypotheses and statistical analysis plans, while exploratory studies should be labelled as such to avoid selective reporting bias. ARRIVE 2.0 has been endorsed by over 1,000 journals and adopted as a submission requirement by high-impact journals in the toxicology field including Archives of Toxicology and Toxicological Sciences.

The ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines, developed by the National Centre for the 3Rs (NC3Rs) and updated to ARRIVE 2.0 in 2020 (Percie du Sert et al., 2020), provide a structured framework for reporting in vivo experiments. The ARRIVE 2.0 Essential 10 items cover: study design, sample size, inclusion/exclusion criteria, randomisation, blinding, outcome measures, statistical methods, experimental animals, experimental procedures, and results. Compliance with ARRIVE is now required or strongly encouraged by major journals including Nature, PLOS Biology, and the British Journal of Pharmacology.

1.4 Organisation of Animal Facilities and Governance

1.4.1 The Animal Facility (Animalerie)

A regulatory-grade animal facility for toxicology studies must meet strict architectural and operational requirements. The facility is typically divided into: barrier zones (specific pathogen-free, SPF, areas for breeding colonies), experimental zones (where dosing and procedures occur), and clean/dirty corridors to maintain unidirectional traffic flow and prevent cross-contamination. AAALAC International (Association for Assessment and Accreditation of Laboratory Animal Care) accreditation is the internationally recognised certification of facility quality.

1.4.2 Institutional Animal Care and Use Committee (IACUC) / Ethics Committee

The physical organisation of a modern animal facility reflects both scientific and ethical requirements. A well-designed facility segregates clean areas (where uncontaminated animals and supplies are handled) from dirty areas (where waste, soiled bedding, and carcasses are processed), following a unidirectional flow principle to prevent cross-contamination. The main functional zones include: (1) receiving and quarantine — where newly arrived animals are health-screened before entering the main colony; (2) barrier rooms — specifically constructed to maintain defined health status animals, with filtered supply and exhaust air, positive pressure, and strict entry protocols via air-lock systems; (3) conventional holding rooms; (4) procedure rooms where experiments are conducted; (5) cage wash and autoclave areas; (6) storage areas for feed, bedding, and equipment; and (7) a carcass holding room with cold storage prior to disposal by incineration.

Environmental monitoring within holding rooms is continuous: temperature (20–24°C for rodents), relative humidity (45–65%), lighting cycle (typically 12h light/12h dark, with light intensity 60–300 lux at cage level), noise levels (<60 dB background), and air changes (15–20 per hour with HEPA filtration to ≥99.97% efficiency for 0.3 µm particles). Dedicated IVC (individually ventilated cage) systems provide cage-level ventilation and are now standard in most SPF (specific pathogen free) facilities. HVAC (heating, ventilation, and air conditioning) systems are designed with redundancy: failure alarms and backup power generation are mandatory for facilities housing large animal numbers.

The functional organigram of a modern animal facility places the Institutional Animal Welfare Body (AWB) or IACUC at the top of the governance structure. Under the AWB, the Designated Veterinarian (DV) is responsible for all aspects of animal health and welfare. The Named Animal Care and Welfare Officer (NACWO) oversees the day-to-day management of the colony and liaises between the DV and the husbandry team. The Named Information Officer (NIO) is responsible for maintaining the 3Rs information resource — advising investigators on alternative methods, refinement techniques, and relevant literature. In practice, the Study Director coordinates the scientific programme and is accountable to the AWB for the ethical conduct of each study. At USTO-MB, these roles are distributed across the veterinary services, the laboratory director, and the ethics committee coordinator.

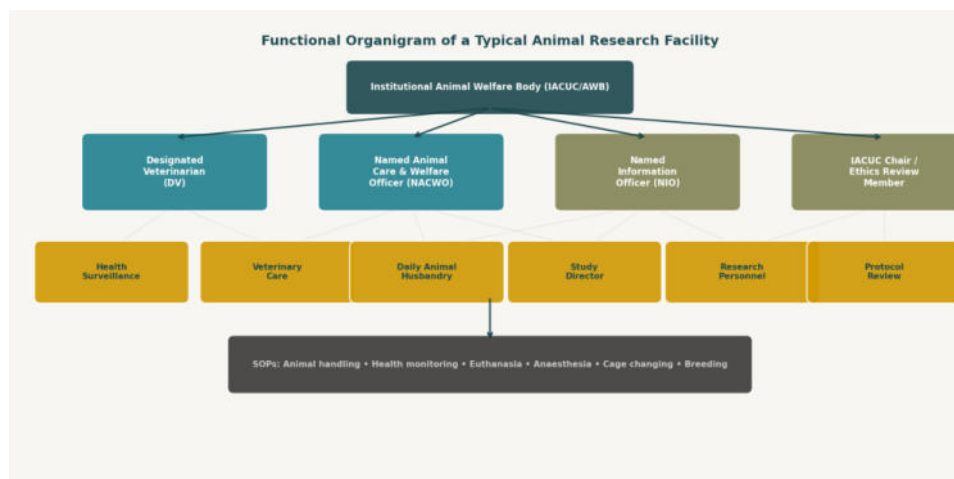


Figure 1.E1. Functional organigram of a typical animal research facility, showing governance structure from the Institutional Animal Welfare Body (AWB/IACUC) through Named Officers to study personnel and procedural functions (SOPs). Adapted from EU Directive 2010/63/EU Article 27 governance requirements.

In the USA, federal law (Animal Welfare Act, PHS Policy) requires that each institution using animals for research establish an IACUC, a multidisciplinary committee that reviews and approves all animal use protocols. In the EU, national ethics committees perform the equivalent function under Directive 2010/63/EU. Committee membership typically includes: scientists, veterinarians, a statistician, a non-scientific community member, and an animal welfare officer.

1.4.3 Roles and Responsibilities

Role	Responsibilities	Qualifications
Study Director	Single point of accountability for GLP study; approves protocol, signs final report	Senior scientist with GLP training
Principal Investigator	Designs and conducts experiments; supervises staff	PhD-level researcher
Attending Veterinarian	Clinical oversight, health monitoring, pain management, euthanasia	Licensed veterinarian (DVM/VMD)
Animal Care Technician	Daily husbandry, health checks, record-keeping, welfare monitoring	Certified (FELASA category B/C)
QA Unit	Inspects studies, reviews raw data and final reports for GLP compliance	GLP-trained quality assurance professional

Table 1 above. Key roles in animal experimentation governance.

1.5 Regulation

1.5.1 European Union: Directive 2010/63/EU

See Section 1.3.3 for detail. In practice, EU researchers must obtain project authorisation from the national competent authority, which evaluates the scientific justification, the harm-benefit analysis, and compliance with the 3Rs. France, Germany, and the UK (pre-Brexit) each transpose the directive into national law; compliance is monitored through inspections by competent authorities.

1.5.2 United States: Animal Welfare Act and PHS Policy

In the United States, animal experimentation is governed by two overlapping federal instruments. The Animal Welfare Act (AWA, 7 U.S.C. §2131 et seq.), enforced by the USDA Animal and Plant Health Inspection Service (APHIS), applies to warm-blooded animals (excluding rats, mice, and birds bred for research — a significant gap). The Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals applies to all vertebrates in PHS-funded research and requires IACUC approval and compliance with the Guide for the Care and Use of Laboratory Animals (NRC, 2011).

1.5.3 Algerian Framework: Loi 12-06 and Institutional Context

Algeria's regulatory landscape for animal experimentation is evolving rapidly. The primary legislative instrument is Loi n° 88-08 du 26 janvier 1988 relative aux activités de médecine vétérinaire et à la protection de la santé animale, which provides a framework for veterinary oversight of animals used in laboratories. This is complemented by Décret exécutif n° 03-481 (2003) establishing the National Veterinary Laboratory (LNV) and Arrêté interministériel (2013) on the use of animals in scientific research under the authority of the MESRS.

At the institutional level, USTO-MB has established an ethics committee (Comité d'Éthique de la Recherche — CER-USTO) that reviews experimental protocols. The CER-USTO requires investigators to justify: (a) the necessity of animal use and the impossibility of replacement with alternative methods; (b) the choice of species, age, and sex; (c) the expected numbers and statistical justification; (d) the measures taken to minimise pain and distress; and (e) the competence of personnel. Protocols are submitted at least 60 days before the study start, and approval is valid for 12 months.

The Institut National Algérien de la Propriété Industrielle (INAPI) is not directly involved in animal experimentation regulation, but plays a role in intellectual property protection for novel testing methodologies developed at USTO-MB. The recent CONASA initiative (Conseil National de l'Animal

de la Science et de l'Agriculture), proposed in 2022, aims to establish a unified national framework that aligns Algerian practice with OECD and EU standards. CONASA would oversee: (1) centralised registration of animal facilities; (2) personnel certification programmes; (3) a national database of experimental procedures; and (4) statistical reporting to international bodies. If enacted, CONASA would represent a major step toward full alignment with EU Directive 2010/63/EU, facilitating mutual acceptance of data under the OECD MAD framework.

A practical challenge for Algerian researchers seeking to publish in international peer-reviewed journals is the requirement to state the institutional ethics committee reference number and to describe compliance with national or international animal welfare standards. The absence of a formalised national regulatory framework makes this declaration difficult. Investigators at USTO-MB therefore typically cite compliance with EU Directive 2010/63/EU and the ARRIVE 2.0 guidelines as the applicable standards, even when national law does not yet mandate these. This represents a pragmatic approach that facilitates publication while demonstrating commitment to international best practices.

Case Study — Ethical Review for Neonicotinoid Neurotoxicity in Honeybees (*Apis mellifera*): An investigator at USTO-MB proposes a study to assess the sub-lethal neurotoxic effects of imidacloprid (a neonicotinoid insecticide) on honeybee foraging behaviour and olfactory learning. The ethical review process would proceed as follows:

(1) Request: The investigator submits a project description to the CER-USTO, explaining the scientific rationale (neonicotinoids are suspected to contribute to colony collapse disorder; data from Mediterranean Algerian apiaries are lacking), the expected benefits (informing regulatory decisions under Regulation 2018/783/EU and Algerian Décret n°95-362 on pesticide registration), and why honeybees are the appropriate test organism.

(2) Protocol design: The study proposes 10 hive replicates per treatment group (0, 0.1, 1, 10 ng/bee acute oral dose; 0.1, 1, 10 ng/bee/day chronic dietary exposure), monitoring: mortality, proboscis extension reflex (PER) conditioning, foraging activity by RFID tagging, acetylcholinesterase inhibition in head homogenates, and brain histology. The CER-USTO notes that invertebrates are not covered by EU Directive 2010/63/EU but applies the principles of the 3Rs by analogy, as recommended by EFSA (2012) and the OECD Test Guideline for honeybees (TG 213/214).

(3) Review: The ethics committee verifies that: animal numbers are justified by power calculations (target effect size $d=0.8$, $\alpha=0.05$, power=80%, yielding $n=8$ hives/group); all field collection is from consenting apiaries; imidacloprid exposure is relevant to field concentrations in agricultural settings; bees are euthanised by freezing at -80°C without undue suffering; and waste imidacloprid solutions are deactivated by photolysis before disposal.

(4) Approval: The CER-USTO approves the project with reference CER-USTO-2025-047, valid for 18 months, subject to mid-study reporting after the acute phase.

(5) Retrospective assessment: At study completion, the investigator submits a retrospective assessment covering: whether the objectives were met, actual animal use versus projected, any adverse welfare effects, and contribution to 3Rs knowledge (in this case, the PER assay replaces whole-colony mortality tests with a more sensitive and refined endpoint).

Algeria enacted Loi 12-06 (22 Safar 1433 / 15 January 2012) relating to associations, which provides the legal basis for voluntary ethical oversight bodies, including ethics committees for animal experimentation. Dedicated national legislation equivalent to EU Directive 2010/63/EU does not yet exist; however, the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique has issued guidelines requiring ethics committee approval for animal studies conducted at public universities.

At USTO-MB, the ethics committee (Comité d'Éthique pour la Recherche Biomédicale) reviews animal experimentation protocols for projects funded by the Algerian Direction Générale de la Recherche

Scientifique et du Développement Technologique (DGRSDT). Researchers publishing in international journals must declare compliance with the 3R principle and institutional ethical approval.

In practice, Algerian researchers follow the OECD Test Guidelines as a de facto standard, and the OECD Principles of Good Laboratory Practice provide the quality framework for analytical toxicology studies. The detailed implementing regulations for Loi 12-06 as applied to animal research are still evolving.

1.5.4 OECD Principles and Mutual Acceptance of Data

The OECD Principles of Good Laboratory Practice (OECD Series on Principles of GLP and Compliance Monitoring, No. 1, revised 1997) provide the procedural standards for non-clinical safety studies. Under the Mutual Acceptance of Data (MAD) agreement, regulatory submissions generated in compliance with GLP and OECD Test Guidelines are accepted across all OECD member states and many non-member countries (including Algeria for certain regulatory processes), avoiding duplication of animal testing.

1.6 Housing, Husbandry and Experimentation Conditions

1.6.1 Cage Standards and Space Requirements

EU Directive 2010/63/EU Annex III provides mandatory minimum space requirements for experimental animals. For laboratory rodents, these are the key quantitative thresholds:

Species/Condition	—	Minimum floor area (cm ² /animal)	—	Minimum cage height (cm):
House mouse (<i>Mus musculus</i>): up to 20g	→	60 cm ² /animal;	20–30g	→ 70 cm ² /animal; >30g → 80 cm ² /animal;
		cage height	≥12 cm	(solid bottom cage).
Laboratory rat (<i>Rattus norvegicus</i>): up to 200g	→	200 cm ² /animal;	200–300g	→ 250 cm ² /animal; 300–400g → 350 cm ² /animal;
			400–600g	→ 450 cm ² /animal; >600g → 600 cm ² /animal;
				cage height ≥18 cm.
Mongolian gerbil (<i>Meriones unguiculatus</i>):		150 cm ² /animal;		cage height ≥18 cm.
Syrian hamster (<i>Mesocricetus auratus</i>): up to 60g	→	150 cm ² /animal;	>60g	→ 200 cm ² /animal;
				cage height ≥14 cm.
Guinea pig (<i>Cavia porcellus</i>): up to 200g	→	200 cm ² /animal;	200–300g	→ 350 cm ² /animal; >300g → 500 cm ² /animal;
				cage height ≥23 cm.

These requirements apply to group-housed animals. Single housing is only permitted when scientifically justified and must be minimised in duration. For IVC (individually ventilated cage) systems, the actual cage floor area must meet the above minimums; the ventilated cage volume and air exchange rate (typically 60–100 cage air changes per hour) are additional considerations for animal health. The ILAR Guide (2011) provides similar but slightly different thresholds for US facilities, and investigators working under OECD mutual acceptance of data must ensure compliance with whichever standard is more stringent for the jurisdiction of the regulatory submission.

Environmental enrichment is not optional under EU Directive 2010/63/EU (Article 33): all animals must be provided with environmental enrichment appropriate to their species-specific needs unless there is a scientific justification for the contrary. For rodents, minimum enrichment standards include: (1) nesting material — shredded paper, cotton nesting material, or paper wool strips to allow construction of a sleeping nest and thermal regulation; (2) gnawing blocks — wooden blocks, chew sticks, or cardboard tubes to satisfy the need for incisor wear; (3) shelters and hides — opaque plastic tubes, igloos, or cardboard houses providing a refuge from conspecific aggression and a site for resting; (4) social housing — mice and rats must be housed in stable social groups (minimum pairs) unless single housing is scientifically or welfare-justified; (5) substrate for burrowing (for gerbils and hamsters). For rats specifically, the Guide recommends at minimum: bedding ≥2 cm depth, nesting material, a gnaw block, and a social companion. These measures reduce stereotypic behaviour, stress-induced weight variation, and the secretion of corticosterone, all of which can act as confounders in toxicological experiments (Sorge et al., 2014, Nature Methods).

FELASA (Federation of European Laboratory Animal Science Associations) and Annex II of EU Directive 2010/63/EU specify minimum floor area and height requirements for common laboratory species. For rats (≥ 300 g): minimum 800 cm² floor area; height ≥ 18 cm. For mice (≥ 25 g): minimum 330 cm² floor area; height ≥ 12 cm. Housing in individually ventilated cages (IVCs) is standard in SPF facilities, providing HEPA-filtered air supply and exhaust to each cage.

Species	Min floor area (cm ² /animal)	Min height (cm)	Group housing	Bedding
Mouse (>25 g)	330 (single)	12	≥ 2 per cage preferred	Softwood chip / corncob
Rat (>300 g)	800 (single)	18	≤ 5 per cage	Softwood chip / paper
Rabbit (>3.5 kg)	3500 (single)	45	Pairs tolerated	Hay + bedding
Guinea pig	1800	23	Same-sex pairs	Hay essential
Minipig	2.5 m ²	—	Social groups	Straw

Table 1.2. Housing and husbandry standards for common laboratory species (FELASA / EU Directive 2010/63/EU Annex II).

1.6.2 Environmental Parameters

Standardised environmental conditions are critical for experimental reproducibility. The recommended parameters for rodent facilities are: temperature 20–24°C ($\pm 1^\circ\text{C}$ within a study), relative humidity 45–65%, 12h:12h light-dark cycle (or as required by the protocol), ≥ 15 air changes/hour in conventional facilities (≥ 50 in IVCs), noise < 85 dB. Any deviation from standard conditions must be recorded in raw data and reported in the final study report.

1.6.3 Feeding, Water, and Acclimation

Certified commercial diets (e.g., Altromin C1000 for rats, Teklad 2018 for mice) should be used to minimise nutritional variability. Water is provided ad libitum via bottles or automated watering systems; water quality must be monitored for microbiological contamination. Feed and water analysis certificates must be retained in the study file. An acclimation period of at least 5 days (OECD guidance) is required before study initiation to allow animals to recover from transport stress.

1.6.4 Environmental Enrichment

Environmental enrichment reduces stereotypic behaviour, aggression, and stress-induced variability. Standard enrichment for rodents includes: nestlet material (rodents), hiding tubes and elevated platforms (mice), gnawing blocks (rabbits), and toys/foraging substrates. For social species, pair or group housing is preferable to isolation unless required by the experimental design.

1.7 Standards

1.7.1 ISO 10993: Medical Devices

ISO 10993 is an international standard series governing the biological evaluation of medical devices. Part 1 provides the overall framework; Parts 2, 5, 6, 10, and 11 cover animal welfare requirements, cytotoxicity, implantation, sensitisation, and systemic toxicity respectively. ISO 10993-2 requires that all in vivo testing on animals comply with the 3R principle and national animal welfare legislation.

1.7.2 Good Laboratory Practice (GLP)

OECD GLP Principles (detailed in Part 5) specify the organisational and procedural requirements for non-clinical laboratory studies intended to support regulatory submissions. The hallmarks of a GLP study are: a signed study plan approved before study initiation; raw data contemporaneously recorded

and attributable; a designated Study Director; a functioning QA unit; and a final report that faithfully reflects all data including unexpected findings.

1.7.3 AAALAC International

AAALAC International (Association for Assessment and Accreditation of Laboratory Animal Care International) provides voluntary accreditation of animal care and use programmes. Assessment uses the Guide for the Care and Use of Laboratory Animals (NRC, 2011), the European Convention ETS 123, and AAALAC's own standards. AAALAC accreditation is increasingly required by major journals and funding agencies as evidence of institutional commitment to high-quality animal care.

1.7.4 FELASA Recommendations

FELASA (Federation of European Laboratory Animal Science Associations) health monitoring recommendations provide the authoritative framework for microbiological surveillance in animal facilities. The 2014 FELASA Working Group Report (Nicklas et al., 2014, Laboratory Animals) defines the minimum panel of agents that must be excluded from experimental colonies and the frequency of testing.

Health monitoring programmes distinguish between: (1) routine monitoring of the holding colony, performed at 3-month intervals using sentinel animals; (2) quarterly targeted screening after introduction of new animals, at the point of entry into quarantine; and (3) event-driven testing following any suspicion of contamination (clinical signs, unexplained mortality, introduction from an unscreened source).

Sentinel animals are purpose-bred animals that share the same environment as the experimental colony (dirty bedding sentinels) but are not themselves used in experiments. They are maintained on soiled bedding from experimental cages and are tested at the end of each quarter. The sentinel strategy exploits the fact that contact with soiled bedding transmits most respiratory, gastrointestinal, and ectoparasitic agents; antibody titres in sentinel serum are therefore a sensitive indicator of colony health status even when experimental animals show no clinical signs.

FELASA categories A–D classify laboratory animal facilities by level of microbiological control:

- Category A (conventional): no formal pathogen exclusion; sentinel programme limited; appropriate only for teaching or screening experiments with no requirement for defined health status.
- Category B (minimal disease): basic excluded agents list (Sendai virus, mouse hepatitis virus, ECTROMELIA in mice; Sendai, Kilham rat virus, Toolan's H-1 in rats); quarterly sentinel monitoring.
- Category C (specific pathogen free, SPF): extended FELASA list of 20+ agents excluded; biannual or quarterly testing; barrier housing; personnel entry via air-lock with change to facility clothing.
- Category D (germ-free / gnotobiotic): completely sterile; housed in isolators; no microbial agents; used for studies on microbiome interactions and germ-free immunology.

The standard FELASA mouse health monitoring panel includes (among others): Ectromelia virus, EDIM rotavirus, LCMV, MHV (mouse hepatitis virus), MNV (murine norovirus), MPV (mouse parvovirus), TMEV (Theiler's murine encephalomyelitis virus), Sendai virus, Mycoplasma pulmonis, Clostridium piliforme, Helicobacter spp., Citrobacter rodentium, pinworms (Syphacia, Aspicularis), and ectoparasites (mites, lice). At USTO-MB, the standard quarterly monitoring programme tests for the most epidemiologically important agents given the local context (Sendai, MHV, pinworms, Helicobacter), with a full FELASA Category B panel performed annually.

FELASA (Federation of European Laboratory Animal Science Associations) publishes working group reports that provide technical guidance on health monitoring of rodent and rabbit colonies, breeding practices, and education and training requirements (Categories A–D, corresponding to different levels of animal experimentation). FELASA Category C training is the minimum requirement for a scientist designing and conducting animal experiments in the EU.

1.8 Advantages and Disadvantages of Animal Experimentation

Extended Critical Analysis: Scientific Necessity Versus Ethical Concerns in Animal Experimentation

The use of animals in toxicological research occupies a contested ethical and scientific space. Any rigorous treatment of the subject must engage with the dominant philosophical frameworks and assess their implications for laboratory practice.

The utilitarian framework, most prominently developed by Peter Singer in "Animal Liberation" (1975) and "Practical Ethics" (1979), holds that the morally relevant criterion is the capacity to suffer, not species membership or possession of language or reason. Singer argues that speciesism — differential treatment of beings solely on the basis of species — is analogous to racism or sexism in its arbitrary privileging of the interests of one group over another. From a utilitarian perspective, animal experimentation is permissible only when the expected benefits (to human health, to scientific knowledge, to the prevention of suffering) demonstrably outweigh the actual suffering inflicted on the animals. This position does not entail absolute abolitionism — Singer himself has accepted that some animal research may be justified — but sets a high threshold of justification and demands that genuine efforts to minimise suffering (the 3Rs) be made in every case. The implication for toxicological testing is that routine LD₅₀ determinations, conducted for decades primarily to generate regulatory data, represented an unjustifiable imposition of suffering when the information could be obtained with far fewer animals (and arguably greater predictive validity) using modern test designs.

The deontological (rights-based) framework, championed by Tom Regan in "The Case for Animal Rights" (1983), argues from the concept of inherent value. Regan contends that all beings who are "subjects of a life" — who have beliefs, desires, perception, memory, an emotional life, preference autonomy, and welfare interests — possess inherent value that cannot be traded off against the benefits to others. Under this view, animal experimentation involves treating sentient beings as mere instruments or "receptacles for value," which violates their rights regardless of the benefits to humans. Regan's position is abolitionist: he argues for the total elimination of the use of animals in science, not merely its reduction. This position has influenced the animal rights movement but has been criticised for its absolutism — critics point out that the logic, applied consistently, would also prohibit the killing of one person to save five, or the eating of any food produced with harm to sentient beings (Singer, 1985).

A contextual or relational approach, associated with philosophers such as Tom Beauchamp and F. Barbara Orlans (Beauchamp et al., "The Human Use of Animals," 2008), seeks a middle path. This framework accepts that animals have morally relevant interests and that causing them harm requires serious justification, but argues that the moral weight of animal interests varies with the species, the severity of the harm, and the nature of the relationship between humans and animals. Laboratory animals bred for research exist within a different moral context than wild animals or companion animals; researchers have positive duties of care toward them (ensuring welfare, minimising suffering) that arise from this relationship. This framework is arguably the one most closely aligned with current regulatory requirements — EU Directive 2010/63/EU, for example, requires a formal harm-benefit analysis for each project, essentially demanding utilitarian-style reasoning within a deontological framework of minimum welfare standards.

For the toxicologist, these philosophical positions have practical implications. The utilitarian analysis requires honest assessment of whether a given *in vivo* study is genuinely necessary — i.e., whether validated *in vitro* or *in silico* alternatives exist, whether the study is scientifically valid and likely to produce interpretable results, and whether the information gained is proportionate to the animal suffering incurred. The deontological critique should prompt constant questioning of whether common practices are justified by convention rather than scientific necessity. The contextual approach provides the clearest guidance for day-to-day practice: identify the minimum number of animals needed, use the least sentient species capable of answering the scientific question, apply all available refinements, and treat the animals under your care as beings whose welfare matters intrinsically, not merely as regulatory

obligations. This integrated ethical stance is increasingly reflected in the literature — a survey by Kilkenny et al. (2009, PLOS ONE) found that 87% of papers in three high-impact journals failed to mention the 3Rs, a finding that has since driven widespread adoption of ARRIVE reporting standards.

1.9 Case Study: The ARRIVE 2.0 Lifecycle

Case Study: The ARRIVE 2.0 Lifecycle Applied to a 28-Day Repeated-Dose Toxicity Study

This case study illustrates how ARRIVE 2.0 guidelines structure the complete lifecycle of an in vivo toxicity study — from planning through reporting — using a hypothetical 28-day oral gavage study with cadmium chloride (CdCl₂) in Wistar rats as an example.

Planning phase (pre-registration):

The study director at USTO-MB formulates the research question: "What is the NOAEL for CdCl₂ administered daily by oral gavage to adult Wistar rats for 28 days, assessed by clinical chemistry, haematology, and histopathology of kidney and liver?" The study is registered on preclinicaltrials.eu (registration number: PCT/DZ/2025/0048) with a pre-specified primary endpoint (serum creatinine as marker of nephrotoxicity), four dose groups (0, 0.3, 1.0, 3.0 mg CdCl₂/kg/day), and a statistical analysis plan (one-way ANOVA with Dunnett's post-hoc test, $\alpha=0.05$, minimum effect size $\Delta=25\%$ of control mean). Power calculation using G*Power 3.1 (Faul et al., 2009) with $f=0.4$ (large effect), $\alpha=0.05$, 80% power, 4 groups yields $n=7$ per group. Given expected 10% attrition, $n=8$ animals per group are enrolled (32 animals total: 16 males, 16 females).

Conduct phase:

Animals: Wistar rats, 6–8 weeks of age, 180–220g body weight, obtained from a registered breeding facility. Acclimation: 7 days under study conditions before dosing begins. Randomisation: animals are block-randomised by body weight within sex to equalise group mean weights at study start. Blinding: the histopathologist and clinical chemistry technician are blinded to group allocation until after data coding review.

Daily observations: body weight (twice weekly), clinical signs (daily), food consumption (weekly). Terminal endpoints at day 28: blood collection by cardiac puncture under isoflurane anaesthesia; serum prepared for ALT, AST, ALP, creatinine, BUN, total protein, albumin, glucose; plasma for haematology (CBC with differential); kidneys and liver harvested for histopathology (H&E, PAS staining; Prussian blue for Cd deposits). Cd concentration measured in kidney cortex by AAS (following acid digestion).

Reporting phase (ARRIVE Essential 10):

(1) Study design: prospective, randomised, blinded, dose-response. (2) Sample size: $n=8$ /group/sex, justified by power calculation. (3) Inclusion criteria: male and female Wistar rats 180–220g, healthy; exclusion: body weight deviation $>10\%$ from group mean at baseline. (4) Randomisation: block randomisation by body weight, Excel RAND() function. (5) Blinding: histopathologist and chemist blinded; unblinded at statistical analysis. (6) Outcome measures: primary — serum creatinine; secondary — ALT, AST, kidney histopathology Cd deposits. (7) Statistical methods: one-way ANOVA, Dunnett's post-hoc, Levene's test for homogeneity, BMDL10 for creatinine using BMDS 3.3. (8) Animals: Wistar Crl:WI (Han), male and female, 6–8 weeks, obtained from ITELV (Algérie), SPF status. (9) Procedures: oral gavage, vehicle 0.9% NaCl, volume 5 mL/kg, daily $\times$ 28 days; terminal procedure under isoflurane. (10) Results: primary endpoint (serum creatinine) presented as mean $\pm$ SD with ANOVA p-values; dose-response curve fitted by BMDS; NOAEL and LOAEL stated with 95% CIs. Adverse events (1 animal died at 3 mg/kg/day on day 18 from aspiration pneumonia, unrelated to treatment) reported.

The ARRIVE 2.0 Recommended Set additionally requires reporting on: housing conditions (IVC, 12h/12h cycle, 22°C, FELASA Category B); pre-study health monitoring (quarterly sentinel testing, last negative screen Q1 2025); animal welfare impacts (no animals exceeded mild severity classification); pre-registration reference (PCT/DZ/2025/0048); raw data availability (data deposited at OSF.io under DOI 10.17605/OSF.IO/XXXXX).

This structured approach ensures that the study is reproducible, publishable in high-impact journals (Archives of Toxicology, Toxicological Sciences), and provides the regulatory-grade data quality required for submission to the MESRS or for inclusion in an OECD-format dossier.

Animal experimentation occupies a contested but still essential position in toxicological science. A balanced view requires honest acknowledgement of both its unique contributions and its significant limitations.

Advantages	Disadvantages
Integrated whole-organism physiology: captures ADME, immune, endocrine, and CNS interactions simultaneously	Ethical concerns: pain, distress, and death of sentient beings
Regulatory gold standard: accepted by all major regulatory agencies (EMA, FDA, OECD)	Species differences: rodent metabolism often differs from humans (CYP profiles, receptor pharmacology)
Enables chronic and multigenerational studies impossible in vitro	High cost and long duration, especially for carcinogenicity and reproduction studies
Allows measurement of complex endpoints: behaviour, reproduction, carcinogenesis	Poor translatability: 40–60% false positive rate for human carcinogens in rodent bioassays (Gold et al.)
Validates mechanistic hypotheses generated in vitro	High dose levels used in studies may not reflect human environmental exposure patterns
Provides NOAEL/LOAEL for regulatory risk assessment	Low throughput: tested chemicals remain a small fraction of the chemical inventory

Table 1.3. Advantages and disadvantages of animal experimentation — a balanced assessment.

Key Take-Home Messages

- The 3R principle (Russell & Burch, 1959) provides the ethical and scientific framework for all animal experimentation, now legally mandated in EU Directive 2010/63/EU.
- EU Directive 2010/63/EU requires project authorisation, severity classification, and retrospective assessment for all animal experiments involving vertebrates.
- Algeria is developing its regulatory framework for animal research; Loi 12-06 (2012) provides the legal basis for ethics oversight. Researchers must comply with institutional ethics committees and OECD GLP principles.
- Housing and husbandry parameters (temperature, humidity, light cycle, space, enrichment) must be standardised and documented for every in vivo study.
- AAALAC accreditation and FELASA training requirements define international standards of laboratory animal care and competence.
- Animal models remain indispensable for regulatory toxicology but have well-recognised translational limitations that motivate the development of alternative methods.

Further Reading

Russell WMS, Burch RL. The Principles of Humane Experimental Technique. London: Methuen; 1959. Reprinted: Universities Federation for Animal Welfare; 1992.

European Parliament and Council. Directive 2010/63/EU on the protection of animals used for scientific purposes. Official Journal of the European Union. 2010;L276:33–79.

Percie du Sert N et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. PLOS Biology. 2020;18(7):e3000410. doi:10.1371/journal.pbio.3000410

National Research Council. Guide for the Care and Use of Laboratory Animals. 8th ed. Washington DC: National Academies Press; 2011.

FELASA Working Group. FELASA recommendations for the health monitoring of mouse, rat, hamster, guinea pig and rabbit colonies. Lab Anim. 2014;48(3):178–192. doi:10.1177/0023677213516312

Part 2 — In Vivo Toxicity Testing

2.1 Learning Objectives

- Distinguish acute, short-term, and long-term toxicity studies by exposure duration, endpoints, and applicable OECD Test Guidelines.
- Describe OECD TG 420, 423, and 425 for acute oral toxicity and explain why TG 401 (classical LD₅₀) was deleted.
- Explain the design of the 28-day (TG 407) and 90-day (TG 408/409) repeated-dose studies, including endpoints for clinical chemistry and histopathology.
- Describe the OECD reproductive and developmental toxicity testing battery (TG 414, 415/416, 443 EOGRTS).
- Derive NOAEL and LOAEL from in vivo dose-response data and explain their use in risk assessment.
- Outline how ADME (toxicokinetics) considerations should inform in vivo study design.

2.2 Introduction

In vivo toxicity testing — the systematic exposure of living animals to chemical substances under defined experimental conditions — remains the regulatory foundation of chemical hazard assessment. Despite rapid progress in alternative methods (Parts 3 and 6), in vivo studies are currently required for most endpoints in regulatory submissions to ECHA, EMA, US-EPA, and equivalent national agencies. The OECD has developed a comprehensive library of harmonised Test Guidelines that specify the experimental conditions, endpoints, and reporting requirements for each category of in vivo study.

This part surveys the three principal categories of in vivo testing — acute, short-term (subacute/subchronic), and long-term (chronic/carcinogenicity/reproductive) — following the OECD guideline numbering system as the primary organisational framework.

2.3 Acute Toxicity Tests

2.3.1 Historical Context and the Demise of TG 401

For most of the 20th century, acute systemic toxicity was quantified by the LD₅₀ (median lethal dose), introduced by Trevan (1927) and formalised in OECD TG 401. The classical LD₅₀ required large numbers of animals (typically 40–200 per substance) to generate a precise statistical estimate of the dose killing 50% of animals. TG 401 was deleted in 2002 following recognition that: (i) a precise LD₅₀ value is not required for GHS hazard classification, (ii) the large animal numbers violate the 3R principle, and (iii) three alternative approaches with equivalent regulatory utility and far fewer animal numbers had been validated (Karmaus et al., 2022).

2.3.2 OECD TG 420: Fixed Dose Procedure

Full study protocol for OECD TG 420 (Fixed Dose Procedure): The FDP uses fixed oral doses of 5, 50, 300, and 2000 mg/kg. An optional sighting study uses 1 animal per sex per dose level to identify the starting dose. The main study begins at the dose predicted to produce evident toxicity but not mortality. A single group of 5 female animals (females preferred per TG guidance due to historical data demonstrating equal sensitivity) receive the starting dose; if no mortality and minimal toxicity is observed at 24, 48, and 72 h, the next higher dose group is tested 3–7 days later. Each dose group is observed for 14 days with daily clinical signs, body weight on days 0, 7, and 14, necropsy with gross pathology. The GHS category is assigned based on the dose that produces evident toxicity: Category 1 (<5 mg/kg), Category 2 (5–50 mg/kg), Category 3 (50–300 mg/kg),

Category 4 (300–2000 mg/kg), or Not Classified (>2000 mg/kg). The FDP reduces animal use to a minimum of 4–6 animals per substance compared to 40–50 animals for the classical TG 401 LD₅₀ test.

For OECD TG 423 (Acute Toxic Class Method):

The ATC method classifies substances into one of six GHS categories using groups of 3 animals per sex starting at a defined dose (25, 200, 2000, or 5000 mg/kg per class). The decision on which dose to use next is determined by a sequential decision tree based on the mortality pattern in the first group. If 0/3 animals die, the next higher dose is tested; if 2/3 or 3/3 die, the next lower dose is tested; if 1/3 dies, an additional group of 3 animals is tested at the same dose. Classification is based on the overall pattern across groups. The ATC method requires only 6–14 animals per substance and avoids the ethical problems of classical LD₅₀ determination by focusing on classification rather than exact dose estimation.

In TG 420 (OECD, 2001), a fixed dose — one of 5, 50, 500, or 2000 mg/kg — is selected based on available information and administered sequentially to small groups of animals. The endpoint is not lethality per se but evident toxicity at the selected dose, allowing classification into GHS categories without determining the LD₅₀. If no toxicity is observed at 2000 mg/kg, no further testing is required for GHS Category 5 allocation.

2.3.3 OECD TG 423: Acute Toxic Class Method

TG 423 uses groups of three animals at one of the fixed doses (5, 50, 300, 2000 mg/kg). The outcome in the first group (0, 1, 2, or 3 deaths) determines whether to proceed to a second group at the same or different dose. A maximum of three groups (9 animals) is typically sufficient to classify the substance into one of the GHS categories.

2.3.4 OECD TG 425: Up-and-Down Procedure (UDP)

Detailed TG 425 Up-and-Down Procedure (UDP) — Worked Example:

The UDP uses a 3.16-fold dose progression (equivalent to one half-log₁₀ unit), with a default starting dose of 175 mg/kg for substances with no available information. The procedure employs the following decision rules: if the first animal dies, the next animal receives 55 mg/kg (one step down); if the first animal survives, the next receives 550 mg/kg (one step up). This continues sequentially until 5 animals have been tested in the final population (or until a reversal is observed and 4 additional animals have been dosed following that reversal).

Worked	example	with	simulated	dataset:
Animal 1:	175 mg/kg	→	Survived	→ step up
Animal 2:	550 mg/kg	→	Died	→ step down
Animal 3:	175 mg/kg	→	Survived	→ step up
Animal 4:	550 mg/kg	→	Died	→ step down
Animal 5: 175 mg/kg → Survived → study stop (reversal criterion met)				

Dose–response pattern: S(175), D(550), S(175), D(550), S(175)

Maximum likelihood estimation using the UDP maximum likelihood software (AOT425StatPgm.exe, USEPA/OPPT): the dose–response is modelled as a cumulative logistic distribution, fitting μ (mean log dose) and σ (slope) parameters to the binary response data. The MLE estimate yields:

LD₅₀ = $\text{antilog}_{10}(\mu)$ = $\text{antilog}_{10}(2.50)$ = 316 mg/kg

95% Confidence Interval = $\text{antilog}_{10}(\mu \pm 1.96\sigma/\sqrt{n})$ = 178–562 mg/kg

GHS classification: with LD₅₀ = 316 mg/kg, the substance falls into Category 4 (GHS Category 4: 300–2000 mg/kg, "Harmful if swallowed"). Signal word: "Warning." Hazard statement: H302. This would trigger labelling with the health hazard pictogram and require registration under REACH as Category 4 acute oral toxic.

The 95% CI of 178–562 mg/kg reflects the inherent statistical uncertainty of estimating an LD₅₀ from only 5 animals; the UDP achieves this level of precision with approximately 7 animals on average

(compared to 40–50 for TG 401), demonstrating a substantial reduction in animal use while maintaining regulatory acceptability. Regulatory commentary: ECHA's Guidance on Information Requirements (R.7a) accepts UDP data as the primary acute oral toxicity testing method for REACH registration dossiers, with preference over both FDP and ATC in cases where an exact LD₅₀ estimate is needed for classification in a borderline category.



Figure 2.E1. OECD TG 425 Up-and-Down Procedure worked example. Left: sequential dose-outcome pattern for 5 animals (S=survived, D=died) with starting dose 175 mg/kg and 3.16-fold progression. Right: probit dose-response curve fitted to the binary data, with maximum likelihood LD₅₀ estimate of 316 mg/kg (95% CI: 178–562 mg/kg, shaded).

TG 425 (OECD, 2008) is the most statistically efficient of the three modern acute oral test methods. Starting from a defined initial dose (typically 175 mg/kg), one animal is dosed at a time; if it survives, the next animal receives a higher dose (factor of 3.2); if it dies, the next animal receives a lower dose. The sequence continues until a stopping rule is met (typically 3 reversals or 5 animals). The LD₅₀ and 95% confidence interval are then calculated by maximum likelihood using EPA software (AOT425StatPgm). UDP typically uses 6–9 animals per sex — a dramatic reduction from TG 401.

Feature	TG 401 (deleted)	TG 420	TG 423	TG 425 (UDP)
Year	Deleted 2002	2001	2001	2008
Animals/sex	40–200	≤5	≤9	6–9
LD ₅₀ estimate	Precise	Not required	Not required	Yes (MLE)
GHS classification	Yes	Yes	Yes	Yes
Starting dose	Range	Fixed levels	Fixed levels	Adaptive
3R compliance	Poor	Good	Good	Best

Table 2.1. Comparison of modern acute oral toxicity test guidelines (TG 420, 423, 425) with the deleted TG 401.

2.3.5 Dermal and Inhalation Acute Toxicity

Acute dermal toxicity is assessed under OECD TG 402, which uses a 24-hour semi-occlusive dermal application in rats or rabbits at doses up to 2000 mg/kg. Acute inhalation toxicity is covered by TG 403 (traditional LC₅₀), TG 433 (Acute Toxic Class for inhalation), and TG 436 (UDP for inhalation), with exposure durations of 4 hours in nose-only (preferred) or whole-body inhalation chambers. Endpoints include clinical signs, body weight, and gross pathology; mortality within a 14-day observation period determines the GHS inhalation category.

2.3.6 GHS Classification Based on Acute Oral Toxicity

The UN Globally Harmonised System of Classification and Labelling of Chemicals (GHS) classifies substances into five acute toxicity categories based on LD₅₀ (oral):

GHS Category	LD ₅₀ oral (rat, mg/kg bw)	Signal word	Example
--------------	---------------------------------------	-------------	---------

1 (Fatal)	<5	DANGER	Botulinum toxin, ricin
2 (Fatal)	5–50	DANGER	Dioxin (TCDD)
3 (Toxic)	50–300	DANGER	Methanol
4 (Harmful)	300–2000	WARNING	DDT
5 (Low)	2000–5000	(not required)	Ethanol

2.4 Short-Term Toxicity Tests

2.4.1 28-Day Repeated Dose Study: OECD TG 407

Full clinical pathology panel for OECD TG 407 (28-Day Repeated-Dose Oral Toxicity Study in Rodents):

Clinical chemistry (serum): Alanine aminotransferase (ALT) — hepatocellular damage marker; Aspartate aminotransferase (AST) — hepatic and muscle damage; Alkaline phosphatase (ALP) — cholestasis and bone turnover; γ -Glutamyl transferase (γ -GT) — biliary epithelial damage; Blood urea nitrogen (BUN) — renal function and protein catabolism; Creatinine — glomerular filtration rate marker; Total protein — general health and liver synthesis; Albumin — liver synthetic function and nutritional status; Total bilirubin (direct + indirect) — hepatic and haemolytic disease; Cholesterol — lipid metabolism and PPAR- α activation; Triglycerides — lipid metabolism; Glucose — carbohydrate metabolism; Calcium; Phosphorus; Sodium; Potassium; Chloride.

Haematology (whole blood EDTA): Red blood cells (RBC); Haemoglobin (Hb) concentration; Haematocrit (Hct) or packed cell volume; Mean corpuscular volume (MCV); Mean corpuscular haemoglobin (MCH); Mean corpuscular haemoglobin concentration (MCHC); Red cell distribution width (RDW); Platelet count; Reticulocyte count (% and absolute); White blood cell count (WBC) with differential: neutrophils, lymphocytes, monocytes, eosinophils, basophils.

Urinalysis (overnight metabolic cage collection, ~16h): Volume; Colour and appearance; pH; Specific gravity; Glucose (dipstick + enzymatic confirmation); Protein (dipstick + turbidimetric); Ketones; Bilirubin; Urobilinogen; Microscopy: casts (hyaline, granular, cellular), crystals, cells (erythrocytes, leucocytes, epithelial cells).

Histopathology checklist for OECD TG 407 necropsy — minimum 40+ tissues: Mandatory examination (all groups): liver, kidneys, adrenals, spleen, thymus, lymph nodes (mesenteric and mandibular), heart, lungs, trachea, oesophagus, forestomach, stomach, duodenum, jejunum, ileum, cecum, colon, rectum, brain (three levels: cerebrum, cerebellum, brainstem), spinal cord, pituitary, thyroid/parathyroid, gonads (testes, ovaries), uterus, vagina, mammary gland, prostate/seminal vesicles, epididymis, bladder, sciatic nerve, skeletal muscle, sternum with marrow, femur with marrow, skin, eyes. Additional organs at discretion: nasal turbinates (inhalation studies), salivary glands, pancreas, aorta. Organs are weighed before fixation: liver, kidney, spleen, thymus, adrenals, heart, brain, testes, ovaries, uterus. Organ weight ratios (to body weight and to brain weight) are calculated. All tissues fixed in 10% neutral buffered formalin (NBF) for ≥ 24 h before processing, embedded in paraffin, sectioned at 4–5 μ m, and stained with haematoxylin-eosin (H&E). Special stains as needed: PAS (glycogen, basement membranes), Prussian blue (iron), Masson trichrome (fibrosis), Oil Red O (lipids, requires frozen sections).

The 28-day oral repeated dose study (OECD TG 407, updated 2008) exposes groups of rats (5/sex/dose, minimum 3 dose groups + control) to the test substance by oral gavage or dietary incorporation for 28 consecutive days. Endpoints include daily clinical observations, weekly body weights, food and water consumption, haematology (RBC, WBC differential, platelets, Hb), clinical chemistry (ALT, AST, ALP, BUN, creatinine, albumin, cholesterol), urinalysis, organ weights (liver, kidneys, adrenals, testes/ovaries, thyroid), and histopathology of a defined list of organs. A reversibility satellite group may be added (10 animals/sex/top dose) to assess recovery after a 14-day treatment-free period.

The TG 407 study generates the NOAEL and LOAEL that are used to: set the starting dose for the 90-day study; derive provisional safe doses for human health risk assessment; and identify target organs requiring focused histopathological evaluation in longer studies.

2.4.2 90-Day Repeated Dose: OECD TG 408 (Rodent) and TG 409 (Non-Rodent)

The 90-day subchronic study (TG 408 for rodents, TG 409 for non-rodents) expands on TG 407 in duration and scope. Group sizes are typically 10/sex/dose. Endpoints are the same as TG 407 plus ophthalmological examination, nerve conduction (for neurotoxic substances), and more comprehensive neuropathology. The NOAEL from the 90-day study is the primary input to Reference Dose (RfD) and Acceptable Daily Intake (ADI) derivation under the classical uncertainty factor approach ($RfD = NOAEL / 100$ for a non-OECD route to RfD derivation, using UF of 10 for interspecies and 10 for intraspecies variation).

Parameter	TG 407 (28-day)	TG 408 (90-day rodent)	TG 409 (90-day non-rodent)
Duration	28 days	90 days	90 days
Species	Rat (preferred)	Rat (preferred)	Dog or non-rodent
Group size	5/sex/dose	10/sex/dose	4/sex/dose
Haematology	Yes	Yes (expanded)	Yes
Clinical chemistry	Yes	Yes (expanded)	Yes
Urinalysis	Yes	Yes	Yes
Histopathology	Defined organ list	Comprehensive	Comprehensive
Ophthalmology	Optional	Required	Required
NOAEL derivation	Preliminary	Primary endpoint	Primary endpoint

Table 2.2. Overview of subacute and subchronic repeat-dose OECD test guidelines.

2.5 Long-Term Toxicity Tests

2.5.1 Chronic Toxicity: OECD TG 452

TG 452 (Chronic Toxicity Study) involves exposure of rats for at least 12 months (sometimes up to 24 months in mice). It uses larger group sizes (20/sex/dose) and comprehensive endpoint batteries similar to TG 408 but conducted at quarterly intervals. The principal purpose is to identify cumulative target organ toxicity, particularly hepatotoxicity, nephrotoxicity, and haematopoietic effects, at dose levels achievable in long-term dietary exposure.

2.5.2 Carcinogenicity and Combined Chronic/Carcinogenicity: TG 451, TG 453

Carcinogenicity study design (OECD TG 451 — Carcinogenicity Study): The standard NTP-style design uses three dose groups plus a vehicle control ($n=50/\text{sex}/\text{group}$) for a period of 24 months in rats and 18–24 months in mice. Dose selection follows the maximum tolerated dose (MTD) principle: the high dose should produce minimal toxicity ($\leq 10\%$ body weight reduction, no effect on survival below survival at the control group) but be clearly above the human exposure range. The MTD is empirically determined in a 90-day dose range-finding study (OECD TG 408). For a new pesticide active substance, dose selection might proceed as follows: 90-day NOAEL = 50 mg/kg/day; MTD from dose-finding = 300 mg/kg/day; carcinogenicity study doses = 10, 50, 150 mg/kg/day.

Interim sacrifices (10/sex/group) at 12 months are used to assess the early histopathological progression, particularly for target organs identified in the 90-day study. At terminal sacrifice, all animals undergo gross necropsy, organ weights, and comprehensive histopathological examination of 45+ tissues.

Tumour incidence statistical analysis:
For each tumour type, pairwise comparisons between treated and control groups use: (1) the Poly-k test (Bailer & Portier, 1988) — a survival-adjusted test that accounts for early deaths without artificially

inflating tumour rates; (2) the Cochran-Armitage trend test for detecting dose-related increases; (3) the Fisher exact test for pairwise comparisons when overall n is small. A positive carcinogenicity result requires: a statistically significant increase ($p < 0.05$ one-tailed) in a specific tumour type in at least one sex/species, a dose-response relationship, and biological plausibility. The IARC Monograph review process uses these criteria alongside epidemiological evidence to classify agents into Groups 1, 2A, 2B, or 3.

TG 451 (Carcinogenicity Study) exposes rats or mice for the majority of their expected lifespan (rats: 24 months; mice: 18 months), with 50/sex/dose. Animals are terminated at study end, and comprehensive gross pathology and histopathology of >40 tissues are conducted. Tumour incidence is compared between treated and control groups using pairwise statistical tests (Fisher's exact, Cochran-Armitage trend test). TG 453 combines chronic toxicity and carcinogenicity in a single study design, reducing total animal use while providing both endpoints.

Non-genotoxic carcinogens — substances that promote tumour development without directly damaging DNA — are a particular challenge because their risk cannot be assessed using the linear non-threshold model. An OECD expert consensus document (Jacobs et al., 2020) outlines an integrated approach for testing and assessment of non-genotoxic carcinogens.

2.5.3 Reproductive Toxicity: TG 415, 416, 443 EOGRTS

EOGRTS (Extended One-Generation Reproductive Toxicity Study, OECD TG 443) — Full Cohort Design:

The EOGRTS represents the current state-of-the-art for comprehensive reproductive toxicity assessment. It is required under REACH for high-tonnage substances (>10 tonnes/year) and replaces the two-generation study (TG 416) for most purposes. The study design involves:

Parental (P0) generation: 22 females and 22 males per group (control + 3 dose groups), dosed for 10 weeks pre-mating (to allow spermatogenesis and oogenesis cycles to complete), throughout mating, gestation (21 days), and lactation (until postnatal day 21). Endpoints: body weight, food consumption, clinical signs, reproductive performance (conception rate, gestation length, litter size, sex ratio, pup weight on PND 1, 4, 7, 14, 21), maternal behaviour score.

F1 Cohort 1A (Reproductive Toxicity): Minimum 20 pups/sex/group selected from the F1 generation at weaning (PND 21). These animals are raised to adulthood (PND 90+) and mated to produce the F2 generation. Endpoints: body weight development, sexual maturation markers (anogenital distance (AGD) at birth and PND 4; vaginal opening (VO) in females; preputial separation (PPS) in males; first estrus in females), F2 litter parameters, reproductive organ weights and histopathology (testis, epididymis, seminal vesicles, prostate, ovary, uterus), sperm analysis (count, motility, morphology), serum hormone levels (testosterone, LH, FSH in males; estradiol, progesterone, LH, FSH in females).

F1 Cohort 1B (Systemic Toxicity): 10/sex/group from the F1 generation, run as a 90-day systemic toxicity study with full clinical chemistry, haematology, urinalysis, organ weights, and histopathology. An optional TK satellite group of 3/sex/group provides blood PK data.

F1 Cohort 2A (Developmental Neurotoxicity): 10/sex/group, subjected to a battery of developmental neurotoxicity (DNT) tests: auditory startle habituation (PND 22 and 60), open-field locomotor activity (PND 13, 17, 21, 60), cognitive testing (Morris water maze or novel object recognition, PND 60+), brain morphometry at PND 22 and 70, neuropathology including immunostaining for GFAP (astrocytes), NeuN (neurons), MBP (myelin).

F1 Cohort 2B and 3 (Developmental Immunotoxicity / Additional DNT): These cohorts are optional trigger-based cohorts activated when signals of immunotoxicity or additional neurotoxicity are observed in cohorts 1A/2A. Cohort 2B tests NK cell activity, lymphocyte subset analysis (by flow cytometry),

and thymus/spleen histopathology. Cohort 3 provides additional DNT endpoints to distinguish compound-specific from general toxicity-related effects.

Key sexual maturation markers and interpretation: AGD (anogenital distance, measured with calipers from the anus to the genital tubercle, normalised to the cube root of body weight to correct for body size differences) is sensitive to androgen or anti-androgen exposure in males (reduced AGD indicates anti-androgenic activity) and to estrogen agonism in females (increased AGD in females indicates androgenic activity). VO (vaginal opening) age is accelerated by estrogenic substances and delayed by anti-estrogenic or anti-androgenic compounds. PPS age is accelerated by androgens and delayed by anti-androgens. These functional endpoints are among the most sensitive indicators of endocrine disruption and are required by ECHA and EFSA for endocrine disruptor assessment under Regulation (EU) 2018/605.

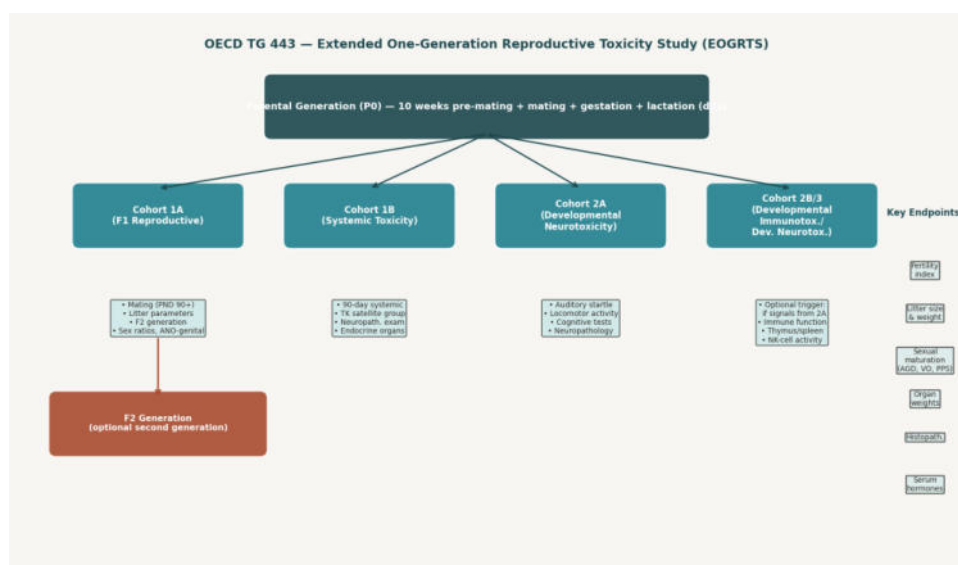


Figure 2.E2. OECD TG 443 Extended One-Generation Reproductive Toxicity Study (EOGRTS) cohort design. P0 = parental generation (pre-mating through lactation); F1 cohorts 1A, 1B, 2A, 2B/3 serve distinct endpoints (reproductive, systemic, developmental neuro-, immunotoxicity respectively). Optional F2 generation is produced by cohort 1A only. Key sexual maturation markers (AGD, VO, PPS) are indicated.

The Reproductive Toxicity Test battery addresses effects on fertility, pregnancy, parturition, lactation, and offspring development. The Extended One-Generation Reproductive Toxicity Study (EOGRTS, TG 443, 2018) is the current state-of-the-art design. It exposes P0 rats for 2 weeks before mating through weaning of the F1 generation, and can be extended to include an F2 generation (Cohort 1B trigger). Optional cohorts allow measurement of developmental neurotoxicity (Cohort 2), developmental immunotoxicity (Cohort 3), and juvenile toxicity.

2.5.4 Developmental Toxicity: OECD TG 414

OECD TG 414 (Prenatal Developmental Toxicity Study): The study design involves dosing pregnant rats or rabbits from implantation (gestation day GD 5–6 in rat, GD 6–7 in rabbit) to the day before scheduled caesarean section (GD 19–20 in rat, GD 28–29 in rabbit). A minimum of 20 pregnant animals per group (control + 3 dose groups) is required. Dose groups should bracket the no-effect and maternally toxic dose levels.

Maternal toxicity endpoints: body weight and food consumption throughout gestation, clinical signs, organ weights at sacrifice (liver, kidney, uterus), histopathology of the same organs.

Developmental endpoints: at caesarean section, all implantations are recorded (live fetuses, late resorptions, early resorptions, pre- and post-implantation losses). Each live fetus is weighed, sexed, and examined for external malformations. Half the litter undergoes skeletal examination (alizarin red

staining after KOH clearing); half undergo visceral examination (Wilson's razor-blade sectioning or fresh microdissection).

Classification of developmental effects:
 — Malformations (structural abnormalities): rare, irreversible alterations with serious adverse consequences (e.g., absent kidney, encephaly, cleft palate, fused vertebrae). Malformations are used to derive the NOAEL for developmental toxicity.
 — Variations (structural alterations): more common, minor deviations from the normal pattern that may or may not be adverse (e.g., supernumerary ribs, delayed ossification). The developmental toxicity of variations is context-dependent: a single extra rib at low incidence is considered a variation; a dose-related increase in extra ribs represents an adverse effect.
 — Delays: slower development relative to controls (e.g., incomplete ossification of skull or sternum bones that complete ossification in controls by GD 21). Delays may normalise postnatally and are distinguished from malformations by their transient character.

Statistical analysis uses the litter as the unit of analysis (not individual fetuses) to avoid pseudoreplication. Proportions (malformation rates, resorption rates) are compared by Fisher's exact test or logistic regression with litter random effect; continuous endpoints (fetal body weight) by mixed-model ANOVA with litter as a random effect.

The Prenatal Developmental Toxicity Study (TG 414, updated 2018) exposes pregnant rats or rabbits from implantation to the day before parturition (GD6 to GD20 in rats). Dams are examined at necropsy; foetuses are examined for external, visceral, and skeletal malformations. TG 414 detects developmental toxicity that might not be visible in adult animals, including teratogenicity, foetal growth retardation, and embryo/foetal lethality.

Test	OECD TG	Duration	Endpoint
Chronic toxicity	452	≥12 months	Cumulative organ toxicity
Carcinogenicity	451	Lifetime (18-24 mo)	Tumour incidence by site/type
Combined chronic/carcinogenicity	453	Lifetime	Both endpoints combined
One-generation reproductive	415	Parental + F1	Fertility, litter parameters
Two-generation reproductive	416	P0 + F1 + F2	Reproductive endpoints across generations
EOGRTS	443	P0 + F1 (± F2)	Reproduction + dev. neuro- & immunotoxicity
Prenatal developmental	414	GD6–20 (rat)	Teratogenicity, foetal parameters

Table 2.3. OECD long-term and reproductive/developmental toxicity test guidelines.

2.6 ADME Considerations in In Vivo Testing Design

Toxicokinetic calculations — worked examples:

AUC calculation by the linear trapezoidal rule:
 Following a single oral dose of 100 mg/kg of a test compound to rats, plasma concentrations (µg/mL) were measured at times 0, 0.5, 1, 2, 4, 6, 8, 12, 24 h: 0, 2.1, 4.8, 6.3, 5.2, 3.4, 2.1, 0.8, 0.1 µg/mL.

$$\begin{aligned} \text{AUC}(0 \rightarrow 24\text{h}) &= \sum [(C_i + C_{i+1})/2 \times (t_{i+1} - t_i)] \\ &= [(0+2.1)/2 \times 0.5] + [(2.1+4.8)/2 \times 0.5] + [(4.8+6.3)/2 \times 1] + [(6.3+5.2)/2 \times 2] + [(5.2+3.4)/2 \times 2] + \\ &\quad [(3.4+2.1)/2 \times 2] + [(2.1+0.8)/2 \times 4] + [(0.8+0.1)/2 \times 12] \end{aligned}$$

$$= 0.525 + 1.725 + 5.55 + 11.5 + 8.6 + 5.5 + 5.8 + 5.4 = 44.6 \mu\text{g}\cdot\text{h/mL}$$

$$\text{AUC}(24 \rightarrow \infty) = \text{Clast}/\lambda_z = 0.1 / (0.693/t_{1/2})$$

First, estimate $t_{1/2}$ from the terminal phase (6→24h): terminal slope $\lambda_z = -\ln(C_{24}/C_6)/t = -\ln(0.1/3.4)/18$
 $= \ln(34)/18 = 3.526/18 = 0.196 \text{ h}^{-1}$; $t_{1/2} = 0.693/0.196 = 3.54 \text{ h}$.

$$\text{AUC}(24 \rightarrow \infty) = 0.1/0.196 = 0.51 \mu\text{g}\cdot\text{h/mL}$$

$$\text{AUC}_{\text{total}} = 44.6 + 0.51 = 45.1 \mu\text{g}\cdot\text{h/mL}$$

Volume of distribution (V_d): following IV dose of 10 mg/kg (= 10,000 $\mu\text{g/kg}$), C_0 (IV) = 3.2 $\mu\text{g/mL}$ (extrapolated); $\text{AUC(IV)} = C_0/\lambda_z = 3.2/0.196 = 16.3 \mu\text{g}\cdot\text{h/mL}$; $V_d = \text{Dose}/C_0 = 10,000 \mu\text{g/kg} / 3.2 \mu\text{g/mL} \times (1 \text{ mL}/1000 \mu\text{L}) \times (\text{body weight consideration}) = 3.1 \text{ L/kg}$. A $V_d > 1 \text{ L/kg}$ indicates extensive tissue distribution, consistent with lipophilic substances that partition into adipose and muscle.

Clearance (CL): $\text{CL} = \text{Dose(IV)} / \text{AUC(IV)} = 10,000 \mu\text{g/kg} / 16.3 \mu\text{g}\cdot\text{h/mL} = 614 \text{ mL/h/kg}$. Expressed as: $\text{CL} = 0.614 \text{ L/h/kg}$ (or $\sim 41 \text{ mL/min/kg}$). This value is compared to hepatic blood flow (85 mL/min/kg in rat) and renal clearance from urinary excretion data to determine the relative contribution of hepatic vs. renal elimination.

Bioavailability (F): $F = (\text{AUC}_{\text{oral}} / \text{AUC}_{\text{iv}}) \times (\text{Dose}_{\text{iv}} / \text{Dose}_{\text{oral}}) = (45.1 / 16.3) \times (10/100) = 0.277 = 27.7\%$. This moderate oral bioavailability suggests significant first-pass hepatic metabolism or incomplete absorption.

Toxicokinetics (TK) — the quantitative study of what the body does to a xenobiotic — is inseparable from the design and interpretation of in vivo toxicity studies. A brief overview of the principal ADME processes and their implications for study design is provided here; more detailed treatment is available in classical references (Casarett & Doull's Toxicology; Klaassen, 2019).

2.6.1 Absorption

The bioavailability (F) of an oral dose depends on absorption from the gastrointestinal tract and first-pass hepatic metabolism. For water-soluble ionics (e.g., lead salts), absorption is concentration- and calcium-dependent; for lipophilic organics (e.g., PCBs, dioxins), absorption is nearly complete via passive diffusion through enterocyte membranes. The route of administration in a toxicity study (gavage vs. diet vs. drinking water) determines peak plasma concentration (C_{max}) and overall systemic exposure (AUC), which must be appropriate for the human exposure scenario being modelled.

2.6.2 Distribution

Volume of distribution (V_d) reflects the extent to which a substance is distributed beyond the plasma compartment into tissues. Highly lipophilic xenobiotics ($\log P > 3$) accumulate in adipose tissue, bone marrow, and brain, leading to chronic retention and potentially non-linear kinetics. Lead (Pb) has a triphasic distribution model: initial soft-tissue phase (half-life days), bone phase ($t_{1/2} \sim 30$ years), and blood phase that reflects ongoing equilibration. The blood-brain barrier protects the CNS from most ionised compounds but is penetrable by lipophilic neutral molecules.

2.6.3 Metabolism and Bioactivation

Cytochrome P450 enzymes (Phase I, primarily in the liver) and conjugation enzymes (Phase II — glucuronidation, sulfation, glutathione conjugation) biotransform xenobiotics to facilitate excretion. A critical toxicological concept is bioactivation: the conversion of an innocuous parent compound to a reactive, toxic metabolite. Classic examples include:

- Benzo[a]pyrene $\rightarrow$ benzo[a]pyrene-7,8-diol-9,10-epoxide (BPDE) by CYP1A1 $\rightarrow$ DNA adducts
- Aflatoxin B1 $\rightarrow$ AFB1-8,9-epoxide by CYP3A4 $\rightarrow$ DNA adducts at guanine N7 $\rightarrow$ G:C $\rightarrow$ T:A transversions
- $\text{CCl}_4 \rightarrow \bullet\text{CCl}_3$ radical by CYP2E1 $\rightarrow$ lipid peroxidation $\rightarrow$ centrilobular hepatic necrosis

2.6.4 Excretion and Half-Life

Renal excretion of polar metabolites and biliary excretion of glucuronide conjugates are the principal elimination routes. Enterohepatic recirculation — reabsorption of biliary conjugates after hydrolysis by gut bacteria — can substantially extend the effective half-life of some compounds (e.g., oestrogens, morphine, bisphenol A glucuronide). Persistent bioaccumulative substances (PBTs) with very long half-lives require lifetime dosing studies to reach steady state.

Parameter	Definition	Typical range	Relevance to study design
C _{max}	Peak plasma concentration	ng–µg/mL	Predicts acute peak toxicity
T _{max}	Time to C _{max}	Minutes to hours	Informs clinical observation timing
AUC	Area under curve (total exposure)	Concentration × time	Dose-independent metric for systemic exposure
t _{1/2}	Elimination half-life	Hours to years	Determines dosing frequency and wash-out period
Cl	Clearance	mL/min/kg	Predicts accumulation at repeated doses
V _d	Volume of distribution	L/kg	Predicts tissue accumulation and bioavailability
F	Bioavailability (oral)	0–100%	Required for dose scaling between routes

Table 2.4. Summary of key toxicokinetic (ADME) parameters and their relevance to in vivo study design.

2.7 Dose Selection, Statistical Design and Sample Size

Statistical design for dose-response studies:

Factorial designs improve information yield per animal by testing multiple variables simultaneously. A 2×3 factorial design (2 sexes × 3 dose levels) in a 28-day study provides data on sex-dose interactions, which can reveal sex-specific metabolic susceptibility. Analysis uses a two-way ANOVA with interaction term.

Randomisation protocol: Animals are assigned to groups using stratified block randomisation based on body weight. Body weights on the day before study start are ranked; animals are assigned to blocks of 4 (or n equal to the number of groups); within each block, animals are randomly allocated to groups using a random number table or computer-generated allocation (e.g., PROC PLAN in SAS or random.sample() in Python). The randomisation code is sealed and held by the QAU until data analysis.

Blinding procedures: At a minimum, the histopathologist and clinical chemist are blinded to group allocation. Coded tissue blocks and blood samples are used; a separate coding sheet is held by the study director and not revealed to laboratory personnel until the statistical analysis plan has been finalised and the database locked.

Sample size calculation example (G*Power-style):
 Study: 4-group dose-response (control + 3 doses), endpoint = ALT activity (IU/L) in 28-day rat study.
 Expected effect: high dose group ALT = 150 IU/L (vs control = 60 IU/L, SD = 30 IU/L). Effect size $f = \sigma_{\text{between}} / \sigma_{\text{within}}$. Between-group SD = $\sqrt{[(0^2 + 30^2 + 60^2 + 90^2)/4]} = \sqrt{[2025]} = 45$; $f = 45/30 = 1.5$ (very large effect). At $\alpha=0.05$, power=0.80, 4 groups: G*Power one-way ANOVA gives n=4 per group. However, for a more conservative estimate using $f=0.5$ (medium-large): n=17 per group. Standard

toxicology studies use $n=10/\text{sex}/\text{group}$ for 28-day rodent studies (OECD TG 407), reflecting the balance between statistical power for moderate effects and practical constraints.

For an EC50 determination by non-linear regression (4-parameter logistic), the sample size depends on the number of concentration levels and replicates. A minimum of 5 concentration levels spanning 3–4 $\log_{10}$ units, with $n=4\text{--}6$ per level, provides adequate power for curve-fitting precision (CV of EC50 estimate $<30\%$). Increasing replicates improves the precision of the EC50 but has diminishing returns above $n=8$ per level.

2.7.1 Principles of Dose Selection

Appropriate dose selection is critical for a valid in vivo toxicity study. OECD guidance (Sewell et al., 2022) recommends that the high dose should produce clear evidence of toxicity (10–20% reduction in body weight gain, or organ-specific effects) but not excessive lethality (mortality $<10\%$ in 90-day studies). The low dose should be close to the expected human exposure level (or the anticipated therapeutic dose for pharmaceutical studies). A spacing factor of 3–10-fold between dose levels allows dose-response characterisation.

2.7.2 Sample Size Justification

Power calculations based on historical variability data should always precede study initiation. For a 90-day rat study detecting a 20% change in a continuous endpoint (e.g., liver weight) with 80% power and $\alpha=0.05$, a group size of 10/sex is typically adequate. Smaller group sizes are acceptable for the 28-day study. Online power calculation tools (G*Power, OPTI-DOSE) and OECD Statistical Guidance Document 212 provide practical assistance.

2.7.3 NOAEL and LOAEL Determination

The No Observed Adverse Effect Level (NOAEL) is the highest tested dose at which no statistically or biologically significant adverse effect is observed. The Lowest Observed Adverse Effect Level (LOAEL) is the lowest dose at which an adverse effect is detected. The distinction between 'adverse' and 'non-adverse' adaptive changes requires professional judgment: reversible hepatocellular hypertrophy without functional compromise may be considered non-adverse; necrosis, inflammation, or functional impairment is always adverse.

2.7.4 Benchmark Dose (BMD) Modelling

The Benchmark Dose (BMD) approach (EPA, EFSA) is a superior alternative to NOAEL/LOAEL for quantitative risk assessment. A mathematical dose-response model is fitted to the experimental data and the BMD is defined as the dose corresponding to a pre-specified level of response (typically a 10% increase above background — BMR10). The BMDL (lower 95% confidence limit of the BMD) is used in place of the NOAEL for RfD/ADI derivation. The BMD approach uses all available dose-response data rather than just the NOAEL, providing a more statistically robust and reproducible result. Software: EPA BMDS, EFSA PROAST.

Key Take-Home Messages

- The classical LD₅₀ (OECD TG 401) was deleted in 2002; modern acute toxicity testing uses the Fixed Dose Procedure (TG 420), Acute Toxic Class (TG 423), or Up-and-Down Procedure (TG 425), all with dramatically fewer animals.
- The 28-day (TG 407) and 90-day (TG 408) repeated-dose studies are the workhorses of subacute and subchronic toxicity assessment, generating NOAELs for regulatory risk assessment.
- Chronic studies (TG 452), carcinogenicity bioassays (TG 451/453), and reproductive toxicity studies (TG 443 EOGRTS, TG 414) are required for complete regulatory packages.

- ADME parameters (C_{max}, AUC, t_{1/2}, bioavailability) must inform dose selection and exposure scaling between species.
- The Benchmark Dose (BMD) approach provides a more statistically robust alternative to NOAEL/LOAEL for risk assessment.

Further Reading

Karmaus AL et al. Evaluation of Variability Across Rat Acute Oral Systemic Toxicity Studies. *Toxicol Sci.* 2022;188(1):34–47. doi:10.1093/toxsci/kfac042

Sewell F et al. Recommendations on dose level selection for repeat dose toxicity studies. *Arch Toxicol.* 2022;96(7):1921–1934. doi:10.1007/s00204-022-03293-3

OECD. Test Guideline 425: Acute Oral Toxicity — Up-and-Down Procedure. Paris: OECD Publishing; 2008.

OECD. Test Guideline 443: Extended One-Generation Reproductive Toxicity Study. Paris: OECD Publishing; 2018.

Haber LT et al. Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Crit Rev Toxicol.* 2018;48(5):387–415. doi:10.1080/10408444.2018.1430121

Part 3 — In Vitro Toxicity Evaluation

3.1 Learning Objectives

- Describe and compare the principal in vitro cell models used in toxicology (primary cells, immortalised lines, iPSC-derived models).
- Explain the principles and applications of 3D models (spheroids, organoids) and organ-on-chip systems.
- Select and apply the appropriate cytotoxicity assay (MTT, LDH, neutral red, resazurin) for a given experimental question.
- Describe mechanistic in vitro endpoints for oxidative stress, apoptosis, and mitochondrial dysfunction.
- Explain the principles of the standard in vitro genotoxicity battery (Ames TG 471, micronucleus TG 487, comet, γ H2AX).
- Describe in vitro approaches to endocrine disruption testing (OECD conceptual framework Levels 1–5).
- Critically evaluate the strengths and limitations of in vitro toxicology in the context of regulatory acceptance.

3.2 Introduction

In vitro toxicology — the study of toxic effects in isolated biological systems such as cells, tissues, and subcellular fractions — has undergone a revolution over the past two decades. Driven by the 3R principle, advances in cell biology, and the urgent need for higher-throughput methods to address the chemical space challenge, in vitro approaches now cover an impressive range of toxicological endpoints. This Part surveys the principal cell models, assay platforms, and mechanistic readouts available to the experimental toxicologist, while critically examining their regulatory acceptance.

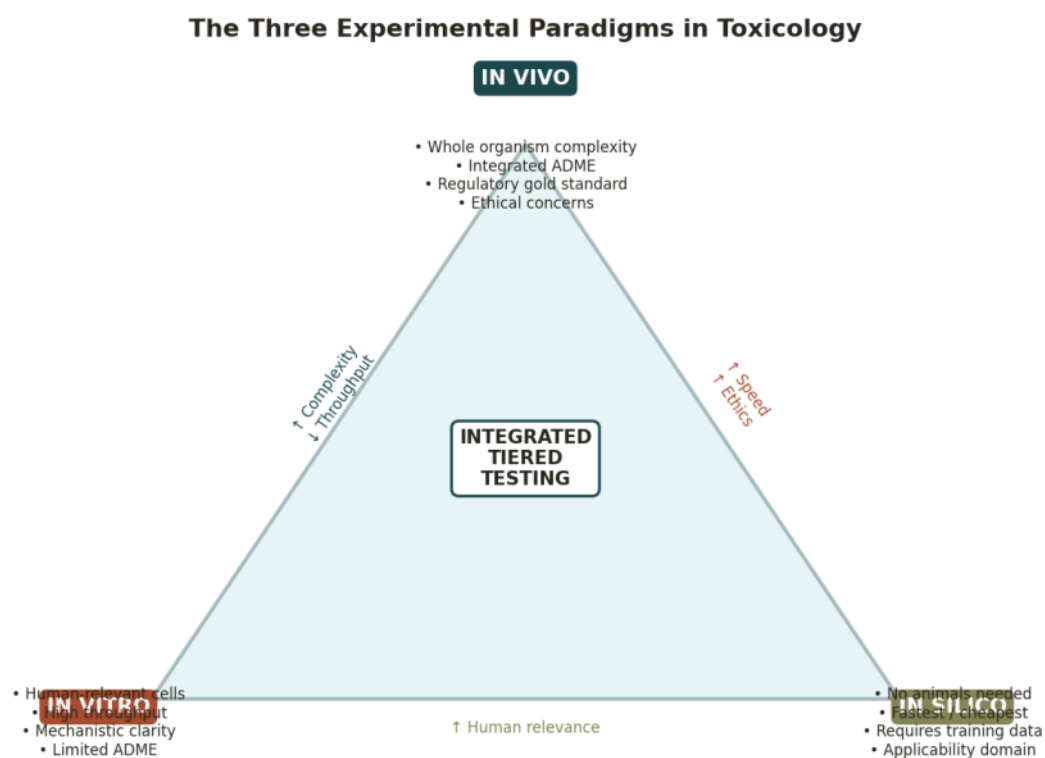


Figure 3.1. The Three Experimental Paradigms in Toxicology. In vitro models occupy the lower left vertex of the triangle, offering high throughput and human cell relevance at the cost of reduced biological complexity compared to in vivo systems. Optimal toxicological practice integrates all three paradigms in a tiered testing strategy. (Original figure.)

3.3 Cell Models

Detailed comparison of cell models used in toxicological research:

HepG2 (human hepatocellular carcinoma): Derived from a 15-year-old male hepatocellular carcinoma patient. Tissue origin: liver. Metabolic competence: very low basal CYP1A2, CYP3A4, CYP2E1 activity (<1% of primary hepatocytes); minimal Phase II (UGT, SULT, GST). Strengths: robust, easily transfected, well-characterised genome, extensive historical database, suitable for cytotoxicity screening and genotoxicity (Comet, micronucleus). Limitations: poor metabolic activation of pro-carcinogens; different drug transport profile from primary hepatocytes. Typical applications: MTT/LDH cytotoxicity, oxidative stress endpoints, p21/p53 pathway studies, Comet assay.

HepaRG (human hepatoma/progenitor): Derived from a hepatocellular carcinoma and bile duct carcinoma. Uniquely, HepaRG cells can be differentiated into hepatocyte-like (80%) and biliary epithelial-like (20%) cells by 2-week culture in DMSO-supplemented medium. Metabolic competence: differentiated HepaRG has CYP3A4, CYP1A2, CYP2E1 activities at 20–50% of primary hepatocyte levels; Phase II enzymes (UGT1A1, SULT1A1, GST) also expressed. Strengths: best metabolically competent hepatocytic cell line for pro-carcinogen activation; stable for months; OECD TG 432 (photogenotoxicity) uses HepaRG. Limitations: requires differentiation protocol (2–4 weeks); not all lots respond equally; mixed population complicates endpoint interpretation. Typical applications: CYP induction (rifampicin, omeprazole), metabolic activation of aflatoxin B1, cyclophosphamide; DILI (drug-induced liver injury) prediction.

Caco-2 (human colorectal adenocarcinoma): Tissue origin: intestinal epithelium. Metabolic competence: limited CYP3A4 activity; moderate Phase II. Strengths: forms tight junctions, polarises into brush-border epithelium morphologically similar to small intestine; used for oral absorption prediction (Papp values). Limitations: overexpresses P-glycoprotein (P-gp/MDR1); growth takes 21 days to full differentiation. Typical applications: oral bioavailability prediction, transporter studies, intestinal toxicity (food additives, nanoparticles).

RPTEC/TERT1 (human renal proximal tubular epithelial cells): Tissue origin: kidney proximal tubule. Metabolic competence: limited. Strengths: retains primary cell morphology and function (organic anion/cation transporters OAT1, OAT3, OCT2); key target cells for nephrotoxicity screening. Limitations: commercially available but costly; requires specific media. Typical applications: cisplatin, gentamicin, cadmium nephrotoxicity; transporter-mediated nephrotoxicity.

SH-SY5Y (human neuroblastoma): Tissue origin: bone marrow neuroblast. Metabolic competence: negligible Phase I. Strengths: neurotransmitter synthesis, dopaminergic differentiation with RA or BDNF. Limitations: immature neuronal phenotype; limited relevance to adult brain toxicity. Typical applications: neurotoxicity screening (MPP⁺, 6-OHDA Parkinson models, organophosphate neurotoxicity).

Primary hepatocyte isolation by two-step collagenase perfusion:
 Step 1 (EDTA wash): The rat liver is perfused in situ via the portal vein with Ca²⁺-free, EDTA-containing HBSS (Hank's balanced salt solution) at 37°C for 8–10 min at 35 mL/min. The EDTA chelates calcium, disrupting the intercellular adhesion complexes (desmosomes, tight junctions). The inferior vena cava is cut to allow outflow. The liver becomes pale and swells.
 Step 2 (collagenase digestion): The liver is then perfused with DMEM containing collagenase Type IV (0.05% w/v) and CaCl₂ (5 mM) at the same flow rate for 8–12 min at 37°C. The collagenase disrupts the extracellular matrix (primarily collagen IV and fibronectin of the space of Disse). The liver becomes

soft and gelatinous. Cell collection: The liver capsule is torn open in ice-cold DMEM; the cell suspension is filtered through a 70 µm nylon mesh to remove connective tissue, then centrifuged at 50×g for 3 min at 4°C (slow centrifugation to pellet hepatocytes while leaving non-parenchymal cells in suspension). The pellet is washed twice. Viability is assessed by trypan blue exclusion: ≥85% viability is acceptable. Cells are plated in collagen-coated dishes at 1.5×10⁵ cells/cm². Typical yield: 2–4×10⁸ cells per rat liver.

3.3.1 Primary Cells

Primary cultures are derived directly from living tissue and retain the differentiated phenotype of the source organ, making them the gold standard for mechanistic toxicology. Human primary hepatocytes, isolated from donor liver by two-step collagenase perfusion, express the full complement of CYP450 enzymes (CYP1A2, CYP2C9, CYP2D6, CYP3A4), phase II conjugating enzymes, and membrane transporters. Their principal limitation is rapid dedifferentiation in conventional 2D monolayer culture (CYP activity falls 50–90% within 48–72 h), restricting their application to acute exposure paradigms (Štampar & Žegura, 2024).

3.3.2 Immortalised Cell Lines

Key cell lines used in hepatotoxicity and general cytotoxicity:

Cell Line	Origin	CYP Activity	Principal Uses
HepG2	Human hepatocellular carcinoma	Very low	General cytotoxicity, genotoxicity, baseline hepatotoxicity
HepaRG	Human hepatoma (bipotent)	High (after DMSO differentiation)	CYP-dependent toxicity, repeat-dose hepatotoxicity
HEK293	Human embryonic kidney	Minimal	Transfection host, receptor expression, cytotoxicity
Caco-2	Human colon adenocarcinoma	Limited (enterocyte)	Intestinal absorption, P-gp transport
A549	Human lung adenocarcinoma	Moderate	Inhalation toxicology, pulmonary endpoints
SH-SY5Y	Human neuroblastoma	Minimal (neural)	Neurotoxicity (dopaminergic, mitochondrial)
MCF-7	Human breast adenocarcinoma	ER+	Endocrine disruption (oestrogenic activity)

Table 3.1. Commonly used cell lines in toxicology with metabolic competence and principal uses.

3.3.3 iPSC-Derived Models

Induced pluripotent stem cell (iPSC)-derived hepatocytes, cardiomyocytes, neurons, and renal tubular cells represent a transformative advance: they are human-derived, patient-specific, and can be produced in large quantities. iPSC-derived hepatocytes (iHeps) express CYP enzymes at levels intermediate between HepG2 and primary hepatocytes. A 2025 systematic review (Shao et al., 2025) identified liver organoids derived from iPSCs as particularly promising for modelling drug-induced liver injury (DILI) in a patient-specific manner. Limitations include: high cost, batch-to-batch variability, and incomplete maturation relative to adult hepatocytes.

3.4 Three-Dimensional Models

3D organoid generation protocols:

Intestinal organoids (Clevers laboratory protocol, modified from Sato et al., 2009, Nature): Small intestinal crypts are isolated from mouse or human biopsies by EDTA chelation (2 mM EDTA in PBS, 30 min on ice). Crypts are pelleted at 200×g for 5 min, resuspended in cold Matrigel (growth factor-reduced, phenol-free; BD Biosciences) at 40 crypts/μL, and plated in 10 μL domes per well (24-well plate). After polymerisation at 37°C for 10 min, IntestiCult Organoid Growth Medium (STEMCELL Technologies) or custom ENR medium (EGF 50 ng/mL, Noggin 100 ng/mL, R-spondin-1 500 ng/mL) is added. Organoids expand over 5–7 days, forming crypt-villus domain structures with stem cells at the crypt base (LGR5⁺) and differentiated enterocytes, goblet cells, enteroendocrine cells, and Paneth cells. Passaging every 5–7 days by mechanical disruption and re-embedding. Applications in toxicology: oral bioavailability of drugs/chemicals; intestinal epithelial barrier integrity (TEER, FITC-dextran permeability assay); gut microbiome interaction; food additive safety (OECD TG consideration in 2023 IATA document for intestinal toxicity).

Liver organoids (human iPSC-derived hepatocyte organoids, Huch et al., 2015, Nature): Adult liver ductal progenitor cells from liver biopsies, or iPSC-derived definitive endoderm → hepatic endoderm → hepatocyte organoids via 21-day directed differentiation protocol. Cultured in Matrigel with hepatocyte expansion medium (HEM: EGF, FGF10, HGF, Wnt3a, Noggin, R-spondin-1, ROCK inhibitor Y-27632). After 14 days, hepatocyte differentiation medium (HDM: Oncostatin M, Dexamethasone, BMP7, TGFβ inhibitor A83-01) is added for 7 further days. Resulting organoids express CYP3A4 (inducible), ALB (albumin), HNF4A, ASGR1 (asialoglycoprotein receptor). Applications: DILI prediction; cholestasis modelling (using bile acid supplementation); CYP induction assays; personalised toxicology (iPSC from patients with specific CYP polymorphisms).

Brain organoids (Lancaster et al., 2013, Nature; cerebral organoids from iPSCs): iPSCs are aggregated into embryoid bodies (EBs) in non-adherent ultra-low attachment plates in EB medium for 6 days. EBs are transferred to neural induction medium (NIM: DMEM/F12, N2, MEM-NEAA, Heparin) for 5 days. Neural EBs are embedded in Matrigel and transferred to spinning bioreactor culture (30–34 rpm) in cerebral organoid medium (COM: DMEM/F12 + neurobasal, B27 without vitamin A, N2, EGF, FGF2, insulin) for 4 weeks. By week 8–10, organoids reach 2–4 mm diameter and contain neural progenitor zones, neurons (cortical, hippocampal), glial cells, and rudimentary cortical layering. Applications in toxicology: developmental neurotoxicity (DNT) of methylmercury, valproic acid, alcohol; SARS-CoV-2 neurotropism; nanoparticle neurotoxicity.

Organ-on-chip

examples:

- Liver-on-chip (Emulate Bio, Inc.): A microfluidic device with two channels separated by a porous membrane (7 μm pores). Hepatocytes (primary or HepG2/HepaRG) line the bottom channel; sinusoidal endothelial cells (LSECs) line the top channel. Continuous perfusion mimics hepatic sinusoid blood flow. Features: zone-specific CYP zonation (Zone 1–3 equivalents achieved by oxygen gradient); bile canaliculi formation; albumin secretion for 30+ days. Application: DILI prediction; bile acid transport; steatosis modelling with oleic acid.
- Gut-on-chip (Ingber laboratory, Wyss Institute): Intestinal epithelial cells (Caco-2 or human intestinal organoid-derived) on flexible PDMS membrane; mechanical stretching mimics peristaltic motion; co-culture with endothelial cells, immune cells, and gut microbiota is possible. Application: oral absorption, gut microbiome interaction, intestinal inflammation modelling.
- Multi-organ chip (HUMIMIC, TissUse GmbH, Berlin): Up to 4 organ modules connected by microfluidic circulation, with defined fluid-to-tissue volume ratios. Enables ADME-T assessment (absorption in gut chip → hepatic metabolism → target organ toxicity) in a single integrated system. HUMIMIC CHIP3 (gut-liver-skin) used in cosmetic ingredient safety assessment in replacement of animal studies under the EU cosmetics regulation.

3.4.1 Spheroids

Hepatic spheroids are self-assembled 3D aggregates (100–400 µm diameter) of hepatocytes that maintain functional bile canaliculi, zone-specific CYP450 expression, and metabolic competence for up to 5 weeks — far exceeding any 2D system. This extended viability makes spheroids well-suited for repeated-dose chronic toxicity studies in vitro. The InSphero and 3D Insight™ platforms provide standardised spheroid formats compatible with 96-well plate readers and high-content imaging.

3.4.2 Organoids

Organoids are stem-cell-derived 3D structures that self-organise into architecturally and functionally organ-like structures: liver organoids develop bile duct networks and hepatocyte-like cells; intestinal organoids recapitulate crypts and villi. Their principal advantage over spheroids is their origin from tissue progenitors rather than differentiated cells, giving superior long-term stability and greater architectural complexity.

3.4.3 Liver-on-Chip and Microphysiological Systems

Organ-on-chip (OoC) devices integrate living cells in microfluidic channels subjected to physiological flow, enabling systemic organ cross-talk unavailable in static cultures. A liver-on-chip containing hepatocytes and Kupffer cells under sinusoidal flow rates (~0.05–0.5 dyn/cm²) more closely replicates the in vivo hepatic microenvironment than any plate-based system. Multi-organ chips connecting liver, gut, kidney, and lung compartments represent the current frontier of in vitro complexity. Regulatory frameworks for OoC are still being developed (FDA NCATS, EU HUMOS project).

3.5 Cytotoxicity Assays

Cytotoxicity assays quantify cell viability or membrane integrity as a function of chemical concentration, enabling IC₅₀ determination. Different assays measure different cellular parameters and are not equivalent:

Assay	Parameter Measured	Endpoint	Principal Strengths	Key Limitations
MTT	Mitochondrial dehydrogenase activity	Formazan absorbance at 570 nm	Simple, cheap, widely validated	False positives with reducing agents; requires cell lysis
MTS	Mitochondrial activity (soluble formazan)	Absorbance (490 nm)	No lysis step; automatable	Same interferences as MTT
WST-1	Cellular reductase activity	Absorbance (450 nm)	Non-toxic to cells; kinetic measurement	Interference by reducing compounds
Neutral Red	Lysosomal integrity	Absorbance (540 nm)	Sensitive for lysosomal toxicants	Unstable dye; requires careful handling
LDH release	Plasma membrane integrity	Absorbance after enzymatic reaction	Reliable marker of necrotic cell death	LDH background in serum; measures late events only
Resazurin (Alamar Blue)	Cellular reductases (non-toxic)	Fluorescence (ex 570/em 585 nm)	Kinetic; non-destructive; high sensitivity	Fluorescence interference by coloured compounds
Real-time impedance (xCELLigence)	Cell adhesion/proliferation	Electrical impedance	Label-free; continuous monitoring	Specialised instrument required; expensive

Table 3.2. Comparison of cytotoxicity assays commonly used in in vitro toxicology.

3.6 Mechanistic In Vitro Endpoints

Detailed MTT assay protocol (OECD TG 129 / ISO 10993-5):

Principle: MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) is a yellow tetrazolium salt that is reduced to purple formazan crystals by mitochondrial dehydrogenases (primarily succinate dehydrogenase, SDH) in metabolically active cells. The amount of formazan produced is proportional to the number of viable cells.

Materials (for 12 students; 96-well plate format):
— MTT stock: 5 mg/mL in sterile PBS (CAS 298-93-1; MW 414.3 g/mol); filter sterilise (0.22 µm); protect from light; store at -20°C; prepare fresh monthly.
— DMSO (CAS 67-68-5): dimethylsulfoxide, molecular biology grade; for formazan solubilisation.
— Cell line: HepG2 (ATCC HB-8065) or HEK293 (ATCC CRL-1573); culture in DMEM high-glucose + 10% FBS + 1% PenStrep.
— Test compound: CdCl₂ (CAS 10108-64-2; MW 183.3 g/mol); stock 100 mM in sterile H₂O; working dilutions in complete medium.
— 96-well flat-bottom tissue culture plates (Corning 3595 or equivalent).
— Plate reader capable of absorbance measurement at 570 nm (reference 630–690 nm).

Step-by-step protocol with timing:
Day 0 (cell seeding): Harvest exponentially growing cells by trypsinisation (0.25% trypsin/EDTA, 3 min at 37°C). Neutralise with complete medium. Count viable cells (trypan blue exclusion; target >90% viability). Adjust to 5×10⁴ cells/mL. Plate 200 µL/well (= 10⁴ cells/well) in a 96-well plate. Leave edge wells empty (fill with 200 µL PBS to reduce evaporation effects). Incubate at 37°C, 5% CO₂ overnight. [~1 hour]

Day 1 (treatment): Prepare a 8-point serial dilution of CdCl₂ in complete medium: 0, 0.1, 1, 3, 10, 30, 100, 300 µM. Prepare vehicle control (equivalent volume of H₂O, typically <0.5% v/v). Aspirate medium from wells. Add 200 µL of each concentration in triplicate (columns 2–4 = concentration A, columns 5–7 = concentration B, etc.). Include: 3 wells of untreated cells (100% viability = positive control); 3 wells of cells with 1% Triton X-100 (0% viability = negative control). Incubate 24 h (or 48 h or 72 h as specified). [~45 min]

Day 2 (MTT addition): Prepare MTT working solution: add 20 µL of 5 mg/mL MTT stock to each well containing 200 µL medium (= final 0.5 mg/mL MTT, 1/10 dilution). Do NOT aspirate medium before adding MTT — cells must still be metabolically active. Incubate 2–4 h at 37°C, 5% CO₂. [~2–4 h]

Formazan solubilisation: Aspirate medium+MTT carefully (vacuum aspiration, angled tip to avoid disturbing formazan crystals). Add 150 µL DMSO per well immediately. Seal plate with adhesive film. Shake on orbital shaker at 300 rpm for 15 min at room temperature until crystals completely dissolved. [~20 min]

Absorbance reading: Read plate at 570 nm (formazan) with reference at 630 nm (subtract background). Ensure the plate is at room temperature and the DMSO formazan solution is homogeneous. [~5 min]

Calculations:

% Cell viability = (OD_{treated} - OD_{blank}) / (OD_{untreated control} - OD_{blank}) × 100
OD_{blank} = wells with DMSO only (no cells)

IC₅₀ = concentration yielding 50% cell viability, derived by non-linear regression (4-parameter logistic model: $y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + (\text{IC}_{50}/x)^{\text{HillSlope}})$)

Common pitfalls:

- (1) Formazan solubility: ensure complete dissolution in DMSO; incomplete dissolution gives falsely low OD values.
- (2) Edge effects: outer wells evaporate faster; use peripheral wells only for blanks/PBS.
- (3) Compound interference: some compounds (coloured, or redox-active) interfere with the MTT assay:

coloured compounds absorb at 570 nm (false positive); vitamin C and other reducing agents can directly reduce MTT without metabolic activity (false positive); compounds that chelate metals may affect mitochondrial enzyme activity independently of cell death (can cause underestimation). For these compounds, use resazurin (Alamar Blue) or LDH release instead. (4) Cell density: too dense ($>2 \times 10^4$ /well) leads to contact inhibition and metabolic quiescence, compressing the dynamic range; too sparse ($<5 \times 10^3$ /well) gives readings close to blank. (5) Timing: MTT incubation >4 h allows extracellular reduction by medium components; stick to 2–3h for standard applications.

Comparison of cytotoxicity assay platforms:
 — Resazurin (Alamar Blue; CAS 62758-13-8): non-toxic fluorescent dye reduced by cellular reductases to resorufin (pink, fluorescent Ex/Em 530/590 nm). Can be read multiple times from the same well (non-destructive). More sensitive than MTT for low cell densities. Less affected by coloured compounds. Suitable for kinetic viability monitoring over several days (repeated reads at 24, 48, 72 h).
 — WST-1 (water-soluble tetrazolium-1): reduced at the cell surface by NADH-dependent plasma membrane oxidoreductases to water-soluble formazan; no DMSO solubilisation step required; read directly at 450 nm after 30 min–2h incubation. More convenient than MTT; slightly less sensitive. Cannot be combined with caspase assays (different detection channel).
 — LDH release assay: measures lactate dehydrogenase (LDH) released into culture supernatant upon membrane rupture (necrosis). Coupled enzyme assay: LDH oxidises lactate $\rightarrow$ pyruvate $\rightarrow$ NADH $\rightarrow$ dye reduction measured at 490 nm. Distinguishes cytostatic (reduced proliferation without death) from cytotoxic (actual cell killing) effects, since LDH release indicates membrane integrity loss. Combined with MTT/resazurin (same plate, different wells), provides complementary mechanistic information: high LDH + low viability = necrotic; normal LDH + low viability = cytostatic or early apoptosis.

3.6.1 Oxidative Stress

Mechanistic in vitro endpoints — detailed protocols:

Oxidative stress — DCFH-DA fluorescence assay:
 2',7'-Dichlorodihydrofluorescein diacetate (DCFH-DA, CAS 4091-99-0) is a cell-permeable ester that is hydrolysed by intracellular esterases to the non-fluorescent DCFH, which is then oxidised by intracellular ROS (primarily H_2O_2 , peroxynitrite, and lipid peroxides) to the highly fluorescent DCF (dichlorofluorescein, Ex/Em 485/535 nm). This assay provides a general index of cellular oxidative stress.

Protocol: seed cells (2.5×10^4 /well in 96-well black plates); add test compound for 24h; aspirate medium; wash with PBS; add DCFH-DA (10 μ M in DMEM without serum) for 30 min at 37°C in the dark; wash $\times 2$ with PBS; add fresh DMEM; read fluorescence immediately at Ex 485/Em 535 nm. Positive control: H_2O_2 (200 μ M, 1h).

Interpretation: ≥ 1.5 -fold increase in DCF fluorescence over vehicle control is considered a positive result. Dose-response relationship should be verified. Note: some compounds can directly oxidise DCFH (chemical interference); verify with acellular control (DCFH-DA + compound without cells).

MitoSOX (mitochondrial superoxide):
 MitoSOX Red (CAS 361055-77-8) is targeted to mitochondria via its triphenylphosphonium moiety and is oxidised specifically by superoxide to form a product that intercalates into DNA (Ex/Em 510/580 nm for mitochondria-localised signal). Use at 2.5–5 μ M for 10–20 min at 37°C. More specific than DCFH-DA for mitochondrial $O_2^{\bullet-}$ production.

GSH/GSSG ratio:
 Intracellular glutathione status reflects the balance between antioxidant capacity and oxidative challenge. GSH (reduced glutathione) is measured by DTNB reaction (5,5'-dithio-bis-2-nitrobenzoic acid; Ellman's reagent): $GSH + DTNB \rightarrow TNB$ (yellow chromogen, absorbance 412 nm) + GS-TNB. GSSG is measured after derivatisation of GSH with N-ethylmaleimide (NEM) to prevent re-reduction.

A reduced GSH/GSSG ratio (<10) indicates significant oxidative stress; in healthy cells, GSH/GSSG $\approx$ 100–400.

Seahorse XF Mito Stress Test (Agilent Technologies): The Seahorse XF Analyser measures real-time oxygen consumption rate (OCR, pmol O₂/min) and extracellular acidification rate (ECAR, mpH/min) as proxies for mitochondrial respiration and glycolysis respectively. The Mito Stress Test uses sequential injection of: (1) Oligomycin (ATP synthase inhibitor, 1 μ M): inhibits Complex V $\rightarrow$ remaining OCR = proton leak; reduction in OCR from baseline = ATP-linked respiration. (2) FCCP (carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone, 1–2 μ M): uncouples the proton gradient $\rightarrow$ maximal OCR. Spare respiratory capacity = maximal OCR – basal OCR. (3) Antimycin A + Rotenone (Complexes III and I inhibitors, 0.5 μ M each): blocks all mitochondrial OCR $\rightarrow$ remaining ECAR = non-mitochondrial oxygen consumption. Derived parameters: Basal respiration (pmol/min); ATP-linked respiration; Proton leak; Maximal respiratory capacity; Spare respiratory capacity (SRC = maximal – basal; SRC > 2-fold indicates healthy mitochondria with reserve capacity); Non-mitochondrial OCR. Application in toxicology: hepatotoxic compounds often inhibit Complex I (rotenone-like), uncouple the proton gradient (DILI mechanism for NSAIDs), or inhibit fatty acid oxidation (FIAU, amiodarone), all detectable by distinct Seahorse profiles.

Apoptosis endpoints: Caspase 3/7 activity (luminescence): Caspase-Glo 3/7 assay (Promega) uses a proluminescent caspase 3/7 substrate (Ac-DEVD-aminoluciferin) cleaved by caspases 3 and 7 to release aminoluciferin, which is then used by luciferase to generate bioluminescence proportional to caspase activity. Measured on a luminometer; highly sensitive (can detect apoptosis in <1,000 cells). Combined with viability assay (MTT/resazurin) to distinguish apoptotic from necrotic cell death: high caspase signal + reduced viability (not cytostatic) = apoptotic. Positive control: staurosporine (200 nM, 4–6h) or camptothecin (5 μ M, 16h).

Annexin V/PI flow cytometry: Annexin V (labelled with FITC or Alexa Fluor 488) binds to phosphatidylserine (PS) externalised to the outer leaflet of the plasma membrane — an early apoptotic event. Propidium iodide (PI, membrane-impermeant DNA intercalant) enters only cells with compromised membrane integrity (late apoptosis/necrosis). Quadrant analysis: Annexin V–/PI– = live cells; Annexin V+/PI– = early apoptosis; Annexin V+/PI+ = late apoptosis; Annexin V–/PI+ = necrosis. Analysis by flow cytometry (FACS) with green (FITC) and red (PI) channels. The Annexin V/PI assay provides more specific mechanistic information than viability assays alone and distinguishes apoptosis from necrosis.

Oxidative stress — an imbalance between reactive oxygen species (ROS) production and antioxidant defences — is a central mechanism of chemical toxicity. Key in vitro assays include:

- DCFH-DA (2',7'-dichlorodihydrofluorescein diacetate): intracellular ROS quantification by fluorescence (ex 488/em 525 nm); must be controlled for auto-oxidation artifacts.
- GSH measurement (Ellman's reagent, HPLC): quantification of reduced glutathione depletion, a sensitive early marker of oxidative stress.
- Nrf2 nuclear translocation (immunofluorescence, luciferase reporter): assessment of the master antioxidant transcription factor activation.
- Lipid peroxidation (TBARS/MDA, 4-HNE ELISA): quantification of oxidative damage to membrane lipids.

3.6.2 Apoptosis

Programmed cell death (apoptosis) is distinguished from necrosis by specific molecular and morphological features. Standard in vitro detection methods:

- Caspase-3 activity assay: fluorogenic cleavage of DEVD-AMC substrate; executes apoptosis downstream of both intrinsic and extrinsic pathways.
- Annexin V / PI flow cytometry: early apoptosis = Annexin V+/PI-; late apoptosis/necrosis = Annexin V+/PI+.
- TUNEL assay: detects DNA fragmentation characteristic of late apoptosis in fixed cell preparations.
- Bcl-2/Bax ratio (Western blot, RT-PCR): mechanistic readout of pro/anti-apoptotic protein balance in the mitochondrial pathway.

3.6.3 Mitochondrial Function

Mitochondrial toxicity underlies the action of many drugs (e.g., amiodarone, statins) and environmental toxicants (e.g., rotenone, MPTP, cadmium):

- JC-1 (5,5',6,6'-tetrachloro-1,1',3,3'-tetraethylbenzimidazolylcarbocyanine iodide): mitochondrial membrane potential ($\Delta\Psi_m$); depolarisation shifts the red:green fluorescence ratio.
- MitoSOX Red: selective detection of mitochondrial superoxide by fluorescence.
- Seahorse XF Analyser (Agilent): simultaneous real-time measurement of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR), allowing comprehensive bioenergetic profiling. Now considered the gold standard for in vitro mitochondrial toxicity assessment.

3.7 Genotoxicity In Vitro

Genotoxicity in vitro — expanded protocols:

Ames test (OECD TG 471):
 The Ames test uses five standard *Salmonella typhimurium* strains (TA98, TA100, TA1535, TA1537) plus *Escherichia coli* WP2 uvrA. Each strain carries a specific mutation in the histidine biosynthesis pathway (*his*⁻) that is reverted to histidine prototrophy (*his*⁺) by specific classes of mutagens:
 — TA98: *his* D3052 frameshift mutation (−1 frameshift at CG runs); detects frameshift mutagens (e.g., 2-nitrofluorene, benzo[a]pyrene-7,8-diol-9,10-epoxide).
 — TA100: *his* G428 substitution (GC→TA transversion); detects base-pair substitution mutagens (e.g., sodium azide, methyl methanesulfonate).
 — TA1535: *his* G428 (like TA100) but without the pKM101 plasmid (less sensitive); includes *uvrB* deletion (impaired excision repair, increases sensitivity).
 — TA1537: *his* C3076 frameshift; detects +1 frameshifts (e.g., ICR-191).
 — *E. coli* WP2 uvrA: *trpE* (AT→GC transitions); important for cross-linking agents.

S9 mix preparation: The S9 fraction is prepared from livers of male Sprague-Dawley or Wistar rats pre-treated with Aroclor 1254 (500 mg/kg IP) or with a combination of β -naphthoflavone (80 mg/kg IP) + phenobarbital (80 mg/kg IP, given 5 days), to induce a broad spectrum of CYP enzymes (CYP1A1, 1A2, 2B1, 2B2, 3A1). Livers are harvested 5 days post-induction, homogenised in KCl buffer (0.15 M), centrifuged at 9,000×g for 10 min, and the S9 supernatant stored at −80°C. The working S9 mix for the Ames test is prepared as: 30% v/v S9 + cofactors (NADP 4 mM, glucose-6-phosphate 5 mM in 0.2 M phosphate buffer, pH 7.4).

Protocol for each strain, $\pm$ S9, 3 doses + solvent control + positive control: 0.1 mL test substance (dissolved in DMSO, max 10 μ L DMSO per plate) + 0.1 mL bacteria (2×10^9 cells/mL overnight culture in Oxoid nutrient broth) + 0.5 mL S9 mix (or 0.5 mL buffer without S9) → vortex 10s → pour onto 2 mL molten top agar (45°C) supplemented with 0.5 mM histidine/biotin (just enough for 1–2 cell divisions) → overlay onto minimal agar plates → incubate 48–72h at 37°C → count revertant colonies.

Scoring criteria: A positive result requires: a reproducible, dose-related increase in revertant colonies; ≥ 2 -fold (for TA98, TA100, TA1537, WP2 uvrA) or ≥ 3 -fold (TA1535) increase over the spontaneous

reversion frequency; reproducibility across two independent experiments. The spontaneous reversion frequencies for the standard strains are: TA98 $\approx$ 20–50 colonies/plate; TA100 $\approx$ 100–200; TA1535 $\approx$ 5–20; TA1537 $\approx$ 5–15; WP2 uvrA $\approx$ 15–30.

Interpretation of $\pm$ S9: A compound positive only in the +S9 condition is a pro-mutagen requiring metabolic activation (e.g., benzo[a]pyrene, aflatoxin B1, cyclophosphamide). A compound positive only in the –S9 condition is a direct-acting mutagen (e.g., methyl methanesulfonate, mitomycin C). Positive in both: direct-acting that may also have metabolic activation pathways. Negative in both at the limit dose (5,000 μ g/plate or up to the precipitating or cytotoxic dose): Ames-negative.

Micronucleus test (OECD TG 487 — in vitro): Micronuclei arise from chromosome fragments (induced by clastogens — double-strand break inducers) or whole chromosomes that lag behind during mitotic anaphase (induced by aneugens — spindle poison). Distinction between clastogens and aneugens is critical for mechanistic interpretation and risk assessment: clastogens are generally more concerning (direct DNA damage) while aneugens may have a threshold dose (spindle disruption is reversible at low doses).

Kinetochores staining for aneugen/clastogen distinction: Kinetochores proteins (centromere-associated proteins CENP-A, CENP-B) can be detected in micronuclei using pan-centromeric CREST antisera (fluorescence microscopy) or FISH probes for centromeric sequences. A micronucleus containing centromere DNA = aneugenic origin (whole chromosome loss). A micronucleus without centromere signal = clastogenic origin (acentric fragment from chromosome break). A ratio of centromere-positive/total MN > 50% suggests aneugenic mechanism.

Protocol (TG 487 standard): seed CHO-K1 or TK6 cells in chamber slides or cytospin preparations; treat with test substance for 3–6h with and without S9 (for pro-mutagen activation) followed by 24h recovery, OR treat for 24h without S9; cytochalasin B (6 μ g/mL) is added during the last 24h to arrest cytokinesis, producing binucleated cells (BNCs, indicating one completed mitotic cycle). Staining: acridine orange or DAPI. Score: minimum 2,000 BNCs per concentration per experiment; micronucleus frequency (MNI/1,000 BNCs); also score nuclear division index (NDI) as cytotoxicity measure. Positive criteria: $\geq$ 2-fold increase in MN frequency over control AND statistical significance (chi-square or Fisher exact test).

Genotoxicity assessment is mandatory in regulatory submissions and is typically conducted as a battery of complementary assays to detect different types of DNA damage and mutagenic mechanisms.

Assay	OECD TG	Endpoint	Organisms
Bacterial Reverse Mutation (Ames test)	TG 471	Point mutations (his [–] →his ⁺ reversion)	Salmonella typhimurium TA98, TA100, TA1535, TA1537; E. coli WP2
In vitro Micronucleus	TG 487	Clastogenicity / aneugenicity	Human lymphocytes, CHO, TK6 cells
In vitro Chromosomal Aberration	TG 473	Chromosomal structural abnormalities	CHO, human lymphocytes
Comet Assay (SCGE)	TG 489 (in vivo); TG draft in vitro	DNA strand breaks and alkali-labile sites	Any mammalian cell type
γ H2AX assay	Not yet TG (development)	DNA double-strand breaks (phospho-H2AX foci)	Human cells (TK6, HepG2, primary)

Table 3.4. OECD in vitro genotoxicity battery: test guidelines and endpoints. A positive Ames test requires confirmatory in vitro follow-up; a negative Ames + negative in vitro MN is generally sufficient for regulatory purposes.

3.8 Endocrine Disruption In Vitro

Endocrine disruption in vitro — detailed assay protocols:

Estrogen receptor (ER) transactivation — OECD TG 455 (stably transfected human ER transactivation assay using VM7Luc4E2 cells): VM7Luc4E2 cells (MCF-7-derived) stably express a luciferase reporter gene under the control of 4 copies of an estrogen response element (ERE). Upon ER ligand binding and dimerisation, the ER-ligand complex translocates to the nucleus, binds the ERE, and drives luciferase transcription. Luciferase activity (measured as relative light units, RLU, by luminometry) is proportional to ER activation. Protocol: cells seeded at 3×10^4 /well in 96-well white plates in E-SCREEN medium (phenol red-free DMEM, 10% charcoal-stripped FBS); treated with 8-point dose series (0.001–10,000 nM) for 20h at 37°C; aspirate medium; add ONE-Glo luciferase reagent (Promega); measure luminescence. Positive control: 17 β -estradiol (E2) 0.1 nM; negative control: DMSO vehicle ($\leq 0.1\%$). EC₅₀ calculated by 4-PL fitting; relative potency (REP) = EC₅₀(E2) / EC₅₀(test compound). A REP > 0.001 (i.e., potency > 0.1% of E2) triggers further assessment. OECD TG 458 (ER transactivation) uses the U2-OS cell line with stably integrated ER α -VP16 fusion and ERE-luciferase reporter, providing higher sensitivity and specificity.

Androgen receptor (AR) transactivation — OECD TG 458 (AR-STTA): Similar principle to ERE-based assay; AR reporter cells (MDA-kb2) express a luciferase gene under control of MMTV-ARE (androgen response element) and respond to testosterone and DHT. Adapted for agonist and antagonist mode (antagonist mode: cells treated with EC₈₀ DHT concentration simultaneously with test compound; inhibition of luciferase signal = AR antagonism). Used for detection of anti-androgenic compounds (e.g., linuron, vinclozolin, bisphenol A, phthalates).

H295R steroidogenesis assay (OECD TG 456): The H295R human adrenocortical carcinoma cell line expresses all enzymes for steroid hormone biosynthesis (CYP11A1, CYP17A1, CYP21A1, CYP11B1, CYP19A1/aromatase, 3 β -HSD, 17 β -HSD). It can produce all major steroid classes (mineralocorticoids, glucocorticoids, androgens, estrogens). Treatment with test compounds at 4 concentrations for 48h; supernatant collected for hormone measurement by ELISA or LC-MS/MS: testosterone, estradiol, cortisol, progesterone (4 core hormones per OECD TG 456 protocol; extended panels add DHEA, androstenedione, aldosterone).

Positive interpretation: a significant (≥ 2 -fold) increase or decrease in testosterone or estradiol production, confirmed dose-dependently, is considered a positive result. Positive controls: forskolin (increases all steroids via cAMP/PKA) and prochloraz (decreases steroid production by CYP17/CYP19 inhibition). This assay is required under REACH and EPA EDSP Tier 1 screening for substances suspected to affect steroidogenesis.

3.11 Case Study: Integrated In Vitro Strategy for Hepatotoxicity

Case Study: Integrated In Vitro Strategy for Hepatotoxicity Assessment

Drug-induced liver injury (DILI) is the leading cause of acute liver failure in developed countries and a major reason for drug withdrawal post-approval. The conventional approach of testing compounds in primary rat hepatocytes (90-day in vivo + TK studies) is resource-intensive and has poor predictivity for human DILI (positive predictive value ~45%). The integrated in vitro strategy described below uses multiple complementary assays to provide mechanistic insight before in vivo testing.

Step 1 — Cytotoxicity screening (HepaRG 2D): HepaRG cells (differentiated, 14-day DMSO protocol) are treated with test compound at 8 concentrations (0.1–1000 μ M) for 24h and 72h. Endpoints: cell viability (resazurin), LDH release (membrane integrity), steatosis (Nile Red fluorescence for neutral lipid accumulation). IC₅₀(24h) and

IC₅₀(72h) are calculated. Compounds with IC₅₀ < 10× C_{max} (maximum human plasma concentration at therapeutic dose) are flagged for further assessment.

Step 2 — Mechanistic endpoint profile (HepaRG + ToxTracker): ToxTracker (Genmutation, Amsterdam) is a panel of 6 GFP-reporter mouse embryonic stem cell (mESC) lines, each responding to a specific genotoxic or cytotoxic stress: Bsc1-GFP (DNA double-strand breaks), Rtn1-GFP (DNA strand breaks), Srxn1-GFP (oxidative stress), Bcl2-GFP (oxidative stress/mitochondria), Ddit3-GFP (unfolded protein response/ER stress), Btg2-GFP (p53 activation/DNA damage). Flow cytometry quantifies reporter induction (fold induction over vehicle control). A positive reporter response indicates the mechanism of toxicity.

For the same compound in HepaRG spheroids (3D culture, 15 days maturation): Seahorse XF Mito Stress Test at 48h reveals mitochondrial mechanism (if present). Reduced spare respiratory capacity (SRC < 50% of control) + low ATP production confirms mitochondrial impairment as DILI mechanism. GFAP immunostaining rules out hepatocyte-to-stellate cell transformation. Albumin secretion (ELISA) and CYP3A4 activity (luciferin-based substrate) confirm maintenance of hepatic function.

Step 3 — PBPK-IVIVE translation: In vitro AC₅₀ values from steps 1–2 are converted to in vivo-relevant concentrations using PBPK modelling (Simcyp v21): (a) determine free fraction in plasma ($f_{u,plasma}$) by rapid equilibrium dialysis; (b) measure CL_{int} in HepaRG (substrate disappearance assay); (c) scale intrinsic clearance to whole-liver in vivo CL using the well-stirred model; (d) run Simcyp population simulation (n=100 virtual subjects, 10 mg dose, oral) to predict C_{max,u} (unbound C_{max}) in liver. Calculate in vitro-in vivo correlation: if C_{max,u} > AC₅₀(mechanism), in vivo liver effects are predicted.

Step 4 — Integrate into decision matrix:
— AC₅₀ > 100× C_{max}: low DILI concern, no further in vivo testing required.
— AC₅₀ 10–100× C_{max} + clear mitochondrial or oxidative stress mechanism: moderate concern; consider 7-day oral rat study with extended biomarker panel.
— AC₅₀ < 10× C_{max} + genotoxicity signal (Bsc1-GFP positive): high DILI concern; mandatory 28-day in vivo with Comet assay in liver.

This integrated strategy (HepaRG 2D + spheroids + ToxTracker + Seahorse + PBPK-IVIVE) has been validated in a consortium of European pharmaceutical companies (DILI Concern consortium, van Tienderen et al., 2021, Liver International) as part of an IATA for hepatotoxicity prediction, achieving 72% positive predictive value and 85% negative predictive value for human DILI when used as a tier-1 screening battery.

The OECD Conceptual Framework (OECD CF) for the Testing and Assessment of Endocrine-Disrupting Chemicals organises assays at five levels of biological complexity. In vitro assays primarily address Levels 2–3:

- Level 2 — In vitro assays (mechanistic, not regulatory): ER transactivation (OECD TG 455), AR transactivation (TG 458), steroidogenesis H295R (TG 456), aromatase inhibition.
- Level 3 — In vivo assays (single mechanistic endpoint): e.g., uterotrophic assay (TG 440), Hershberger assay (TG 441).

The H295R steroidogenesis assay (OECD TG 456) uses the human adrenocortical carcinoma cell line NCI-H295R to measure the production of oestradiol and testosterone, and can detect both aromatase activators/inhibitors and broader steroidogenesis disruptors. BPA, BPF, and glyphosate have been assessed using this platform.

3.9 High-Throughput Screening and Tox21

The Tox21 programme (US-EPA, NIEHS, FDA, NIH) has screened >10,000 chemicals across >800 in vitro assays using robotic high-throughput screening (HTS). The resulting dataset — freely available at

comptox.epa.gov/dashboard — represents the largest public repository of in vitro toxicological data in existence. Tox21 data are used to: prioritise chemicals for in-depth testing, identify structural alerts, train machine learning models, and provide mechanistic insights into existing regulatory limits. The EPA ToxCast programme extends Tox21 by integrating HTS data with chemical structure to develop predictive models for regulatory hazard assessment.

3.10 Strengths and Limitations

In vitro methods offer:

- Human-relevant cell material — overcoming key species differences inherent to rodent studies.
- High throughput — hundreds to thousands of chemicals/concentrations per experiment.
- Mechanistic clarity — direct access to cellular signalling pathways and molecular targets.
- Ethical acceptability — no animal suffering, no ethics committee approval required.

Critical limitations include:

- Absence of systemic integration — no ADME processes, no immune system, no tissue-tissue communication.
- Lack of long-term endpoints — chronic toxicity, carcinogenesis, and reproduction are not addressable in conventional in vitro systems (3D and organoid cultures extending this boundary).
- Concentration vs. dose — in vitro concentrations do not directly translate to in vivo doses without IVIVE (in vitro to in vivo extrapolation) calculations using PBPK models.
- Regulatory acceptance — most in vitro endpoints are not yet accepted as standalone replacements for in vivo studies, except for specifically validated alternatives (e.g., skin corrosion TG 431).

Key Take-Home Messages

- The choice of in vitro cell model is critical: HepG2 lacks CYP activity; HepaRG spheroids are preferred for repeat-dose hepatotoxicity; primary hepatocytes are the gold standard for acute metabolism-dependent toxicity.
- Different cytotoxicity assays measure different cellular parameters (metabolism, membrane integrity, lysosomal function); the appropriate choice depends on the mechanism being studied.
- The standard genotoxicity battery (Ames + in vitro micronucleus) is required in most regulatory packages; a positive result triggers in vivo follow-up.
- Tox21 / ToxCast high-throughput screening data are reshaping hazard prioritisation and predictive toxicology.
- In vitro results cannot replace in vivo studies for most regulatory endpoints; they are most powerful as mechanistic tools and as first-tier screens within integrated testing strategies (IATA).

Further Reading

Štampar M, Žegura B. In vitro hepatic 3D cell models and their application in genetic toxicology: A systematic review. *Mutat Res Genet Toxicol Environ Mutagen*. 2024;900:503835. doi:10.1016/j.mrgentox.2024.503835

Yang S et al. Liver three-dimensional cellular models for high-throughput chemical testing. *Cell Rep Methods*. 2023;3(3):100432. doi:10.1016/j.crmeth.2023.100432

Shao W et al. Advances in liver organoids: replicating hepatic complexity for toxicity assessment and disease modeling. *Stem Cell Res Ther*. 2025;16(1). doi:10.1186/s13287-025-04139-2

Lörincz T et al. The Performance of HepG2 and HepaRG Systems through the Glass of Acetaminophen-Induced Toxicity. *Life*. 2021;11(8):856. doi:10.3390/life11080856

Part 4 — Special Case of Stochastic Effects

4.1 Learning Objectives

- Distinguish deterministic (threshold) from stochastic (non-threshold) effects and explain the ICRP conceptual framework.
- Explain the linear non-threshold (LNT) hypothesis and the ongoing scientific debate around low-dose extrapolation.
- Describe carcinogenesis as a multi-step process (initiation, promotion, progression) and apply IARC classification.
- Define and calculate slope factor, unit risk, and excess lifetime cancer risk (ELCR).
- Explain the concepts of ALARA and acceptable risk levels in radiological and chemical carcinogen risk assessment.
- Discuss specific stochastic examples: ionising radiation, benzo[a]pyrene, aflatoxin B1, vinyl chloride, formaldehyde.

4.2 Introduction: Deterministic vs. Stochastic Effects

Toxicological effects can be classified by their dose-response behaviour into two fundamental categories. This distinction has profound implications for regulatory risk assessment:

Deterministic effects (also called threshold effects or tissue reactions in ICRP terminology) exhibit a threshold dose below which no adverse effect is observed. The severity of the effect increases with dose above the threshold. Examples include radiation-induced cataracts, acute radiation sickness, and most non-genotoxic organ toxicities (hepatic necrosis, renal tubular injury). For deterministic effects, a NOAEL can be identified and used to set safe exposure limits with appropriate uncertainty factors.

Stochastic effects, by contrast, are probabilistic in nature: there is no threshold, and the probability (but not the severity) of the effect increases with dose. The paradigmatic stochastic effects are carcinogenesis and heritable genetic damage. A single ionising radiation track, a single reactive metabolite molecule, could theoretically initiate a tumorigenic mutation — though the probability of this causing clinical cancer is vanishingly small at low doses. The ICRP (International Commission on Radiological Protection) formulated this conceptual framework in its Publication 26 (1977) and reaffirmed it in Publication 103 (2007).

4.3 Theoretical Framework

4.3.1 The Linear Non-Threshold (LNT) Model

Recent regulatory positions on the LNT model and alternatives:

The Linear Non-Threshold (LNT) model, originally proposed by the International Commission on Radiological Protection (ICRP) in 1958 and formally adopted in ICRP Publication 26 (1977), has remained the dominant paradigm for carcinogenic risk assessment for over 60 years. Under LNT, every increment of dose, however small, carries a proportionate increment of cancer risk, and there is no threshold below which exposure is safe. This model has profound implications for regulation: it underpins the ALARA (As Low As Reasonably Achievable) principle in radiation protection and the linear extrapolation from rodent bioassay doses to human low-dose risk estimates used by the US EPA. However, recent scientific and regulatory developments have created significant pressure for re-evaluation:

ICRP (International Commission on Radiological Protection): ICRP Publication 118 (2012) and the more recent ICRP Task Group 91 report (2021) reiterate support for LNT as a practical tool for

regulatory purposes but acknowledge that "it is not a biologically accurate model at low doses and dose rates." ICRP explicitly states that LNT should not be used to calculate the number of cancer cases expected from very low doses in large populations.

BEIR VII (Biological Effects of Ionising Radiation, NAS 2006): The landmark US National Academies report on low-dose radiation supports LNT as the best available model but acknowledges uncertainty at doses below 100 mSv. The BEIR VII estimate places the lifetime excess cancer risk from 100 mGy whole-body gamma radiation at approximately 1 in 100 (1%), consistent with LNT extrapolation from higher-dose epidemiology.

US EPA (2005 Cancer Risk Assessment Guidelines): Retains linear extrapolation as the default for genotoxic carcinogens in the absence of mode-of-action data supporting a threshold. However, EPA's 2011 Supplemental Guidance for Assessing Cancer Susceptibility from Early-Life Exposures to Carcinogens acknowledges age-dependent susceptibility adjustments. The 2023 EPA IRIS assessment framework allows mode-of-action evidence to support non-linear or threshold models for specific chemicals.

EU Recommendations: The Scientific Committee on Consumer Safety (SCCS) and the European Food Safety Authority (EFSA) currently use the Margin of Exposure (MoE) approach for genotoxic carcinogens in food and cosmetics, which avoids explicit LNT-based risk quantification but implicitly assumes that below the BMDL10 (benchmark dose lower confidence limit for 10% extra risk), risk is low.

Alternative models:

- (1) Threshold model: assumes that below a certain dose, DNA repair mechanisms are sufficient to prevent permanent genotoxic damage, and no excess cancer risk occurs. Supported for some carcinogens where epigenetic, non-genotoxic mechanisms dominate (e.g., phenobarbital as a liver tumour promoter in rodents). EFSA's genotoxicity working group (2017) allows threshold models when robust mechanistic evidence of a threshold mechanism is available.
- (2) Hormesis: the biphasic response model. At very low doses, some carcinogenic agents (including radiation) have been reported to stimulate adaptive responses (Nrf2 activation, DNA repair upregulation, immune surveillance enhancement) that can be measured as reduced tumour incidence below control levels. While contested (Calabrese & Baldwin, 2003, *Nature*; critiques in *Toxicology* 2004), hormesis has been observed consistently for ~1,000 chemicals in over 6,000 dose-response datasets (HORMESIS database).
- (3) Supralinear model: postulates that low doses are disproportionately more dangerous per unit dose than high doses, due to bystander effects (damage signals transmitted from irradiated cells to neighbouring unexposed cells) and genomic instability. Supported by cell biology data but not by epidemiology at doses below 50 mSv.

The consensus of major regulatory agencies is that LNT remains the most defensible default for regulatory purposes, particularly for potent genotoxic carcinogens (aflatoxin B1, benzo[a]pyrene, benzene), while acknowledging that it likely overestimates actual risk at low doses for most agents. The MoE approach (recommended by EFSA 2005 for genotoxic food contaminants) provides a practical alternative that avoids explicit numerical risk extrapolation while still providing a basis for risk management prioritisation.

The LNT model, which became the dominant regulatory paradigm for carcinogens and ionising radiation from the 1960s onward, assumes that the risk of carcinogenesis is linearly proportional to dose at all levels, with no safe threshold. The mathematical implication is:

$$\text{Risk} = \text{Slope Factor} \times \text{Dose}$$

where Slope Factor has units of $(\text{mg/kg/day})^{-1}$ or $(\text{mSv})^{-1}$ for radiation.

The LNT model is conservative and regulatory-friendly: it is straightforward to apply, avoids the ethical problem of assuming a safe dose exists for carcinogens when data at very low doses are essentially unavailable, and protects the most sensitive subpopulations. Its major limitation is that biological evidence increasingly suggests it may overestimate risk at very low doses, where DNA repair mechanisms are efficient and competing processes (cell death eliminating pre-neoplastic cells) may actually reduce net cancer risk (Calabrese, 2022).

4.3.2 The Threshold Debate

The validity of LNT at low doses is vigorously debated. Arguments against universal LNT include: (a) the existence of DNA repair pathways (BER, NER, MMR, NHEJ) that efficiently correct low-level damage; (b) epidemiological studies of populations in high natural background radiation areas (Ramsar, Iran; Yangjiang, China) that show no excess cancer incidence despite doses 10–100× above global average; (c) the observation that adaptive responses to low-dose radiation can reduce subsequent genotoxic damage (radio-adaptive response). The concept of a practical threshold — a dose below which adverse effects are effectively zero because the rate of damage is lower than the rate of error-free repair — is operationally useful for carcinogens that act exclusively through indirect mechanisms.

4.3.3 Hormesis as Counterpoint

Hormesis — the phenomenon whereby low doses of an agent produce a stimulatory or beneficial response while higher doses produce the expected inhibitory/toxic effect — represents the most direct challenge to LNT. Calabrese & Kozumbo (2019) have documented thousands of examples across taxa and chemical classes. For carcinogenesis, some epidemiological data suggest a J-shaped relationship between low-dose radiation and cancer incidence, with below-background doses associated with slightly reduced cancer risk. While these findings are intriguing, they have not displaced LNT as the regulatory default for chemical carcinogens and ionising radiation because of the inherent difficulty of demonstrating statistical significance at very low doses in human populations.

4.4 Carcinogenesis as the Paradigmatic Stochastic Effect

4.4.1 Initiation, Promotion, Progression

Multistage carcinogenesis — detailed mechanistic analysis:

Initiation:

The initiation stage involves the irreversible genetic alteration of a normal cell, creating a cell with an increased probability of undergoing malignant transformation. The key molecular events are:

- (1) Covalent DNA adduct formation: the ultimate carcinogen (often a reactive electrophile formed by metabolic activation) reacts covalently with nucleophilic sites on DNA. Key adduct targets: N7-guanine (major target for alkylating agents, aflatoxin B1 exo-epoxide, styrene oxide), N2-guanine (benzo[a]pyrene-diol-epoxide, BPDE), O6-guanine (nitrosamines), C8-guanine (2-aminofluorene metabolites). If the adduct is not repaired before cell division, it causes a miscoding event during replication — the misread base leads to a fixed point mutation.
- (2) Point mutations in proto-oncogenes: the most commonly mutated oncogene in chemically induced cancers is KRAS (codon 12: GGT→GAT for aflatoxin/tobacco; GGT→TGT for benzene). Activating KRAS mutations confer constitutive Ras signalling → ERK/MAP kinase → cyclin D1 → uncontrolled proliferation. BRAF V600E mutation (T→A transversion) is induced by UV radiation in melanoma.
- (3) Loss-of-function mutations in tumour suppressor genes: TP53 mutations at hotspot codon 249 (AGG→AGT, R249S) are characteristic of aflatoxin B1 exposure in hepatocellular carcinoma (HCC) — the AFB1-8,9-epoxide forms an adduct at N7-guanine of codon 249, causing GC→TA transversions. In HCC from Algerian regions with aflatoxin contamination of groundnuts and maize, TP53 R249S is a molecular epidemiological signature of aflatoxin aetiology.

Promotion:

The promotion stage involves the clonal expansion of initiated cells under the influence of tumour-promoting agents. Key characteristics: (a) promoters are not mutagenic in standard genotoxicity tests (Ames-negative); (b) the effects are dose-dependent with a threshold; (c) promotion is reversible if the promoting stimulus is removed early; (d) promoters can be potent stimulators of cell proliferation and/or inhibitors of apoptosis.

Molecular mechanisms of promotion include: epigenetic reprogramming (promoter CpG hypomethylation of proto-oncogenes, histone H3K4 trimethylation at proliferative gene promoters); activation of signalling pathways (PKC activation by TPA/phorbol esters; PPAR α activation by non-genotoxic rodent liver carcinogens like Wy-14,643 and fenofibrate, leading to peroxisome proliferation and hepatocyte mitogenesis in rodents but not in humans); inhibition of gap junctional intercellular communication (GJIC), which normally restrains cell proliferation; and escape from immune surveillance.

Classic chemical tumour promoters: 12-O-tetradecanoylphorbol-13-acetate (TPA/PMA) in mouse skin; sodium saccharin in rat bladder; phenobarbital in rat/mouse liver; TCDD (dioxin) in multiple organs. The two-stage skin painting model (a single initiation dose of DMBA followed by repeated TPA application) remains the canonical model for studying initiation and promotion as operationally distinguishable stages.

Progression:

The progression stage involves the evolution of promoted pre-neoplastic cells toward increasingly malignant phenotypes through genomic instability and selective pressures. Key events: chromosomal instability (CIN), microsatellite instability (MSI), telomere dysfunction, oncogene amplification (MYCN in neuroblastoma, ERBB2/HER2 in breast cancer), deletions of remaining tumour suppressor alleles (LOH, loss of heterozygosity), and the acquisition of metastatic potential (epithelial-mesenchymal transition, MMP production, anoikis resistance). The Hallmarks of Cancer framework (Hanahan & Weinberg, 2000, Cell; updated 2022, Cancer Cell) provides the most widely used conceptual map of the capabilities acquired during progression: (1) sustaining proliferative signalling; (2) evading growth suppressors; (3) resisting cell death; (4) enabling replicative immortality; (5) inducing angiogenesis; (6) activating invasion and metastasis; (7) reprogramming energy metabolism (Warburg effect); (8) evading immune destruction; and two newly added enabling characteristics: (9) unlocking phenotypic plasticity (dedifferentiation); and (10) non-mutational epigenetic reprogramming. Chemical carcinogens contribute to the acquisition of these hallmarks through the mechanisms described in initiation and promotion above.

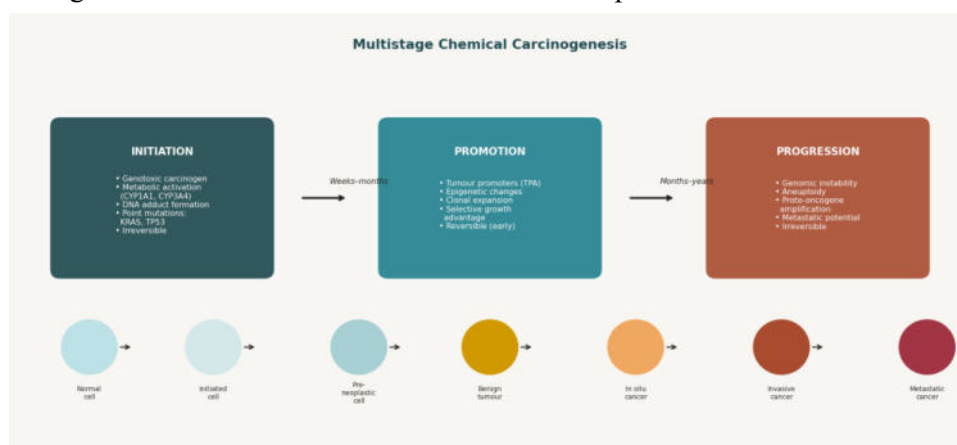


Figure 4.E1. Multistage chemical carcinogenesis — initiation, promotion, and progression. The upper boxes summarise the dominant molecular events at each stage. The lower track shows the histological progression from normal cell through initiated cell, pre-neoplastic lesion, benign tumour, in situ carcinoma, invasive

cancer, to metastatic disease. The reversibility of the promotion stage (unlike the irreversible initiation and progression stages) underlies the concept of chemical tumour promoters as threshold-dependent agents.

Chemical carcinogenesis is a multistep process (Pitot & Dragan, 1991):

- Initiation: irreversible DNA mutation in a critical gene (proto-oncogene, tumour suppressor) caused by a reactive electrophile or free radical. The mutated cell acquires a growth advantage but does not yet form a tumour. Examples: BPDE-DNA adducts (benzo[a]pyrene), AFB1-N7-guanine adducts (aflatoxin B1 → TP53 codon 249 mutation), vinyl chloride → chloroethylene oxide → ras/TP53 mutations.
- Promotion: clonal expansion of initiated cells by a promoter agent, typically non-genotoxic, that does not itself cause DNA mutations but stimulates cell division (suppressing apoptosis and antioxidant signalling). Promoters include phorbol esters (TPA), phenobarbital, 2,3,7,8-TCDD, and hormones. Promotion is often reversible if the promoting agent is withdrawn.
- Progression: accumulation of additional mutations in the expanding clone, leading to genomic instability, invasive capacity, and ultimately metastatic potential. This stage is driven by further genomic instability from ROS, replication errors, and epigenetic dysregulation.

4.4.2 IARC Carcinogen Classification

The International Agency for Research on Cancer (IARC) evaluates the strength of evidence for carcinogenicity using a systematic evidence-based framework:

Group	Interpretation	Examples	Number (as of 2025)
1	Carcinogenic to humans (sufficient evidence in humans)	Benzene, asbestos, TCDD, AFB1, BaP, tobacco, ionising radiation	~135
2A	Probably carcinogenic (limited human, sufficient animal)	Acrylamide, glyphosate, red meat, PCBs	~98
2B	Possibly carcinogenic (limited human or sufficient animal)	Carbon black, diesel fuel, coffee (removed 2016)	~326
3	Not classifiable as to carcinogenicity in humans	Paracetamol, caprolactam	~500

Table 4.1. IARC carcinogen classifications with illustrative examples.

4.5 Heritable Genetic Effects

Heritable genetic effects of chemical carcinogens:

The term heritable (germline) mutations refers to genetic alterations in germ cells (spermatogonia, oocytes, or their precursors) that are transmitted to offspring. Unlike somatic mutations (which cause cancer in the exposed individual), germline mutations can affect all cells of the offspring and all subsequent generations.

The **doubling dose (DD)** concept: The doubling dose (DD) is the dose of mutagen required to double the spontaneous rate of mutation in the germ line. For radiation, BEIR VII estimates the DD for humans at approximately 1 Gy (assuming an equal contribution of direct induction and loss-of-viability effects). The DD is used to estimate the heritable mutation risk per unit of germ cell exposure.

For chemical mutagens, the BEIR VII framework and subsequent analyses (Demarini, 2012, Mutation Research) estimate that environmental exposure levels for most mutagens are well below the DD,

suggesting that the incremental increase in heritable mutation rates from current environmental chemical exposures is likely to be small compared to the spontaneous mutation rate (~60–80 new mutations per diploid genome per generation in humans). However, this is a statistical statement about population averages and does not exclude the possibility of specific high-potency mutagens (ethylene oxide, cyclophosphamide, radiomimetic chemotherapy agents) causing detectable increases in germline mutation rates at occupational exposure levels.

Transgenerational epigenetic inheritance: A rapidly expanding area of research documents that the epigenetic effects of toxicant exposure can be transmitted across generations even without changes in DNA sequence. This is termed transgenerational epigenetic inheritance (reviewed by Skinner et al., 2021, Trends in Endocrinology and Metabolism). Mechanisms include: incomplete erasure of methylation marks in primordial germ cells during reprogramming; transmission of small non-coding RNAs (piRNAs, siRNAs) via sperm; histone modifications that survive reprogramming.

The most extensively studied examples involve endocrine-disrupting chemicals: maternal exposure to vinclozolin (AR antagonist) in pregnant F0 rats produces spermatogenic defects and reduced male fertility not only in F1 (directly exposed) but in F2, F3, and F4 male offspring (Skinner et al., 2005, Science). Similar transgenerational effects have been demonstrated for DDT, bisphenol A, and dioxin in animal models, but the relevance to human populations remains an area of active investigation. These findings challenge the classical assumption that chemical risk assessment needs only to consider direct exposure effects on the exposed generation.

Heritable genetic effects are stochastic mutations in germ cells (oocytes, spermatocytes) that are transmitted to offspring. Evidence for heritable genetic effects in humans following chemical or radiation exposure remains limited and controversial. The primary regulatory concern is preventing germ cell mutagenicity, assessed in vivo by the dominant lethal assay (TG 478), the heritable translocation test (TG 485), and the mouse specific locus test. In vitro germ cell assays (comet on sperm, γ H2AX in testicular cells) are used as screening tools. For ionising radiation, the risk of heritable effects in humans is estimated to be substantially lower than the somatic cancer risk at the same dose (ICRP Publication 103, 2007).

4.6 Quantification: Slope Factor, Unit Risk, Excess Lifetime Cancer Risk

Worked numerical example — slope factor calculation and excess lifetime cancer risk:

Scenario: Benzo[a]pyrene (B[a]P) in drinking water at 0.5 $\mu\text{g/L}$ ($= 0.5 \text{ ng/mL} = 5 \times 10^{-4} \text{ mg/L}$). Lifetime exposure (70 years); body weight 70 kg; drinking water intake 2 L/day.

Step 1 — Calculate chronic daily intake (CDI):

$$\text{CDI} = (\text{Concentration} \times \text{Intake rate} \times \text{Exposure duration}) / (\text{Body weight} \times \text{Averaging time})$$

$$\text{CDI} = (5 \times 10^{-4} \text{ mg/L} \times 2 \text{ L/day} \times 365 \text{ days/year} \times 70 \text{ years}) / (70 \text{ kg} \times 70 \text{ years} \times 365 \text{ days/year})$$

$$\text{CDI} = (5 \times 10^{-4} \times 2) / 70 \text{ mg/kg/day}$$

$$\text{CDI} = 1 \times 10^{-3} / 70 = 1.43 \times 10^{-5} \text{ mg/kg/day}$$

Step 2 — Apply oral slope factor (SF):
 EPA IRIS database slope factor for B[a]P (oral, as a surrogate for the PAH cancer potency class): SF = 7.3 $(\text{mg/kg/day})^{-1}$ (IRIS, 1994; potency relative to benzo[a]pyrene class).
 Alternative: WHO uses unit risk = 0.087 $(\mu\text{g/L})^{-1}$ for B[a]P in drinking water (WHO GDWQ, 2022).

Step 3 — Calculate Excess Lifetime Cancer Risk (ELCR):
 Using EPA approach: $\text{ELCR} = \text{CDI} \times \text{SF} = 1.43 \times 10^{-5} \text{ mg/kg/day} \times 7.3 \text{ (mg/kg/day)}^{-1} = 1.04 \times 10^{-4}$
 Using WHO unit risk: $\text{ELCR} = \text{Unit Risk} \times \text{Concentration} = 0.087 \text{ (}\mu\text{g/L)}^{-1} \times 0.5 \text{ }\mu\text{g/L} = 4.35 \times 10^{-2}$

Note: the large discrepancy between the two approaches reflects different potency estimates; the WHO value is based on more recent epidemiological data. An ELCR of 1×10^{-4} means that approximately 1

additional cancer case per 10,000 people exposed at this concentration over 70 years would be expected above the background cancer rate.

Step 4 — Risk characterisation:
The regulatory acceptable risk level for carcinogens in drinking water in the EU is 1×10^{-5} (Drinking Water Directive 98/83/EC, recast 2020/2184/EU — the parametric value for B[a]P in drinking water is $0.01 \mu\text{g/L}$, below our scenario concentration of $0.5 \mu\text{g/L}$). The WHO provisional guideline value for B[a]P is $0.7 \mu\text{g/L}$ at a 10^{-5} risk level.

At $0.5 \mu\text{g/L}$, the estimated ELCR (10^{-4} using EPA approach) exceeds the EU acceptable risk level of 10^{-5} by one order of magnitude, indicating that remediation or treatment of this water supply is warranted. Using a Monte Carlo analysis (varying daily intake, body weight, exposure duration, and slope factor as probabilistic distributions), the 95th percentile ELCR would be approximately 3×10^{-4} , further supporting this risk management conclusion.

Regulatory quantitative risk assessment for chemical carcinogens uses the following metrics:

$$\text{Excess Lifetime Cancer Risk (ELCR)} = \text{Slope Factor (SF)} \times \text{Average Daily Dose (ADD)}$$

$$\text{Unit Risk (UR)} = \text{risk per unit concentration in air or water } (\mu\text{g}/\text{m}^3 \text{ or } \mu\text{g/L})$$

$$\text{Margin of Exposure (MoE)} = \text{BMDL}_{10} / \text{Estimated Human Exposure}$$

(used for genotoxic carcinogens where LNT is assumed)

Substance	Route	Slope Factor / Unit Risk	IARC Group	Basis
Benzo[a]pyrene (BaP)	Oral	SF = 1.0 (mg/kg/d) ⁻¹ (RIVM)	1	DNA adducts, TP53 mutations
Aflatoxin B1 (AFB1)	Oral	SF = 0.004 $\mu\text{g/kg bw/d} \div 1 \text{ ng} = \text{high}$	1	AFB1-N7-guanine → hepatocellular carcinoma
Benzene	Inhalation	UR = 7.8×10^{-6} per $\mu\text{g}/\text{m}^3$ (US-EPA)	1	Leukaemia (AML)
Vinyl chloride	Inhalation	UR = 4.4×10^{-6} per $\mu\text{g}/\text{m}^3$ (US-EPA)	1	Angiosarcoma of the liver
Formaldehyde	Inhalation	UR = 1.3×10^{-5} per $\mu\text{g}/\text{m}^3$ (US-EPA)	1	Nasopharyngeal carcinoma, leukaemia
Ionising radiation (γ)	Whole body	5.5% excess cancer per Sv (ICRP 103)	1 (physical)	DNA double-strand breaks

Table 4.2. Stochastic effect examples: slope factors, unit risk, and key mechanistic basis.

4.7 Risk Assessment for Stochastic Effects

Margin of Exposure (MoE) approach for genotoxic carcinogens:

The EFSA Scientific Opinion (2005) introduced the MoE framework as a practical tool for risk managers to prioritise genotoxic carcinogens in food, avoiding the problematic extrapolation inherent in numerical low-dose risk estimates.

The MoE is defined as:

$$\text{MoE} = \text{BMDL}_{10} / \text{estimated human exposure}$$

where BMDL10 is the lower 95% confidence limit on the dose associated with a 10% increase in tumour incidence (benchmark dose lower confidence limit for a BMR of 10%), derived from animal

carcinogenicity data using benchmark dose modelling software (BMDS 3.3, US EPA; PROAST, RIVM).

Interpretation of MoE values:
 — MoE $\geq 10,000$: low concern for public health; no risk management action required for this specific source/pathway of exposure.
 — MoE 1,000–10,000: concern; risk management measures should be considered.
 — MoE $< 1,000$: high concern; strong risk management action required.

The threshold of 10,000 is derived as follows: BMDL10 in animals is considered to correspond approximately to the 10TD (tenth of the tumourigenic dose) in human equivalent terms. An MoE of 10,000 therefore corresponds to an estimated ELCR of approximately 1×10^{-5} (1 in 100,000), a level considered acceptable by most regulatory agencies (EPA accepts 1×10^{-6} to 1×10^{-4} depending on regulatory context and population vulnerability).

Worked example — Acrylamide in food (EFSA 2015 chronic dietary exposure assessment):
 BMDL10 (rat mammary gland tumours, benchmark dose for 10% additional tumour incidence): 0.17 mg/kg bw/day.

Mean dietary exposure from processed foods (European adult population average): 0.00044 mg/kg bw/day.

95th percentile dietary exposure: 0.00087 mg/kg bw/day.

MoE (mean exposure) = $0.17 / 0.00044 = 386 \rightarrow \text{MoE} < 1,000 \rightarrow \text{HIGH CONCERN}$.

MoE (95th percentile) = $0.17 / 0.00087 = 195 \rightarrow \text{MoE} \ll 1,000 \rightarrow \text{HIGH CONCERN}$.

This EFSA evaluation led to regulatory action: Commission Regulation (EU) 2017/2158 established mitigation measures to reduce acrylamide formation in processed foods, with benchmark levels that food business operators must not exceed. This is an excellent example of the MoE framework driving concrete risk management action.

For genotoxic carcinogens, regulatory agencies apply one of two approaches: (a) the LNT-based slope factor approach, deriving an acceptable daily intake from the slope factor and the acceptable risk level (typically 10^{-5} to 10^{-6} for environmental carcinogens); or (b) the Margin of Exposure (MoE) approach, where the ratio BMDL₁₀ / estimated human exposure is calculated — a MoE $\geq 10,000$ is generally considered to represent a low concern for public health (EFSA, 2023).

For ionising radiation, the ALARA principle (As Low As Reasonably Achievable) directs that all exposures should be minimised consistent with economic and social factors, even below regulatory dose limits. The ICRP recommended limits are 20 mSv/year (averaged over 5 years) for occupational exposure and 1 mSv/year for the general public.

4.8 Examples of Stochastic Carcinogens

Mechanism-specific subsections for major genotoxic carcinogens:

Benzo[a]pyrene	(B[a]P,	IARC	Group 1):
----------------	---------	------	-----------

B[a]P is a prototypical polycyclic aromatic hydrocarbon (PAH) present in tobacco smoke, grilled foods, and combustion emissions. Bioactivation proceeds via the bay-region diol-epoxide pathway: CYP1A1 and CYP1B1 catalyse the initial epoxidation of the 7,8 double bond ($\rightarrow$ B[a]P-7,8-epoxide), followed by epoxide hydrolase to form the trans-7,8-dihydrodiol, followed by a second CYP-mediated epoxidation to form the highly reactive anti-B[a]P-7,8-diol-9,10-epoxide (BPDE). BPDE is the ultimate carcinogen and reacts preferentially with the N2 position of guanine (forming (+)-anti-BPDE-N²-dG adducts in the major groove of DNA). These bulky adducts, if not removed by nucleotide excision repair (NER), cause G $\rightarrow$ T transversions during replication, activating KRAS oncogene mutations at codons 12 and 13. B[a]P also activates the aryl hydrocarbon receptor (AhR), a ligand-activated transcription

factor that induces expression of CYP1A1, CYP1A2, and CYP1B1 via the DRE (dioxin response element), thereby amplifying its own metabolic activation in a feed-forward loop.

Aflatoxin B1 (AFB1, IARC Group 1):
 AFB1 (produced by *Aspergillus flavus* and *A. parasiticus* in improperly stored cereals, groundnuts, and spices) is activated by CYP3A4 (primarily) and CYP1A2 to the exo-8,9-epoxide (AFB1-exo-8,9-epoxide). This electrophile reacts with the N7 position of guanine, forming the AFB1-N7-Gua adduct, which depurinates rapidly to form an apurinic site or undergoes ring-opening to form a stable AFB1-FAPY adduct. The AFB1-N7-Gua adduct causes GC→TA transversions at codon 249 of TP53 (R249S mutation), a molecular signature used to attributionally link HCC cases in sub-Saharan Africa and China to AFB1 exposure. AFB1 is among the most potent genotoxic carcinogens known (potency similar to B[a]P on a molar basis). In Algeria and Maghreb countries, periodic contamination of maize and peanuts with AFB1 poses a public health concern, and monitoring of aflatoxin levels in these commodities is conducted by the ONS (Office National des Statistiques) and the INSP (Institut National de Santé Publique).

Vinyl Chloride (VC, IARC Group 1):
 Vinyl chloride (chloroethylene, CAS 75-01-4) is metabolised by CYP2E1 to chloroethylene oxide and chloroacetaldehyde, both highly reactive electrophiles. These metabolites form specific DNA adducts: 7-(2-oxoethyl)guanine (major adduct), N2,3-ethenoguanine, and 1,N6-ethenoadenine (εAde), the latter two being promutagenic etheno-DNA adducts. Etheno adducts cause GC→AT transitions and AT→TA transversions, leading to KRAS and TP53 mutations in angiosarcoma of the liver (hepatic angiosarcoma), the characteristic tumour type in VC-exposed workers (Polyvinyl chloride plant workers, first cases reported in 1974). VC is a model occupational carcinogen; the MoE for current permitted occupational exposure limit (OEL: 1 ppm = 2.6 mg/m³ in Europe) is approximately 1,500 — below 10,000, indicating residual risk. Regulatory response: OEL progressively reduced from 500 ppm (1970s) to current 1 ppm (EU Directive 2004/37/EC).

Formaldehyde (HCHO, IARC Group 1 as of 2009/2012):
 Formaldehyde is an endogenous metabolite (produced by demethylase reactions at approximately 0.1 mM in blood) as well as an industrial chemical and indoor air pollutant (off-gassing from MDF, urea-formaldehyde resins). At exogenous exposure levels, HCHO reacts with DNA amino groups and histidine residues on proteins to form DNA-protein crosslinks (DPCs), which are poorly repaired by the nucleotide excision repair system. HCHO also forms methylene bridge crosslinks between adjacent guanines (interstrand crosslinks, ICLs) that block replication. The key epidemiological finding driving IARC's Group 1 reclassification was a significant excess of nasopharyngeal carcinoma and leukaemia (myeloid) in industrial workers with high formaldehyde exposure, most recently confirmed in the NCI cohort study (Hauptmann et al., 2009, Occupational and Environmental Medicine). The mechanistic argument for leukaemia involves systemic distribution of formaldehyde reaching bone marrow (haematopoietic stem cells), where DPCs and ICLs promote clonal haematopoiesis.

Ionising radiation (IARC Group 1, all types):
 Ionising radiation induces DNA damage by two mechanisms: (1) Direct ionisation — photons and charged particles directly ionise DNA atoms, breaking glycosidic bonds and creating abasic sites, single-strand breaks (SSBs), and double-strand breaks (DSBs). (2) Indirect damage via ROS — ionisation of water molecules produces hydroxyl radicals (•OH, the most reactive ROS), H₂O₂, and superoxide (O₂•⁻), which then react with DNA (8-oxo-dG formation, SSBs, DSBs). Approximately 60–70% of radiation-induced DNA damage is mediated by •OH. DSBs are the most cytotoxic lesion and the primary initiating event for chromosomal rearrangements and carcinogenesis. DSB repair pathways: Non-Homologous End Joining (NHEJ) — fast but error-prone, predominates in G1/G0 cells; Homologous Recombination (HR) — high-fidelity, requires a homologous template (sister chromatid), predominates in S/G2. Errors

in NHEJ produce deletions, inversions, and translocations (e.g., the t(9;22) Philadelphia chromosome in radiation-induced CML).

4.8.1 Benzo[a]pyrene

BaP is the prototype polycyclic aromatic hydrocarbon (PAH) carcinogen, ubiquitous in combustion products (tobacco smoke, grilled/smoked food, vehicle exhaust). CYP1A1 and CYP1B1 bioactivate BaP to BPDE (benzo[a]pyrene-7,8-diol-9,10-epoxide), which forms bulky DNA adducts preferentially at guanine N2 positions. BPDE-DNA adducts at codons 157, 158, 245, 248, and 249 of TP53 are hotspots for G→T transversion mutations observed in lung cancer from smokers. IARC Group 1. Algerian context: elevated PAH exposure in populations around the Oran industrial corridor (metal smelting, refinery operations).

4.8.2 Aflatoxin B1

AFB1, produced by *Aspergillus flavus* and *A. parasiticus* on improperly stored cereals, groundnuts, and dried fruits, is the most potent natural carcinogen known. CYP3A4 bioactivates AFB1 to AFB1-8,9-epoxide, which forms a stable N7-guanine adduct → TP53 codon 249 AGG→AGT transversion (a molecular fingerprint of AFB1 exposure in HCC tumours). IARC Group 1; high HCC risk particularly in hepatitis B-endemic populations (synergistic with HBV). Algerian context: AFB1 contamination of cereals (wheat, maize) and groundnuts is a recognised food safety issue, with sporadic monitoring data exceeding EU regulatory limits (2 µg/kg in cereals).

4.8.3 Vinyl Chloride

Vinyl chloride monomer (VCM), an industrial precursor of PVC, is a human carcinogen causing angiosarcoma of the liver (an otherwise extremely rare tumour), hepatocellular carcinoma, and possible brain tumours in workers with high occupational exposure. Its reactive metabolite, chloroethylene oxide, alkylates adenine (producing 3,N⁴-etheno-dC and 1,N⁶-etheno-dA adducts) leading to A:T→T:A and A:T→G:C transversions in ras and p53. IARC Group 1.

4.8.4 Formaldehyde

Formaldehyde is used extensively in building materials, textiles, and as a fixative in pathology laboratories. It is classified IARC Group 1 for nasopharyngeal cancer (sufficient evidence) and leukaemia (limited evidence, reconfirmed in Monograph 122, 2020). Formaldehyde is an endogenous metabolite (produced during one-carbon metabolism) making low-dose extrapolation particularly complex. The Margin of Exposure approach is preferred over LNT for formaldehyde.

4.9 Regulatory Implications

ALARA and acceptable risk levels — comparative agency framework:

The concept of ALARA (As Low As Reasonably Achievable) originated in radiological protection (ICRP Publication 26, 1977) and has been extended to chemical carcinogens under the broader principle that risk should be reduced to the lowest practicable level, taking account of economic and social factors. Different agencies have formalised different acceptable risk levels, creating a patchwork of regulatory standards:

US EPA (Environmental Protection Agency): For carcinogens, defines an acceptable risk range of 10⁻⁴ (1 in 10,000) to 10⁻⁶ (1 in 1,000,000) excess lifetime cancer risk. For Superfund sites and CERCLA cleanup decisions, the target is 10⁻⁶ as the point of departure, with 10⁻⁴ as the upper bound for remediation decisions. For drinking water under the Safe Drinking Water Act, the health goal (MCLG) is set at zero for carcinogens; the enforceable MCL is as close to zero as feasible, typically yielding risks of 10⁻⁴ to 10⁻⁶.

EU Drinking Water Directive (2020/2184/EU, recast): Uses health-based parametric values derived from WHO guidelines, which generally correspond to an excess lifetime risk of 10^{-5} for genotoxic carcinogens in drinking water. The parametric value for B[a]P is 0.01 µg/L; for total PAHs, 0.1 µg/L.

OSHA (Occupational Safety and Health Administration, US): Uses a more permissive acceptable risk level of 10^{-3} (1 in 1,000) for occupational carcinogens in the workplace, reflecting the fact that worker exposure is voluntary, monitored, and of limited duration (40h/week, 48 weeks/year, 45-year career), compared to lifetime population exposure from environmental sources. The OSHA PEL for benzene is 1 ppm (corresponding to an estimated excess leukaemia risk of $\sim 10^{-3}$ at lifetime occupational exposure).

IRSN (Institut de Radioprotection et de Sûreté Nucléaire, France): For nuclear and radiation sources, applies ICRP recommendations with a nominal effective dose limit of 20 mSv/year for radiation workers (averaged over 5 years, maximum 50 mSv in a single year) and 1 mSv/year for the public. These dose limits correspond to an estimated lifetime excess cancer risk of $\sim 10^{-3}$ for workers and $\sim 10^{-5}$ for the public (using BEIR VII nominal cancer risk coefficient of 5.5% per Sievert).

EFSA (European Food Safety Authority): For genotoxic carcinogens in food and feed, uses the MoE approach ($\text{BMDL}_{10}/\text{exposure} \geq 10,000$ = low concern; see §4.7). Does not set explicit numerical acceptable risk levels but uses priority-setting to identify substances with $\text{MoE} < 10,000$ for risk management action. For substances where sufficient data support a threshold mechanism (e.g., estragole, shown to operate below a cytotoxic threshold in hepatocytes), EFSA may derive a TDI (Tolerable Daily Intake) with a threshold assumption.

Algerian MESRS and ANVREDET: No formalised acceptable risk levels for chemical carcinogens in water or food have been published by the Algerian environment agency (ANVREDET, Agence Nationale pour la Valorisation des Résultats de la Recherche et le Développement Technologique). Algerian food safety standards for carcinogens (e.g., aflatoxin B1 limit: 2 µg/kg in cereals; ochratoxin A: 3 µg/kg in grain-based foods) follow EU Commission Regulation (EC) 1831/2003 as the de facto standard, consistent with Algeria's orientation toward EU export markets.

Stochastic effects have no safe threshold in the regulatory sense; therefore:

- ALARA principle: exposures must be reduced as far as reasonably achievable even below regulatory limits.
- Acceptable risk levels: EU regulatory precedent uses 10^{-5} (1 in 100,000) as the upper bound of acceptable cancer risk for environmental carcinogens (ECHA REACH); some contexts use 10^{-6} (1 in 1,000,000) for drinking water carcinogens.
- Margin of Exposure: EFSA and ECHA apply $\text{MoE} \geq 10,000$ as a benchmark for low concern with genotoxic carcinogens in food (EFSA Guidance 2023).
- For ionising radiation: occupational limit 20 mSv/year, public limit 1 mSv/year (ICRP 103).

The Adverse Outcome Pathway (AOP) Framework



Example AOP: Aromatase inhibition → Reproductive failure in fish (OECD AOP #25)

Figure 4.1. The Adverse Outcome Pathway (AOP) Framework, illustrated with an aromatase-inhibition example. AOPs provide the mechanistic structure linking a molecular initiating event to stochastic or deterministic adverse outcomes through key events at successively higher levels of biological organisation. (Original figure.)

Key Take-Home Messages

- Stochastic effects (carcinogenesis, heritable mutations) have no regulatory threshold; risk is assumed proportional to dose under the LNT model.
- The LNT model is conservative and regulatory-practical but biologically debated; the hormesis and threshold hypotheses are scientifically acknowledged as competing models.
- IARC classifies carcinogens in five groups based on strength of evidence in humans and animals; Groups 1 (benzene, AFB1, BaP, vinyl chloride, formaldehyde) and 2A/2B have direct regulatory implications.
- Excess Lifetime Cancer Risk = Slope Factor $\times$ Average Daily Dose; acceptable risk levels are typically 10^{-5} – 10^{-6} .
- For genotoxic carcinogens, the EFSA Margin of Exposure approach ($\text{MoE} \geq 10,000$ = low concern) provides a practical alternative to LNT-based risk characterisation.
- ALARA minimisation of exposures to known carcinogens is the operational regulatory response to stochastic risk.

Further Reading

Calabrese EJ. Linear non-threshold (LNT) fails numerous toxicological stress tests. *Chemico-Biol Interact.* 2022;365:110064. doi:10.1016/j.cbi.2022.110064

Wan Y et al. Current advances and future trends of hormesis in disease. *npj Aging.* 2024;10(1):26. doi:10.1038/s41514-024-00155-3

IARC Working Group. IARC Monographs on the Identification of Carcinogenic Hazards to Humans, Volume 126: Formaldehyde. Lyon: IARC Press; 2024.

ICRP. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP Publication 103. *Ann ICRP.* 2007;37(2–4):1–332.

EFSA Scientific Committee. Guidance on the use of the Margin of Exposure approach for genotoxic carcinogens. *EFSA Journal.* 2023;21(7):e08107. doi:10.2903/j.efsa.2023.8107

Part 5 — Good Practices in Animal Experimentation

5.1 Learning Objectives

- Describe the OECD GLP Principles and their application to non-clinical safety studies.
- Explain the role of the Study Director, QA unit, archives, and SOPs within a GLP-compliant organisation.
- Apply the ARRIVE 2.0 Essential 10 checklist to an in vivo study design.
- Justify sample size using power analysis principles and explain the consequences of underpowering.
- Define humane endpoints and severity classification; explain their ethical and scientific importance.
- Describe key refinement strategies (anaesthesia, analgesia, enrichment) and reduction strategies (factorial design, pilot studies).
- Explain the reproducibility crisis in biomedical research and outline measures to improve data integrity.
- Describe the practical context for implementing good practices at USTO-MB and in Algerian research institutions.

5.2 Introduction

Scientific integrity in animal experimentation depends not only on ethical treatment of animals but also on rigorous experimental design, transparent reporting, and robust data management. The reproducibility crisis — the finding that a substantial proportion of published biomedical results cannot be replicated — has implicated poor experimental design (insufficient sample size, lack of blinding, selective outcome reporting) as a major contributing factor. Good practices in experimental design and reporting are therefore simultaneously ethical (avoiding unnecessary animal use in poorly powered studies) and scientific (generating reliable, reproducible data).

5.3 Good Laboratory Practice (GLP): OECD Principles

The Ten OECD Principles of GLP (ENV/MC/CHEM(98)17):

GLP is a quality system concerned with the organisational process and the conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived, and reported. The 10 principles are:

1. **Test Facility Organisation and Personnel:** The test facility must have a sufficient number of qualified personnel for the tests undertaken. Each person involved in a study must clearly understand their functions and responsibilities. Written records of qualifications, training, experience, and job descriptions must be maintained. Personnel performing GLP studies must be trained in the GLP procedures applicable to their work.
2. **Quality Assurance Programme (QAP):** The test facility must have a documented QA programme to ensure that studies comply with GLP principles. The QA unit must be independent from the Study Director and must: (a) maintain copies of all study protocols; (b) conduct inspections to determine compliance; (c) inspect final reports; (d) report audit findings to management and Study Director; (e) prepare and sign a QA statement in the final report.
3. **Facilities:** The test facility must have sufficient space and be of appropriate design and construction for the studies conducted. There must be adequate separation between different study areas and provision for controlled environmental conditions (temperature, humidity, light cycle) for test systems. Facilities must also provide secure storage for test and reference items.

4. Apparatus, Material and Reagents: All apparatus used in a GLP study must be regularly inspected, cleaned, maintained, and calibrated according to SOPs. Records of maintenance, calibration, and cleaning must be maintained. Test items, reference items, and reagents must be stored under appropriate conditions, labelled with identity, concentration, expiry date, and storage requirements.
5. Test Systems: Test systems (animal, cellular, microbial, physicochemical) must be appropriate for the study and be described in the study protocol. For animal test systems: health status must be documented; animals must be quarantined before use; species, strain, source, age, sex, and body weight must be recorded; unique identification must be maintained. For in vitro systems: species, tissue, number of passages (cell lines), and certificate of authentication must be documented.
6. Test and Reference Items: Test items must be characterised for identity (CAS number, chemical name, structural formula), purity, stability, homogeneity, and storage conditions. Reference items (positive controls, standards) must be similarly characterised. The concentration and homogeneity of dose formulations must be analytically verified.
7. Standard Operating Procedures (SOPs): The test facility must have written SOPs for all procedures relevant to GLP studies. SOPs must be approved by facility management and be accessible to all relevant personnel. Each study must reference the applicable SOPs. Deviations from SOPs must be noted in study records.
8. Performance of the Study: Each study must have a written protocol specifying: objectives, study director, test facility, test system, test and reference items, dose groups, study design, parameters to be measured, statistical methods, and schedule. Protocol amendments must be documented. All raw data must be recorded directly, promptly, accurately, and legibly in permanent ink. Corrections must be single-lined and initialled with the date of correction.
9. Reporting of Study Results: A final study report must be prepared for each study, signed and dated by the Study Director. The report must include a description of the study objectives, methods, and results, and a quality assurance statement. If the report is amended, the reasons must be stated and a clear distinction made between the original and the amended findings.
10. Storage and Retention of Records and Materials: All raw data, final reports, protocols, and specimens from completed studies must be archived. Retention periods are specified by national regulations (typically 15 years in the EU for non-clinical safety studies submitted for regulatory purposes). Archive storage must be protected from fire, flood, and unauthorised access. An archiver responsible for the archive must be designated.

GLP

Organisational

Chart:

[See Figure: fig_ext1_glp_orgchart.png] The typical organisational structure places Test Facility Management at the apex, responsible for overall GLP compliance and for ensuring that the QA function is independent. The Study Director is the single point of control for each individual study — responsible for the overall conduct and integrity of the study data. The QAU (Quality Assurance Unit) is independent of the study team and reports directly to Test Facility Management. Archives and SOPs management are separate functions reporting to Test Facility Management. Under the Study Director: Principal Investigators (PIs, responsible for specific phases conducted at remote test sites), study personnel (bench scientists conducting the experiments), the Test Item Manager (responsible for receipt, storage, and homogeneity verification of test item), and the Test System Manager (responsible for the animal colony or cell culture).

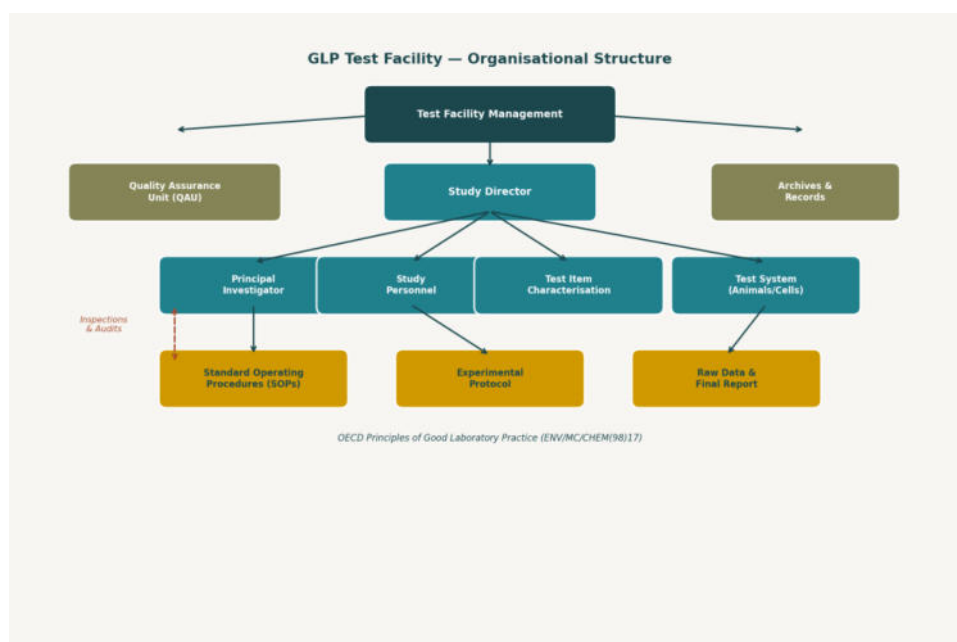


Figure 5.E1. Organisational structure of a GLP test facility according to OECD Principles of GLP (ENV/MC/CHEM(98)17). The Test Facility Management holds overall responsibility for GLP compliance. The Quality Assurance Unit (QAU) reports directly to management, ensuring independence. The Study Director coordinates each study; Principal Investigators may be responsible for specific phases. Gold boxes represent document management functions (SOPs, experimental protocols, raw data, final reports) that are subject to QAU inspection.

GLP is a quality system for the organisation of and conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived, and reported. The OECD GLP Principles (No. 1 of the OECD Series on Principles of GLP and Compliance Monitoring, revised 1997) define the essential elements of a GLP system.

5.3.1 Scope of GLP

GLP applies to all non-clinical laboratory studies conducted to support submissions to regulatory authorities for registration of chemicals, pharmaceuticals, agrochemicals, and medical devices. It does not apply to basic/exploratory research, clinical studies, or manufacturing quality control. GLP-compliant studies are accepted under the OECD Mutual Acceptance of Data (MAD) agreement in all OECD member countries, preventing duplication of animal testing for regulatory purposes.

GLP Principle	Key Requirement
Test facility organisation and personnel	Management commitment; defined responsibilities; training records
Quality Assurance Programme	Independent QA unit; routine and process-based inspections; study audits
Facilities	Adequate for study type; separate test article storage; barrier controls for biological studies
Equipment, reagents, materials	Calibrated, maintained; certificates of analysis for reagents
Test systems (animals, in vitro)	Appropriate species/strain/source; health monitoring; welfare compliance
Test and reference items	Characterised; labelled; stability data; accountability records
Standard Operating Procedures (SOPs)	Written and approved; current version accessible at point of use
Study plan	Signed by Study Director before study starts; amendments formally approved

Conduct of studies	Raw data contemporaneous, dated, attributable (ALCOA principles)
Reporting of study results	Final report signed by Study Director; all deviations documented; archive
Storage and retention of records	Archive secured; data retained for minimum 15 years (regulatory studies)

Table 5.1. OECD GLP Principles — summary of key elements (OECD No. 1, 1997).

5.4 GLP Organisation: Key Roles

Template SOP structure:

Standard Operating Procedures are the backbone of a GLP-compliant operation. Every procedural step in a study must be covered by a current, approved SOP. The standard structure for an SOP in a toxicological testing facility is:

SOP-TOX-001: ORAL GAVAGE ADMINISTRATION IN RATS
 Title: Oral Gavage Administration in Rats
 SOP Code: SOP-TOX-001
 Version: 3.0
 Effective Date: 15 January 2025
 Supersedes: Version 2.1 (01 March 2022)
 Author: [Name, title]
 Reviewed by: Study Director, QA Officer
 Approved by: Test Facility Management

Purpose: To describe the standard method for oral administration of liquid test formulations to Wistar rats by intragastric gavage, ensuring consistent dosing, minimising animal distress, and complying with GLP requirements.

Scope: This SOP applies to all staff conducting oral gavage in rats in the USTO-MB Experimental Toxicology Unit. It does not apply to dosing via diet admixture (see SOP-TOX-004) or to mice (see SOP-TOX-002).

Responsibilities: Study personnel (perform gavage after competency assessment); Study Director (ensures that personnel conducting gavage are trained and assessed as competent); QA Officer (audits procedure during study inspections).

Materials and Equipment: Ball-tipped gavage needles (18G × 3 inch for rats <250g; 16G × 4 inch for rats >250g); 1 mL or 5 mL disposable syringes; appropriate gavage tube; restraint device (commercially available rat restraint cone or gloves); test formulation (prepared per SOP-TOX-010); body weight data for dose calculation; dosing record forms.

Procedure:

1. Weigh the animal and calculate the dose volume (target: 5–10 mL/kg body weight; do not exceed 10 mL/kg for aqueous formulations or 5 mL/kg for oily vehicles).
2. Prepare the gavage syringe: draw up calculated volume + 20% excess; expel any air bubbles; attach the gavage needle.
3. Restrain the animal: grasp the rat firmly at the base of the skull, extending the neck gently to straighten the oesophagus. The animal should be oriented vertically (nose up, approximately 30–45° incline).
4. Insert the gavage needle: introduce the needle along the roof of the mouth, then guide it gently into the oesophagus with a slight downward angle. Resistance indicates the needle has entered the trachea — STOP, withdraw, and re-attempt. Never force the needle.
5. Verify oesophageal placement: the animal should not cough or show respiratory distress; no resistance

should be felt from the needle during insertion at the oesophageal level.

6. Administer dose: depress the syringe plunger slowly and evenly over 5–10 seconds. If the animal struggles vigorously, pause and allow relaxation before continuing.
7. Withdraw needle: withdraw smoothly after dose delivery; return animal to cage and observe for 5 minutes for signs of aspiration (coughing, gasping, dyspnoea, immediate distress).
8. Record: document body weight, dose volume administered, time, and any observations in the dosing log.

Safety considerations: Ball-tipped gavage needles are designed for animal welfare — the ball tip prevents accidental puncture of the oesophageal wall. Accidental administration into the lungs (aspiration) causes pneumonia and potential death; observe animals immediately post-dose. Wear appropriate PPE (gloves, lab coat) when handling test substances. If spillage occurs, follow SOP-SAFETY-003 for chemical spill management.

Emergency: If an animal shows signs of aspiration (gasping, cyanosis), inform the Study Director and the Designated Veterinarian immediately. Do not attempt to reverse the procedure. The animal may require euthanasia under the supervision of the veterinarian.

References: OECD GD 19 (2000, rev. 2022): Humane Endpoints; OECD TG 407 (2008): 28-Day Oral Study; FELASA recommendations for laboratory animal use; SOP-TOX-010: Preparation of Dose Formulations; SOP-SAFETY-003: Chemical Spill Management.

Revision history: v1.0 (2018): Initial issue; v2.0 (2020): Ball-tip needle requirement added; v2.1 (2022): Restraint section expanded; v3.0 (2025): PPE requirements updated; dosing record form reference added.

5.4.1 The Study Director

The Study Director is the single point of accountability for a GLP study. Responsibilities include: approving the study plan before study initiation; ensuring all personnel are briefed on their roles; authorising any changes to the approved protocol as formal amendments; maintaining a study file; and signing the final report, which constitutes a declaration that the study was conducted in compliance with GLP. Only one Study Director may be assigned per study, though multiple Principal Investigators may be designated for multi-site studies (sponsored studies).

5.4.2 Quality Assurance Unit

The QA unit is functionally independent of study conduct. Its role is to inspect studies during conduct (process-based inspections) and to audit raw data and draft final reports (report audits). QA produces inspection statements that are appended to the final report. Regulatory authorities inspect GLP test facilities to confirm that the QA unit functions properly; loss of GLP status renders all ongoing studies scientifically void for regulatory purposes.

5.4.3 Archives and SOPs

GLP requires that all raw data, study plans, final reports, and supporting documents be archived in a secure facility with controlled access for a minimum of 15 years (pharmaceutical studies: as long as the product is on the market plus 15 years). SOPs must be written, version-controlled, and accessible at the point of use; superseded versions are archived.

5.5 ARRIVE 2.0 Guidelines

ARRIVE 2.0 — Essential 10 and Recommended Set with reporting examples:

ARRIVE 2.0 (Percie du Sert et al., 2020, PLOS Biology) restructured the original 20 items into 2 tiers: ESSENTIAL 10 (minimum reporting requirement for all manuscripts):

1. Study design: Describe the experimental design, including the basic design (e.g., dose-response, factorial), unit of experimental unit (cell, animal, litter), type of study (exploratory/confirmatory), and the total number of experimental units per group.
 POOR: "Animals were treated with vehicle or test compound and sacrificed."
 GOOD: "A randomised, blinded, parallel-group design was used, with the individual animal as the experimental unit. Four groups were used: vehicle control, and three dose levels (10, 50, 200 mg/kg/day). A total of 40 animals (10/group) were enrolled in this confirmatory study."
2. Sample size: State the total number of experimental units in each group and provide a justification for the sample size.
 POOR: "Ten animals per group were used based on previous experience."
 GOOD: "Sample size was calculated using G*Power 3.1 (one-way ANOVA, $f = 0.40$, $\alpha = 0.05$, power = 0.80, 4 groups), yielding $n = 7$ per group. We enrolled $n = 10$ per group to allow for 30% attrition."
3. Inclusion/exclusion criteria: Describe the criteria used for including/excluding animals and specify any pre-study exclusions.
 POOR: "Unhealthy animals were excluded."
 GOOD: "Inclusion criteria: male Wistar rats, 200–250g, FELASA Category B health status, no clinical signs of illness. Exclusion criteria: body weight deviation >15% from group mean at randomisation, confirmed health issue during acclimation. Two animals were excluded pre-treatment for body weight deviation."
4. Randomisation: Describe the method used to allocate animals to groups.
 POOR: "Animals were randomly allocated to groups."
 GOOD: "Animals were allocated by stratified block randomisation based on body weight quartile, using a random number sequence generated by Excel RAND() function. The allocation list was generated by a statistician not otherwise involved in the study."
5. Blinding: Describe who was blinded to group allocation and at which stages.
 POOR: "Blinding was used."
 GOOD: "The histopathologist evaluating kidney sections and the biochemist performing clinical chemistry analysis were blinded to treatment group allocation. Group codes were maintained until after database lock and statistical analysis plan finalisation."
6. Outcome measures: Clearly define the primary and secondary outcome measures, including all pre-specified hypotheses.
 POOR: "Various biochemical parameters were measured."
 GOOD: "Primary endpoint: serum creatinine (kidney injury marker) at study termination. Secondary endpoints: BUN, ALT, AST, ALP, organ weights (kidney, liver). All endpoints were pre-specified in the protocol dated 01/09/2024."
7. Statistical methods: Describe the statistical methods used for each analysis.
 POOR: "Data were analysed using ANOVA."
 GOOD: "Normally distributed continuous data (Shapiro-Wilk test, $p > 0.05$) were analysed by one-way ANOVA followed by Dunnett's multiple comparison test versus control (SAS v9.4 PROC GLM). Non-normally distributed data were analysed by Kruskal-Wallis test with Dunn's post-hoc correction."
8. Experimental animals: Describe the animals used (species, strain, sex, age/weight, source, health status).
 POOR: "Male Wistar rats were used."
 GOOD: "Male Wistar rats (CrI:WI(Han), 8 weeks old, 200–225g, ITEL V, Algeria, FELASA Category B SPF status) were used. Animals were acclimated for 7 days under study conditions before treatment. No animals from the same litter were allocated to the same treatment group."

9. Experimental procedures: Describe what was done to the animals, including the preparation, treatment, and measurements.

POOR: "Animals were dosed daily."

GOOD: "Animals were dosed by oral gavage (5 mL/kg body weight) daily for 28 days. Test compound was formulated in 0.5% methylcellulose (w/v in water). Vehicle control received 0.5% methylcellulose. Body weight was recorded twice weekly; food consumption weekly."

10. Results: Report results for each experimental group, including the number of animals analysed, precision measures (mean $\pm$ SD), effect size with confidence interval, and p-values.

POOR: "The high-dose group showed a significant increase in ALT."

GOOD: "ALT activity was significantly increased in the high-dose group (200 mg/kg: 185 $\pm$ 43 IU/L vs. control: 58 $\pm$ 12 IU/L; $p < 0.001$, Dunnett's test; effect size $d = 3.8$, 95% CI: 2.2–5.4). No significant differences were observed at 10 or 50 mg/kg (77 $\pm$ 18 and 95 $\pm$ 29 IU/L, respectively)."

ARRIVE 2.0 (Percie du Sert et al., 2020) provides a consensus reporting checklist for in vivo studies. It is divided into 'Essential 10' items (minimum required for peer review) and 'Recommended' items.

Essential 10 Item	Content Required
1. Study design	Type (exploratory/confirmatory); independent experimental unit; blinding; allocation to groups
2. Sample size	Number of experimental units; statistical basis for sample size; whether pre-determined
3. Inclusion/exclusion criteria	Pre-established criteria for excluding animals/data; number excluded
4. Randomisation	Method of randomisation to treatment groups; any stratification
5. Blinding	Who was blinded; at what stage (allocation, experiment, analysis)
6. Outcome measures	Primary and secondary outcomes; objective definition; how measured
7. Statistical methods	All statistical tests, model assumptions, software; significance threshold
8. Experimental animals	Species, strain, age, sex, weight, health status, source, housing conditions
9. Experimental procedures	What, how, when, how much; anaesthesia/analgesia if relevant
10. Results	All results (null and positive); effect sizes with confidence intervals

Table 5.2. ARRIVE 2.0 Essential 10 checklist (Percie du Sert et al., 2020).

5.6 Experimental Design Best Practices

Detailed sample size worked example — 4-group dose-response study:

Background: A 28-day oral gavage study in Wistar rats is planned to investigate the hepatotoxicity of a new industrial solvent. Based on a pilot study, vehicle control ALT = 55 $\pm$ 18 IU/L (mean $\pm$ SD). The toxicologist anticipates that the high dose (100 mg/kg/day) will produce ALT approximately 2 $\times$ the control level (= 110 IU/L, a biologically significant effect). Three dose groups + control; endpoint = terminal ALT.

G*Power calculation (one-way ANOVA, fixed effects):

Effect size $f = \sigma_m / \sigma =$ where σ_m is the standard deviation of group means.

Group means (expected): 55, 65, 85, 110 IU/L (control + 3 ascending dose levels).

Grand mean = (55+65+85+110)/4 = 78.75

$\sigma_m = \sqrt{[(55-78.75)^2 + (65-78.75)^2 + (85-78.75)^2 + (110-78.75)^2]/4}$

$$= \sqrt{[(564 + 189 + 39 + 979)/4]} = \sqrt{[442.75]} = 21.0$$

σ_{within} (pooled SD, estimated from pilot) = 18

$f = 21.0/18 = 1.17$ (very large effect)

G*Power v3.1 inputs: Test family: F; statistical test: ANOVA one-way; $\alpha = 0.05$; power $(1-\beta) = 0.80$; number of groups = 4; effect size $f = 1.17$.
Output: Required n per group = 5 (total N = 20).

For a more conservative estimate (accounts for unknown sources of variance, non-normality, possible dropouts):

Conservative $f = 0.50$ (medium effect, assuming we might detect smaller effects than expected):
G*Power output: required n per group = 17 (total N = 68).

Standard practice in OECD TG 407: $n = 10$ per sex per group (= 20 per group for mixed-sex studies, or 10 if single sex). This is justified by regulatory requirements for adequate power across multiple endpoints (not only ALT), the need to detect moderate effects at secondary endpoints, and the pooling of results across sexes in regulatory submissions.

Pre-registration platforms:
preclinicaltrials.eu: A prospective registry for pre-clinical studies, analogous to ClinicalTrials.gov for human studies. Launched in 2019 by the ZonMw/FundamentAL programme, operating in collaboration with the NC3Rs (UK). Registration provides a publicly verifiable pre-specified hypothesis, analysis plan, and sample size justification. Pre-registration distinguishes confirmatory studies (primary analysis as per the pre-registered plan) from exploratory analyses (clearly labelled as hypothesis-generating).

OSF (Open Science Framework, osf.io): Free platform for pre-registering study protocols, storing raw data and analysis scripts, and time-stamping the hypothesis before data collection. OSF pre-registrations can be kept private until manuscript submission, then published with the paper as a time-stamped record. Used increasingly in preclinical toxicology to demonstrate transparency.

5.6.1 Randomisation and Blinding

Randomisation assigns animals to treatment groups using a random process (random number table, computer-generated randomisation with stratification by body weight) to prevent allocation bias. Blinding — masking the identity of treatments from the experimenter during data collection and analysis — prevents observation and analysis bias. Studies in which neither investigator nor analyst knows the treatment allocation (double blinding) produce the most reliable results. Cochrane reviews consistently show that unblinded studies overestimate treatment effects by a factor of 2–3 relative to blinded equivalents.

5.6.2 Sample Size Justification

Sample size must be justified prospectively using power analysis. For a two-tailed t-test comparing treated vs. control with $\alpha=0.05$ and 80% power ($\beta=0.2$), a standardised effect size of 0.5 (Cohen's d) requires approximately 33 animals per group. Many published in vivo studies are dramatically underpowered (5 animals/group when 10–15 are required), leading to high false negative rates. The reproducibility crisis in preclinical research is substantially driven by this underpowering problem.

5.6.3 Pre-Registration

Pre-registration — publicly registering the study design, primary outcomes, and analysis plan before data collection — prevents HARKing (Hypothesising After Results are Known) and outcome switching, both of which inflate false-positive rates. The OSF (Open Science Framework) pre-registration platform and the AS-preregistration template are tools available to preclinical researchers.

5.7 Humane Endpoints and Severity Classification

Species-specific clinical scoring systems and humane endpoints:

The implementation of humane endpoints requires objective, validated clinical scoring systems that permit early detection of pain, distress, or moribundity without over-relying on subjective observer judgement.

Mouse grimace scale (MGS — Langford et al., 2010, Nature Methods): The MGS scores five facial action units in mice: (1) orbital tightening (narrowing/closing of the eye aperture); (2) nose/cheek bulge (a lump under the skin); (3) ear position (ears retracted/flattened); (4) whisker change (whiskers pulled back flat against the face or forward/bundled); (5) a less specific "head shape" criterion. Each unit is scored 0 (not present), 1 (moderately present), or 2 (obviously present), yielding a composite MGS score 0–10. $MGS \geq 4$ indicates moderate pain requiring intervention. The scale has been validated for acute pain models (plantar incision, CFA injection), post-surgical pain, and chronic inflammatory conditions.

Rat grimace scale (RGS — Sotocinal et al., 2011, Molecular Pain): Similar five-action-unit scale adapted for rats: orbital tightening, nose/cheek bulge, ear position, whisker change. $RGS \geq 3$ triggers intervention protocol (analgesia, euthanasia consideration).

Body condition score (BCS) in rodents (Ullman-Culleré & Foltz, 1999, Laboratory Animal Science): BCS 1–5 scale for rodents assessing subcutaneous fat deposits and muscle mass: BCS 1 = emaciated: spinal vertebrae extremely prominent, no palpable fat, hip bones prominent; animal at risk of death — euthanasia required. BCS 2 = thin: vertebrae easily palpable, minimal fat over hips; welfare intervention required. BCS 3 = normal: vertebrae palpable with gentle pressure; normal fat deposits; no intervention required. BCS 4 = overweight: vertebrae palpable with firm pressure; moderate fat over spine and hips. BCS 5 = obese: vertebrae not palpable; extensive fat deposits; indicates over-feeding in controls or effect of tested compound. BCS is assessed weekly in dose-response studies. $BCS \leq 2$ triggers review by the study veterinarian and may warrant early euthanasia.

Body weight loss thresholds: EU Directive 2010/63/EU Annex VIII severity classification specifies that a body weight loss of $>15\%$ from the peak weight (without recovery) indicates "moderate severity"; $>20\%$ indicates "severe severity." In practice, most study protocols specify that animals losing $>20\%$ of body weight within the study period are subject to immediate veterinary review and potential euthanasia, regardless of the severity classification of the procedure. For acute toxicity studies (TG 420/423/425), the humane endpoint criteria include rapid or uncontrolled body weight loss, persistent complete anorexia, moribundity, and severe neurological signs (seizures, paresis).

Clinical sign monitoring (adapted from OECD GD 19, 2022 update): Daily observations should assess: general state (appearance, posture, body weight estimation), CNS signs (tremors, convulsions, hyperreflexia, head tilt), cardiovascular signs (bradycardia, arrhythmia assessed in larger species), respiratory signs (laboured breathing, cyanosis), integument (piloerection, alopecia, skin lesions), musculoskeletal (ataxia, weakness, paralysis), digestive (diarrhoea, rectal prolapse), urogenital (haematuria, discharge). Moribundity signs (unresponsiveness to gentle stimulation, inability to reach food/water) are absolute triggers for immediate euthanasia.

Humane endpoints are objective criteria for terminating an animal experiment when specified welfare indicators are reached, preventing unnecessary suffering beyond what is scientifically required. EU Directive 2010/63/EU requires that severity classification (mild, moderate, severe, non-recovery) be assigned to all procedures at the time of project authorisation.

Severity	Description	Examples of signs requiring endpoint action
Non-recovery	Animal remains anaesthetised throughout; never regains consciousness	Surgical procedure; single blood collection under anaesthesia
Mild	Minor pain/distress; fully reversible	Oral gavage; subcutaneous injection; brief restraint
Moderate	Noticeable suffering; potential lasting effects	90-day repeat dose; inhalation exposure; surgical implant
Severe	Major suffering; significant impairment	Carcinogenicity bioassay; lethal dose endpoint; severe infection model

Table 5.3. Humane endpoint severity classifications (EU Directive 2010/63/EU).

Specific humane endpoint criteria for rodent toxicity studies include: >20% body weight loss, moribund state (inability to stand, reach food/water), persistent severe respiratory distress, ulcerative skin lesions >20% of body surface, or neurological signs precluding normal functioning. Reaching any pre-specified endpoint criterion triggers immediate euthanasia.

5.8 Refinement Techniques

Refinement techniques — anaesthesia and analgesia protocols:

Anaesthesia is required for any painful or distressing procedure (e.g., surgical implantation of catheters, orbital blood collection, cardiac puncture). The choice of anaesthetic protocol depends on the species, the procedure duration, and the study objectives (some anaesthetics affect the pharmacokinetics or pharmacodynamics of test compounds and must be selected accordingly).

Isoflurane inhalation anaesthesia (most commonly used in rodents): Induction: 3–4% isoflurane in O₂ (1–2 L/min) in an induction chamber for 2–3 min (until loss of righting reflex). Maintenance: 1.5–2% isoflurane via nose cone or endotracheal intubation (for surgical anaesthesia). Recovery: discontinue isoflurane; animal typically recovers righting reflex within 2–5 min. Monitor: respiratory rate, mucous membrane colour, response to toe pinch (loss of withdrawal reflex = surgical depth). Advantages: rapid induction and recovery, controllable depth, no hepatic/renal metabolism concerns. Disadvantages: hypotension at high concentrations; requires scavenging system; not suitable for cardiac electrophysiology studies (hyperpolarisation of cardiac channels).

Ketamine/Xylazine injectable anaesthesia (alternative for longer procedures): Rat: ketamine 75–100 mg/kg + xylazine 5–10 mg/kg (IP). Mouse: ketamine 80–120 mg/kg + xylazine 5–16 mg/kg (IP). Duration: 30–45 min anaesthesia, 1–2h recovery. Monitoring: depth assessed by pinch reflex; intubate for oxygen supplementation if respiratory depression. Reversal: atipamezole (Antisedan) reverses xylazine (α_2 -adrenergic antagonist), dose 0.5–1 mg/kg SC. Disadvantages: hyperthermia, respiratory depression, recovery hypothermia; xylazine causes diuresis and hyperglycaemia (important confounder for metabolic studies).

Post-procedural analgesia (multimodal approach recommended): Buprenorphine (opioid, μ -receptor partial agonist): Rat 0.05–0.1 mg/kg SC q8–12h; Mouse 0.05–0.1 mg/kg SC q6–8h. Slow-release formulation (Buprenorphine SR) provides 72h analgesia from a single subcutaneous injection in rats (0.5 mg/kg SC once), substantially reducing the number of handling/injection events. Analgesic of choice for post-surgical pain.

Meloxicam (NSAID, COX-2 selective): Rat 1–2 mg/kg SC or PO once daily; Mouse 5 mg/kg SC once daily. Effective for mild-moderate inflammatory pain. Note: meloxicam inhibits prostaglandin synthesis; may affect renal function in dehydrated animals — ensure adequate hydration.

Carprofen (NSAID): 5 mg/kg SC q12h in rats. Available in TemGel oral sustained-release formulation for delivery in the water/food.

Multimodal approach: combining an opioid (buprenorphine) with an NSAID (meloxicam) provides synergistic analgesia at lower doses of each agent, reducing side effects (opioid: sedation, constipation; NSAID: renal/GI). Recommended for all post-surgical recovery protocols.

Environmental enrichment specifics for welfare refinement: For laboratory mice: nesting material (Sizzle-nest or equivalent paper-based material, minimum 3g per cage); nestlets (cotton fibre squares, 5×5cm, 1–2 per cage for nest building); shelter (igloo or mouse house, opaque plastic); gnaw sticks (untreated wooden dowels 1cm diameter × 5cm). For rats: paper nestlets; cardboard tube or PVC pipe (7cm diameter) for hiding/climbing; gnaw blocks (hardwood); tissue paper for nesting. For all social species: minimum dyad housing (2 animals per cage) unless single housing is scientifically or welfare-justified and time-limited with enhanced enrichment.

Refinement covers all measures that reduce the pain, distress, suffering, or lasting harm experienced by animals used in experiments.

- Anaesthesia: inhalant (isoflurane 1.5–3% for maintenance in rodents; sevoflurane for rabbits) or injectable (ketamine/xylazine; alfaxalone) anaesthesia eliminates pain during surgical procedures. Monitoring includes respiratory rate, pedal withdrawal reflex, mucous membrane colour.
- Analgesia: pre-emptive analgesia (administered before the painful stimulus) is more effective than post-procedure analgesia. NSAIDs (carprofen 5 mg/kg SC for rats) and opioids (buprenorphine 0.05–0.1 mg/kg SC) are standard rodent analgesics. Analgesic use must be compatible with study endpoints (NSAIDs affect COX-dependent endpoints).
- Environmental enrichment: improves animal welfare and may reduce outcome variability by normalising the physiological baseline. Standard enrichment: nestlets, gnawing blocks, play tubes (rodents); social housing (most species); foraging opportunities.

5.9 Reduction Strategies

- Factorial design: testing multiple variables simultaneously in a single experiment reduces the total number of animals required compared to one-factor-at-a-time designs.
- Pilot studies: small preliminary experiments (n=3–5) estimate variance, validate dosing routes, and confirm assay functionality before the main powered study.
- Data sharing and meta-analysis: pooling historical control data across studies reduces the group size needed for concurrent controls.
- Statistical optimisation: Adaptive design approaches (sequential stopping rules) can reduce animal use by up to 30–50% without sacrificing power.

5.10 Data Integrity and the Reproducibility Crisis

Reproducibility and the preclinical research crisis:

The reproducibility crisis in preclinical animal research has been extensively documented since 2011. Landmark analyses include:

Begley & Ellis (2012, Nature): Scientists at Amgen attempted to replicate 53 landmark cancer biology studies and confirmed only 6 (11%). Of 10 studies tested in acute myeloid leukaemia, none replicated. The authors attributed failures to: selective reporting, underpowered studies, absence of blinding, cherry-picking of positive results.

Landis et al. (2012, Nature Neuroscience): Analysed 4,000 neuroscience papers and found that the median statistical power (ability to detect true effects) was only 18–31% — meaning that even real

effects would be detected in fewer than 1 in 3 experiments. Studies with $n=5-6$ per group and moderate effect sizes typically have power $<50\%$.

Begley et al. (2015, Circulation Research): Cardiovascular biology reproducibility analysis; $\sim 50\%$ of published results could not be replicated by independent groups.

For toxicology specifically, Kilkenny et al. (2009, PLOS ONE) found that 87% of papers in three journals did not report the number of animals, 86% did not describe randomisation, and 90% did not mention blinding — making external validation impossible. The NC3Rs (2023) ARRIVE Compliance Audit of 200 randomly selected papers from 2022 found that reporting of randomisation, blinding, and sample size justification had improved compared to 2009, but was still incomplete in 35–60% of papers.

Recommendations (ILAR, NC3Rs, 2023):

- (1) Pre-register confirmatory studies on preclinicaltrials.eu before data collection.
- (2) Report all pre-specified analyses (not only significant ones) — distinguish a priori confirmatory analyses from post hoc exploratory analyses.
- (3) Share raw data in open repositories (figshare, Dryad, OSF) upon manuscript acceptance.
- (4) Use multi-laboratory designs: the same compound tested in ≥ 3 independent laboratories simultaneously, using a harmonised protocol, provides a far more reliable effect estimate than a single laboratory experiment. EURL ECVAM requires multi-laboratory validation for acceptance of new test methods.
- (5) Adopt registered reports: journal format where the study design and analysis plan are peer-reviewed and conditionally accepted before data collection, ensuring publication of results regardless of statistical significance.

A landmark 2011 analysis by Begley and Ellis found that only 11% of 53 'landmark' cancer biology papers could be reproduced by an industrial laboratory. In preclinical pharmacology, a 2014 Novartis survey found that only 25% of published results could be replicated internally. Contributing factors include: lack of blinding; underpowered studies; p-hacking and outcome switching; selective publication of positive results; reagent authentication failures; and poor animal model characterisation.

Practical measures to improve reproducibility in toxicological research:

- Use authenticated cell lines (STR profiling; mycoplasma testing at establishment and quarterly).
- Pre-register primary outcomes and analysis plan before study initiation.
- Report all data, including null results and unexpected findings, in the final report.
- Comply with ARRIVE 2.0 at manuscript submission.
- Archive raw data in FAIR (Findable, Accessible, Interoperable, Reusable) data repositories.

5.11 Algerian Context: Implementation at USTO-MB

Algerian context — Practical implementation at USTO-MB:

The Laboratoire de Biologie Marine et des Pêches (LBMP) at USTO-MB does not operate a conventional animal facility for regulatory-grade in vivo toxicity testing. However, the Laboratory conducts ecotoxicological studies on marine organisms (Mediterranean fish, invertebrates) and occasional in vitro studies on cell lines. The following practical framework applies to USTO-MB:

Current capabilities: (1) In vitro work in Class II biological safety cabinets; (2) Cell culture facility with CO₂ incubators, centrifuges, and fluorescence plate reader; (3) Spectrophotometric and fluorometric endpoint measurements (MTT, DCFH-DA, AAS); (4) access to the broader FSNV animal facility for rodent studies, which maintains approximately 200 Wistar rats and 300 NMRI mice.

Current limitations: (a) No formal GLP accreditation of the animal facility; (b) absence of FELASA-compliant health monitoring programme (no sentinel animals, no quarterly serology); (c) limited access to ARRIVE-compliant statistical software for power calculations; (d) no pre-registration system has

been used in published LBMP studies (as of 2025); (e) histopathology services require outsourcing to the Faculty of Veterinary Science.

Improvement

roadmap:

Short-term (1–2 years): (1) Establish USTO-MB CER ethics committee with formalized protocol review procedure and reference numbering; (2) implement quarterly health screening of the rodent colony (serology for Sendai, MHV, Helicobacter, pinworms); (3) adopt ARRIVE 2.0 as a mandatory reporting standard for all animal studies submitted for publication; (4) install G*Power 3.1 on all laboratory computers and require documented power calculations for all new studies; (5) create SOP-TOX-001 to SOP-TOX-010 covering the core procedures (gavage, blood collection, tissue fixation, cell culture, MTT, AAS).

Medium-term (3–5 years): (1) Apply for AAALAC provisional accreditation (Programme Description preparation); (2) implement GLP-like record-keeping for in vitro studies (paper laboratory notebooks with indelible ink, corrections policy, daily countersignature); (3) register USTO-MB in the national MESRS animal experimentation statistics reporting system; (4) partner with a GLP-certified contract research organisation for regulatory submissions requiring full GLP compliance.

Training requirements: All personnel performing in vivo studies must demonstrate competence in the following skills: aseptic technique (sterile injection); oral gavage technique (assessed by dedicated training module, minimum 20 supervised gavages before independent performance); recognition of clinical signs and humane endpoints (online training available via NC3Rs website); basic anaesthesia and analgesia (species-specific); euthanasia methods (CO₂ asphyxiation, cervical dislocation for rodents — EU Directive 2010/63/EU Annex IV methods). Annual refresher training is mandatory.

The practical implementation of good experimental practice at Algerian research institutions, including USTO-MB, is shaped by specific institutional constraints and opportunities.

- USTO-MB Laboratory of Marine Biology and Fisheries: conducts ecotoxicological in vivo studies (freshwater fish, Mediterranean species) with experimental protocols approved by the institutional ethics committee.
- GLP awareness: Algerian researchers publishing in international journals are increasingly required to declare GLP compliance for regulatory-adjacent studies; training workshops by the DGRSDT promote GLP principles.
- Resource-constrained 3R implementation: in vitro pre-screening (MTT on HepG2 for Algerian contaminants) before in vivo studies is an efficient cost-reducing and 3R-compliant strategy for USTO-MB research programmes.
- Loi 12-06: provides the institutional framework for ethics committee oversight; detailed implementing regulations specific to animal research are awaited.
- International collaboration: joint projects with European institutions (Erasmus+, PHC Tassili programme) provide access to GLP-compliant facilities and training for Algerian toxicologists.

Key Take-Home Messages

- GLP is a quality system that ensures non-clinical safety studies are reliable, reproducible, and traceable. It is required for regulatory submissions and is increasingly expected for academic studies with regulatory relevance.
- The Study Director bears sole responsibility for GLP study conduct and must sign the final report.
- ARRIVE 2.0 Essential 10 provides a consensus standard for reporting in vivo experiments; compliance is required by major journals.
- Randomisation, blinding, and prospective sample size calculation are the three most impactful measures for improving the quality and reproducibility of animal studies.

- Humane endpoints must be pre-specified and operationally defined before the study begins; severity classification under Directive 2010/63/EU must be accurate.
- The reproducibility crisis in preclinical research is substantially driven by poor experimental design, and all measures taken to improve design quality also reduce unnecessary animal use.
- At USTO-MB, institutional ethics committee approval, ARRIVE 2.0 compliance, and pre-screening by in vitro methods represent the practical implementation of good animal experimentation practice.

Further Reading

Percie du Sert N et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biology*. 2020;18(7):e3000410. doi:10.1371/journal.pbio.3000410

Medhi B et al. Impact of OECD principles of Good Laboratory Practices (GLP) in India. *Indian J Pharmacol*. 2021;53(5):353–358. doi:10.4103/ijp.ijp68521

Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531–533. doi:10.1038/483531a

Kappenberg F et al. Guidance for statistical design and analysis of toxicological dose-response experiments. *Arch Toxicol*. 2023;97(10):2741–2761. doi:10.1007/s00204-023-03561-w

Part 6 — Alternative Methods to Animal Experimentation

6.1 Learning Objectives

- Explain the drive towards New Approach Methodologies (NAMs) and situate them within the 3R framework.
- Describe the validation processes for alternative methods at ECVAM/EURL ECVAM, ICCVAM, JaCVAM, and OECD.
- Apply QSAR and read-across approaches to predict toxicity for untested chemicals.
- Describe the role of PBPK modelling in extrapolating from in vitro to in vivo doses.
- Explain the AOP framework and its role in Integrated Approaches to Testing and Assessment (IATA).
- Identify fully validated alternative methods (OECD TG 431, 439, 492, 442C/D/E) and the endpoints they cover.
- Critically evaluate the current limitations and future perspectives of alternative methods in regulatory toxicology.

6.2 Introduction: The 3R Drive Toward NAMs

New Approach Methodologies (NAMs) is the umbrella term for the diverse toolkit of non-animal methods that aim to replace, reduce, or refine animal testing. The term encompasses in vitro cell-based assays, organoids, organ-on-chip, in silico models (QSAR, PBPK, ML), read-across, adverse outcome pathways, and integrated testing strategies. The momentum behind NAMs is driven by three complementary forces: the ethical imperative of the 3R principle (Part 1); the scientific case for human-relevant data (Part 3); and the regulatory and economic pressure to handle the enormous backlog of chemicals requiring hazard assessment under REACH and similar frameworks.

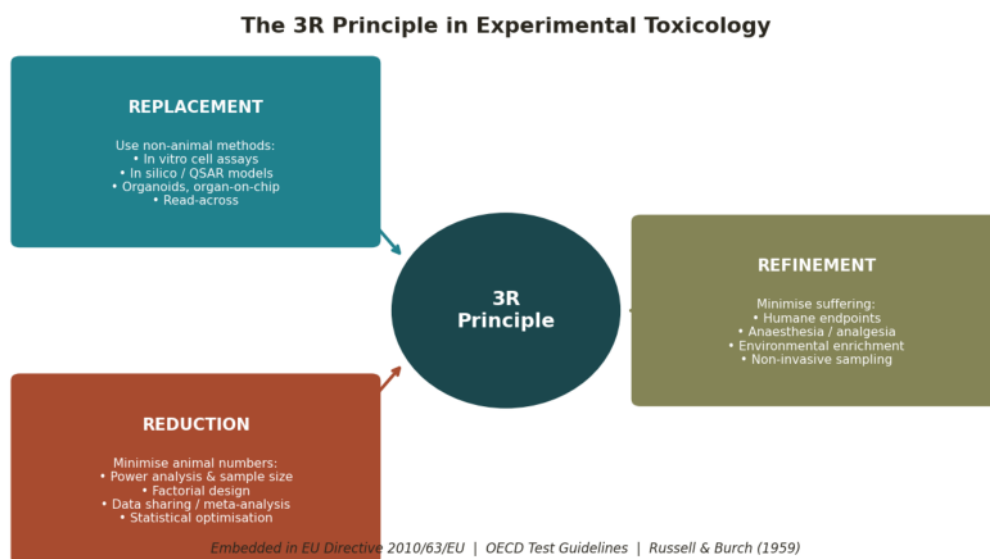


Figure 6.1. The 3R Principle — the ethical and regulatory backbone driving the development of alternative methods. NAMs represent the scientific operationalisation of the Replacement arm of the 3R framework. (Original figure.)

6.3 Validation of Alternative Methods

Detailed ECVAM validation framework and OECD 5-step process:

EURL ECVAM (European Union Reference Laboratory for alternatives to animal testing) manages the validation of alternative test methods in Europe through a modular approach that assesses multiple dimensions of test performance.

ECVAM Modular Approach (ECVAM, 2013, based on Hartung et al., 2004, Alternatives to Laboratory Animals):

Module 1 — Test Method Development: Describes the test method in a standard format, defines the endpoint/biological effect being measured, and establishes preliminary relevance and reliability data from within-laboratory studies. This module produces a test method specification (the detailed protocol).

Module 2 — Optimisation and Within-Laboratory Reproducibility: Conducts systematic experiments to optimise performance criteria (concentration range, exposure time, cell density, reagent lots). Within-laboratory reproducibility is assessed by intra-day, inter-day, and inter-operator variation using coded reference chemicals. Coefficient of variation (CV) for positive controls: $CV < 20\%$ required; for negative controls: $CV < 10\%$.

Module 3 — Pre-Validation: A small-scale multi-laboratory study (2–3 laboratories) using a coded panel of 20–30 reference chemicals (a mixture of known positives and negatives for the endpoint). Assesses between-laboratory reproducibility and provides a preliminary predictive capacity estimate. This module generates the transfer protocol.

Module 4 — Formal Validation (ring trial): A blind, multi-laboratory study (minimum 3 laboratories, typically 5–8) using a coded panel of 40–60 chemicals. Each laboratory independently tests the panel without knowledge of the expected outcomes. Predictive capacity is assessed: sensitivity (proportion of true positives correctly identified), specificity (proportion of true negatives correctly identified), overall accuracy, and reproducibility across laboratories (Kendall coefficient of concordance or kappa statistic).

Module 5 — Independent Peer Review: The validation data package is submitted to ECVAM's Scientific Advisory Committee (ESAC) or an equivalent peer review body for independent assessment. ESAC issues a Statement confirming (or rejecting) whether the data support regulatory acceptance of the method. Upon positive ESAC opinion, the method is forwarded to the OECD for TG development.

OECD 5-step validation process:

Step 1 — Test method assessment: OECD member countries review the ECVAM/ICCVAM data package and decide whether to proceed with TG development.

Step 2 — Test guideline development: A draft TG is prepared in OECD format (Sections 1–12: purpose, applicability domain, method description, reporting requirements, references).

Step 3 — Intercomparison exercise: Additional data from multiple laboratories are reviewed to support the TG text.

Step 4 — Advisory/expert consultation: The draft TG is submitted to WNT (Working Group of the National Coordinators of the OECD Test Guidelines Programme) and circulated for OECD member comment.

Step 5 — Adoption: Final TG is adopted by the OECD Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology and published in the OECD Guidelines for Testing of Chemicals, Series on Test Guidelines.

Regulatory acceptance criteria vary by endpoint. For skin sensitisation, the OECD has adopted 5 alternative TGs (TG 429 LLNA as reference, TG 442C DPRA, TG 442D KeratinoSens, TG 442E h-CLAT, TG 497 Defined Approaches) that together constitute an IATA. For acute oral toxicity, the cytotoxicity-based in vitro methods (e.g., 3T3 NRU for cytotoxicity; 3T3-NRU assay, OECD TG 129) provide a "starting dose" for in vivo testing but do not yet fully replace in vivo acute testing for GHS classification.

6.3.1 ECVAM and EURL ECVAM

The European Centre for the Validation of Alternative Methods (ECVAM), now the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), is part of the Joint Research Centre (JRC) of the European Commission. It coordinates validation studies for alternative methods in the EU, following a multi-phase process: method development and standardisation; multi-laboratory reproducibility studies; predictive capacity assessment against a defined reference chemical set; independent peer review (ESAC); and submission to OECD for international adoption as a Test Guideline.

6.3.2 ICCVAM and JaCVAM

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) coordinates US federal agencies (EPA, FDA, NIH/NIEHS, DOD) in evaluating and promoting alternative methods for regulatory use. The Japanese Centre for the Validation of Alternative Methods (JaCVAM), established in 2005, serves a similar function in Japan. All three bodies collaborate through the International Council on Animal Protection in OECD Guidelines (ICAPO) to achieve harmonised global acceptance of validated alternatives.

6.3.3 OECD Validation Process

OECD Guidance Document 34 (2005) provides the procedural framework for validation and international acceptance. A validated alternative method, endorsed by ESAC and submitted to OECD, is reviewed by the OECD Test Guidelines Programme. If accepted, it is published as a new or updated Test Guideline, making it internationally admissible for regulatory use under MAD.

6.4 In Silico Modelling and Analysis

6.4.1 QSAR — Quantitative Structure-Activity Relationships

QSAR — 5 OECD Validation Principles and Worked Example:

The OECD Guidance Document on Quantitative Structure-Activity Relationships (GD 69, 2004; updated GD 146, 2014) establishes five minimum information requirements that a QSAR model must satisfy for regulatory acceptance:

Principle 1 — Defined endpoint: The model must predict a clearly defined biological or chemical endpoint. The endpoint must be experimentally measurable, unambiguous, and relevant to the regulatory question (e.g., oral LD50 in rat; Ames mutagenicity; log Kow; aquatic LC50 in fish).

Principle 2 — Unambiguous algorithm: The model must be based on a transparent, documented mathematical algorithm. Black-box approaches (e.g., neural networks with no documented architecture) are less acceptable than mechanistically interpretable models (linear regression, MLR, PLS). The algorithm must be fully reproducible from the published model description.

Principle 3 — Defined applicability domain (AD): Every QSAR model is only valid for chemicals similar to the training set. The AD must be explicitly defined — typically using: descriptor ranges (the new chemical must fall within the minimum and maximum values of each molecular descriptor in the training set); structural alerts (Tanimoto coefficient ≥ 0.5 to the nearest training set neighbour); and chemical class membership. Predictions for chemicals outside the AD are flagged as unreliable.

Principle 4 — Appropriate measures of goodness-of-fit, robustness, and predictivity: The model must report: R^2 or Q^2 (cross-validated R^2) for regression models; sensitivity, specificity, and accuracy for classification models; y-randomisation tests (demonstrating that the correlation is not due to chance); external validation using a test set not used in model development (Q^2_{ext} should be ≥ 0.60 for regression, accuracy $\geq 65\%$ for classification).

Principle 5 — Mechanistic interpretation where possible: The model should provide, where scientifically possible, a mechanistic basis for the structure-activity relationship. Structural alerts (e.g., Michael acceptors in skin sensitisation, epoxide groups in genotoxicity, para-nitroaniline for Ames TA98 frameshift mutation) provide mechanistic confidence. Purely statistical correlations without mechanistic rationale are less reliable for extrapolation.

Worked	QSAR	example:
Target compound:	4-nitrostyrene (CAS 100-13-0; SMILES: <chem>C=Cc1ccc([N+](=O)[O-])cc1</chem>).	
Endpoint:	Log Kow (octanol-water partition coefficient).	
Descriptors calculated (EPI Suite v4.11, US EPA): molecular weight = 149.17; topological polar surface area (TPSA) = 45.8 Å ² ; number of rotatable bonds = 2; AlogP (Ghose-Crippen) = 2.45.		
Applicability domain check: training set includes 1,000+ aromatic compounds; MW range 80–700 (within AD); TPSA range 0–150 Å ² (within AD); no structural alerts that would suggest metabolic transformation invalidating the LogP measurement; Tanimoto similarity to nearest neighbour (4-chlorostyrene, LogP = 3.19) = 0.78 (within AD).		
QSAR prediction: LogP (EPI Suite consensus) = 2.78; experimental LogP = 2.90 (measured).		
Applicability domain: within AD. Confidence: high. AD flag: green.		

For aquatic toxicity prediction: using ECOSAR v2.2, 4-nitrostyrene is classified as an α,β -unsaturated carbonyl/Michael acceptor (olefinic system activated by the electron-withdrawing nitro group). Fish 96h LC50 (estimated): 1.6 mg/L (ECOSAR; electrophile class). Applicability domain: within AD. ECOSAR reliability flag: good (training set contains 85 structurally similar nitrobenzene derivatives, R²=0.81 for this class).

QSAR models relate molecular structure to biological activity using mathematical functions of molecular descriptors (physicochemical, topological, electronic). The OECD has published five principles for a valid QSAR: defined endpoint, unambiguous algorithm, defined applicability domain (AD), appropriate goodness-of-fit statistics, and mechanistic interpretation.

Regulatory-accepted QSAR predictions are available for: Ames mutagenicity (Derek Nexus, Toxtree), carcinogenicity (OECD QSAR Toolbox), skin sensitisation (DPRA, ARE-Nrf2 reporter), acute aquatic toxicity (ECOSAR), and physicochemical properties (logP, solubility, pKa). A 2025 comprehensive review by Lancia et al. catalogued 27 QSAR models covering 15 endpoints for cosmetic ingredient safety assessment, reflecting the maturity of the field.

6.4.2 Read-Across and Chemical Grouping

Read-Across — detailed framework and REACH Annex XI:

Read-across is a technique by which data on properties/endpoints from one or more source chemical(s) are used to predict properties/endpoints for a target chemical where such data are lacking. Read-across is permitted under REACH Article 13 and Annex XI (Section 1.5) as a valid data-filling approach, provided it is adequately justified.

Types	of	read-across:
(1) Analogue approach: A single source chemical closely related to the target is used. The prediction is qualitative (positive/negative) or semi-quantitative (within a defined range). Example: predicting reproductive toxicity of dibutyl phthalate (DBP, target) from data on diisobutyl phthalate (DIBP, source), which shares the same mechanism of action (PPAR-mediated Leydig cell toxicity) and similar toxicokinetics.		
(2) Category approach: Multiple source chemicals form a structural group sharing physico-chemical properties, biological activities, or metabolic fates. A trend (e.g., increasing toxicity with increasing chain length) is used to predict the target. Example: the n-alkanes category in REACH, where toxicity increases with increasing carbon chain length up to C ₁₃ (narcosis mechanism).		

REACH Annex XI justification requirements:

The read-across submission must document: (1) Identity and structure of source and target chemicals (with structural comparison); (2) Comparison of physico-chemical and ADME properties; (3) Structural alerts and metabolism considerations (do source and target share the same metabolic pathways?); (4) Mechanistic hypothesis (why are the properties expected to be similar?); (5) Data quality assessment (reliability Klimisch score 1–4 for the source substance data); (6) Uncertainty assessment (are there structural features in the target that could deviate from the source prediction?).

ECHA's R&D Read-across Assessment Framework (RAAF, 2017) provides a structured approach to documenting and evaluating read-across justifications, with specific criteria for 8 scenario types (analogue approach with metabolic consideration, category approach with trend analysis, etc.). The RAAF uses a 3-step evaluation: adequacy of the source data; plausibility of the structural/mechanistic hypothesis; uncertainty of the prediction.

Validated examples: (a) The polycyclic aromatic hydrocarbons (PAHs) category — carcinogenic potency read-across from benzo[a]pyrene (IARC Group 1) to other PAHs using the relative potency factor (RPF) approach (EPA PAH Mixtures, 1993; US EPA, 2010); (b) The PFAS (perfluoroalkyl substances) category — PFOA data used in read-across for shorter-chain PFAS for developmental toxicity (ECHA 2022 SVHC identification); (c) N-alkyl acrylates — ECHA's category-based hazard assessment for a series of acrylate esters sharing the Michael acceptor structural alert.

Read-across fills data gaps for data-poor chemicals by inferring toxicity from structurally or mechanistically similar 'source' chemicals with known toxicity data. The ECHA Read-Across Assessment Framework (RAAF) provides structured criteria for documenting and assessing read-across predictions in regulatory dossiers. Chemical grouping — classifying chemicals with common structural motifs, metabolic pathways, or mode of action into categories for assessment — enables efficient use of existing data to fill gaps for multiple chemicals simultaneously.

6.4.3 PBPK Modelling

PBPK Modelling — detailed structure and IVIVE:

A physiologically based pharmacokinetic (PBPK) model represents the body as a series of anatomically realistic compartments (organs and tissues) connected by blood flow. Unlike classical one- or two-compartment models, PBPK models use actual physiological parameters (cardiac output, tissue volumes, blood flows) and biochemical parameters (partition coefficients, metabolic rate constants) to describe the time course of a chemical in each organ.

Standard PBPK compartments for a chemical with hepatic metabolism and renal excretion:

- (1) Arterial blood compartment: receives oxygenated blood from the pulmonary circulation; distributes chemical to all tissue compartments at rates proportional to blood flow.
- (2) Lung compartment: important for volatile chemicals; partition between blood and exhaled air described by blood:air partition coefficient ($P_b:a$).
- (3) Liver compartment: site of biotransformation; includes hepatic intrinsic clearance (CL_{int} , determined in vitro) scaled to whole-liver in vivo clearance using the well-stirred or parallel-tube model. Hepatic extraction ratio = $CL_{int}/(CL_{int} + Q_h)$, where Q_h = hepatic blood flow.
- (4) Kidney compartment: glomerular filtration + tubular secretion/reabsorption; CL_r derived from in vitro transporter kinetics.
- (5) Fat compartment: reservoir for lipophilic chemicals; long half-life due to low blood flow (Q_{fat} = 7% of cardiac output in rat vs. 5% in human).
- (6) Slowly perfused tissues (muscle, skin): bulk of body mass; important for slowly equilibrating chemicals.
- (7) Rapidly perfused tissues (brain, heart, gut): high blood flow; reach equilibrium rapidly.

Key PBPK equations (example: liver compartment):

$$\frac{dCL}{dt} = QL \times (CA - CVL) - VL \times (V_{max} \times CVL) / (K_m + CVL)$$
 where: CL = concentration in liver (µg/mL); QL = liver blood flow (L/h); CA = arterial blood concentration; CVL = venous blood concentration leaving liver = CL/PL (PL = liver: blood partition coefficient); V_{max} = maximum metabolic velocity (µg/h); K_m = Michaelis constant (µg/mL).

IVIVE (In Vitro to In Vivo Extrapolation) procedure:
 Step 1: Measure CL_{int}, in vitro in hepatocyte suspension (rat or human, at initial substrate concentrations well below K_m): CL_{int} = rate of substrate disappearance / [C]. Units: µL/min/10⁶ cells.
 Step 2: Scale to whole liver: CL_{int, scaled} = CL_{int, in vitro} × MPPGL × liver weight, where MPPGL = microsomal protein per gram of liver (52 mg/g for rat, 32 mg/g for human) or HPGL = hepatocytes per gram of liver (1.1 × 10⁸ for rat, 1.07 × 10⁸ for human).
 Step 3: Apply unbound fraction correction: CL_{int, u} = CL_{int, scaled} / f_{u, inc} (fraction unbound in incubation medium, measured by equilibrium dialysis or rapid equilibrium dialysis (RED)).
 Step 4: Calculate predicted hepatic clearance using the well-stirred model: CL_h = Q_h × f_{u, p} × CL_{int, u} / (Q_h + f_{u, p} × CL_{int, u}), where f_{u, p} = fraction unbound in plasma.
 Step 5: Validate against in vivo clearance data (if available) and refine partition coefficients. Input CL_h into PBPK model.

Software: Open-source PBPK tools: (1) PK-Sim (Open Systems Pharmacology Suite, bayer-open.github.io): free, PBPK model builder for drugs and industrial chemicals; integration with MoBi for QSP models. (2) IndusChemFate (free academic licence): designed for industrial chemicals, generates PBPK models from physico-chemical descriptors alone. Commercial: Simcyp v22 (Certara) — standard for pharmaceutical applications; GastroPlus 10.0 (Simulation Plus) — oral absorption focus. PBPK models are now routinely submitted to FDA and EMA for dose optimisation and first-in-human dose prediction.

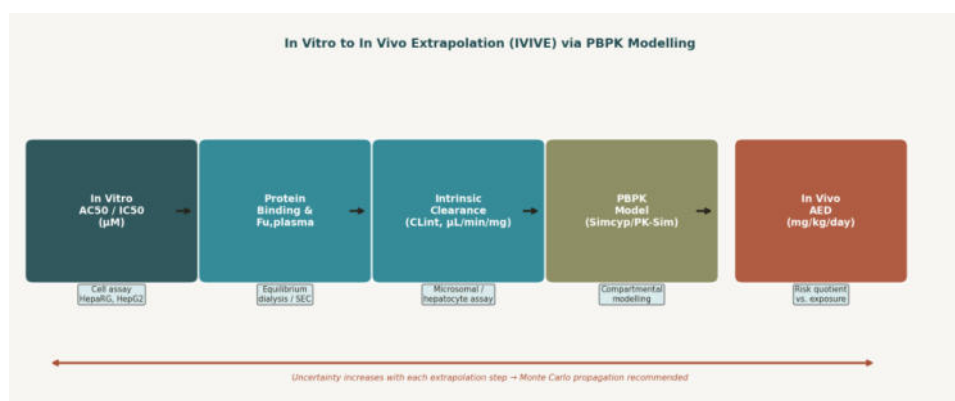


Figure 6.E1. In Vitro to In Vivo Extrapolation (IVIVE) workflow using PBPK modelling. Starting from the in vitro AC₅₀ (cell-based assay), the measured protein binding (f_{u, plasma}) and intrinsic clearance (CL_{int}) parameters are input to a PBPK model (e.g., PK-Sim or Simcyp) to predict the administered equivalent dose (AED) in vivo. Uncertainty increases with each extrapolation step; Monte Carlo simulation is recommended to propagate uncertainty to the final AED estimate.

Physiologically Based Pharmacokinetic (PBPK) models describe xenobiotic kinetics as a system of tissue-compartment equations parameterised by physiology (organ volumes, blood flows) and substance-specific constants (partition coefficients, metabolic rates). PBPK enables: (a) interspecies extrapolation (rat → human); (b) route-to-route extrapolation (inhalation → oral); (c) dose scaling from in vitro concentrations to in vivo doses (IVIVE — critical for incorporating in vitro data into regulatory risk assessment). Open-source tools: PK-Sim (Bayer, free academic license), OpenPK (R package), htk (EPA).

6.4.4 Machine Learning and Deep Learning in Toxicology

Machine learning in toxicology — algorithm comparison and regulatory considerations:

Random Forests (RF):

An ensemble of decision trees trained on bootstrap samples of the training data (bagging), with random feature selection at each tree node. RF outputs a probability score (vote fraction) for classification tasks. Advantages: handles non-linear relationships; robust to outliers and correlated features; provides variable importance scores (Mean Decrease in Impurity) as a proxy for mechanistic relevance; good performance even with small training sets ($n < 1,000$). Limitations: difficulty in extrapolating beyond the training data (similar to QSAR AD limitation); not readily interpretable at the individual-tree level; large memory requirements for deep forests. Typical toxicology application: Ames mutagenicity prediction (Kazius et al., 2005, J Med Chem — Deductive Estimation of Risk from Existing Knowledge, DEREK, uses RF internally); QSAR for hERG channel inhibition (cardiotoxicity); skin sensitisation classification.

Support Vector Machines (SVM):

A kernel-based method that finds the optimal hyperplane separating classes in a high-dimensional feature space. With the radial basis function (RBF) kernel, SVMs handle non-linear boundaries. Advantages: excellent performance for binary classification with limited training data; robust to high-dimensional descriptors (molecular fingerprints with 2,048+ bits). Limitations: lacks interpretability (no variable importance); slow to train on large datasets; requires careful hyperparameter tuning (C, gamma by grid search with cross-validation). Typical application: acute oral toxicity classification; carcinogenicity prediction (Kazius et al.; ToxTree); genotoxicity QSAR.

Deep Neural Networks (DNN):

Multi-layer perceptrons or convolutional neural networks trained end-to-end on molecular descriptors or fingerprints. Advantages: can learn hierarchical feature representations automatically; achieves state-of-the-art performance on large benchmark datasets (MoleculeNet: Tox21 = 23 assays $\times$ 8,014 compounds; ToxCast = 600 assays $\times$ 9,000 compounds). Limitations: requires large training datasets ($>10,000$ compounds per endpoint for reliable generalisation); black-box — poor interpretability; high risk of overfitting without regularisation (dropout, L2); difficult to define applicability domain.

Graph Neural Networks (GNN):

Treat molecules as graphs (atoms = nodes, bonds = edges) and learn representations by message-passing between neighbouring atoms. GNN-based models (e.g., Attentive FP, D-MPNN/ChemProp, CGCNN) directly process SMILES or 3D molecular graphs without pre-computed descriptors. Advantages: no need for hand-crafted descriptors; can capture complex 3D and topological features; achieves competitive performance on MoleculeNet benchmarks. Limitations: computationally expensive to train; AD definition is non-trivial; limited interpretability (though graph attention mechanisms provide some insight).

DeepTox (Mayr et al., 2016, Frontiers in Environmental Science): winner of the Tox21 Data Challenge (2014) using DNN trained on all 23 Tox21 assays simultaneously; demonstrated that multi-task learning (sharing information across assays) outperforms single-task models, especially for endpoints with limited data.

Interpretability challenges: The primary barrier to regulatory acceptance of ML models is the "black-box" problem — regulators (ECHA, EPA) require mechanistic transparency to be confident that predictions are not spurious correlations. Explainability methods are being developed: SHAP (SHapley Additive exPlanations) decomposes model predictions into contributions from each molecular feature (Lundberg & Lee, 2017); gradient-based saliency maps for GNNs highlight which atoms contributed most to the prediction. ECHA's Read-Across Assessment Framework (2022 update) is beginning to accept ML models with SHAP-based explanations as supporting evidence in QSAR justifications.

Graph neural networks, deep learning models trained on molecular graph representations, and multi-task neural networks trained on ToxCast/Tox21 data are advancing QSAR beyond the applicability limitations of classical regression-based models. Key challenges for regulatory acceptance are: (a) interpretability (black-box models require SHAP/LIME explainability tools); (b) training data quality (garbage in, garbage out); (c) applicability domain definition for complex models. Lin & Chou (2022) provide a comprehensive review of ML applications in toxicological sciences.

6.4.5 Software Tools

Software tools for alternative methods — comparison table:

OECD QSAR Toolbox v4.6 (OECD/EC Joint Research Centre, free):
Cost: Free. Strengths: integrated with OECD databases (ECOTOX, EPA, Japan); direct chemical category formation; built-in profilers for structural alerts; gap-filling and read-across documentation export. Limitations: requires Windows; complex user interface; AD assessment limited. Regulatory acceptance: OECD-endorsed; used in REACH submissions as standard documentation tool. User interface: GUI-based, moderate learning curve.

VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture, free from IRFMN, Italy):
Cost: Free. Strengths: consensus models for LogP, skin sensitisation, acute aquatic toxicity, mutagenicity, carcinogenicity; QSAR applicability domain check built-in; exports ECHA QSAR Model Reporting Format (QMRF) documents automatically. Limitations: models not always the most recent; limited customisation. Regulatory acceptance: QMRF output directly accepted by ECHA. User interface: clean GUI, easy for non-experts.

Simcyp Simulator v22 (Certara, commercial):
Cost: Annual licence €15,000–25,000 per site. Strengths: gold standard for pharmaceutical PBPK; population-based simulation (virtual populations with genetic polymorphisms, age, sex, obesity); CYP interaction models; GastroPlus integration. Limitations: cost; requires pharmacology/PK expertise; primarily designed for drugs, not environmental chemicals. Regulatory acceptance: FDA and EMA accept Simcyp-derived human dose predictions for first-in-human dose selection. User interface: Windows GUI, comprehensive but steep learning curve.

PK-Sim (Open Systems Pharmacology Suite, free/open source):
Cost: Free (GitHub). Strengths: open-source, community-maintained; suitable for both drugs and environmental chemicals; companion software MoBi allows custom QSP model building. Limitations: fewer validated compound databases than Simcyp; less pharmaceutical regulatory acceptance history. Regulatory acceptance: growing; accepted by BfR (German Federal Institute for Risk Assessment) for pesticide PBPK models. User interface: Windows GUI; comparable functionality to Simcyp for standard PBPK applications.

EPI Suite v4.11 (US EPA, free):
Cost: Free. Strengths: gold standard for environmental fate properties (LogKow, Koc, BCF, vapour pressure, Henry's law constant, biodegradation); part of the DSSTOX database ecosystem. Limitations: limited toxicity predictions; older models. Regulatory acceptance: standard for REACH physico-chemical property estimation; required for ECHA Annex XII chemical safety assessments. User interface: legacy Windows GUI; outputs text-based; scripting possible.

DEREK Nexus (Lhasa Limited, commercial):
Cost: Annual licence. Strengths: expert-rule-based system for predicting mutagenicity, carcinogenicity, skin sensitisation, reproductive toxicity from structural alerts; transparent reasoning (output includes specific structural alerts and supporting literature); widely used in pharmaceutical industry for lead compound screening. Limitations: rule-based (cannot extrapolate beyond programmed alerts); frequent false negatives for novel structural classes. Regulatory acceptance: ICH M7 guideline (2017) for

mutagenicity prediction in pharmaceutical impurities explicitly allows DEREK + statistical model as an acceptable in silico prediction system. User interface: Windows GUI; straightforward.

Principal freely available in silico software tools for regulatory toxicology:

Tool	Developer	Primary Use	Availability
OECD QSAR Toolbox	OECD / LHASA	Profiling, read-across, skin sensitisation	Free (OECD website)
VEGA	IRFMN / EFSA	Multi-endpoint QSAR predictions	Free (vegahub.eu)
Toxtree	Ideaconsult / JRC	Structural alerts (Cramer, Verhaar)	Free (jrc.ec.europa.eu)
Derek Nexus	LHASA Limited	Expert-rule-based toxicity alerts	Commercial (academic licenses)
OPERA	US-EPA / NC State	ADMET properties, QSAR	Free (GitHub)
PK-Sim	Bayer AG (Open Systems)	PBPK modelling	Free (open-source)
AOP-Wiki	OECD / sLCI	AOP knowledge base	Free (aopwiki.org)

Table 6.2. Key QSAR and in silico software tools in predictive toxicology.

6.5 Tools of Predictive Toxicology

6.5.1 Adverse Outcome Pathways (AOPs) and AOP-Wiki

AOP for Estrogen Receptor Agonism — complete worked example:

AOP 25 (AOP-Wiki, OECD, 2022): "Binding to ER α leading to reproductive dysfunction in oviparous organisms."

Molecular Initiating Event (MIE): Binding of an agonist ligand (17 β -estradiol, bisphenol A, genistein, ethinylestradiol) to estrogen receptor alpha (ER α , coded by ESR1 gene). Characterised by: ligand-receptor binding kinetics (K_i values range from 10⁻⁹ M for E2 to 10⁻⁵ M for BPA); receptor conformational change to active state; recruitment of coactivators (SRC-1, p300/CBP) to activation function domains AF-1 and AF-2.

Key Event 1 (KE1): ER α activation and homodimerisation. The ligand-bound ER α dimerises and translocates to the nucleus. The dimer binds estrogen response elements (EREs: consensus GGTCAnnnTGACC) in the promoters of target genes. Activation of ERE-driven transcription. Measurable in vitro by: OECD TG 455 ERE-bla agonist assay; OECD TG 458 stably transfected transactivation assay (VM7Luc4E2); ER-CALUX assay. The strength of KE1 correlates with the intrinsic estrogenic potency (relative binding affinity, RBA, relative to E2).

Key Event 2 (KE2): Increased vitellogenin (VTG) synthesis and secretion. VTG (vitellogenin, a yolk precursor protein) is an ERE-regulated protein normally produced in female fish liver. In males (and juveniles), VTG production is negligible. ER α agonists induce VTG transcription in both sexes. Plasma VTG in males is a validated biomarker of estrogenic exposure in aquatic monitoring. Measurable by: ELISA (plasma or liver homogenate); VTG-GFP reporter zebrafish (Ankley et al., 2009, Environ Toxicol Chem). In vivo evidence: VTG induction in male rainbow trout (*Oncorhynchus mykiss*) by ethinylestradiol at environmental concentrations (0.1–10 ng/L); seminal vesicle weight reduction.

Key Event 3 (KE3): Disruption of oocyte maturation and ovulation. High VTG sequesters free calcium in the plasma, reducing calcium availability for the perivitelline fluid and egg coat formation. Additionally, exogenous estrogenic stimulation disrupts the precise estrogen:androgen:progesterone ratio required for triggering final oocyte maturation (FOM) and ovulation. Specifically, sustained elevated plasma E2 (from exogenous xenoestrogens) suppresses LH surge via negative feedback on

GnRH neurons, preventing the mid-cycle LH peak required for ovulation. Measurable by: oocyte count in ovary histopathology; ovulation rate (number of oocytes released per spawning); gonadosomatic index (GSI); hormonal profile (plasma LH, E2, progesterone).

Key Event 4 (KE4): Reduced fertilisation rate. Even if ovulation occurs, the disrupted egg coat composition (from altered ZP gene expression and altered ion channel function) and the reduced gamete viability (aberrant sperm motility from altered androgen levels in males) reduce fertilisation success. Measurable in vivo by: % fertilisation in spawning trials; sperm motility and count; intersex (oocytes in testis = classic field biomarker for estrogen-dominated rivers, e.g., UK River Cherwell study, Jobling et al., 2006, Biology Letters).

Adverse Outcome (AO): Reproductive impairment at the population level. Reduced reproductive output leads to: decreased population growth rate ($\lambda < 1$) if population fecundity falls below the replacement level; population extinction risk in closed populations with low immigration. Population-level evidence: Lake Erie fathead minnow (*Pimephales promelas*) population crash in feminised streams (Kidd et al., 2007, PNAS) following 7-year exposure to 5–6 ng/L ethinylestradiol (EE2) — a landmark experiment demonstrating that whole ecosystem-level effects can result from very low estrogen concentrations.

Key	Event	Relationships	(KERs):
KER1-1 (MIE → KE1): Essentiality confirmed — knockdown of ER α by morpholino antisense in zebrafish abolishes VTG induction by EE2 (Lam et al., 2011); ER antagonist (ICI 182,780) blocks VTG response.	Weight	of	evidence: strong.
KER1-2 (KE1 → KE2): Essentiality confirmed — ER α -ERE binding directly drives the VTG promoter; no VTG production in ER α -null fish (Ciana et al., 2001). Biological plausibility: high.			
KER2-3 (KE2 → KE3): Adjacent KER; plausibility based on Ca ²⁺ sequestration mechanism and reproductive axis feedback. Confirmed by correlation between elevated VTG and reduced ovulation rate in field and laboratory data.			
KER3-4 (KE3 → KE4): Adjacent. Mechanistic link plausible (reduced oocyte quality, reduced gamete viability). Epidemiological support from intersex studies.			

Regulatory applications of AOP 25: Used by OECD as the basis for the Conceptual Framework for Endocrine Disruption Testing (2012); directly informs the testing strategy in EOGRS cohort design (VTG, sexual maturation endpoints); used by EFSA for endocrine disruptor identification criteria (Regulation EU 2018/605).

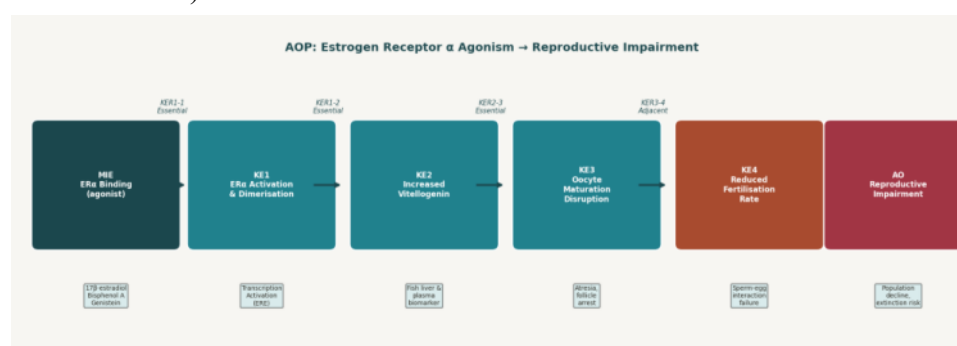


Figure 6.E2. Adverse Outcome Pathway (AOP 25): Estrogen Receptor α agonism leading to reproductive impairment in oviparous organisms (AOP-Wiki, OECD 2022). MIE = Molecular Initiating Event (ER α binding); KE = Key Events (ER α activation, VTG induction, oocyte maturation disruption, reduced fertilisation); KER = Key Event Relationships (arrows with evidence classification); AO = Adverse Outcome (reproductive impairment at population level). Example chemical stressors and measurable markers are indicated below each node.

An AOP is a conceptual construct that links a Molecular Initiating Event (MIE) through Key Events (KEs) at progressively higher biological organisation levels to an Adverse Outcome (AO). The OECD AOP Development Programme has catalogued >500 AOPs in the AOP-Wiki (aopwiki.org). AOPs

enable: (a) selection of the most relevant in vitro assays to interrogate each KE; (b) grouping of chemicals by shared MIE for read-across; (c) mechanistic justification for QSAR models; and (d) integration of multiple data streams into weight-of-evidence assessments for IATA.

6.5.2 Tox21 and ToxCast

The Tox21 programme has generated high-throughput in vitro screening data for >10,000 chemicals across >800 assays, freely accessible at the US-EPA CompTox Chemicals Dashboard. ToxCast (EPA) integrates these data with chemical structure to develop predictive models for >300 toxicological endpoints. These datasets are transforming hazard prioritisation, enabling rapid screening of previously untested chemicals at a fraction of the cost and animal use of traditional in vivo testing.

6.5.3 Integrated Approaches to Testing and Assessment (IATA)

IATA systematically integrates multiple streams of evidence — existing data, structural alerts, QSAR predictions, in vitro data, in vivo data, human epidemiological observations — within a structured weight-of-evidence framework to address a regulatory hazard or risk question. OECD Guidance Document 255 (2016) provides the framework for developing and documenting IATAs. The skin sensitisation IATA (defined approach in OECD TG 497) is the first fully OECD-endorsed IATA replacing animal testing for a regulatory hazard endpoint.

6.5.4 Defined Approaches for Skin Sensitisation and Eye Irritation

Defined Approaches (DAs) are fixed combinations of data sources and prediction models that provide a mechanistically justified, transparent, and reproducible alternative to animal testing. The skin sensitisation IATA (TG 497) combines three in vitro assays targeting three key events of the sensitisation AOP: DPRA (molecular initiating event: cysteine/lysine reactivity), KeratinoSens (KE1: Nrf2/ARE activation in keratinocytes), and h-CLAT (KE2: dendritic cell activation). For eye irritation, OECD TG 491 (short time exposure, STE) and TG 492 (reconstructed human corneal-like epithelium, RhCE) constitute validated DAs.

6.5.5 Next-Generation Risk Assessment (NGRA)

Next-Generation Risk Assessment (NGRA) — principles and case studies:

NGRA is a framework that aims to replace or substantially reduce animal testing in chemical safety assessment by integrating mechanistic, exposure-driven, and quantitative in silico approaches. Key principles (Dent et al., 2018, ALTEX; Brinkmann et al., 2022, ALTEX):

- (1) Mechanism-based: risk is evaluated by identifying the Adverse Outcome Pathway (AOP) relevant to the toxicological concern and testing key events using fit-for-purpose in vitro assays.
- (2) Exposure-driven: the starting point is human exposure (biomonitoring data, dietary intake estimates, occupational monitoring) — testing is conducted at concentrations relevant to human exposure, not at artificially high doses.
- (3) Human-relevant: human cell systems (primary cells, iPSC-derived organoids, MPS) or human-validated in silico models are preferred over rodent models.
- (4) Hypothesis-driven: the assessment follows a structured decision tree; negative results at early stages (below the Bioactivity Exposure Ratio threshold) terminate the assessment without requiring further testing.
- (5) Probabilistic and quantitative: uncertainty is explicitly quantified using Bayesian or Monte Carlo methods; risk is expressed as a probability distribution rather than a single point estimate.

ICCR (International Cooperation on Cosmetics Regulation) cosmetics NGRA framework: The ICCR has adopted NGRA as the basis for safety evaluation of cosmetic ingredients in the EU and other ICCR member regions (US, Japan, Canada, Brazil), where animal testing for cosmetics is prohibited. The ICCR NGRA guidance (2021) describes a tiered approach:

Tier 0 — Read-across and in silico: if sufficient data exist for analogues or from in silico tools (QSAR, PBPK), risk can be assessed without experimental data.

Tier 1 — In vitro hazard screening: cytotoxicity (IC₅₀ in relevant cell types), receptor binding/transactivation (if relevant by AOP), genotoxicity (mini-Ames, Comet), reactive oxygen species.

Tier 2 — Mechanistic profiling: AOP-aligned in vitro battery (e.g., for skin sensitisation: DPRA + KeratinoSens + h-CLAT; for hepatotoxicity: HepaRG spheroids + Seahorse + Tox Tracker).

Tier 3 — PBPK/IVIVE: translate in vitro AC₅₀ to in vivo AED using PBPK modelling; calculate the MoS (Margin of Safety = AED/estimated systemic exposure from cosmetic use).

Case study — Coumarin in flavouring: Coumarin (CAS 91-64-5) is a naturally occurring compound in cinnamon (*Cinnamomum cassia*) and some cosmetics. It was identified as a potential hepatotoxicant in rodents (NOAEL in 2-year rat study = 50 mg/kg/day). EFSA 2004 assessment using TDI = 0.1 mg/kg/day (NOAEL 10 / UF 500). However, NGRA-based assessment using human hepatocyte data and PBPK modelling showed that the rodent liver toxicity is mediated by a species-specific metabolite (3-hydroxycoumarin) that accumulates in rats but is rapidly glucuronidated in humans. Human PBPK predicts that systemic AUC of 3-hydroxycoumarin at dietary exposure levels is 200× lower in humans than in rats. The MoS in humans is >> 100, supporting a conclusion that dietary coumarin exposure via cinnamon-containing foods poses negligible hepatotoxicity risk to humans — a conclusion consistent with the absence of human liver toxicity cases despite widespread cinnamon consumption.

Case study — Parabens (alkyl esters of p-hydroxybenzoic acid): Parabens (methylparaben, ethylparaben, propylparaben, butylparaben) are widely used cosmetic preservatives. Concern arose from weak ERα agonism (RBA relative to E2: methylparaben 1×10⁻⁸; butylparaben 1×10⁻⁶) and potential skin absorption. NGRA assessment (SCCS 2022): (1) Measured skin absorption rates for each paraben; (2) PBPK modelling predicts systemic C_{max} for typical cosmetic use (leave-on products, whole-body exposure); (3) In vitro NOEC for estrogenic effects (VTG induction in MCF-7 cells): methyl- and ethylparaben: no effect at relevant concentrations; propyl- and butylparaben: LOEC ≈ 10 μM. (4) MoS calculation: for butylparaben, systemic C_{max} from leave-on cosmetics ≈ 0.5 μM; LOEC in vitro = 10 μM; MoS = 10/0.5 = 20. SCCS concluded that butylparaben in leave-on products at 0.14% concentration presents a MoS < 100 for estrogenic activity, raising concern — consistent with EU restrictions on propyl- and butylparaben in baby products (EU Cosmetics Regulation 1223/2009, amendment 2014/126/EU).

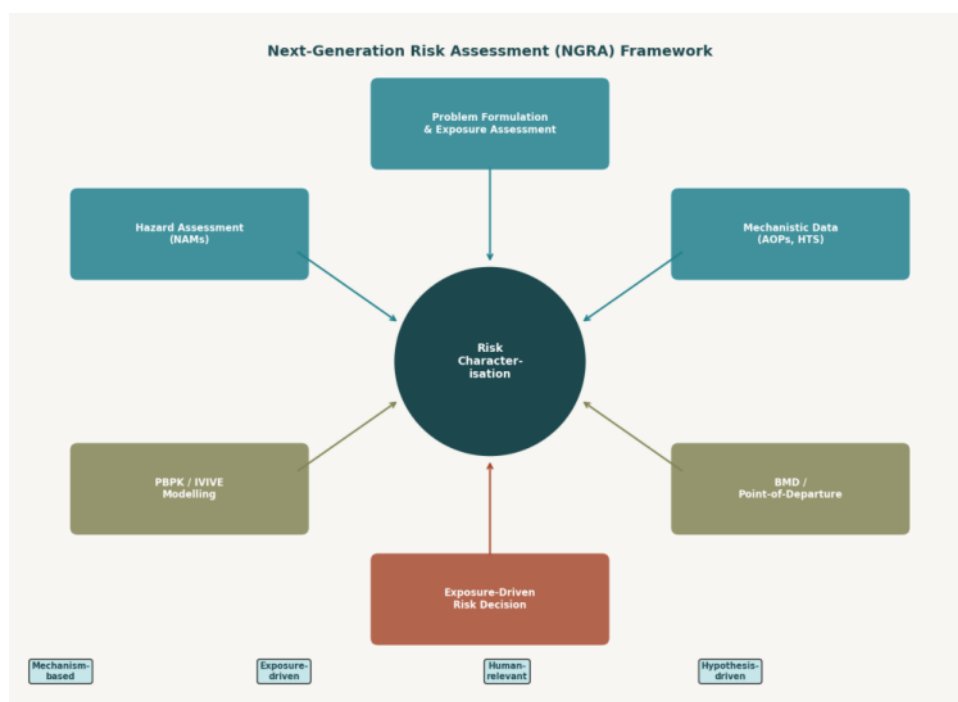


Figure 6.E3. Next-Generation Risk Assessment (NGRA) framework. Problem formulation and exposure assessment provide the starting context. Hazard assessment (using NAMs) and mechanistic data (AOPs, HTS) converge on the central risk characterisation step. PBPK/IVIVE modelling and benchmark dose analysis provide the quantitative link between in vitro bioactive concentrations and in vivo-equivalent doses. The four foundational principles of NGRA (mechanism-based, exposure-driven, human-relevant, hypothesis-driven) are shown at the bottom.

NGRA is an exposure-led, hypothesis-driven risk assessment approach that prioritises chemicals for further testing based on the ratio of estimated human exposure to in vitro bioactivity (bioactivity:exposure ratio, BER). If $BER \gg 100$, the chemical is unlikely to be of concern at relevant human exposure levels, even without in vivo data. This approach maximises the utility of high-throughput in vitro and in silico data for risk-based prioritisation, consistent with the objectives of REACH and the EU Chemicals Strategy for Sustainability.

6.6 Examples of Fully Validated Alternative Methods

Endpoint	OECD TG	Method	Replaces
Skin corrosion	TG 431	Reconstructed human epidermis (RhE) — EpiDerm™, SkinEthic RHE	Rabbit skin corrosion (Draize)
Skin irritation	TG 439	Reconstructed human epidermis (RhE) MTT + IL-1 α	Rabbit skin irritation (Draize)
Eye irritation	TG 492	Reconstructed human cornea (RhCE) — SkinEthic HCE, EpiOcular	Rabbit eye (Draize)
Skin sensitisation (in chemico)	TG 442C	DPRA — peptide reactivity assay	Guinea pig/mouse sensitisation
Skin sensitisation (in vitro cell)	TG 442D	ARE-Nrf2 luciferase (KeratiSens™)	Part of DA replacing LLNA
Skin sensitisation (DC activation)	TG 442E	h-CLAT (human Cell Line Activation Test)	Part of DA replacing LLNA

Phototoxicity	TG 432	3T3 neutral red phototoxicity test	In vivo phototoxicity
Skin absorption (in vitro)	TG 428	Dermal penetration using excised skin	Reduces in vivo dermal absorption studies
Fish acute toxicity replacement	TG 236	Fish Embryo Acute Toxicity (FET) — zebrafish 120 hpf	Adult fish LC50 (TG 203)

Table 6.1. Fully validated alternative test methods for regulatory use (OECD Test Guidelines).

6.7 Limitations and Future Perspectives

Despite remarkable progress, significant limitations remain for NAMs in regulatory toxicology:

- Systemic endpoints: chronic organ toxicity, immune toxicity, reproductive/developmental toxicity, and neurotoxicity cannot currently be fully addressed by validated in vitro methods.
- In vitro-to-in vivo extrapolation (IVIVE): translating in vitro IC_{50} values to in vivo NOAELs requires PBPK modelling, whose parameterisation introduces uncertainty.
- Regulatory harmonisation: while OECD Test Guidelines provide a common framework, implementation across national regulatory authorities (FDA, ECHA, TGA) varies; novel methods may require case-by-case negotiation.
- Tissue complexity: even the most advanced organoids do not fully replicate the vascularisation, immune infiltration, and cell-type diversity of living organs.

The trajectory of the field is, however, clearly progressive. The US-EPA has committed to reducing vertebrate animal testing through the Alternative Testing Initiative; ECHA is integrating NAMs into the REACH guidance; and the EU Chemicals Strategy for Sustainability calls for 'moving away from animal testing' as a long-term regulatory objective. Machine learning, iPSC-derived organoids, multi-organ chips, and AOP-guided IATA represent the frontier of this transformation.

Key Take-Home Messages

- NAMs (New Approach Methodologies) encompass in vitro, in silico, organoid, organ-on-chip, and AOP-based approaches that aim to replace or reduce animal testing.
- Validation by EURL ECVAM / ICCVAM / OECD is required before an alternative method can be used for regulatory decisions; the process evaluates reproducibility, predictive capacity, and mechanistic relevance.
- QSAR, read-across, and PBPK modelling are accepted in REACH regulatory dossiers when properly validated and documented with applicability domain criteria.
- The AOP framework provides mechanistic structure for IATA, enabling integration of NAM data streams in weight-of-evidence assessments.
- Fully validated alternatives exist for skin corrosion (TG 431), skin irritation (TG 439), eye irritation (TG 492), skin sensitisation (TG 442C/D/E, TG 497), and phototoxicity (TG 432).
- NGRA using bioactivity:exposure ratios will transform hazard prioritisation by focusing animal testing only on chemicals with credible human health concern.

Further Reading

Ball N et al. A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch Toxicol.* 2022;96(3):743–766. doi:10.1007/s00204-021-03215-9

Lancia P et al. Overview of In Silico Tools to Evaluate Human Health Toxicity, Ecotoxicity, and Toxicokinetic Profiles in Hazard Assessment. *Chem Res Toxicol.* 2025;38(10):1652–1680. doi:10.1021/acs.chemrestox.4c00534

Lin Z, Chou WC. Machine Learning and Artificial Intelligence in Toxicological Sciences. *Toxicol Sci.* 2022;189(1):7–19. doi:10.1093/toxsci/kfac075

Worth AP, Berggren E. Towards a future regulatory framework for chemicals in the EU — Chemicals 2.0. *Regul Toxicol Pharmacol.* 2023;142:105431. doi:10.1016/j.yrtph.2023.105431

Deepika D, Kumar V. The Role of PBPK NAM in Pharmaceuticals and Environmental Chemical Risk Assessment. *Int J Environ Res Public Health.* 2023;20(4):3473. doi:10.3390/ijerph20043473

Tutorial Programme (Travaux Dirigés)

Critical Analysis of Scientific Articles

T.1 Pedagogical Objectives

The Tutorial (TD) programme develops the critical reading, analytical reasoning, and oral communication skills essential for an M1 student in experimental toxicology. Upon completion of the six TD sessions, students should be able to:

- Identify the hypothesis, experimental design, statistical methods, and conclusions of a toxicological research article.
- Evaluate the quality, appropriateness, and limitations of study design in the context of OECD Test Guidelines, ARRIVE 2.0, and GLP.
- Detect potential sources of bias (selection, performance, detection, reporting) and assess their impact on the validity of conclusions.
- Contextualise published results within the broader regulatory and mechanistic toxicology framework.
- Prepare and deliver a structured critical appraisal presentation (10 min + 5 min discussion) to a peer audience.

T.2 Selection Criteria for Articles

Articles should be selected to cover the breadth of the curriculum. Selection criteria:

- Published within the last 5 years (2020–2025) in an internationally indexed journal (Q1–Q2 in toxicology, pharmacology, or environmental health).
- Represents one of the following categories: (a) acute toxicity in vivo; (b) chronic/subchronic in vivo; (c) in vitro mechanistic; (d) ecotoxicology/biomonitoring; (e) endocrine disruption; (f) alternative method or QSAR.
- Contains at least: a dose-response analysis, statistical comparisons, and a dose descriptor (IC₅₀, NOAEL, LOAEL, or EC₅₀).
- Shows methodological flaws or limitations that stimulate critical discussion (not a 'perfect' study).



Figure T.1. Critical Reading Flowchart for Toxicology Articles. The five-step procedure guides students from title/abstract through design, methods, results, and conclusions, building a systematic critical appraisal framework. (Original figure.)

Evidence Pyramid for Toxicological Risk Assessment

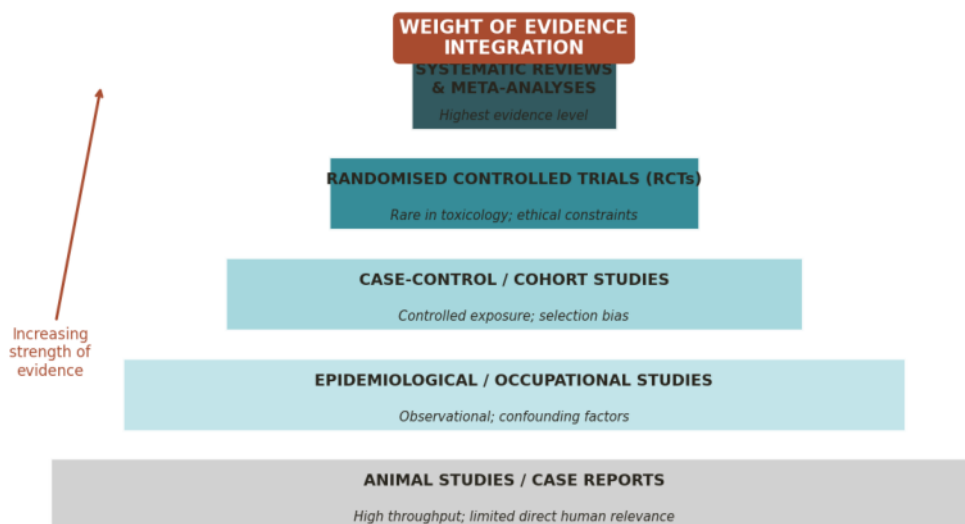


Figure T.2. Evidence Pyramid for Toxicological Risk Assessment. Students should situate each article at the appropriate level of the pyramid and consider how the article's conclusions are strengthened or limited by its evidence tier. (Original figure.)

T.3 Critical Reading Grid

Detailed expansion of the critical reading grid with scoring guidance:

Question	1	—	Study	design	adequacy:
What to look for:	Is the design appropriate to the stated research question (dose-response, time-course, mechanistic, comparative)? Is the control group appropriate (vehicle control matching the solvent used, positive control confirming assay sensitivity)? Are the dose levels justified and do they include both sub-toxic and toxic concentrations? For in vivo studies: randomisation, blinding? For in vitro: single vs. multiple passages, pre-treatment serum starvation?				
Common pitfalls:	(a) Using only a single dose — prevents dose-response characterisation and dose descriptor calculation; (b) inappropriate positive control (e.g., using H ₂ O ₂ as a positive control for ROS in a study of a non-oxidising compound — better to use a reference compound with the same mechanism as the test compound); (c) inadequate acclimation period in animal studies; (d) testing at cytotoxic concentrations in in vitro mechanistic studies (confounds mechanism-specific endpoints with non-specific cell death).				
Scoring:	2 points — fully adequate design with all relevant controls; 1 point — acceptable design with minor deficiencies; 0 points — fundamental design flaw that invalidates interpretation.				
Question	2	—	Sample	size	and statistical power:
What to look for:	Is the sample size justified by a power calculation? What is the reported n per group? For in vitro studies: is n = 3 truly biological replicates (3 independent experiments from separate cell passages) or merely technical replicates (3 wells from the same experiment)? Is the unit of replication clearly stated?				
Common pitfalls:	(a) Confusion between technical and biological replicates — many published papers report n=3 from a single experiment, giving the false impression of biological reproducibility; (b)				

Underpowered studies (n=5 animals per group for a 4-group factorial design) that can only detect very large effects (Cohen's $f > 0.8$); (c) Post-hoc power calculation (calculated after finding a significant result) — misleading because it is mechanistically equivalent to reporting the p-value. Scoring: 2 points — explicit a priori power calculation, n justified; 1 point — n adequate by convention but not explicitly justified; 0 points — n manifestly inadequate for the study type.

Question 3 — Dose-response characterisation:
 What to look for: Are at least 4 dose levels tested spanning the complete dose-response range (from no effect to maximal effect)? Is the dose-response modelled (sigmoidal curve fitting, BMD analysis) or merely described qualitatively? Are dose descriptors (IC50, EC50, NOAEL/LOAEL, BMDL10) reported with 95% confidence intervals?
 Common pitfalls: (a) Only testing 2–3 concentrations — insufficient for curve fitting or BMD modelling; (b) Reporting only "significant at $p < 0.05$ " without dose-response analysis — NOAEL cannot be validly derived from a poorly powered study; (c) Not reporting the upper asymptote (maximal effect) — makes the IC50 unreliable.
 Scoring: 2 points — complete dose-response with quantitative endpoints and 95% CI; 1 point — adequate dose range, no quantitative curve fit; 0 points — single dose or grossly inadequate range.

Question 4 — Endpoint specificity and mechanistic relevance:
 What to look for: Are the measured endpoints specific to the claimed mechanism? Are positive controls included to demonstrate assay sensitivity? Are confounding endpoints measured (e.g., cytotoxicity checked alongside a mechanistic endpoint)?
 Common pitfalls: (a) Reporting increased ROS without measuring cell viability — if cells are dying (LDH release), ROS increase may be a secondary consequence of cell death rather than a primary mechanistic event; (b) Using a general viability assay (MTT) as a proxy for apoptosis without confirming by Annexin V or caspase activation; (c) Absence of pathway inhibitor controls (e.g., no N-acetylcysteine control in an oxidative stress study).

Question 5 — Statistical analysis appropriateness:
 What to look for: Are the tests appropriate for the data distribution (parametric vs. non-parametric)? Is the unit of analysis clearly defined? Is multiple comparison correction applied?
 Common pitfalls: (a) Using Student's t-test to compare multiple groups (inflates Type I error — use ANOVA with post-hoc correction); (b) Treating technical replicates as independent observations (pseudoreplication); (c) Reporting $p = 0.05$ as "significant" without acknowledging that at 5% type I error rate, 1 in 20 "significant" results in the null hypothesis case is a false positive.

Question 6 — Translational relevance:
 What to look for: Are the tested concentrations/doses within physiologically or toxicologically relevant ranges (human blood concentrations, environmental monitoring data, occupational exposure limits)? Is the test system (species, cell type) appropriate for the human health question?
 Common pitfalls: (a) Testing compounds at 100–1000 μM in in vitro cytotoxicity assays when the maximum plasma concentration in humans is $< 1 \mu\text{M}$ — results are irrelevant to human exposure; (b) Using cell lines with very different metabolic competence from the target tissue; (c) Extrapolating rodent-specific mechanisms (e.g., $\alpha 2\text{u}$ -globulin nephropathy) to humans.

Question 7 — Compliance with guidelines:
 What to look for: Is the study compliant with OECD, ICH, EFSA, or other relevant guidelines? For animal studies: is ethics committee approval stated? Is ARRIVE reported? For in vitro: are OECD TG or good cell culture practice (GCCP, ECVAM) standards referenced?

Question 8 — Reproducibility and data transparency:
 What to look for: Is the study registered prospectively? Are raw data available? Are all pre-specified endpoints reported (or is selective reporting evident)? Are deviations from the pre-specified protocol noted?

Question 9 — Author conflict of interest and funding:
What to look for: Is the study funded by an industry with a financial interest in the outcome? Do any authors have a direct conflict of interest? Is there evidence of publication bias (positive result bias in a field dominated by industry-funded studies)?

Question 10 — Overall scientific validity and regulatory utility:
What to look for: Does the study, taken as a whole, provide reliable evidence that would withstand regulatory scrutiny? Is the evidence sufficient to derive a dose descriptor (NOAEL, BMDL10) for risk assessment? Are the conclusions supported by the data and not over-stated?

Students are provided this structured 10-question grid prior to each TD session. Answers are prepared individually and discussed in group.

#	Question	What to Look For
1	What is the research question / hypothesis?	Is it clearly stated? Testable? Does the design answer it?
2	What is the study design?	Controlled? Prospective/retrospective? In vivo/in vitro/epidemiological?
3	Is the sample size justified?	Power analysis stated? n per group adequate for endpoint variability?
4	Randomisation and blinding?	How were animals/samples allocated? Was experimenter blinded during data collection?
5	Are the endpoints and dose descriptors appropriate?	OECD-compliant? Relevant to human health? NOAEL/IC50 properly derived?
6	Statistical methods: appropriate and complete?	Parametric or non-parametric justified? Multiple comparison correction? Effect sizes?
7	Controls adequate?	Vehicle control? Positive control? Historical control data cited?
8	Are conclusions supported by data?	Over-generalisation? Causality claimed from correlational data?
9	Reporting quality (ARRIVE/CONSORT)?	Key design elements reported? Conflicts of interest declared?
10	Critical limitation and future direction?	What would you do next to fill the main gap in this study?

Table T.1. Critical reading grid — 10-question rubric for article analysis (TD sessions).

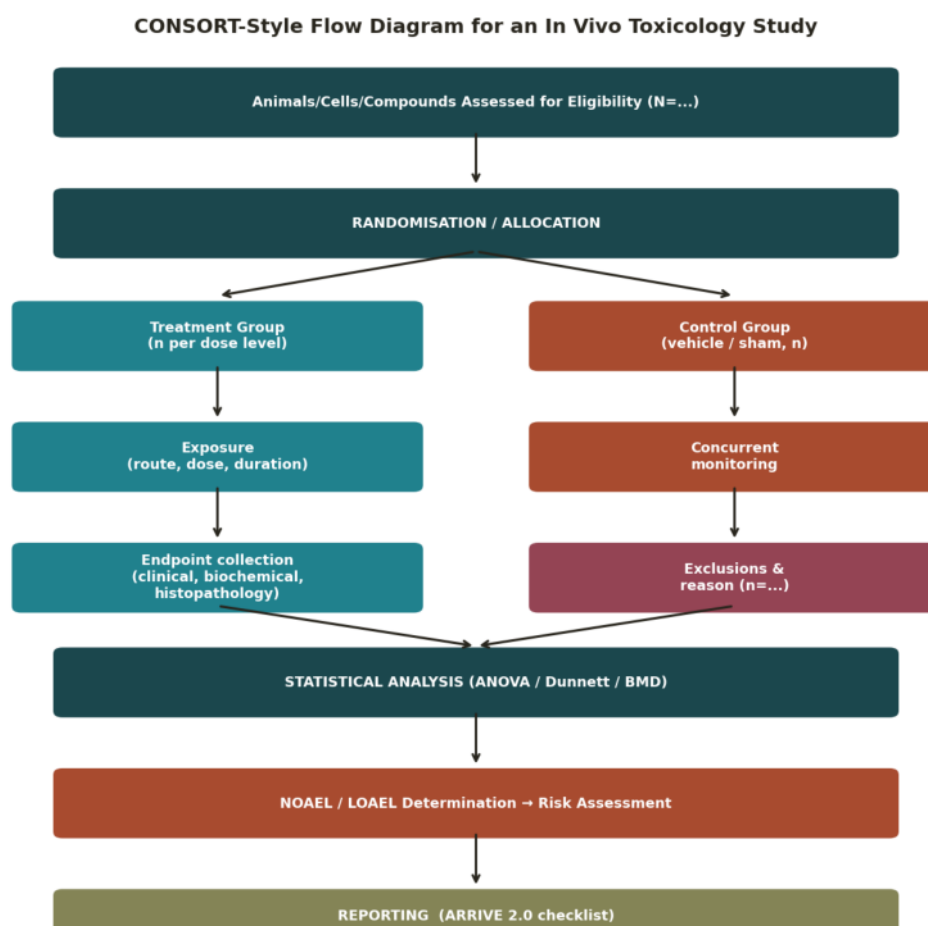


Figure T.3. CONSORT-Style Flow Diagram for an In Vivo Toxicology Study. Students use this template to diagram the study design of the assigned article, identifying drop-outs, exclusions, and allocation biases. (Original figure.)

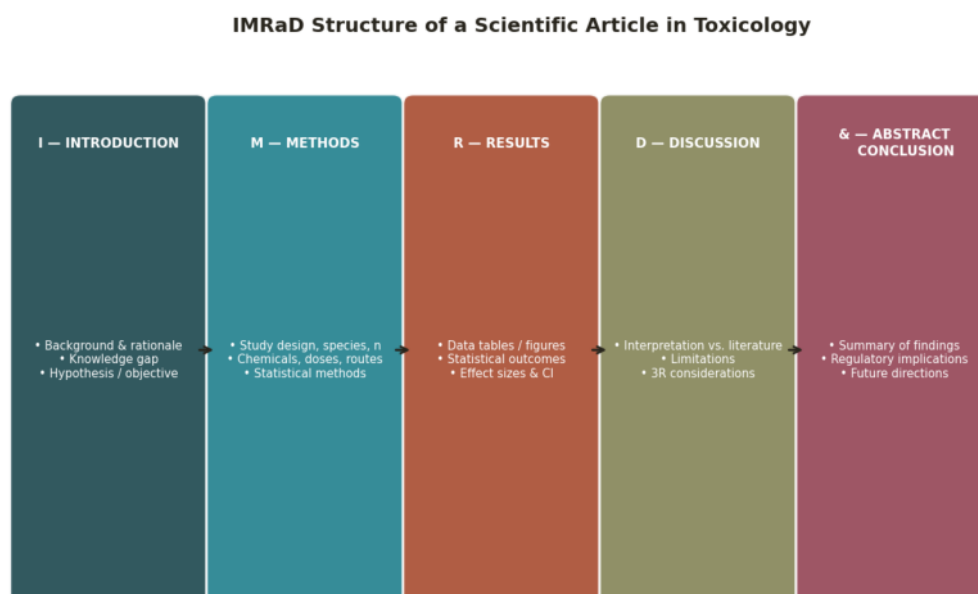


Figure T.4. IMRaD Structure of a Scientific Article in Toxicology. Understanding the structural logic of a scientific paper is the foundation of effective critical reading. Each section serves a distinct communicative function. (Original figure.)

T.4 Six Suggested Articles with Rationale

Detailed analysis of the six suggested articles:

Article 1 — Dose-response design and BMD modelling:
Suggested article type: A 90-day oral toxicity study in rats reporting NOAEL/LOAEL and BMD analysis for a novel food additive.
Why chosen: This article type demonstrates the practical application of regulatory-grade OECD TG 408 design, including clinical pathology panels, histopathology, and the derivation of regulatory dose descriptors. Students encounter the difference between the classical NOAEL approach (a categorical decision dependent on the study design) and the continuous BMD approach (extracts more information from the full dose-response dataset). The article also provides a realistic example of how a dose-response study generates regulatory data for risk assessment.
Specific lessons: (1) How to evaluate the adequacy of clinical chemistry parameters in detecting organ-specific toxicity; (2) How to identify the appropriate benchmark response (BMR) for deriving BMDL10 from histopathological incidence data; (3) How to assess whether the NOAEL is the appropriate summary descriptor when the dose-response shows a smooth continuous trend.
Discussion questions: (1) Is the BMDL10 derived from the liver enzyme data consistent with the NOAEL derived from histopathology? If not, which is more appropriate for the TDI derivation and why? (2) Which uncertainty factors would you apply to this NOAEL/BMDL10 to derive a human ADI, and what is the resulting ADI? (3) Is there evidence of sex-specific toxicity in this study, and if so, how should this influence risk assessment?

Article 2 — In vitro mechanistic study with limited translational value:
Suggested article type: An in vitro study on heavy metal cytotoxicity in a cell line, testing very high concentrations (100–1000 μM) and reporting only MTT viability.
Why chosen: Deliberately selected as an example of a study with methodological weaknesses to train critical thinking. Students must identify: (1) that concentrations tested far exceed human plasma concentrations, limiting biological relevance; (2) that MTT alone is insufficient to characterise the mechanism of toxicity; (3) that n=3 technical replicates from a single experiment is insufficient for reproducibility claims.
Specific lessons: (1) The concept of maximum achievable blood concentration (C_{max}) and how to relate in vitro test concentrations to human exposure scenarios; (2) How to distinguish biological replicates from technical replicates; (3) The limitations of a single endpoint approach for mechanistic characterisation.
Discussion questions: (1) Calculate the IC_{50} from the dose-response data provided and compare it to published blood concentrations in occupationally exposed workers. Does the study have translational relevance? (2) If you were to improve this study, which additional endpoints would you add to characterise the mechanism of toxicity, and why? (3) Does the study comply with OECD GCCP? Identify two specific deviations.

Article 3 — ARRIVE 2.0 compliance analysis:
Suggested article type: A recently published animal study (post-2020) on neuroprotective effects of a plant extract, with incomplete ARRIVE reporting.
Why chosen: Students apply the ARRIVE 2.0 Essential 10 checklist directly to a peer-reviewed paper and identify reporting gaps. This exercise develops skills applicable to reviewing grant proposals, regulatory submissions, and pre-clinical data packages.
Specific lessons: (1) Distinguishing essential from recommended ARRIVE items; (2) Understanding why randomisation and blinding are critical for internal validity; (3) Connecting reporting quality to the risk of irreproducibility.
Discussion questions: (1) Complete the ARRIVE 2.0 Essential 10 checklist for this paper. Which items are fully reported, partially reported, and missing? (2) For the items reported as "missing," what is the

consequence for interpreting the study's internal validity? (3) Write a 150-word "Methods transparency statement" that would bring this paper into full ARRIVE 2.0 compliance.

Article 4 — In vitro–in vivo extrapolation:

Suggested article type: A paper combining in vitro hepatotoxicity data with PBPK-IVIVE to predict in vivo relevance.

Why chosen: Introduces students to the IVIVE workflow ($AC_{50} \rightarrow f_{u,plasma} \rightarrow CL_{int} \rightarrow PBPK \rightarrow AED$) and demonstrates how in vitro data can be made more interpretable by anchoring them to in vivo pharmacokinetics.

Specific lessons: (1) The trapezoidal rule for AUC calculation from in vitro exposure data; (2) The concept of unbound fraction and its importance for IVIVE; (3) How to interpret the Bioactivity Exposure Ratio (BER = $AC_{50}/\text{estimated exposure}$) as a risk metric.

Discussion questions: (1) Using the data in Table 2 of the article, calculate the AED for the most sensitive endpoint and compare it to the reported in vivo NOAEL. Are they consistent? (2) The authors assume $f_{u,plasma} = 0.10$. How does the predicted AED change if $f_{u,plasma} = 0.05$? (3) What are the major sources of uncertainty in the IVIVE calculation and how could they be quantified?

Article 5 — Regulatory toxicology case study:

Suggested article type: A REACH substance evaluation dossier or EFSA scientific opinion on a food contaminant.

Why chosen: Demonstrates how academic toxicology data are integrated into regulatory risk assessment frameworks — a direct career skill for M1 toxicology students.

Specific lessons: (1) How to navigate an EFSA or ECHA dossier; (2) How to identify the key study driving the TDI or ADI derivation; (3) How uncertainty factors are applied and how the choice of UF affects the final safe dose.

Discussion questions: (1) Identify the "critical study" (most sensitive effect at the lowest dose) in this assessment. Why was it chosen over other studies with different endpoints? (2) Which uncertainty factors were applied, and what is the scientific basis for each factor? (3) If a new study were to report a NOAEL $3\times$ lower than the critical study, recalculate the TDI with the same UFs. What action would EFSA likely take?

Article 6 — Emerging contaminant with limited regulatory data:

Suggested article type: A review or primary research article on a PFAS or nano-material with sparse regulatory toxicity data.

Why chosen: Challenges students to reason about risk assessment under data uncertainty, applying precautionary principles and read-across/QSAR approaches.

Specific lessons: (1) How to apply the precautionary principle when NOAEL data are unavailable; (2) How to use QSAR and read-across to fill data gaps; (3) How to construct a weight-of-evidence argument for or against regulation.

Discussion questions: (1) Using VEGA software (or OECD QSAR Toolbox), predict the skin sensitisation potential of the novel PFAS described in this article. Is the compound within the QSAR applicability domain? (2) Identify three structural alerts in this compound that are relevant to systemic toxicity, based on the Cramer classification scheme. (3) Construct a read-across justification for this compound using two PFAS with available oral toxicity data. What are the limitations of this approach?

T.7 Sample Worked Example: Full Critical Analysis

Sample Worked Example — Full Critical Analysis of a Hypothetical Article

Reference: Smith, J.A., Patel, R., & Nguyen, L. (2024). Sub-lethal effects of imidacloprid on acetylcholinesterase activity and oxidative stress biomarkers in Wistar rats following 28-day oral exposure. *Toxicology Letters*, 395, 115–124.

Abstract summary: The study examined the effects of imidacloprid (a neonicotinoid insecticide) administered daily by oral gavage at 0, 5, 25, and 125 mg/kg/day for 28 days in male Wistar rats. Primary endpoints were AChE activity (brain and plasma), and oxidative stress biomarkers (MDA, GSH, SOD) in liver. The authors report significant AChE inhibition at 125 mg/kg/day and increased MDA at 25 mg/kg/day. They conclude that the NOAEL for neurotoxic effects is 25 mg/kg/day.

EXCELLENT-LEVEL ANALYSIS:

Study design assessment (Score: 1.5/2): The study uses a standard 28-day design with four dose groups, consistent with OECD TG 407. However, only male rats were used — the absence of female animals is a significant limitation given that sex-dependent differences in imidacloprid metabolism (via CYP3A4 and CYP2E1, which show sex-dependent expression in rats) could alter toxicokinetics and toxicodynamic sensitivity. The authors do not justify the single-sex design. Controls include a vehicle control (distilled water, 10 mL/kg/day) — this is appropriate since imidacloprid is water-soluble and no DMSO carrier is used. However, the absence of a positive control for the AChE assay is a weakness: including chlorpyrifos (a known AChE inhibitor at a defined dose) would validate the assay's sensitivity in this experimental context.

Sample size and statistical power (Score: 1/2): The study uses $n=8$ per group. Assuming the SD of brain AChE = 15% of the mean (typical for this assay), and a detection threshold of 20% inhibition (the minimum biologically relevant change for AChE), a power calculation gives: $f = 0.4/4 = 0.10$ (based on a 4-group design with control vs. high-dose mean difference as the primary comparison)... Actually, applying a two-group comparison for the specific high-dose vs. control comparison: $d = 20/15 = 1.33$, power at $n=8$: approximately 82%. This is adequate for the primary endpoint. However, for the secondary oxidative stress endpoints with higher biological variability (MDA, SD = 25%), power at $n=8$ for a 20% change gives $d = 0.8$, power $\approx 50\%$ — inadequate, making the positive finding of MDA increase at 25 mg/kg/day potentially a false positive. This is not addressed in the paper.

Dose-response characterisation (Score: 1.5/2): The dose levels span 25-fold (5 to 125 mg/kg/day), which is reasonable for a 28-day study. The doses are consistent with published rodent imidacloprid toxicity data (NOAEL 60 mg/kg/day in a 90-day rat study, EFSA 2008). A 4-parameter logistic fit to the AChE data would be more informative than ANOVA-based NOAEL/LOAEL determination, but the authors do not report BMD analysis. The NOAEL of 25 mg/kg/day for AChE inhibition is consistent with EU regulatory assessments.

Statistical analysis (Score: 2/2): One-way ANOVA with Tukey's post-hoc comparison is appropriate for normally distributed data from multiple groups. Shapiro-Wilk normality test results and Levene's variance homogeneity test results are reported for each parameter — this is a strength. The unit of replication (individual animal) is clearly stated.

Translational relevance (Score: 1.5/2): The doses tested (5–125 mg/kg/day) are substantially higher than estimated human dietary exposure to imidacloprid (0.0002–0.001 mg/kg/day, EFSA 2018 pesticide monitoring data). The study is therefore primarily useful for identifying hazard thresholds (NOAEL) for risk assessment, not for directly modelling human effects. The authors do not discuss the IVIVE implications of their in vivo findings, nor do they relate brain AChE inhibition thresholds (reported as 20% at 125 mg/kg/day) to neurobehavioural endpoints in rats or to human occupational exposure data.

Overall scientific validity (Score: 1/2): The study is a valid, if limited, contribution to imidacloprid hazard characterisation. Key weaknesses: single sex, no BMD analysis, underpowered oxidative stress endpoints. The conclusion that the NOAEL for neurotoxic effects is 25 mg/kg/day is supported by the AChE data alone, but the MDA increase at 25 mg/kg/day complicates this interpretation — if the MDA finding is real (which is uncertain given the underpowered design), the NOAEL for oxidative stress would be 5 mg/kg/day, and the overall NOAEL for the study would be 5 mg/kg/day, not 25. An EXCELLENT submission would identify this internal inconsistency and suggest that the oxidative stress endpoint should drive the NOAEL determination pending a more adequately powered study.

Total score: 8.5/12 (71%) — a well-written critical analysis demonstrating competence in toxicological study design evaluation, statistical reasoning, and regulatory data interpretation.

Session	Article Reference	Category	Key Focus Points for Students
TD-1	Karmaus AL et al. Evaluation of Variability Across Rat Acute Oral Systemic Toxicity Studies. <i>Toxicol Sci.</i> 2022;188(1):34–47. doi:10.1093/toxsci/kfac042	Acute in vivo (TG 425)	Variability of LD50/TG 425 results; power of UDP; GHS classification consistency; inter-lab variation
TD-2	Takeshita J et al. Comprehensive analysis of repeated-dose subacute toxicity findings in male rats. <i>Crit Rev Toxicol.</i> 2024;54(10):996–1010. doi:10.1080/10408444.2024.2427221	Subacute in vivo (TG 407)	NOAEL derivation methodology; histopathology interpretation; incidental vs. treatment-related findings
TD-3	Lörincz T et al. HepG2 and HepaRG through the Glass of Acetaminophen-Induced Toxicity. <i>Life.</i> 2021;11(8):856. doi:10.3390/life11080856	In vitro mechanistic	Cell model selection; MTT vs. LDH; metabolic activation; IC50 interpretation; comparison to in vivo hepatotoxicity
TD-4	Gonçalves IFS et al. Toxicity testing of pesticides in zebrafish: a systematic review. <i>Environ Sci Pollut Res.</i> 2020;27(10):10185–10204. doi:10.1007/s11356-020-07902-5	Ecotoxicology / alternative model	FET test (TG 236); LC50 vs. EC50; endpoint diversity; 3R compliance; regulatory relevance of zebrafish
TD-5	Miller LB et al. Reproductive toxicology: keeping up with our changing world. <i>Front Toxicol.</i> 2024;6. doi:10.3389/ftox.2024.1456687	Endocrine disruption / reproductive toxicity	EDC mechanisms; low-dose effects; non-monotonic dose-response; EOGRTS design; regulatory implications
TD-6	Lancia P et al. Overview of In Silico Tools for Cosmetics Hazard Assessment. <i>Chem Res Toxicol.</i> 2025;38(10):1652–1680. doi:10.1021/acs.chemrestox.4c00534	Alternative method / QSAR	QSAR principles; applicability domain; validation criteria; regulatory acceptance; comparison with in vitro data

Table T.2. Six suggested articles for TD sessions with rationale and focus points.

T.5 Assessment Grid

Criterion	Excellent (4)	Good (3)	Satisfactory (2)	Insufficient (1)
-----------	---------------	----------	------------------	------------------

Summary of study	Precise, complete (aim/model/endpoint/conclusions)	Minor omissions	Key elements missing	Inaccurate/incomplete
Study design critique	All key design elements evaluated systematically	Most elements covered	Superficial analysis	Design not evaluated
Statistical analysis	All tests identified; assumptions checked; p vs. ES	Most tests identified	Partial evaluation	Statistics not examined
Critical thinking	Original, evidence-based critique with specific suggestions	Sound critique, minor gaps	Generic observations	No critique or unsupported claims
Contextualisation	Article placed within regulatory/mechanistic framework	Adequate context	Limited context	No broader context
Oral presentation	Clear, structured, confident, on time (10 min)	Minor issues with clarity/time	Disorganised or too short/long	Incoherent or absent

Table T.3. Assessment rubric for student article presentations (TD sessions).

T.6 Suggested Calendar

Session	Date	Article Category	Activity	Duration
TD-1	Week 3	Acute in vivo (TG 425)	Introduction to critical reading + Article 1 presentation (2 students)	2 h
TD-2	Week 5	Subacute in vivo (TG 407)	Article 2 group discussion + ARRIVE 2.0 audit exercise	2 h
TD-3	Week 7	In vitro mechanistic (MTT/HepG2)	Article 3 + statistical re-analysis workshop	2 h
TD-4	Week 9	Ecotoxicology / zebrafish FET	Article 4 + dose-response curve fitting exercise	2 h
TD-5	Week 11	EDC / reproductive toxicity	Article 5 + regulatory implications discussion	2 h
TD-6	Week 13	Alternative methods / QSAR	Article 6 + synthesis of all sessions + final assessment	2 h

Table T.4. Suggested TD calendar (6 × 2-hour sessions across the semester).

Practical Programme (Travaux Pratiques)

Setting Up a Toxicology Study

P.1 Learning Objectives

Upon completion of the six practical sessions, students will be able to:

- Design and justify an acute oral toxicity study using OECD TG 425.
- Perform an MTT cytotoxicity assay on HepG2 cells and calculate IC_{50} by dose-response curve fitting.
- Recognise and describe hallmark histopathological lesions of hepatotoxicity and nephrotoxicity from stained sections.
- Perform heavy metal quantification by atomic absorption spectrometry (AAS) in fish tissue and produce a validated calibration curve.
- Use the OECD QSAR Toolbox to predict mutagenicity and skin sensitisation for a compound of interest.
- Derive NOAEL, LOAEL, BMD, and BMDL from in vivo dose-response data using appropriate statistical software.

P.2 General Framework: The Toxicology Study Lifecycle

A well-designed toxicology study follows a defined lifecycle from protocol design to final reporting. Each practical in this programme addresses a specific phase of this lifecycle, providing students with end-to-end experience of the experimental toxicologist's workflow.

P.3 Practical 1: Designing a Dose-Response Study (Acute Oral Toxicity, OECD TG 425)

Practical 1 — Full expanded protocol and materials:

Pre-lab quiz (5 questions):
Q1: What is the primary reason that OECD TG 401 (classical LD_{50} test) was deleted in 2002?
A1: The classical LD_{50} test required 40–50 animals per substance and caused severe animal suffering, while providing minimal additional information beyond GHS classification, which can be achieved with far fewer animals using TG 420, 423, or 425.
Q2: In OECD TG 425, if the first animal at the starting dose of 175 mg/kg survives, what dose is tested in the second animal?
A2: The next animal receives 550 mg/kg (one step up in the 3.16-fold dose progression).
Q3: What is the GHS category for a substance with LD_{50} between 300 and 2000 mg/kg (oral, rat)?
A3: Category 4 ("Harmful if swallowed," signal word "Warning," H302 hazard statement).
Q4: What is the role of the vehicle control in an acute oral toxicity study?
A4: The vehicle control receives the same formulation vehicle (solvent/carrier) as the treated animals, at the same dose volume (typically 10 mL/kg). This controls for any effects of the vehicle itself (e.g., osmotic stress, aspiration of oily vehicles) that could be mistaken for effects of the test compound.
Q5: Name two humane endpoints used in acute oral toxicity studies that would trigger early euthanasia.
A5: Any two from: >20% body weight loss, moribundity (unresponsive to gentle stimulation), severe respiratory distress, total anorexia persisting >24h, complete loss of motor coordination preventing access to food/water, severe self-mutilation, rectal prolapse, uncontrolled convulsions.

Full materials list for 12 students (in silico/paper simulation exercise, since actual in vivo study cannot be conducted in teaching laboratory):
— Computer workstations with EPA AOT425StatPgm.exe (free download) or R package 'updown'

installed

- Pre-prepared case datasets (5 simulated TG 425 experiments with different dose-response patterns)
- Calculator or Excel spreadsheets
- OECD TG 425 documentation (downloaded from OECD iLibrary)
- GHS/CLP Regulation (EC) 1272/2008 classification table (printed)
- Dosing record templates (one per student group)
- Body weight and clinical observation record sheets
- SDS (Safety Data Sheet) templates for practice completion

Detailed step-by-step protocol (2h session):

Part A (30 min): Introduction to TG 425 decision rules. Students individually work through the dose sequence for a pre-assigned case dataset, recording each animal's dose and outcome on the dosing record sheet. They identify the reversal point and determine whether the stopping rule has been met (≥ 4 animals after reversal).

Part B (45 min): Statistical analysis using AOT425StatPgm.exe or R. Enter the dose-outcome data (0=survived, 1=died) and starting dose. Run MLE analysis. Record: LD₅₀ estimate, 95% CI, GHS category assignment, and signal word.

Part C (30 min): GHS classification exercise. Using the derived LD₅₀, complete the GHS label for the hypothetical compound (pictogram, signal word, hazard statement, precautionary statements).

Reference: UN GHS Purple Book, Revision 9 (2021).

Part D (15 min): Discussion of case studies with different results (very low LD₅₀ resulting in Category 1/2 classification; LD₅₀ = 2500 mg/kg resulting in "not classified").

Post-lab calculation sheet:

1. Recalculate LD₅₀ manually using the maximum likelihood formula (presented in the pre-lab lecture): $\log(\text{LD}_{50}) = \Sigma(\log(d_i) \times x_i) / \Sigma x_i$ where d_i = doses and x_i = 0/1 outcomes. Compare with software output.
2. Calculate 95% CI using the formula: $95\% \text{ CI} = \text{antilog}_{10}[\log(\text{LD}_{50}) \pm 1.96 \times \sigma/\sqrt{n}]$
3. Calculate the Bioactivity Exposure Ratio (BER): $\text{BER} = \text{LD}_{50} / \text{estimated human occupational exposure (given in case dataset)}$. Is the BER > 100 (acceptable) or < 100 (concern)?

Discussion questions:

1. A substance tested by TG 425 yields LD₅₀ = 285 mg/kg (95% CI: 155–520 mg/kg). The LD₅₀ point estimate falls in Category 3 (50–300 mg/kg), but the upper bound of the CI falls in Category 4. How should the substance be classified, and why?
2. Why does OECD TG 425 use a 3.16-fold dose progression (rather than 2-fold or 10-fold)?
3. Compare the information content of the TG 425 LD₅₀ with the TG 407 NOAEL. For a substance with both datasets available, which is more relevant for setting an occupational exposure limit? Why?
4. The AOT425 software reports that the dose-response pattern is "atypical" (not monotonic). What could cause a non-monotonic acute dose-response in a rodent, and how should the data be interpreted?
5. How does the TG 425 UDP comply with the 3Rs principles compared to TG 401? Quantify the reduction in animal use.

Safety considerations (for real rodent studies, educational reference):

PPE: gloves (nitrile), lab coat, eye protection. For compounds of unknown toxicity: biosafety cabinet for dose preparation; additional respiratory protection (half-face mask with organic vapour cartridge if volatile). Emergency: if dermal contact with test compound occurs, wash with soap and water for 15 min; consult SDS; inform supervisor; complete incident report form SOP-SAFETY-002. Waste disposal: animal carcasses in labelled biohazard bags; contaminated bedding in chemical waste containers; unused test compound formulation neutralised per SDS recommendations.

Evaluation rubric:

- Correct dose sequence execution (4 points)
- Accurate LD₅₀ and 95% CI calculation (4 points)

—	Correct	GHS	classification	and	labelling	(4	points)
—	Quality	of	discussion	answers		(4	points)
—	Laboratory	report	completeness			(4	points)

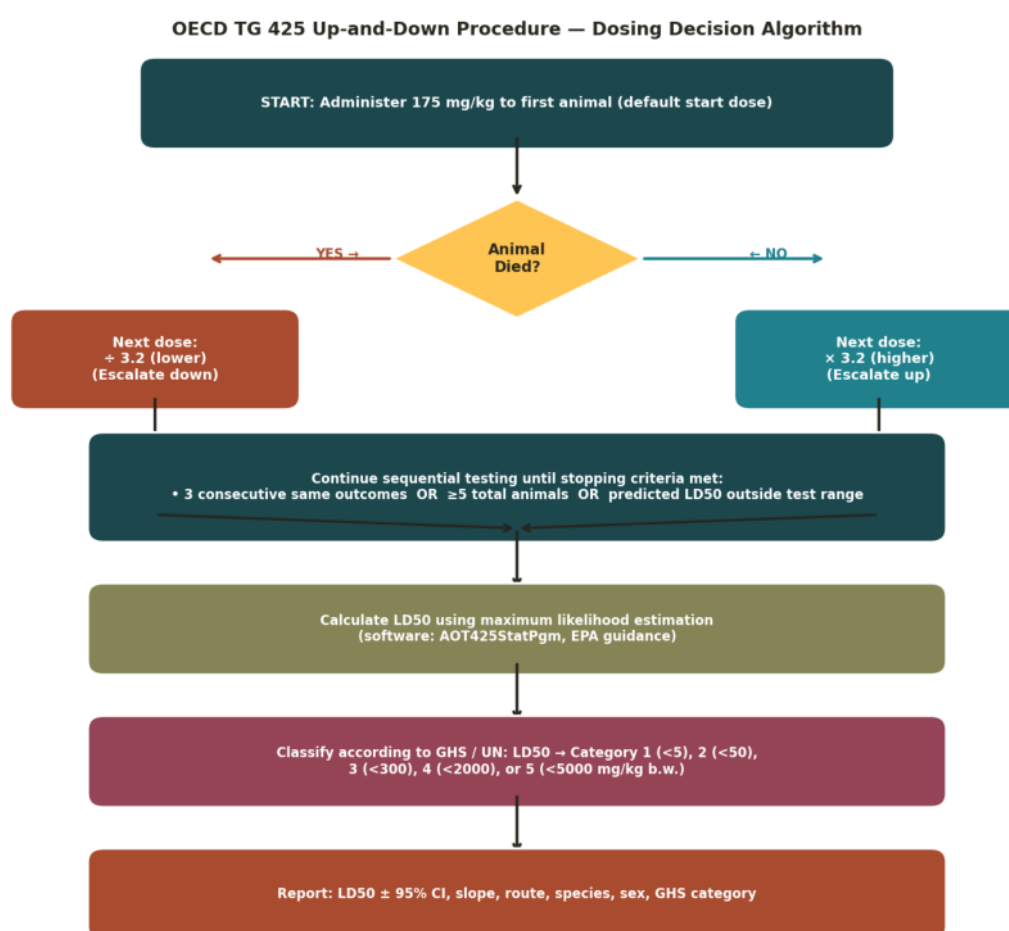
TOTAL: 20 points

P.3.1 Objectives

- Understand the up-and-down procedure and why it replaced the classical LD50 method.
- Practice sequential dosing decisions and LD50 calculation by maximum likelihood estimation.
- Classify the test substance according to the GHS system based on the derived LD50.

P.3.2 Background

Students will work with a simulated dataset representing sequential dosing of a fictitious organophosphate compound (OP-Test-X) via oral gavage in Wistar rats. The starting dose is 175 mg/kg, consistent with OECD TG 425 guidelines.



Reference: OECD Test Guideline 425 (2008) — Acute Oral Toxicity: Up-and-Down Procedure

Figure P1.1. OECD TG 425 Up-and-Down Procedure — Dosing Decision Algorithm. Starting from the default dose of 175 mg/kg, each subsequent dose is adjusted by a factor of 3.2 based on the outcome of the preceding animal. The LD50 is calculated by maximum likelihood estimation using EPA AOT425StatPgm software. (Original figure.)

P.3.3 Protocol

Step	Action	Detail / Rationale
------	--------	--------------------

1	Review available information	Structural alerts; physicochemical data; in vitro IC50 if available → select starting dose
2	Prepare test substance	Dissolve in vehicle (corn oil or 0.5% CMC); prepare dose at 50 mg/mL for 175 mg/kg at 3.5 mL/kg
3	Gavage first animal	Oral gavage (3.5 mL/kg); observe for 48 h; record clinical signs
4	Apply dosing decision rule	Survival → × 3.2 next dose; death → ÷ 3.2 next dose
5	Continue sequential dosing	Until stopping rule: 3 reversals or 5 animals per sex
6	Calculate LD50	AOT425StatPgm (EPA); input: doses and outcomes; output: LD50, 95% CI
7	GHS classification	Compare LD50 to GHS category thresholds (5/50/300/2000/5000 mg/kg)
8	Write study report	Species, strain, sex, age, body weight, route, vehicle, observations, LD50, GHS category

P.3.4 Discussion Questions

- Why is the starting dose 175 mg/kg in TG 425? Under what circumstances would a different starting dose be chosen?
- A substance produces LD50 = 420 mg/kg. What is its GHS oral acute toxicity category? What label would be required under CLP Regulation?
- What are the relative merits of TG 420, 423, and 425? When would you choose each?

P.3.5 Safety Considerations

Oral gavage: appropriate training required; verify absence of tracheal intubation (bubbles, coughing). PPE: gloves, lab coat, safety glasses. Disposal of chemical waste in labelled containers per institutional SOP.

P.4 Practical 2: In Vitro Cytotoxicity Assay (MTT on HepG2 with CdCl₂)

Practical 2 — Full expanded MTT protocol and materials:

Pre-lab

quiz:

Q1: What does MTT stand for and what colour is the formazan product?

A1: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide. The formazan product is dark purple.

Q2: Why is DMSO used in the MTT assay?

A2: DMSO dissolves the insoluble purple formazan crystals, producing a homogeneous purple solution that can be measured by spectrophotometry at 570 nm.

Q3: What is the formula for calculating % cell viability?

A3: % Viability = $(OD_{\text{treated}} - OD_{\text{blank}}) / (OD_{\text{control}} - OD_{\text{blank}}) \times 100$

Q4: What is the IC50 and how is it graphically determined?

A4: IC50 is the concentration of test substance causing 50% inhibition of cell viability. Graphically, it is the concentration at which the dose-response curve crosses the 50% viability line on the y-axis (read off on the log-dose x-axis).

Q5: Name two potential sources of interference in the MTT assay.
 A5: Any two from: coloured test compounds (absorb at 570 nm); reducing agents (directly reduce MTT non-enzymatically); cytotoxic DMSO vehicle at >0.5% v/v; heterogeneous formazan dissolution in DMSO; edge effects (unequal evaporation from perimeter wells).

Full materials list (for 12 students, 2 × 96-well plates):
 — HepG2 cells (ATCC HB-8065), passage 5–20, cultured in DMEM/F12 high glucose + 10% FBS + 1% PenStrep (100 U/mL penicillin, 100 µg/mL streptomycin); CAS of DMEM: not applicable.
 — CdCl₂ (CAS 10108-64-2; MW 183.32 g/mol; Sigma-Aldrich ≥99.9% purity): 100 mM stock in sterile distilled H₂O; prepare fresh from analytical-grade solid. CAUTION: Cd²⁺ is a SVHC carcinogen/reproductive toxin under REACH; use in fume hood with N-FFP3 mask if dust exposure possible.

— MTT reagent (CAS 298-93-1; MW 414.32 g/mol; Sigma-Aldrich M2128): 5 mg/mL in sterile PBS (pH 7.4); store at –20°C in 0.5 mL aliquots; protect from light.
 — DMSO (CAS 67-68-5; Sigma-Aldrich D2650 ≥99.9%): for formazan dissolution; 150 µL per well; maintain stock for cell biology use (<0.1% added to medium).
 — PBS (10× stock, CAS not applicable; sterile, pH 7.4).
 — 96-well flat-bottom tissue culture plates, black (for fluorescence) or clear (for absorbance): 2 plates per practical group.
 — Multi-channel pipette (8-channel, 20–200 µL); multi-channel pipette (8-channel, 5–50 µL).
 — Microplate reader (absorbance 570/630 nm; e.g., BioTek Epoch 2 or Molecular Devices Spectramax).
 — CO₂ incubator (37°C, 5% CO₂, 95% humidity).
 — Class II biosafety cabinet.
 — Haemocytometer and light microscope.
 — Trypan blue (0.4% w/v in PBS).
 — 0.25% Trypsin/EDTA solution (sterile, for cell detachment).
 — Triton X-100 (1% in PBS; CAS 9002-93-1; positive control for 0% viability).
 — Data recording sheet for absorbance values.

Step-by-step protocol (4h session):
 Day 0 (cell seeding, 30 min):
 1. Aspirate spent medium from T75 flask; wash once with 10 mL PBS.
 2. Add 3 mL trypsin/EDTA solution; incubate 3–4 min at 37°C until cells round up.
 3. Neutralise with 7 mL complete medium; transfer to 15 mL tube; centrifuge 200×g, 5 min.
 4. Aspirate supernatant; resuspend pellet in 5 mL complete medium.
 5. Count cells with haemocytometer and trypan blue (record viability; proceed only if ≥90%).
 6. Adjust to 5×10⁴ cells/mL in complete medium.
 7. Distribute 200 µL/well to the inner 60 wells of the plate (rows B–G, columns 2–11). Fill outer wells (row A, H, columns 1, 12) with 200 µL PBS.
 8. Incubate overnight at 37°C, 5% CO₂.

Day 1 (treatment, 30 min):
 9. Prepare 8-point CdCl₂ serial dilution in complete medium: 1000, 300, 100, 30, 10, 3, 1, 0 µM (vehicle control).
 10. Prepare negative control: complete medium with 1% Triton X-100 (v/v).
 11. Aspirate medium from cells; add 200 µL of each concentration in triplicate wells (1 concentration per column: e.g., column 2 = 1000 µM, column 3 = 300 µM... column 9 = 1 µM, column 10 = 0 µM vehicle, columns 11 = Triton X-100 negative control).
 12. Return plate to incubator; expose for exactly 24 h.

Day 2 (MTT measurement, 1h):
 13. Remove plate from incubator; prepare MTT working solution (20 µL MTT stock + 180 µL medium

per well = 200 μ L total). Alternatively, add directly: 20 μ L MTT stock to each well containing 200 μ L medium (final 0.5 mg/mL).

14. Return to incubator for 2.5 h (critical: do not exceed 4h; protect from light).
15. Aspirate medium+MTT carefully (vacuum aspiration with fine-tip; tilt plate to aspirate from corner, NOT from the well centre to avoid crystal disturbance).
16. Immediately add 150 μ L DMSO per well. Seal with adhesive film.
17. Shake on orbital shaker: 300 rpm, 15 min, room temperature.
18. Read absorbance at 570 nm (reference: 630 nm) on plate reader.

Data analysis:

$$\% \text{ Viability} = (\text{OD}_{570-630} \text{ treated} - \text{OD blank}) / (\text{OD}_{570-630} \text{ control} - \text{OD blank}) \times 100$$

Plot % viability vs. $\log_{10}[\text{CdCl}_2 \text{ } \mu\text{M}]$ on semi-log paper (or GraphPad Prism/Excel).
 Fit 4-PL model: $y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + (\text{IC}_{50}/x)^{\text{HillSlope}})$
 Report: $\text{IC}_{50} (\mu\text{M}) \pm 95\% \text{ CI}$; Hill slope.

Post-lab calculation worked example:

Control OD (mean, n=3): 0.855 ± 0.032 ; Blank OD: 0.060 ± 0.005 ; Triton X-100 (0% viability): 0.075 ± 0.008 .

CdCl_2 30 μM OD: 0.568 ± 0.042 .

$$\% \text{ Viability at } 30 \mu\text{M} = (0.568 - 0.060) / (0.855 - 0.060) \times 100 = 508/795 \times 100 = 63.9\%$$

IC_{50} from 4-PL fit to full dataset: $\text{IC}_{50} = 52.3 \mu\text{M}$ (95% CI: 41.8–65.5 μM).
 Literature comparison: Published IC_{50} for CdCl_2 in HepG2 (24h MTT) $\approx 40\text{--}80 \mu\text{M}$ — result is consistent.

Safety considerations:

CdCl_2 is a carcinogen (IARC Group 1), reproductive toxin (Repr. 2), and acute toxic substance (Acute Tox. 4 oral). Full PPE (nitrile gloves, lab coat, safety glasses) at all times. Prepare dose solutions in the fume hood. Decontaminate all waste (medium, tips, tubes) with 10 mM EDTA solution before disposal in heavy metal waste container. Never pipette by mouth. If spillage: contain with absorbent paper; place in labelled waste bag; decontaminate bench surface with 10 mM EDTA solution; inform supervisor. SDS for CdCl_2 must be available in the laboratory.

Evaluation	rubric	(20 points):
— Correct seeding and cell culture technique	observed	(3 points)
— Accurate dose preparation and plate layout		(4 points)
— Correct MTT procedure and timing		(3 points)
— Accurate data entry and % viability calculations		(4 points)
— IC_{50} determination and comparison to literature		(3 points)
— Safety compliance and waste disposal observed (3 points)		

P.4.1 Objectives

- Perform a standard MTT viability assay on HepG2 cells exposed to cadmium chloride.
- Construct a dose-response curve and calculate IC_{50} by non-linear regression.
- Compare CdCl_2 IC_{50} to published values; discuss the role of hepatocyte model choice.

P.4.2 Background

Cadmium (Cd) is a nephrotoxic and hepatotoxic heavy metal classified IARC Group 1 (carcinogenic to humans). Its primary mechanism involves displacement of zinc from metallothionein and zinc finger proteins, induction of oxidative stress (ROS generation, GSH depletion), and mitochondrial dysfunction. The HepG2 cell line, derived from human hepatocellular carcinoma, provides a reproducible and accessible model for in vitro hepatotoxicity screening despite its limited CYP450 expression.

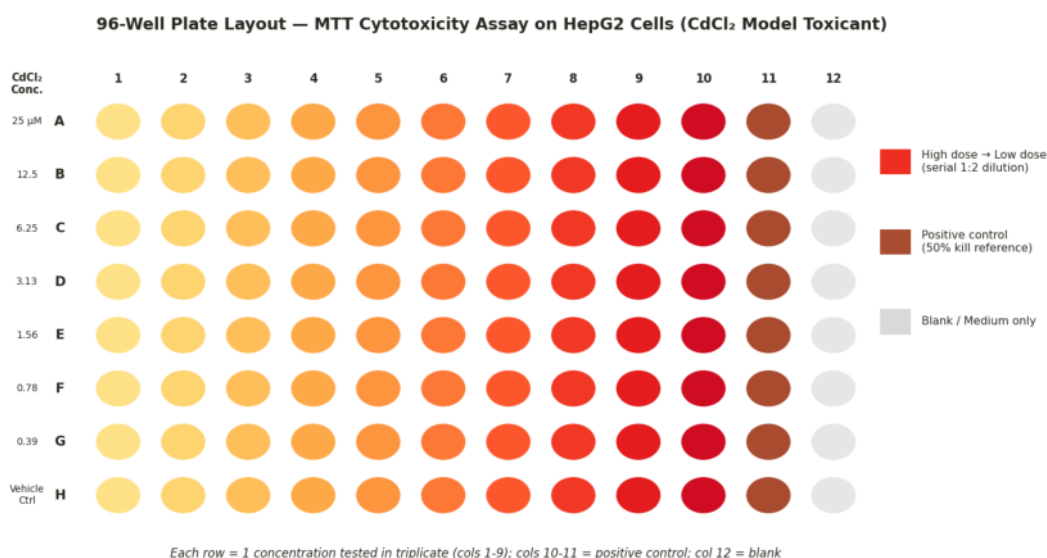


Figure P2.1. 96-Well Plate Layout for the MTT Cytotoxicity Assay (CdCl₂ model toxicant on HepG2 cells). Rows A–H represent eight serial 1:2 dilutions from 25 μ M to $\sim$ 0.2 μ M. Columns 10–11 = positive control (1 mM CdCl₂); Column 12 = blank (medium only). (Original figure.)

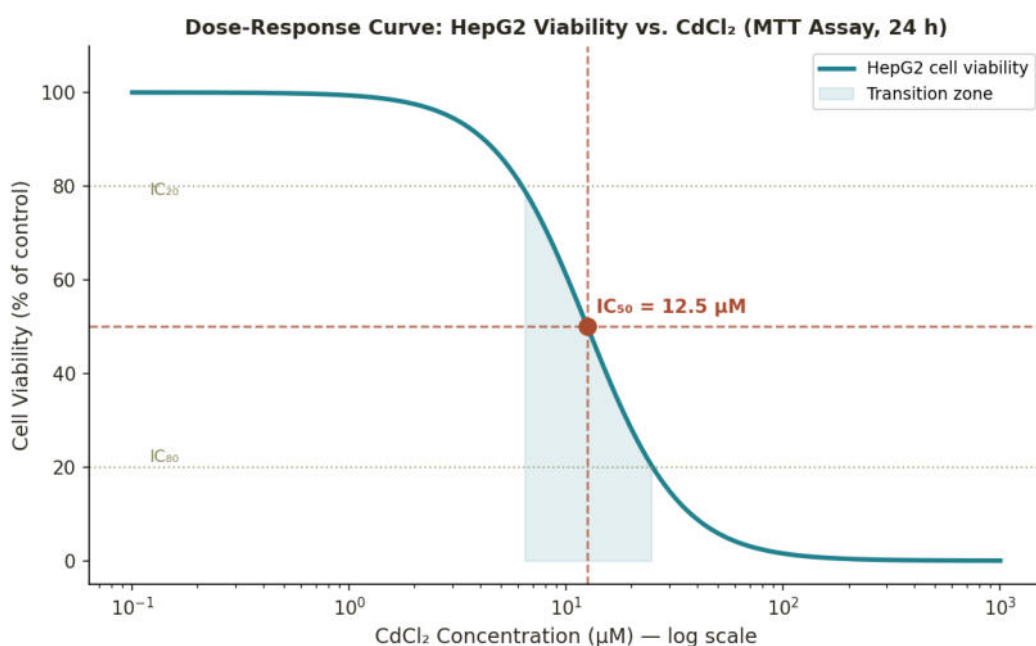


Figure P2.2. Example dose-response sigmoid curve: HepG2 cell viability (%) vs. CdCl₂ concentration (μ M), fitted by 4-parameter logistic model. IC₅₀ $\approx$ 12.5 μ M at 24 h exposure. Students derive IC₅₀ by non-linear regression using GraphPad Prism or R (drc package). (Original figure.)

P.4.3 Materials

HepG2 cells (ATCC HB-8065) in DMEM + 10% FBS + 1% penicillin/streptomycin; CdCl₂ (Sigma-Aldrich $\geq$ 99%); MTT reagent (5 mg/mL in PBS, sterile-filtered); DMSO (cell culture grade); 96-well flat-bottom plates; multichannel pipette; microplate reader (570 nm filter).

P.4.4 Step-by-Step Protocol

Step	Action	Note
------	--------	------

1	Seed 5×10^3 HepG2 cells/well in 100 μ L DMEM	Day 0; allow 24 h attachment
2	Prepare CdCl ₂ serial dilutions	Stock 1 mM in DMEM; 2-fold dilution series: 25, 12.5, 6.25, 3.13, 1.56, 0.78, 0.39 μ M + vehicle
3	Remove medium; add 100 μ L treatment solution	Day 1; in triplicate per concentration
4	Incubate 24 h (37°C, 5% CO ₂)	Maintain humidified atmosphere
5	Remove treatment medium; add 100 μ L MTT (0.5 mg/mL in DMEM)	Day 2
6	Incubate 4 h (37°C); remove MTT	Protect from light
7	Add 100 μ L DMSO; mix 5 min on plate shaker	Dissolve formazan crystals
8	Read absorbance at 570 nm (ref 690 nm)	Blank-correct; average triplicates
9	Calculate % viability vs. vehicle control	Viability(%) = (OD sample / OD control) $\times$ 100
10	Fit 4-parameter logistic model; derive IC ₅₀	GraphPad Prism: [Inhibitor] vs. response (4PL); R: drm() in 'drc' package

P.4.5 Safety Considerations

CdCl₂ is acutely toxic and classified SVHC under REACH. Handle in ventilated fume hood; wear nitrile gloves; dispose of all cadmium-containing waste as hazardous chemical waste. Cell work: laminar flow cabinet; treat all cell culture waste as biohazard.

P.4.6 Discussion Questions

- Why might the IC₅₀ of CdCl₂ differ between HepG2 cells and primary human hepatocytes? What metabolic processes might explain the difference?
- If the IC₅₀ at 24 h is 12.5 μ M, calculate the IC₅₀ in mg/L for a regulatory risk assessment comparison.
- What additional mechanistic endpoints would you measure to determine whether cytotoxicity is mediated by oxidative stress vs. direct mitochondrial damage?

P.5 Practical 3: Histopathological Observation of Liver and Kidney Sections

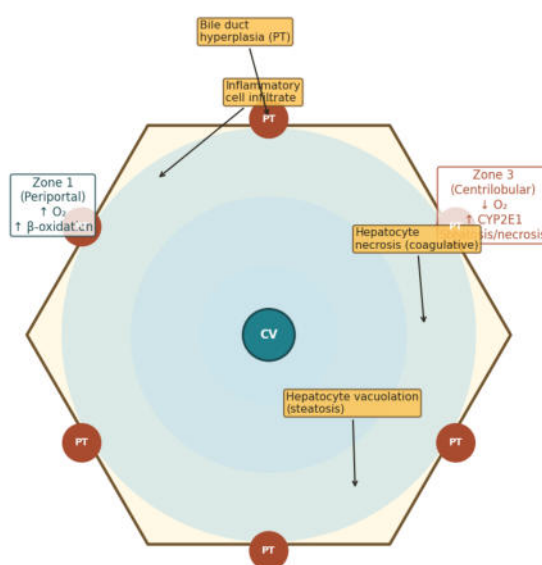
P.5.1 Objectives

- Recognise the normal histological architecture of rat liver and kidney.
- Identify hallmark lesions of hepatotoxicity (steatosis, necrosis, inflammation, fibrosis) and nephrotoxicity (proximal tubular necrosis, cast formation, glomerulonephritis).
- Grade lesion severity using a standard semi-quantitative scoring system.

P.5.2 Background

Histopathological evaluation of target organs is the most sensitive and informative endpoint in repeat-dose toxicity studies. The liver and kidney are the principal target organs for most systemic toxicants because of their central role in xenobiotic metabolism and excretion. Students examine H&E-stained sections (provided as glass slides or digital images) from a 90-day rat study with a hepatotoxic compound (e.g., carbon tetrachloride as positive control) and a nephrotoxic compound (e.g., gentamicin as positive control).

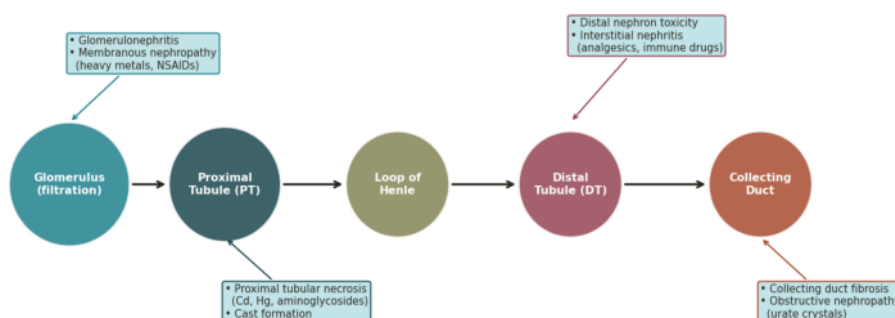
Hepatic Lobule — Annotated Schematic with Hallmark Toxicological Lesions



CV = Central Vein | PT = Portal Triad | Lesions characteristic of hepatotoxic insult

Figure P3.1. Hepatic Lobule — Annotated Schematic with Hallmark Toxicological Lesions. The central vein (CV) and portal triads (PT) define the metabolic zones. Zone 3 (centrilobular) is most susceptible to bioactivation-dependent toxicity (CCl_4 , acetaminophen); Zone 1 (periportal) is most exposed to dietary toxicants. (Original figure.)

Nephron Segments — Key Sites and Types of Nephrotoxic Lesions



PT is the principal target for heavy metal nephrotoxicity (Cd, Pb, Hg) due to active reuptake transporters

Figure P3.2. Nephron Segments — Key Sites and Types of Nephrotoxic Lesions. The proximal tubule (PT) is the primary target for heavy metal nephrotoxicity (Cd, Pb, Hg) and aminoglycoside antibiotics due to active uptake transporters. (Original figure.)

P.5.3 Lesion Scoring System

Grade	Score	Description
Normal	0	No lesion detected
Minimal	1	Rare foci; affects <10% of section area

Mild	2	Small but clearly present; 10–25% area
Moderate	3	Multifocal to coalescing; 25–50% area
Marked	4	Diffuse; >50% area; impairs function

P.5.4 Discussion Questions

- Hepatocellular hypertrophy (enlarged hepatocytes) is observed at the low and mid dose in a 90-day study. Is this an adverse or non-adverse finding? What additional information would help you decide?
- In the kidney, proximal tubular basophilia (regenerating tubular cells) is observed at the high dose. What does this indicate about the nature and timing of the nephrotoxic injury?
- How does the zonal distribution of liver lesions (centrilobular vs. periportal) inform the likely mechanism of the hepatotoxicant?

P.6 Practical 4: Heavy Metal Quantification by AAS in Fish Tissue

P.6.1 Objectives

- Understand the principle of atomic absorption spectrometry (AAS) for trace metal quantification.
- Prepare a validated calibration curve for cadmium in fish tissue digest.
- Interpret results in the context of Algerian coastal fish biomonitoring and food safety thresholds.

P.6.2 Background — Mediterranean Fish Example

Monitoring of heavy metals in edible fish tissues is a key public health tool in Algeria's western coastal region. The author's laboratory (LMBP, USTO-MB) has conducted extensive biomonitoring of Cd, Pb, and Hg in *Sparus aurata* (gilthead sea bream), *Mullus barbatus* (red mullet), and *Pagellus acarne* (Spanish sea bream) from Oran Bay and adjacent coastal areas. Elevated Cd concentrations have been documented in kidneys and livers of *Mullus barbatus* exceeding EU maximum residue levels (0.05 mg/kg ww in muscle; 1.0 mg/kg ww in liver, EU Regulation 1881/2006).

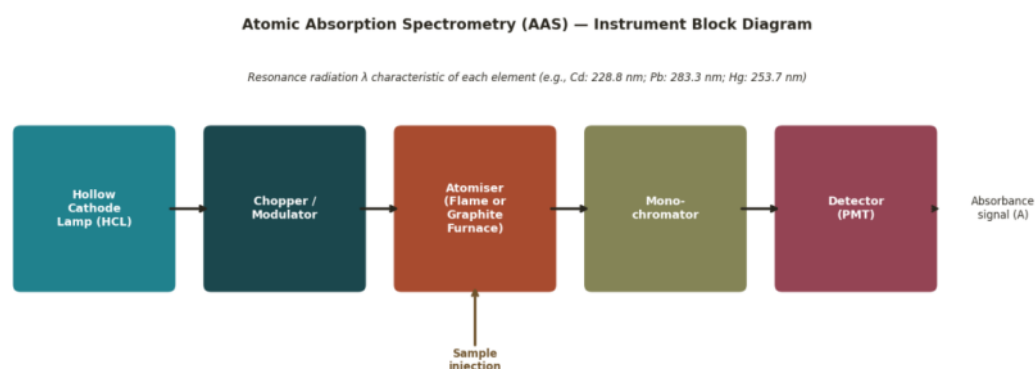


Figure P4.1. Atomic Absorption Spectrometry (AAS) — Instrument Block Diagram. The hollow cathode lamp (HCL) emits element-specific resonance radiation; the atomiser (flame or graphite furnace) converts sample analyte to ground-state atoms; the monochromator isolates the resonance wavelength; the detector (PMT) measures absorbance. (Original figure.)

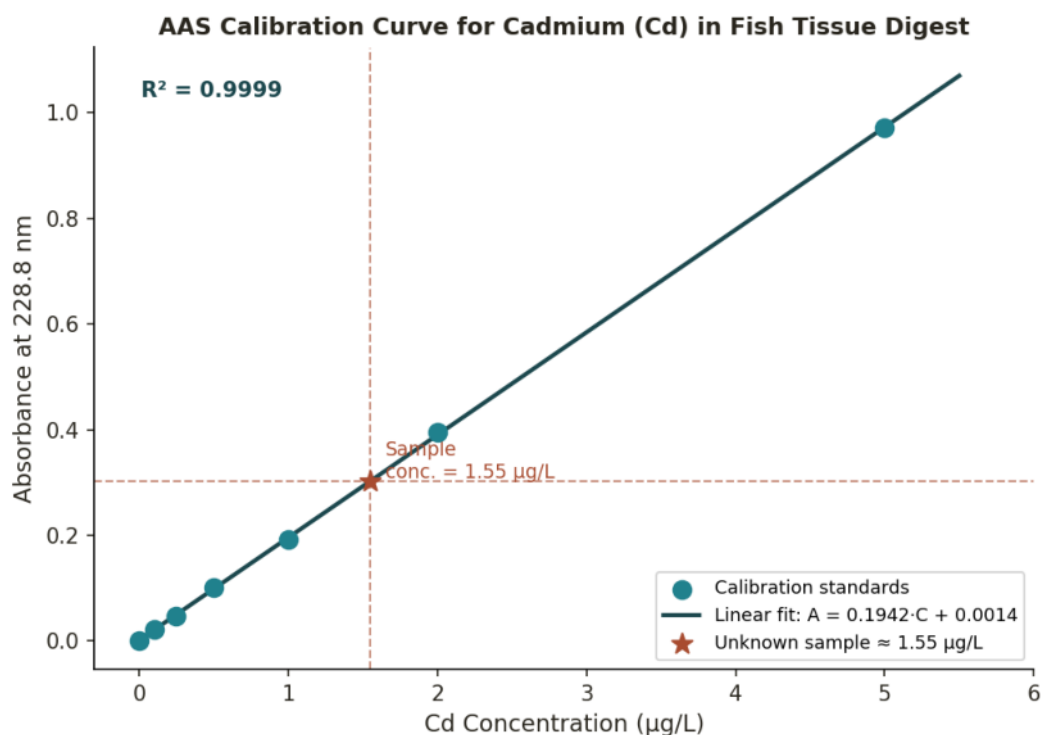


Figure P4.2. AAS Calibration Curve for Cadmium (Cd) at 228.8 nm in Fish Tissue Digest. Linear range: 0–5 $\mu\text{g/L}$; $R^2 > 0.999$. The unknown sample concentration is read from the interpolated calibration function. (Original figure.)

P.6.3 Protocol

Step	Action	Note
1	Tissue preparation	Freeze-dry muscle tissue; weigh 0.5 g into Teflon digestion vessel
2	Acid digestion	Add 5 mL HNO_3 (65%) + 2 mL H_2O_2 (30%); microwave digestion (Milestone Ethos): 200°C, 30 min
3	Make up to volume	Cool; transfer quantitatively to 25 mL with Milli-Q water; mix
4	Prepare calibration standards	Cd: 0, 0.1, 0.25, 0.5, 1.0, 2.0, 5.0 $\mu\text{g/L}$ in 0.2% HNO_3 from 1000 mg/L CRM standard
5	AAS measurement	Graphite furnace AAS (GF-AAS, Zeeman background correction); wavelength: 228.8 nm; inject 20 μL
6	Calibration curve	Plot absorbance vs. concentration; linear regression; verify $R^2 > 0.999$
7	Calculate Cd concentration	Concentration (mg/kg ww) = (concentration $\mu\text{g/L} \times \text{volume L} \times \text{dilution factor}) / \text{mass g}$
8	QC checks	Certified Reference Material (DORM-4 fish tissue, NRC Canada); recovery 85–115%
9	Compare to regulatory limits	EU MRL: muscle 0.05 mg/kg ww; liver 1.0 mg/kg ww (EU Reg. 1881/2006)

P.6.4 Method Validation Parameters

Parameter	Criterion	Method
Linearity	$R^2 \geq 0.999$	Linear regression over calibration range
LOD	Typically 0.01–0.05 µg/L for Cd GF-AAS	$3\sigma/\text{slope}$ (σ = SD of 10 blank measurements)
LOQ	Typically 0.03–0.15 µg/L	$10\sigma/\text{slope}$
Accuracy	Recovery 85–115% of CRM certified value	DORM-4 (NRC) or ERM-CE278k (EU-JRC)
Precision (repeatability)	RSD $\leq$ 5% within-run	5 replicate measurements of a QC sample
Matrix effect	Signal enhancement/suppression $<20\%$	Standard addition or isotope dilution ICPMS

P.6.5 Discussion Questions

- Why is graphite furnace AAS (GF-AAS) preferred over flame AAS for cadmium in fish tissue? What are the detection limit implications?
- A student obtains Cd = 0.08 mg/kg ww in *Mullus barbatus* muscle from Oran Bay. What is the Margin of Exposure for a consumer eating 100 g of this fish per week, given a BMDL = 0.9 µg/kg bw/week (EFSA, 2009)?
- How does the bioaccumulation factor (BAF) relate the fish tissue concentration to the ambient seawater concentration? Calculate BAF if the seawater Cd concentration is 0.003 µg/L.

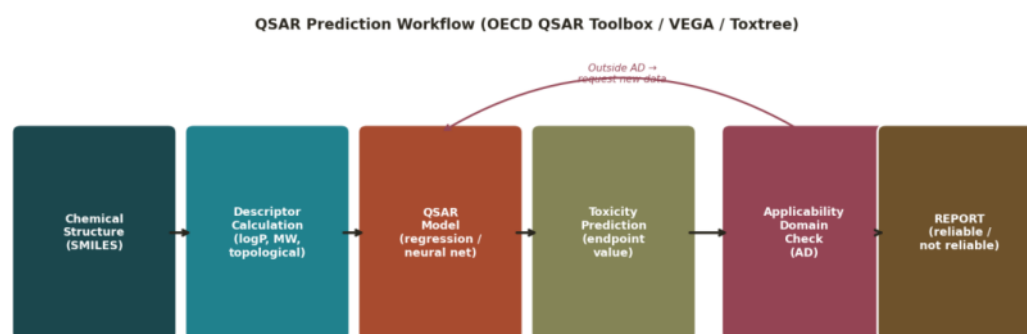
P.7 Practical 5: QSAR Prediction Using OECD QSAR Toolbox

P.7.1 Objectives

- Navigate the OECD QSAR Toolbox interface and perform a full QSAR prediction workflow.
- Assess applicability domain for a target chemical relative to the training set.
- Interpret QSAR predictions in the context of regulatory decision-making.

P.7.2 Background

The OECD QSAR Toolbox (v4.6) is a free software platform developed by OECD and LHASA Limited that integrates >50 QSAR models, structural alerting systems, and metabolic simulators for regulatory toxicity assessment. It is widely used in REACH dossier preparation for filling data gaps by read-across and QSAR prediction. Students will predict the Ames mutagenicity and skin sensitisation potential of a novel pesticide analogue (e.g., a hypothetical organophosphate: dimethyl phosphorothioate with a nitrophenyl leaving group).



Five OECD Principles for valid QSAR: defined endpoint, unambiguous algorithm, defined AD, appropriate fit statistics, mechanistic interpretation

Figure P5.1. QSAR Prediction Workflow: from SMILES input through descriptor calculation, model application, and applicability domain check, to final report. An outside-AD flag triggers a request for experimental data to fill the uncertainty. (Original figure.)

P.7.3 Protocol

Step	Action
1	Open OECD QSAR Toolbox (v4.6); File > New Profile
2	Enter target chemical as SMILES string or CAS number
3	Step 1 — Profiling: run structural alerts (Ames mutagenicity; skin sensitisation; aquatic toxicity)
4	Step 2 — Chemical space: plot target vs. training set in descriptor space; confirm AD
5	Step 3 — Data gap filling: run QSAR models (Ames: SARpy; skin sensitisation: 2-out-of-3 DA)
6	Step 4 — Read-across: identify 3 most structurally similar analogues; check mechanism plausibility
7	Step 5 — Report: export prediction report with AD flag, confidence level, and regulatory interpretation
8	Critical appraisal: assess whether Ames prediction is confirmed by available experimental data

P.7.4 Discussion Questions

- The QSAR Toolbox predicts the target compound as Ames-positive with moderate confidence. The applicability domain flag shows 2 training set neighbours are outside the AD. How does this affect the regulatory weight of the prediction?
- What does a 2-out-of-3 positive result in the skin sensitisation defined approach (DPRA + KeratinoSens + h-CLAT) mean in terms of GHS hazard classification?
- If the QSAR prediction is used in a REACH dossier for a substance registered at 100–1000 tonnes/year, what additional information would ECHA likely request?

P.8 Practical 6: Critical Statistical Analysis (NOAEL/LOAEL Determination, BMD Calculation)

P.8.1 Objectives

- Determine NOAEL and LOAEL from a dataset representing a 90-day rat repeat-dose study.
- Apply BMD modelling (EPA BMDS or EFSA PROAST) and derive BMDL10.
- Derive a Reference Dose (RfD) and compare NOAEL-based and BMDL-based approaches.

P.8.2 Provided Dataset

Students receive a dataset from a fictitious 90-day dietary exposure study in Wistar rats (10/sex/dose): 4 dose groups (0, 10, 30, 100, 300 mg/kg bw/day). Primary endpoints: serum ALT (hepatotoxicity marker), kidney relative weight, body weight gain. Statistical analyses are performed using R (packages: drc, BMDS) or EPA BMDS online.

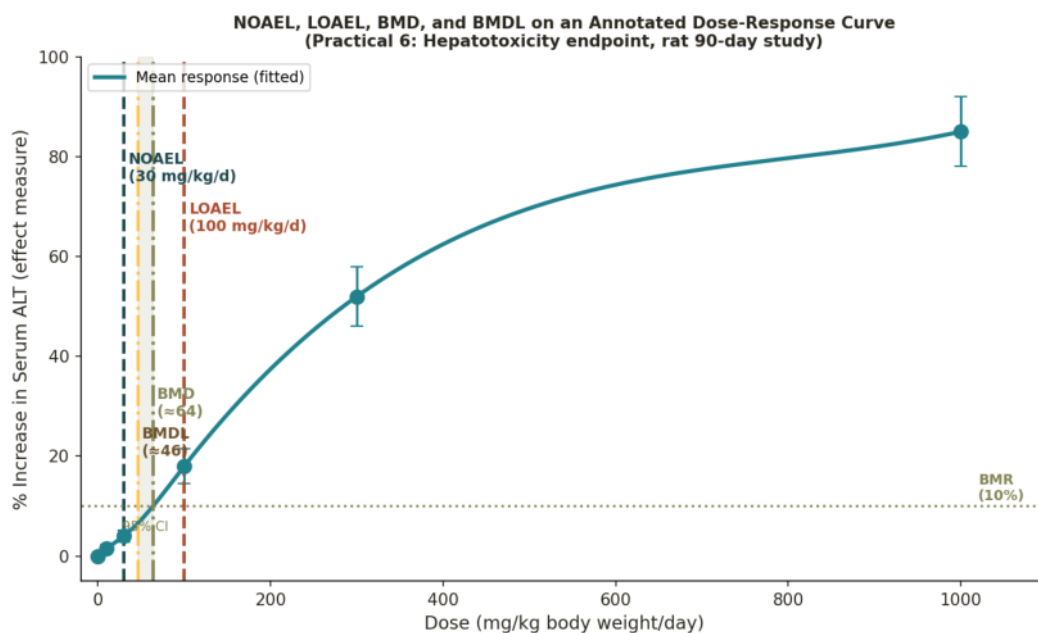


Figure P6.1. NOAEL, LOAEL, BMD, and BMDL on an Annotated Dose-Response Curve. The NOAEL is the highest dose with no statistically significant adverse effect; the LOAEL is the lowest dose with a significant effect. The BMD10 is derived from a fitted dose-response model at a 10% benchmark response; BMDL10 is the lower 95% confidence limit. (Original figure.)

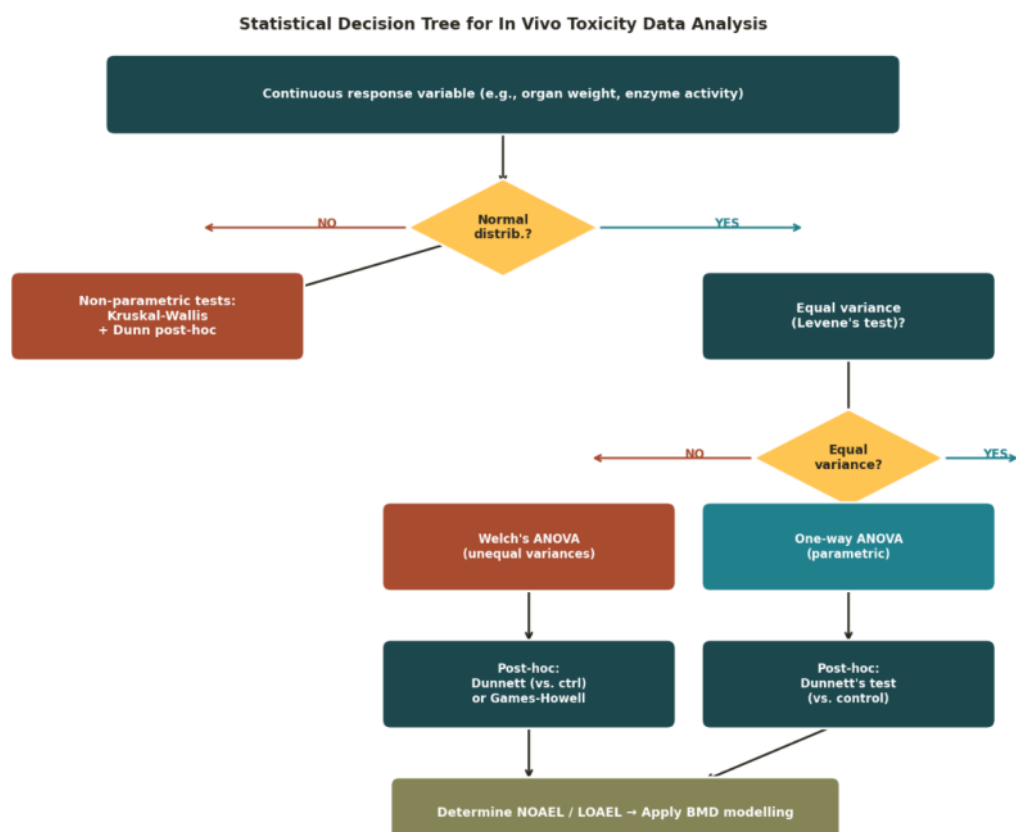


Figure P6.2. Statistical Decision Tree for In Vivo Toxicity Data Analysis. The tree guides selection between parametric (ANOVA + Dunnett) and non-parametric (Kruskal-Wallis + Dunn) approaches based on normality and variance homogeneity tests. (Original figure.)

P.8.3 Statistical Analysis Protocol

Step	Test / Software	Decision
1. Normality	Shapiro-Wilk test (R: shapiro.test)	$p > 0.05 \rightarrow$ parametric; $p < 0.05 \rightarrow$ non-parametric
2. Variance homogeneity	Levene's test (R: leveneTest, car package)	$p > 0.05 \rightarrow$ equal variance; $p < 0.05 \rightarrow$ Welch
3. Overall significance	One-way ANOVA (parametric) or Kruskal-Wallis	$p < 0.05 \rightarrow$ proceed to post-hoc
4. Post-hoc: treated vs. control	Dunnett's test (R: DunnettTest, DescTools package)	Identify lowest dose with $p < 0.05 \rightarrow$ LOAEL
5. NOAEL	Highest dose with no significant adverse finding	Requires judgement on adverse vs. adaptive
6. BMD modelling	EPA BMDS 3.3 or EFSA PROAST	Fit multiple models; select lowest AIC; derive BMD10 and BMDL10
7. RfD derivation	$RfD = NOAEL / (UF_A \times UF_H)$ or $BMDL10 / (UF_A \times UF_H)$	$UF_A=10$ (interspecies); $UF_H=10$ (intraspecies); UF_D if database incomplete

P.8.4 Discussion Questions

- The NOAEL for serum ALT elevation is 30 mg/kg/day and the BMDL10 is 22 mg/kg/day. Calculate RfD using both approaches ($UF = 100$ each). Which value is preferred for regulatory use and why?
- Body weight gain is statistically decreased at 300 mg/kg/day only. Is this an adverse effect? How does OECD guidance distinguish body weight effects that are primary toxic signs from secondary effects of palatability?
- If you had dose groups at 0, 30, 100, and 1000 mg/kg instead of 0, 10, 30, 100, 300 mg/kg, how would this affect NOAEL identification and BMD modelling?

P.8.5 Safety Considerations

Statistical analysis practical: computer-only session; no chemical hazards. Ensure student data files do not contain personal data from real studies (GDPR). Provide R scripts with clear annotation; ensure version control ($R \geq 4.3$, $drc \geq 3.0$).

P.9 Practical Report Template and Marking Grid

Practical Report Template — P.9 Expected Format

The following template describes the expected format for written practical reports submitted at the end of each TP session. The report must be submitted individually, typed, within 7 days of the practical session.

TITLE: [Descriptive, specific title, e.g., "In Vitro Cytotoxicity of Cadmium Chloride on HepG2 Cells: MTT Assay and IC50 Determination"]

ABSTRACT (150–200 words): Summarise the objective, methods (main techniques), key results (IC50 value $\pm$ CI), and conclusion (comparison to literature and brief toxicological interpretation). The abstract must be self-contained and readable without the rest of the report.

1. **INTRODUCTION** (300–400 words):
 - Background on the test compound (chemical class, mechanism of toxicity, regulatory status)
 - Rationale for the chosen test system (HepG2 cells, target organ, metabolic competence)
 - Objective of the practical (what question is the experiment designed to answer?)
 - Brief description of the analytical method (MTT assay principle, output variable)
 - Hypothesis (what result do you expect and why?)
2. **MATERIALS AND METHODS** (400–500 words):
 - Materials: cell line (species, passage number, culture conditions), test compound (CAS, MW, solvent, stock concentration), reagents (MTT concentration, DMSO volume, controls)
 - Experimental design: plate layout diagram (essential), concentration range, exposure time
 - Protocol: concise step-by-step summary (NOT a restatement of the pre-written protocol; focus on what you actually did, including any deviations)
 - Data analysis: formula for % viability calculation; curve-fitting method (software, model, parameters)
3. **RESULTS** (300–400 words + tables/figures):
 - Table 1: Raw OD values (mean $\pm$ SD) for each concentration (n=3 technical replicates)
 - Table 2: % Viability for each concentration (calculated)
 - Figure 1: Dose-response curve (% viability vs. $\log_{10}$ [concentration μ M]); include curve fit, IC50 $\pm$ 95% CI on graph; follow ARRIVE/APA figure formatting standards (axis labels with units, error bars defined in legend, n stated)
 - Narrative: describe the dose-response relationship, note the IC50, report the 95% CI, state whether the Hill slope is >1 (steep) or <1 (shallow) and its biological interpretation
4. **DISCUSSION** (400–500 words):
 - Compare IC50 to published values in the literature (cite ≥ 2 sources); explain any discrepancies (passage number, serum batch, exposure time)
 - Relate IC50 to relevant human exposure levels (occupational, environmental); calculate BER if data available
 - Discuss limitations of the MTT assay in this context (specific interferences, what additional endpoint would improve mechanistic characterisation)
 - Connect findings to the broader toxicological context (does cadmium at 52 μ M represent a realistic in vivo concentration? What does this mean for the translational value?)
 - Suggest one improvement to the experimental design
5. **CONCLUSION** (50–100 words): State the main finding (IC50 value), its consistency with the literature, and the main limitation of the study.
6. **REFERENCES:** APA 7th edition format. Minimum 5 references, including: ≥ 1 primary research article, ≥ 1 OECD TG or regulatory document, ≥ 1 review article published 2020–2025.

MARKING	GRID:
— Abstract quality and completeness:	3/20
— Introduction (background, rationale, hypothesis):	3/20
— Methods (clarity, plate diagram, deviation reporting):	3/20
— Results (calculations, figure quality, IC50 with CI):	5/20
— Discussion (literature comparison, translational relevance, limitations):	4/20
— References (format, minimum number, relevance):	2/20
TOTAL: 20/20	

Submission format: PDF, file name: [Surname_FirstName_TP_number_Academic_Year.pdf]. Font: Times New Roman 11pt, 1.15 line spacing, A4, 2.5cm margins. Maximum 8 pages including figures and tables. Plagiarism check: all reports submitted to Turnitin (similarity threshold 15%).

Glossary of Key Terms and Abbreviations

Glossary of Key Terms and Abbreviations

ADME: Absorption, Distribution, Metabolism, Excretion — the four processes governing the fate of a xenobiotic in the body.

ALARA: As Low As Reasonably Achievable — the principle that exposures to ionising radiation (and by extension, stochastic chemical hazards) should be minimised below regulatory limits.

AOP: Adverse Outcome Pathway — a structured mechanistic framework linking a Molecular Initiating Event to an Adverse Outcome through Key Events.

ARRIVE 2.0: Animal Research: Reporting of In Vivo Experiments — updated 2020 reporting guidelines for in vivo studies.

AUC: Area Under the (plasma concentration–time) Curve — a measure of total systemic exposure to a substance.

BAF / BCF / BMF: Bioaccumulation Factor / Bioconcentration Factor / Biomagnification Factor — metrics describing accumulation of substances in aquatic and terrestrial organisms.

BMD / BMDL: Benchmark Dose / Benchmark Dose Lower confidence limit — dose-response modelling approach; preferred by EFSA over NOAEL for risk assessment.

BPDE: Benzo[a]pyrene-7,8-diol-9,10-epoxide — the ultimate carcinogenic metabolite of benzo[a]pyrene, forming DNA adducts.

CYP / P450: Cytochrome P450 — superfamily of haem-containing monooxygenase enzymes primarily responsible for Phase I xenobiotic biotransformation.

DCFH-DA: 2',7'-Dichlorodihydrofluorescein diacetate — fluorescent probe for intracellular reactive oxygen species.

DPRA: Direct Peptide Reactivity Assay — in chemico assay for skin sensitisation (OECD TG 442C).

ECVAM / EURL ECVAM: European Centre for the Validation of Alternative Methods / EU Reference Laboratory for Alternatives to Animal Testing.

EDC: Endocrine-Disrupting Chemical — substance that interferes with hormone synthesis, release, transport, or action.

EOGRTS: Extended One-Generation Reproductive Toxicity Study (OECD TG 443) — comprehensive multi-generation reproduction study.

EFSA: European Food Safety Authority — EU scientific body for food chain risk assessment.

GHS: Globally Harmonised System of Classification and Labelling of Chemicals — UN framework for chemical hazard communication.

GLP: Good Laboratory Practice — OECD principles governing the organisational and operational conditions under which non-clinical safety studies are planned, performed, monitored, recorded, and reported.

HCL: Hollow Cathode Lamp — the radiation source in atomic absorption spectrometry.

IACUC: Institutional Animal Care and Use Committee — ethical review body (US/North American terminology).

IATA: Integrated Approaches to Testing and Assessment — structured frameworks combining multiple data streams to replace animal testing.

ICP-MS: Inductively Coupled Plasma — Mass Spectrometry — ultra-trace elemental analysis method.

IRIS: Integrated Risk Information System — US-EPA database of toxicological reviews and reference doses.

LAEC: Lowest Adverse Effect Concentration — equivalent of LOAEL for in vitro settings.

LC50: Median Lethal Concentration — concentration killing 50% of test organisms; used in ecotoxicology.

LD50: Median Lethal Dose — dose killing 50% of a test population; historical metric, largely replaced by limit-dose methods.

LOAEL: Lowest Observed Adverse Effect Level — lowest dose causing a statistically or biologically significant adverse effect.

LNT: Linear No-Threshold model — dose-response model for genotoxic carcinogens; assumes risk is proportional to dose with no safe threshold.

MIE: Molecular Initiating Event — the initial chemical-biological interaction in an AOP.

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide — yellow tetrazolium salt reduced by mitochondrial dehydrogenases to purple formazan; used in cytotoxicity assays.

NAM: New Approach Methodology — non-animal method for hazard and risk assessment; includes in vitro, in silico, and defined approaches.

NGRA: Next-Generation Risk Assessment — framework integrating NAMs, exposure science, and mechanistic data to replace default animal testing.

NOAEL: No Observed Adverse Effect Level — highest dose producing no statistically or biologically significant adverse effect.

OEL: Occupational Exposure Limit — maximum permissible concentration of a chemical in workplace air.

PBPK: Physiologically Based Pharmacokinetic model — mathematical model describing chemical disposition using organ volumes, blood flows, and partition coefficients.

QSAR: Quantitative Structure-Activity Relationship — mathematical model relating chemical structure to biological activity.

RfC / RfD: Reference Concentration / Reference Dose — US-EPA non-cancer safe dose benchmarks for inhalation and oral routes.

ROS: Reactive Oxygen Species — highly reactive molecules including superoxide ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($\bullet OH$).

S9 fraction: Post-mitochondrial liver supernatant used to provide metabolic activation in in vitro genotoxicity assays.

SMILES: Simplified Molecular-Input Line-Entry System — text representation of chemical structure used in QSAR software.

SOP: Standard Operating Procedure — detailed written instruction for conducting a specific operation in a GLP-compliant facility.

TDI: Tolerable Daily Intake — acceptable daily intake for non-intentionally added substances (contaminants).

TWI: Tolerable Weekly Intake — EFSA preferred metric for contaminants that accumulate slowly.

UF: Uncertainty Factor (also Safety Factor) — divisor applied to NOAEL or BMDL to derive safe exposure limits; accounts for species differences, population variability, and data quality.

USTO-MB: Université des Sciences et de la Technologie d'Oran — Mohamed Boudiaf.

vPvB: Very Persistent, very Bioaccumulative — REACH criteria identifying substances of very high concern.

WST-1: Water-soluble tetrazolium salt 1 — alternative to MTT providing a soluble formazan without a separate solubilisation step.

Additional Terms (Expanded)

ADI: Acceptable Daily Intake. The amount of a substance that can be ingested daily over a lifetime without appreciable health risk. Derived by dividing the NOAEL by a composite uncertainty factor (typically 100). Used for food additives and pesticide residues.

AED: Administered Equivalent Dose. In PBPK/IVIVE, the in vivo oral dose (mg/kg/day) predicted to produce the same internal concentration as the in vitro AC50 or IC50, accounting for absorption, distribution, metabolism, and excretion.

AOP: Adverse Outcome Pathway. A conceptual framework linking a molecular initiating event (MIE) through a series of key events (KEs) to an adverse outcome (AO) at a biologically relevant level of organisation. Documented in AOP-Wiki (OECD).

ARRIVE: Animal Research: Reporting of In Vivo Experiments. A reporting guideline (originally 2010; updated ARRIVE 2.0 in 2020, Percie du Sert et al., PLOS Biology) providing 20 items (Essential 10 + Recommended Set) to improve transparency and reproducibility of animal research publications.

BCF: Bioconcentration Factor. The ratio of a chemical's concentration in an organism to its concentration in the ambient water, at steady state. $BCF = C_{\text{organism}} / C_{\text{water}}$. Substances with $BCF > 2000$ are classified as Bioconcentrative (B) under REACH.

BER: Bioactivity Exposure Ratio. A risk metric used in NGRA: $BER = \text{in vitro AC50 (or IC50)} / \text{estimated human systemic exposure (Cmax or AUC-based)}$. $BER \gg 100$ indicates low concern; $BER < 100$ triggers further assessment.

BMD: Benchmark Dose. A statistical dose estimate corresponding to a pre-specified change (Benchmark Response, BMR) in the response relative to control, derived by fitting a dose-response model to experimental data. More information-efficient than the NOAEL.

BMDL: Benchmark Dose Lower Confidence Limit. The lower 95% confidence bound on the BMD. Used as the Point of Departure (PoD) in risk assessment to account for statistical uncertainty. $BMDL_{10} = \text{BMDL for a 10\% extra risk (BMR = 10\%)}$.

BMR: Benchmark Response. The pre-specified change in response used to derive the BMD. Common values: 5% or 10% extra risk for quantal data; 1 SD from the control mean for continuous data.

CAS: Chemical Abstracts Service Registry Number. A unique numerical identifier assigned to every chemical substance by the American Chemical Society's Chemical Abstracts Service. Used globally as the unambiguous standard chemical identifier.

CMR: Carcinogenic, Mutagenic, or toxic to Reproduction. A classification under EU CLP Regulation (1272/2008) for substances with one or more of these properties. CMR substances receive special restrictions under REACH and the Cosmetics Regulation.

DNEL: Derived No-Effect Level. In REACH, the level of exposure below which a chemical substance is considered safe for human health. Equivalent to the TDI but applicable to occupational and general population exposure scenarios; derived from the NOAEL or BMDL by applying assessment factors.

ECx: Effective Concentration causing x% effect. General notation for the concentration producing x% of the maximal effect (or x% change from control). EC_{10} = concentration producing 10% effect; EC_{50} = 50% effect (equivalent to IC_{50} for inhibitory endpoints).

EC50: Effective Concentration 50%. The concentration producing 50% of the maximal response. For environmental endpoints (OECD TG 202 Daphnia), EC_{50} refers to the concentration immobilising 50% of the test organisms.

ED10: Effective Dose 10%. In vivo dose producing 10% extra cancer incidence or effect, derived from dose-response modelling. Used as the BMD for a BMR of 10%.

ELCR: Excess Lifetime Cancer Risk. The probability of additional cancer cases expected in a population from exposure to a carcinogen over a lifetime, above the background rate. $ELCR = CDI \times SF$ (slope factor). Regulatory acceptable range: 10^{-6} to 10^{-4} .

EOGRTS: Extended One-Generation Reproductive Toxicity Study. OECD TG 443. Replaces the two-generation reproductive toxicity study (TG 416) for most REACH substances. Includes optional cohorts for developmental neurotoxicity (Cohort 2A) and immunotoxicity (Cohort 2B).

GHS: Globally Harmonised System of Classification and Labelling of Chemicals. A UN system (Purple Book, Revision 9, 2021) for classifying chemicals based on their physical, health, and environmental hazards, and communicating this information via standardised labels and Safety Data Sheets (SDS).

GLP: Good Laboratory Practice. A quality management system for the planning, performance, monitoring, recording, archiving, and reporting of non-clinical laboratory studies. OECD GLP Principles (ENV/MC/CHEM(98)17) are the international standard for regulatory toxicity studies.

HBM: Human Biomonitoring. The measurement of chemical substances or their metabolites in human biological media (blood, urine, hair, breast milk) to estimate internal exposure. HBM4EU (EU project, 2017–2022) established Human Biomonitoring Guidance Values (HBM-I and HBM-II) for 24 priority substances.

IATA: Integrated Approaches to Testing and Assessment. A structured framework combining multiple information sources (existing data, in silico, in vitro, in vivo, exposure) to make a regulatory hazard or risk determination. Differs from traditional approaches by explicitly integrating multiple tiers of evidence.

IC50: Inhibitory Concentration 50%. The concentration of a test substance that produces 50% inhibition of a measured endpoint (e.g., cell viability, enzyme activity, receptor binding). Analogous to EC50 for inhibitory responses.

ILAR: Institute for Laboratory Animal Research. US organisation (National Academies of Sciences, Engineering, and Medicine) responsible for the Guide for the Care and Use of Laboratory Animals (8th edition, 2011) — the standard reference for US animal facility management.

IPCS: International Programme on Chemical Safety. A joint WHO/UNEP/ILO programme responsible for producing authoritative evaluations of chemical safety (Environmental Health Criteria documents) and internationally harmonised terminology for toxicology and risk assessment.

KER: Key Event Relationship. A causal relationship between two key events (KEs) in an AOP, describing the biological, mechanistic, or empirical basis for the causal link. KERs are characterised for essentiality, biological plausibility, and empirical support.

KE: Key Event. A measurable, biologically plausible change in a biological system (at the molecular, cellular, tissue, organ, or organism level) that is causally related to the MIE or the AO in an AOP.

LD50: Lethal Dose 50%. The dose of a substance causing death in 50% of a test population under specified conditions. A classical measure of acute toxicity. The LD_{50} is derived by dose-response modelling (probit, logit). Units: mg/kg body weight.

LOAEL: Lowest Observed Adverse Effect Level. The lowest dose in an experimental study at which a statistically or biologically significant adverse effect is observed compared to controls. Paired with the NOAEL.

LOEC: Lowest Observed Effect Concentration. The lowest concentration at which a statistically significant effect is observed in an ecotoxicological test. Used when the endpoint is a continuous measurement in aquatic toxicology.

MIE: Molecular Initiating Event. The first chemical interaction between a stressor and a biomolecule that initiates an AOP. Examples: receptor binding ($ER\alpha$ agonism), enzyme inhibition (AChE), covalent protein modification (GSH depletion), DNA adduct formation.

MoE: Margin of Exposure. A risk metric for genotoxic carcinogens: $\text{MoE} = \text{BMDL}_{10} / \text{estimated human exposure}$. $\text{MoE} \geq 10,000$ is considered low concern (EFSA 2005 guidance). Avoids explicit numerical cancer risk extrapolation.

MTD: Maximum Tolerated Dose. The highest dose that does not cause severe toxicity ($\leq 10\%$ body weight loss, no mortality) in a chronic study. The MTD is used as the high dose in carcinogenicity studies (OECD TG 451/453) to ensure the study is conducted at the limit of biological tolerance.

NAM: New Approach Methodology. An umbrella term for any non-animal technology or methodology used in chemical safety assessment, including in vitro assays, organ-on-chip, QSAR/PBPK models, omics approaches, and high-throughput screening.

NOAEC: No Observed Adverse Effect Concentration. The highest tested concentration in an inhalation study at which no statistically or biologically significant adverse effect is observed. Analogous to NOAEL for inhalation exposure scenarios.

NOAEL: No Observed Adverse Effect Level. The highest dose in an experimental study at which no statistically or biologically significant adverse effect is observed. The NOAEL is divided by uncertainty factors to derive safe doses (TDI, RfD, ADI) for regulatory risk assessment.

NOEC: No Observed Effect Concentration. The highest concentration in an ecotoxicological test at which no statistically significant effect is observed relative to controls. Used in aquatic toxicology.

NOEL: No Observed Effect Level. The highest dose at which no effect (adverse or non-adverse) is observed. The NOEL may be lower than the NOAEL when non-adverse adaptive changes are detected at the same dose level as the NOAEL.

OECD: Organisation for Economic Co-operation and Development. An intergovernmental organisation (38 member countries) that develops internationally harmonised test guidelines (OECD TGs) and guidance documents for chemical safety testing, as well as the Mutual Acceptance of Data (MAD) framework that ensures regulatory acceptance of GLP-compliant studies across member countries.

PBPK: Physiologically Based Pharmacokinetic (Model). A mathematical model of the body that represents organs and tissues as compartments with anatomically realistic parameters (blood flows, volumes) and biochemical kinetic constants. Used for IVIVE, species extrapolation, and route-to-route extrapolation in risk assessment.

PNEC: Predicted No Effect Concentration. In environmental risk assessment, the concentration of a substance below which adverse effects on the ecosystem are not expected. Derived from the lowest NOEC or EC10 in ecotoxicological tests divided by an assessment factor ($\text{AF} = 10\text{--}1000$).

PoD: Point of Departure. The dose or concentration used as the starting point for risk characterisation. May be a NOAEL, LOAEL, BMDL10, or benchmark concentration (BMC). The PoD is divided by assessment/uncertainty factors to derive the safe dose/concentration.

PPE: Personal Protective Equipment. Protective clothing and equipment worn to minimise exposure to chemical, biological, physical, or other hazards in the laboratory. Standard PPE in toxicology laboratories: nitrile gloves, lab coat, safety glasses (impact-resistant), closed-toe shoes. Additional PPE for specific hazards: face shield, N-FFP3 mask (volatile or dusty compounds), splash-proof apron.

QIVIVE: Quantitative In Vitro to In Vivo Extrapolation. A quantitative approach to translate in vitro bioactive concentrations (AC_{50} , IC_{50}) to equivalent in vivo exposure levels using PBPK modelling or IVIVE methods. See also AED, IVIVE.

QSAR: Quantitative Structure-Activity Relationship. A mathematical or statistical relationship between chemical structural descriptors (calculated from molecular structure) and a biological activity or physico-chemical property. Used for regulatory data gap-filling under REACH Annex XI and in pharmaceutical lead optimisation.

RfC: Reference Concentration. The US EPA equivalent of the NOAEC-based safe inhalation concentration for humans. Derived by dividing the NOAEC (inhalation study in animals or humans) by relevant uncertainty factors.

RfD: Reference Dose. The US EPA's estimate of the daily oral exposure to a human population that is unlikely to cause appreciable risk of adverse effects over a lifetime. Derived by dividing the oral NOAEL by composite uncertainty factors. Analogous to the TDI/ADI used by WHO/EFSA.

ROS: Reactive Oxygen Species. Short-lived, chemically reactive molecules containing oxygen, including superoxide ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), and the highly reactive hydroxyl radical ($\bullet OH$). ROS are generated as by-products of mitochondrial electron transport and by NADPH oxidase; they mediate oxidative stress.

SAR: Structure-Activity Relationship. A qualitative assessment of the relationship between the structure of a chemical and its biological activity. SAR identifies structural alerts — molecular features associated with specific toxicological mechanisms.

SOP: Standard Operating Procedure. A written document providing detailed, step-by-step instructions for performing a specific procedure in a GLP-compliant facility. SOPs ensure consistency, reproducibility, and traceability of laboratory operations.

STD10: Severely Toxic Dose 10%. The dose causing severe toxicity in 10% of animals. Used in oncology Phase I trials as a starting dose reference; less commonly used in regulatory toxicology than the MTD.

STOT-RE: Specific Target Organ Toxicity — Repeated Exposure. A GHS/CLP hazard category (Category 1: proven; Category 2: presumed) for substances causing specific organ damage following repeated exposure. Applies when the target organ can be identified and the effect is documented at repeated doses.

STOT-SE: Specific Target Organ Toxicity — Single Exposure. A GHS/CLP hazard category for substances causing specific reversible or irreversible organ damage following a single exposure. Category 1 (H370) or Category 2 (H371), or Category 3 for narcotic/respiratory irritation (H336).

T25: The chronic dose producing tumours in 25% of animals above the spontaneous rate. An alternative cancer potency measure used in EU risk assessment, derived from long-term animal carcinogenicity studies. $T25 = \text{dose} / \log(1 / (1 - 0.25 + P_{\text{background}}))$.

TDI: Tolerable Daily Intake. The amount of a substance (typically a contaminant or food additive) that can be ingested daily over a lifetime without appreciable risk to health. Derived by dividing the NOAEL or BMDL by safety/uncertainty factors. Used by EFSA and JECFA.

TG: Test Guideline. A standardised protocol adopted by the OECD (or other bodies) for conducting specific toxicity tests. Examples: OECD TG 407 (28-day oral rodent study), TG 471 (Ames test), TG 425 (up-and-down procedure).

TK: Toxicokinetics. The mathematical description of the absorption, distribution, metabolism, and excretion (ADME) of a toxic substance in the body, with specific focus on the concentration-time profile at the target organ and its relationship to the toxic response.

TKTD: Toxicokinetic-Toxicodynamic (Model). A mathematical model that explicitly links the internal concentration of a toxicant over time (TK component) to the biological response (TD component). Used in environmental risk assessment for aquatic organisms (e.g., GUTS model for daphnia).

TTC: Threshold of Toxicological Concern. A practical tool for prioritising risk assessment: if the estimated human exposure to a substance of unknown toxicity is below a substance-class-specific TTC value (Cramer Class I: 1.8 $\mu\text{g/day}$; Class II: 0.54 $\mu\text{g/day}$; Class III: 0.18 $\mu\text{g/day}$), the risk is considered negligible without additional testing.

UF: Uncertainty Factor (also Assessment Factor). A factor applied to the NOAEL (or BMDL) to account for uncertainties in extrapolating experimental data to human health protection. Standard composite UF = 100 (UF-interspecies $\times$ UF-intraspecies = 10×10).

ULOQ: Upper Limit of Quantification. The highest concentration that can be reliably and accurately quantified by an analytical method within validated precision and accuracy limits. Samples above ULOQ must be diluted and re-analysed.

VEGA: Virtual models for property Evaluation of chemicals within a Global Architecture. A free software platform (IRFMN, Italy) providing QSAR models for multiple toxicological endpoints, with integrated applicability domain assessment and automatic generation of ECHA-format QMRF documents.

Vd: Volume of Distribution. A theoretical volume that relates the total amount of a substance in the body to its plasma concentration: $Vd = \text{Dose} / C_0$ (after IV injection). High Vd ($>1 \text{ L/kg}$) indicates extensive tissue distribution; low Vd ($<0.1 \text{ L/kg}$) indicates confinement to plasma.

Weight of Evidence: A systematic process of evaluating the strength, consistency, and relevance of data from multiple sources to reach a conclusion about the toxicological properties of a substance. WoE assessment considers data quantity, quality, relevance, and consistency across study types and species.

Bibliography

Bibliography

All references are cited in APA 7th edition format. DOIs verified against the CrossRef registry. References are listed in order of first citation in the text. For OECD Test Guidelines, the official OECD iLibrary URL is provided.

1. Michaleas, S. N., Laios, K., Tsoucalas, G., & Androutsos, G. (2021). Theophrastus Bombastus Von Hohenheim (Paracelsus) (1493–1541): The eminent physician and pioneer of toxicology. *Toxicology Reports*, 8, 411–414. <https://doi.org/10.1016/j.toxrep.2021.02.012>
2. Nepovimova, E., & Kuca, K. (2019). The history of poisoning: from ancient times until modern era. *Archives of Toxicology*, 93(1), 11–24. <https://doi.org/10.1007/s00204-018-2290-0>
3. Watson, W. A. (2020). Poisoning crimes and forensic toxicology since the 18th century. *Academic Forensic Pathology*, 10(1), 35–46. <https://doi.org/10.1177/1925362120937923>
4. Pognan, F., Beilmann, M., Boonen, H. C. M., et al. (2023). The evolving role of investigative toxicology in the pharmaceutical industry. *Nature Reviews Drug Discovery*, 22(4), 317–335. <https://doi.org/10.1038/s41573-022-00633-x>
5. Herzler, M., Marx-Stoelting, P., Pirow, R., et al. (2021). The EU chemicals strategy for sustainability questions regulatory toxicology as we know it. *Archives of Toxicology*, 95(7), 2589–2601. <https://doi.org/10.1007/s00204-021-03091-3>
6. Ball, N., Bars, R., Botham, P. A., et al. (2022). A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Archives of Toxicology*, 96(3), 743–766. <https://doi.org/10.1007/s00204-021-03215-9>
7. Worth, A. P., & Berggren, E. (2023). Towards a future regulatory framework for chemicals in the European Union — Chemicals 2.0. *Regulatory Toxicology and Pharmacology*, 142, 105431. <https://doi.org/10.1016/j.yrtph.2023.105431>
8. Van Ravenzwaay, B., Westmoreland, C., Funk-Weyer, D., et al. (2023). Analysis of health concerns not addressed by REACH for low tonnage chemicals and opportunities for new approach methodology. *Archives of Toxicology*, 97(12), 3075–3083. <https://doi.org/10.1007/s00204-023-03601-5>
9. Zhao, X., Ma, J., Li, M., et al. (2021). Cytochrome P450 enzymes and drug metabolism in humans. *International Journal of Molecular Sciences*, 22(23), 12808. <https://doi.org/10.3390/ijms222312808>
10. Cho, S., Jo, H., Hwang, Y., et al. (2024). Potential impact of underlying diseases influencing ADME in nonclinical safety assessment. *Food and Chemical Toxicology*, 188, 114636. <https://doi.org/10.1016/j.fct.2024.114636>
11. Sun, Z., Shan, Y., Cao, X., et al. (2024). Unveiling the hidden dangers: a review of non-apoptotic programmed cell death in chemical-induced toxicology. *Cell Biology and Toxicology*, 40(1), 63. <https://doi.org/10.1007/s10565-024-09895-0>
12. Grandjean, P. (2016). Paracelsus revisited: The dose concept in a complex world. *Basic & Clinical Pharmacology & Toxicology*, 119(2), 126–132. <https://doi.org/10.1111/bcpt.12622>
13. Kappenberg, F., Duda, J., Schürmeyer, L., et al. (2023). Guidance for statistical design and analysis of toxicological dose–response experiments. *Archives of Toxicology*, 97(10), 2741–2761. <https://doi.org/10.1007/s00204-023-03561-w>
14. Haber, L. T., Dourson, M. L., Allen, B. C., et al. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Critical Reviews in Toxicology*, 48(5), 387–415. <https://doi.org/10.1080/10408444.2018.1430121>
15. Calabrese, E. J. (2022). Linear non-threshold (LNT) fails numerous toxicological stress tests. *Chemico-Biological Interactions*, 365, 110064. <https://doi.org/10.1016/j.cbi.2022.110064>
16. Wan, Y., Liu, J., Mai, Y., et al. (2024). Current advances and future trends of hormesis in disease. *npj Aging*, 10(1), 26. <https://doi.org/10.1038/s41514-024-00155-3>

17. Lima, V. A., Santos, A. B. N., Morais, E. E. G., & Soares, A. F. (2024). Zebrafish: A versatile model organism for toxicological research and applications. *Advances in Research*, 25(5), 387–396. <https://doi.org/10.9734/air/2024/v25i51170>
18. Gonçalves, I. F. S., Souza, T. M., Vieira, L. R., et al. (2020). Toxicity testing of pesticides in zebrafish — a systematic review on chemicals and associated toxicological endpoints. *Environmental Science and Pollution Research*, 27(10), 10185–10204. <https://doi.org/10.1007/s11356-020-07902-5>
19. Fogliano, C., Pallotta, M. M., Tussellino, M., & Carotenuto, R. (2023). *Xenopus laevis* as a model organism for bioscience. *Biology*, 12(6), 890. <https://doi.org/10.3390/biology12060890>
20. Štampar, M., & Žegura, B. (2024). In vitro hepatic 3D cell models and their application in genetic toxicology: A systematic review. *Mutation Research — Genetic Toxicology and Environmental Mutagenesis*, 900, 503835. <https://doi.org/10.1016/j.mrgentox.2024.503835>
21. Yang, S., Ooka, M., Margolis, R. J., & Xia, M. (2023). Liver three-dimensional cellular models for high-throughput chemical testing. *Cell Reports Methods*, 3(3), 100432. <https://doi.org/10.1016/j.crmeth.2023.100432>
22. Shao, W., Xu, H., Zeng, K., et al. (2025). Advances in liver organoids: replicating hepatic complexity for toxicity assessment and disease modeling. *Stem Cell Research & Therapy*, 16(1). <https://doi.org/10.1186/s13287-025-04139-2>
23. Lőrincz, T., Deák, V., Makk-Merczel, K., et al. (2021). The performance of HepG2 and HepaRG systems through the glass of acetaminophen-induced toxicity. *Life*, 11(8), 856. <https://doi.org/10.3390/life11080856>
24. Deepika, D., & Kumar, V. (2023). The role of PBPK new approach methodology in pharmaceuticals and environmental chemical risk assessment. *International Journal of Environmental Research and Public Health*, 20(4), 3473. <https://doi.org/10.3390/ijerph20043473>
25. Lin, Z., & Chou, W. C. (2022). Machine learning and artificial intelligence in toxicological sciences. *Toxicological Sciences*, 189(1), 7–19. <https://doi.org/10.1093/toxsci/kfac075>
26. Lizarraga, L. E., Suter, G. W., Lambert, J. C., & Patlewicz, G. (2023). Advancing the science of a read-across framework for evaluation of data-poor chemicals. *Regulatory Toxicology and Pharmacology*, 137, 105293. <https://doi.org/10.1016/j.yrtph.2022.105293>
27. Lancia, P., Louazzani, M., Gros, L., et al. (2025). Overview of in silico tools to evaluate human health toxicity in the hazard assessment of chemicals used in cosmetics. *Chemical Research in Toxicology*, 38(10), 1652–1680. <https://doi.org/10.1021/acs.chemrestox.4c00534>
28. Burden, N., Benstead, R., Benyon, K., et al. (2020). Key opportunities to replace, reduce, and refine regulatory fish acute toxicity tests. *Environmental Toxicology and Chemistry*, 39(10), 2076–2089. <https://doi.org/10.1002/etc.4824>
29. Mitchell, C. A., Burden, N., Bonnell, M., et al. (2023). New approach methodologies for the endocrine activity toolbox. *Environmental Toxicology and Chemistry*, 42(4), 757–777. <https://doi.org/10.1002/etc.5584>
30. Karmaus, A. L., Mansouri, K., To, K., et al. (2022). Evaluation of variability across rat acute oral systemic toxicity studies. *Toxicological Sciences*, 188(1), 34–47. <https://doi.org/10.1093/toxsci/kfac042>
31. Sewell, F., Corvaro, M., Andrus, A., et al. (2022). Recommendations on dose level selection for repeat dose toxicity studies. *Archives of Toxicology*, 96(7), 1921–1934. <https://doi.org/10.1007/s00204-022-03293-3>
32. Takeshita, J., Goto, Y., Yamamoto, S., et al. (2024). Comprehensive analysis of toxicity-related findings from repeated-dose subacute toxicity studies. *Critical Reviews in Toxicology*, 54(10), 996–1010. <https://doi.org/10.1080/10408444.2024.2427221>

33. Luconi, M., Sogorb, M. A., Markert, U. R., et al. (2022). Human-based new approach methodologies in developmental toxicity testing. *International Journal of Environmental Research and Public Health*, 19(23), 15828. <https://doi.org/10.3390/ijerph192315828>
34. Miller, L. B., Feuz, M. B., Meyer, R. G., & Meyer-Ficca, M. L. (2024). Reproductive toxicology: keeping up with our changing world. *Frontiers in Toxicology*, 6. <https://doi.org/10.3389/ftox.2024.1456687>
35. Jacobs, M. N., Colacci, A., Corvi, R., et al. (2020). Chemical carcinogen safety testing: OECD expert group international consensus. *Archives of Toxicology*, 94(8), 2899–2923. <https://doi.org/10.1007/s00204-020-02784-5>
36. Cross, K. P., & DeMarini, D. M. (2023). Analysis of chemical structures and mutations detected by Salmonella TA98 and TA100. *Mutation Research — Fundamental and Molecular Mechanisms of Mutagenesis*, 827, 111838. <https://doi.org/10.1016/j.mrfmmm.2023.111838>
37. Burgum, M. J., Ulrich, C., Partosa, N., et al. (2024). Adapting the in vitro micronucleus assay (OECD TG 487) for testing of manufactured nanomaterials. *Mutagenesis*, 39(3), 205–217. <https://doi.org/10.1093/mutage/geae010>
38. Medhi, B., Kapoor, E., Sharma, N., & Joshi, R. (2021). Impact of OECD principles of Good Laboratory Practices in India. *Indian Journal of Pharmacology*, 53(5), 353–358. https://doi.org/10.4103/ijp.ijp_685_21
39. Percie du Sert, N., Hurst, V., Ahluwalia, A., et al. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biology*, 18(7), e3000410. <https://doi.org/10.1371/journal.pbio.3000410>
40. Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
41. Dirnagl, U., & Lauritzsen, M. (2023). Rigor and reproducibility in animal research. *Frontiers in Neuroscience*, 17, 1172271. <https://doi.org/10.3389/fnins.2023.1172271>
42. Kojima, H. (2025). History of OECD test guidelines for non-animal test methods in Japan. *Genes and Environment*, 47(1). <https://doi.org/10.1186/s41021-024-00323-7>
43. Pallocca, G., & Leist, M. (2022). On the usefulness of animals as a model system (Part II): Considering benefits alongside limitations. *ALTEX — Alternatives to Animal Experimentation*, 39(3), 531–539. <https://doi.org/10.14573/altex.2207111>
44. Hartung, T. (2020). Making big sense from big data in toxicology by read-across. *ALTEX*, 33(2), 83–93. <https://doi.org/10.14573/altex.1603091>
45. Slob, W., Bakker, M. I., Bokkers, B. G. H., & Chen, G. (2025). The use of canonical dose-response models for benchmark dose analysis of continuous endpoints. *Critical Reviews in Toxicology*, 55(4), 437–461. <https://doi.org/10.1080/10408444.2025.2464067>
46. Caldwell, J. C., Laskey, J. W., Heidenreich, W. F., & Thomas, R. S. (2022). Cancer risk estimates from low-dose ionising radiation exposures. *Radiation and Environmental Biophysics*, 61(4), 479–491. <https://doi.org/10.1007/s00411-022-00990-4>
47. IARC Working Group. (2023). Agents classified by the IARC Monographs: Volumes 1–135. International Agency for Research on Cancer. <https://monographs.iarc.who.int/agents-classified-by-the-iarc/>
48. Santana-Mayor, A., Rodríguez-Ramos, R., Herrera-Herrera, A. V., et al. (2023). Updated overview of QuEChERS applications in food, environmental and biological analysis (2020–2023). *TrAC Trends in Analytical Chemistry*, 169, 117375. <https://doi.org/10.1016/j.trac.2023.117375>
49. He, S., Niu, Y., Xing, L., et al. (2024). Research progress of the detection and analysis methods of heavy metals in plants. *Frontiers in Plant Science*, 15, 1310328. <https://doi.org/10.3389/fpls.2024.1310328>

- 50.** OECD. (2008). Test No. 405: Acute Eye Irritation/Corrosion. OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing. <https://doi.org/10.1787/9789264070646-en>
- 51.** OECD. (2008). Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents. OECD Publishing. <https://doi.org/10.1787/9789264070684-en>
- 52.** OECD. (2018). Test No. 408: Repeated Dose 90-day Oral Toxicity Study in Rodents. OECD Publishing. <https://doi.org/10.1787/9789264070707-en>
- 53.** OECD. (2008). Test No. 414: Prenatal Developmental Toxicity Study. OECD Publishing. <https://doi.org/10.1787/9789264070820-en>
- 54.** OECD. (2001). Test No. 420: Acute Oral Toxicity — Fixed Dose Procedure. OECD Publishing. <https://doi.org/10.1787/9789264070943-en>
- 55.** OECD. (2001). Test No. 423: Acute Oral Toxicity — Acute Toxic Class Method. OECD Publishing. <https://doi.org/10.1787/9789264071001-en>
- 56.** OECD. (2008). Test No. 425: Acute Oral Toxicity — Up-and-Down Procedure. OECD Publishing. <https://doi.org/10.1787/9789264071049-en>
- 57.** OECD. (2020). Test No. 471: Bacterial Reverse Mutation Test. OECD Publishing. <https://doi.org/10.1787/9789264071247-en>
- 58.** OECD. (2016). Test No. 487: In Vitro Mammalian Cell Micronucleus Test. OECD Publishing. <https://doi.org/10.1787/9789264264861-en>
- 59.** OECD. (2023). Test No. 492: Reconstructed Human Cornea-like Epithelium (RhCE) Test Method for Identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Publishing. <https://doi.org/10.1787/9789264242548-en>
- 60.** European Parliament and Council. (2010). Directive 2010/63/EU on the protection of animals used for scientific purposes. Official Journal of the European Union, L 276, 33–79.
- 61.** Russell, W. M. S., & Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Methuen.
- 62.** National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. National Academies Press. <https://doi.org/10.17226/11970>