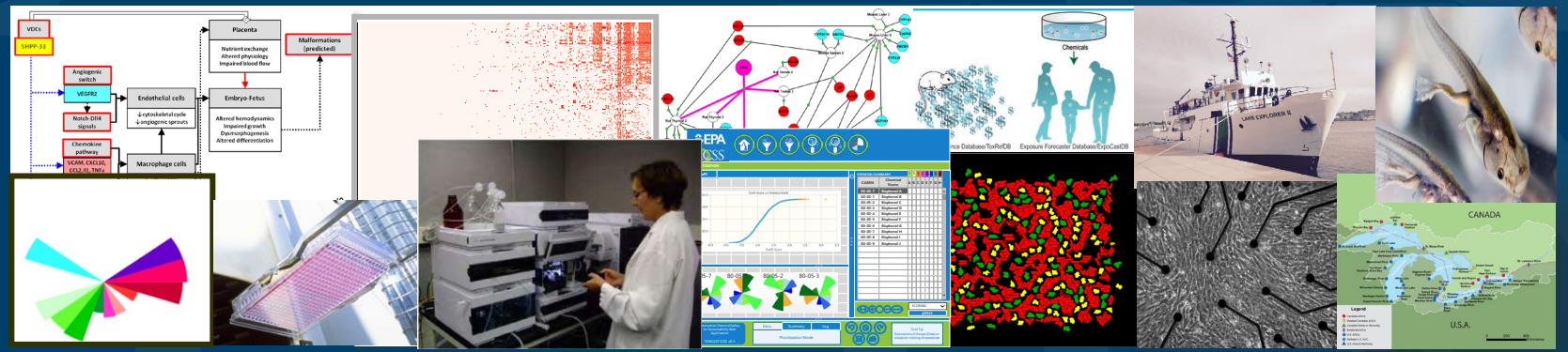
EPA Overview for the NAS Committee on Variability and Relevance of Current Laboratory Mammalian Toxicity Tests and Expectations for NAMs for use in Human Health Risk Assessment



NAS Committee Public Meeting

July 28, 2022

Rusty Thomas
Director
Center for Computational Toxicology and Exposure

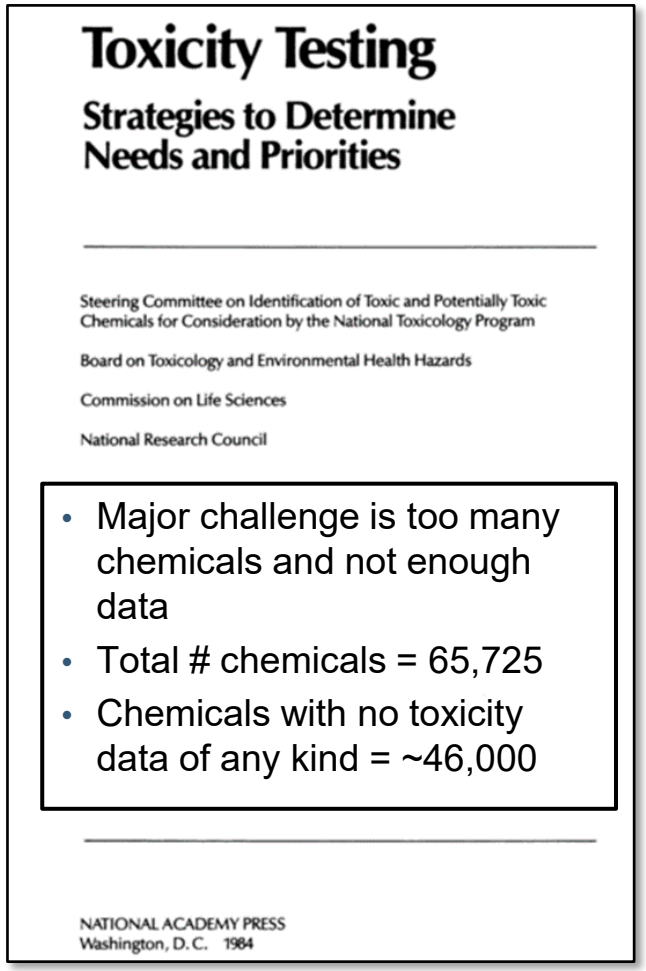
The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. EPA

Key Components in the NAS Committee Statement of Task

- Comprehensive literature review on the variability and human relevance of current laboratory mammalian toxicity tests as well as approaches to validation and establishing scientific confidence in using NAMs.
  - The variability considered in terms of reliability, qualitative and quantitative reproducibility
  - Relevance considered in terms of biological relevance and overall concordance of the results in humans.
- Recommendations on:
  - Variability of laboratory mammalian toxicity tests and concordance with human adverse responses.
  - How the variability in traditional mammalian toxicity test results and concordance with adverse effects in humans can be used to inform benchmarks in evaluating the scientific quality of NAMs.
  - Key components in a fit-for-purpose validation paradigm or scientific confidence framework for NAMs where there is no existing standard test, the standard test is not relevant to the human response, or the standard test has not been benchmarked against human responses.

The Lack of Toxicity Testing Data is a Historical and Current Challenge

1984 NAS Report



Toxicity Testing

Strategies to Determine Needs and Priorities

Steering Committee on Identification of Toxic and Potentially Toxic Chemicals for Consideration by the National Toxicology Program

Board on Toxicology and Environmental Health Hazards

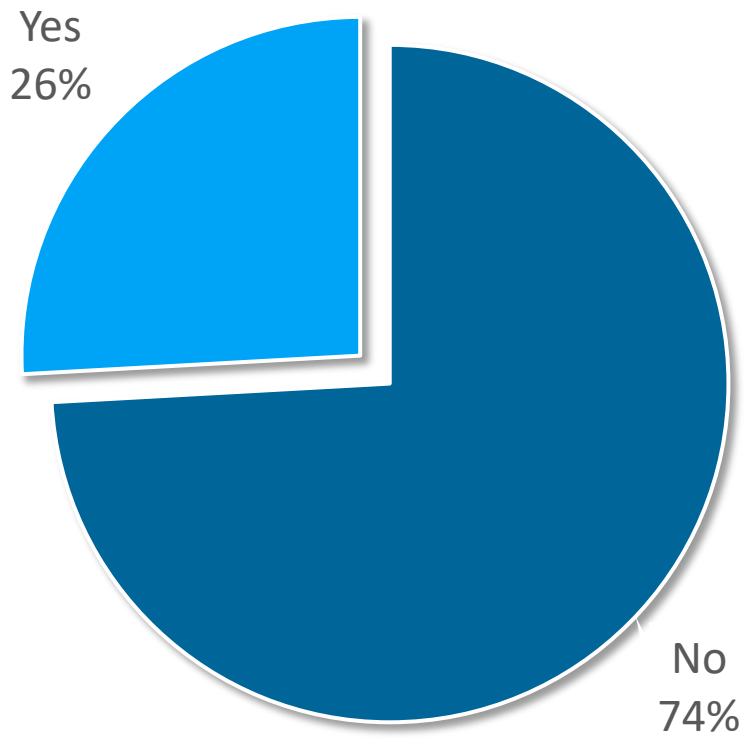
Commission on Life Sciences

National Research Council

- Major challenge is too many chemicals and not enough data
- Total # chemicals = 65,725
- Chemicals with no toxicity data of any kind = ~46,000

NATIONAL ACADEMY PRESS
Washington, D.C. 1984

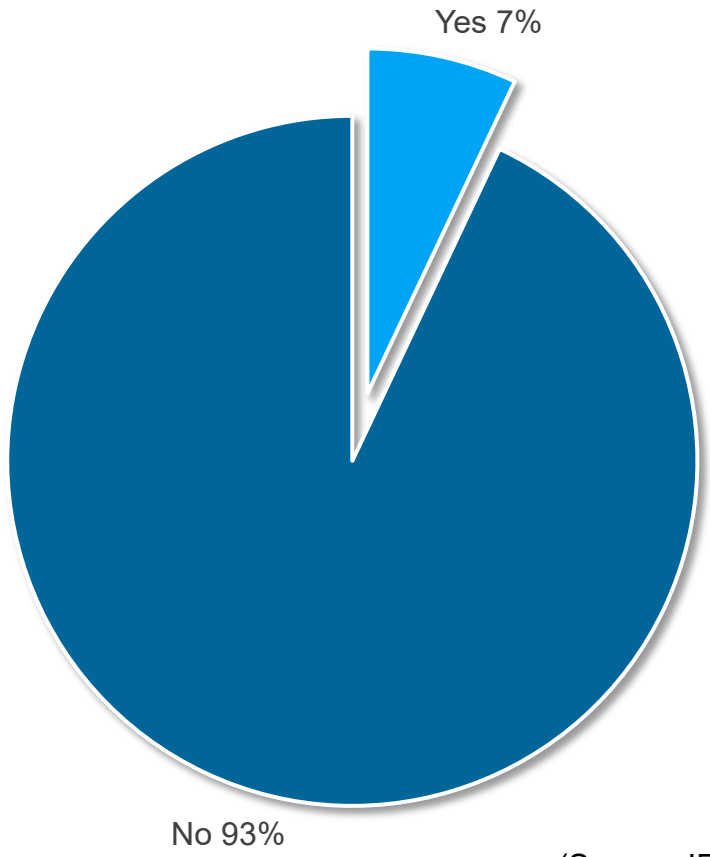
Percentage of Non-Confidential Active TSCA Inventory with Repeat Dose Toxicity Study in 2019



(Source: ToxValDB, 2019)

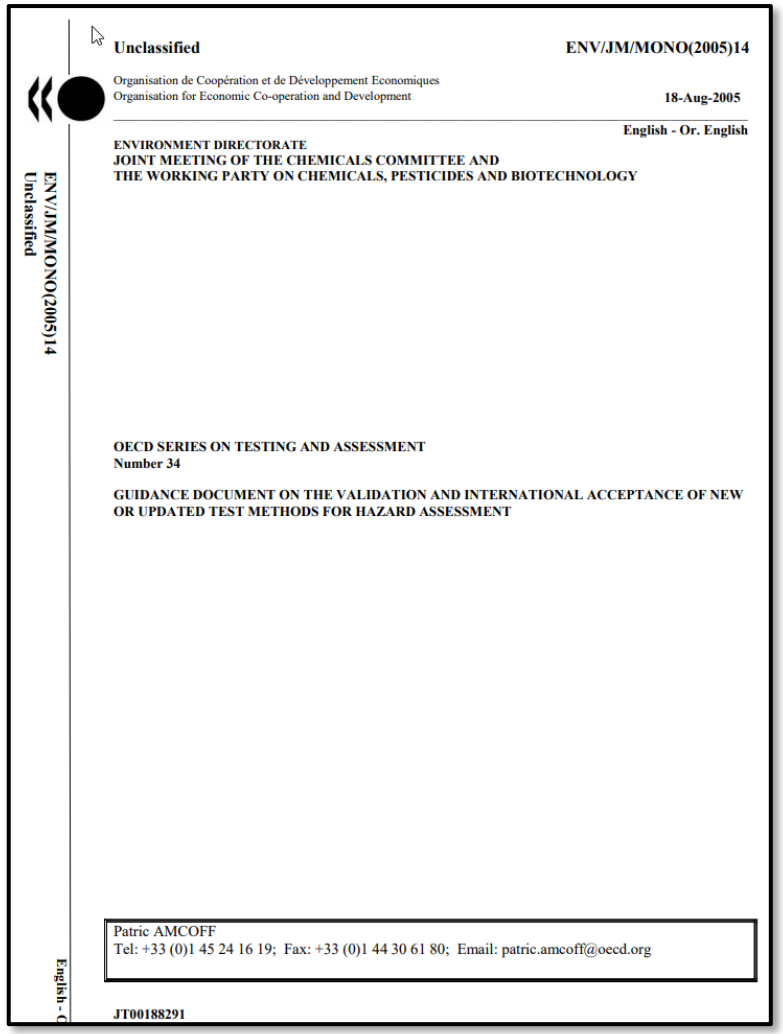
Current Chemical Assessments Frequently Rely on Mammalian *In Vivo* Toxicity Tests

Percentage of IRIS Non-Cancer Chemical Assessments with Human-Derived Critical Effect



(Source: IRIS, 2022)

Understanding the Reliability and Relevance is a Practical and Statutorial Requirement



Unclassified ENV/JM/MONO(2005)14

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

18-Aug-2005

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/MONO(2005)14
Unclassified

OECD SERIES ON TESTING AND ASSESSMENT
Number 34

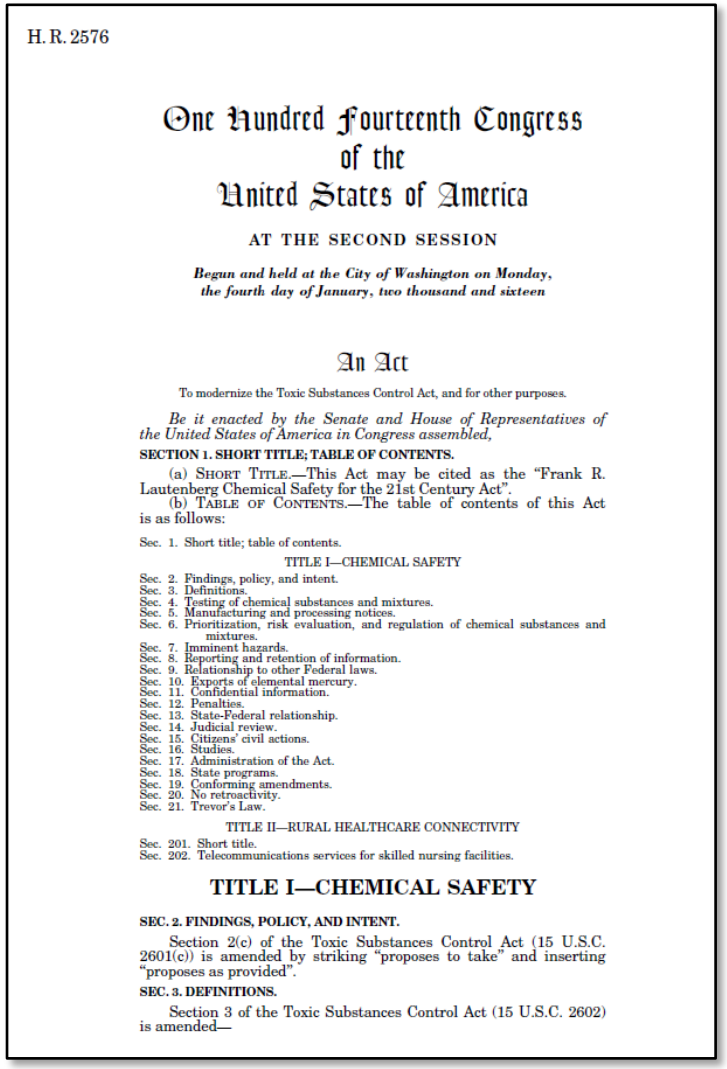
GUIDANCE DOCUMENT ON THE VALIDATION AND INTERNATIONAL ACCEPTANCE OF NEW OR UPDATED TEST METHODS FOR HAZARD ASSESSMENT

Patric AMCOFF
Tel: +33 (0)1 45 24 16 19; Fax: +33 (0)1 44 30 61 80; Email: patric.amcoff@oecd.org

JT00188291

- OECD definition of "validation" as a *"process based on scientifically sound principles (5)(6) by which the reliability and relevance of a particular test, approach, method, or process are established for a specific purpose. Reliability is defined as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardised protocol. The relevance of a test method describes the relationship between the test and the effect in the target species and whether the test method is meaningful and useful for a defined purpose, with the limitations identified."*
- But… also suggested that "*the validation process should be flexible and adaptable*", performance must be "*demonstrated using a series of reference chemicals*", and "*evaluated in relation to existing relevant toxicity data*."

Understanding the Reliability and Relevance is a Practical and Statutorial Requirement



H. R. 2576

One Hundred Fourteenth Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Monday, the fourth day of January, two thousand and sixteen

An Act

To modernize the Toxic Substances Control Act, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Frank R. Lautenberg Chemical Safety for the 21st Century Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

TITLE I—CHEMICAL SAFETY

SEC. 2. FINDINGS, POLICY, AND INTENT.

Section 2(c) of the Toxic Substances Control Act (15 U.S.C. 2601(c)) is amended by striking "proposes to take" and inserting "proposes as provided".

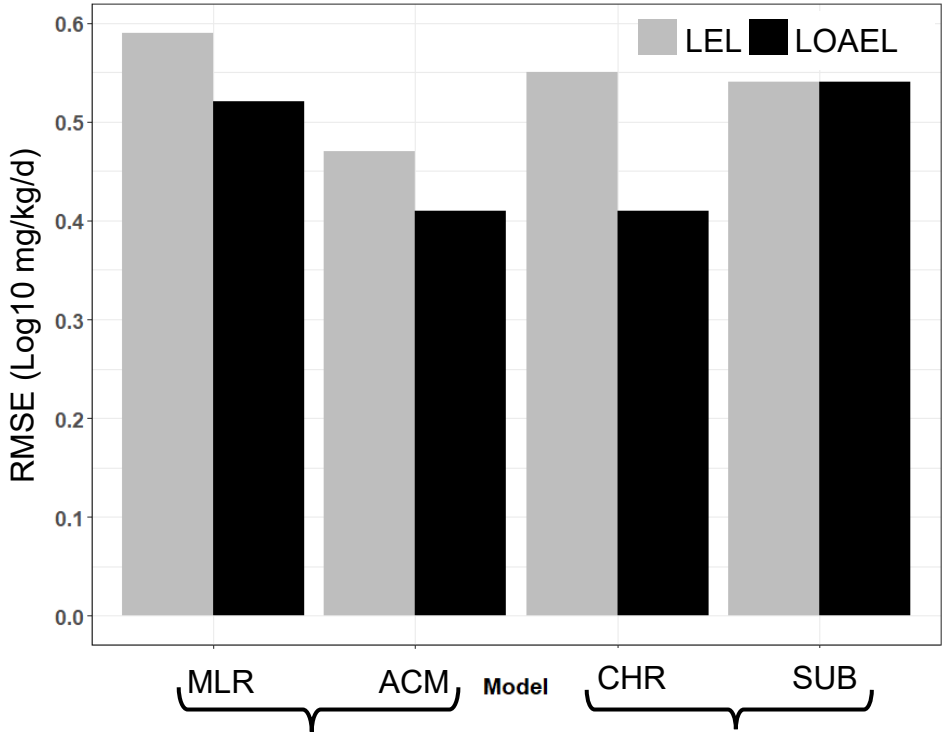
SEC. 3. DEFINITIONS.

Section 3 of the Toxic Substances Control Act (15 U.S.C. 2602) is amended—

- Section 4(h) in amended TSCA law requires the identification of – "…alternative test methods or strategies the Administrator has identified that do not require new vertebrate animal testing and are scientifically reliable, relevant, and capable of providing information of equivalent or better scientific reliability and quality to that which would be obtained from vertebrate animal testing"
- In order to evaluate alternative test methods or strategies based on this standard, EPA must characterize the reliability and relevance of vertebrate animal testing models.

EPA Has Been Using Highly Curated Databases of Legacy Toxicity Studies to Evaluate Variability

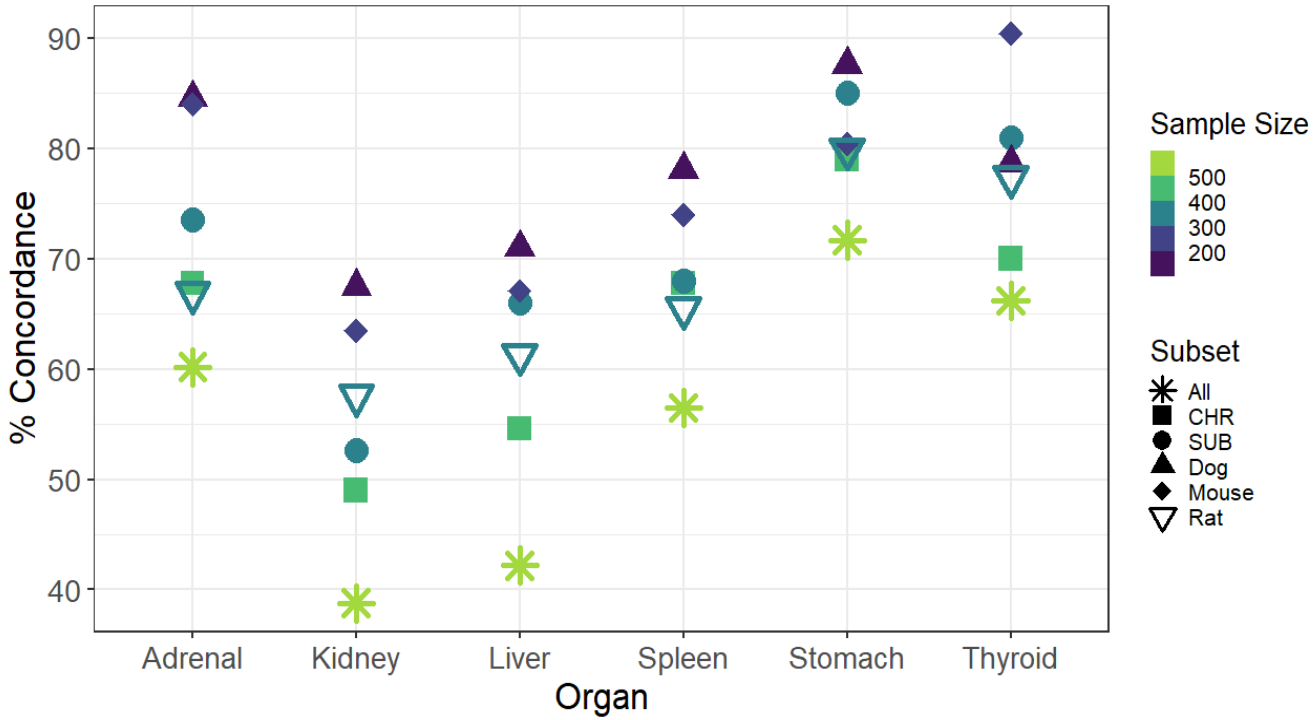
Evaluating LEL/LOAEL Variability in Traditional Toxicity Studies to Set Expectations for NAMs



Using an RMSE=0.59, the 95% Prediction Interval of an LEL/LOAEL is +/- 10-fold (e.g., 1 mg/kg/day, 0.07 – 14)

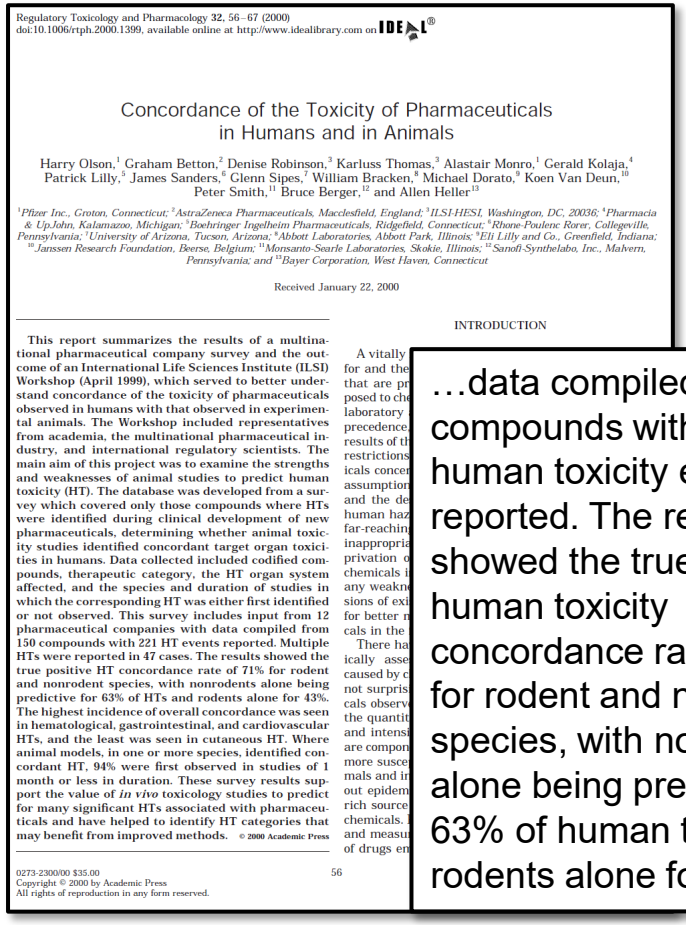
Pham et al., Comp Toxicol., 2020

Evaluating Qualitative Concordance of Organ Toxicity



Paul-Friedman, Unpublished

Studies in Drug Development Provide an Estimate of Concordance Between *In Vivo* Toxicity Studies and Human Responses



Regulatory Toxicology and Pharmacology 32, 56–67 (2000)
doi:10.1006/rtph.2000.1399, available online at http://www.idealibrary.com on IDEAL®

Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals

Harry Olson,[1] Graham Betton,[2] Denise Robinson,[3] Karluss Thomas,[3] Alastair Monro,[1] Gerald Kolaja,[4] Patrick Lilly,[5] James Sanders,[6] Glenn Sipes,[7] William Bracken,[8] Michael Dorato,[9] Koen Van Deun,[10] Peter Smith,[11] Bruce Berger,[12] and Allen Heller[13]

[1]*Pfizer Inc., Groton, Connecticut;* [2]*AstraZeneca Pharmaceuticals, Macclesfield, England;* [3]*ILSI-HESI, Washington, DC, 20036;* [4]*Pharmacia & UpJohn, Kalamazoo, Michigan;* [5]*Boehringer Ingelheim Pharmaceuticals, Ridgefield, Connecticut;* [6]*Rhone-Poulenc Rorer, Collegeville, Pennsylvania;* [7]*University of Arizona, Tucson, Arizona;* [8]*Abbott Laboratories, Abbott Park, Illinois;* [9]*Eli Lilly and Co., Greenfield, Indiana;* [10]*Janssen Research Foundation, Beerse, Belgium;* [11]*Monsanto-Searle Laboratories, Skokie, Illinois;* [12]*Sanofi-Synthelabo, Inc., Malvern, Pennsylvania; and* [13]*Bayer Corporation, West Haven, Connecticut*

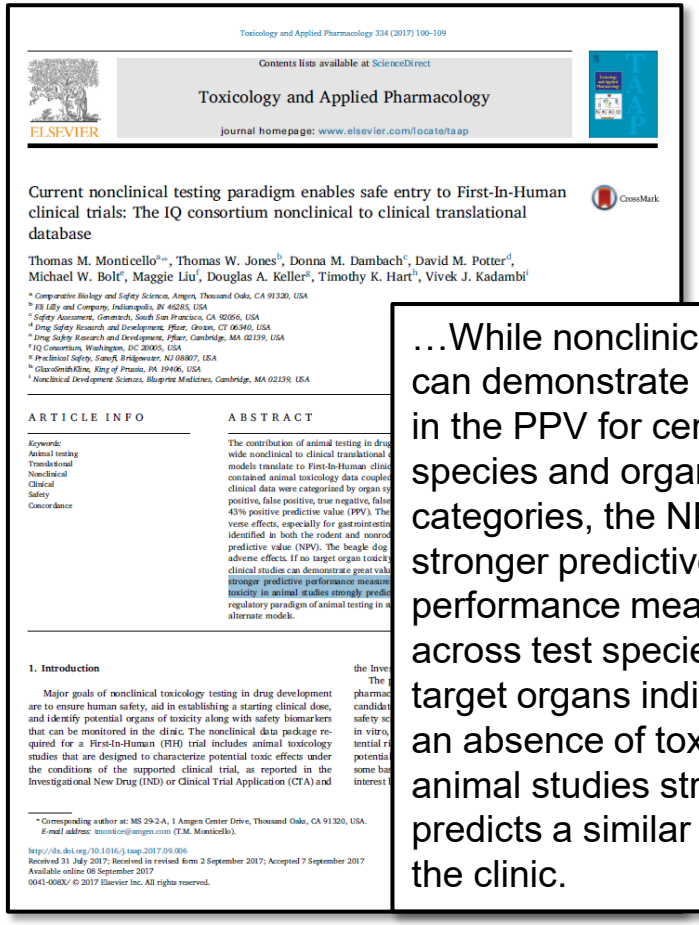
Received January 22, 2000

This report summarizes the results of a multinational pharmaceutical company survey and the outcome of an International Life Sciences Institute (ILSI) Workshop (April 1999), which served to better understand concordance of the toxicity of pharmaceuticals observed in humans with that observed in experimental animals. The Workshop included representatives from academia, the multinational pharmaceutical industry, and international regulatory scientists. The main aim of this project was to examine the strengths and weaknesses of animal studies to predict human toxicity (HT). The database was developed from a survey which covered only those compounds where HTs were identified during clinical development of new pharmaceuticals, determining whether animal toxicity studies identified concordant target organ toxicities in humans. Data collected included codified compounds, therapeutic category, the HT organ system affected, and the species and duration of studies in which the corresponding HT was either first identified or not observed. This survey includes input from 12 pharmaceutical companies with data compiled from 150 compounds with 221 HT events reported. Multiple HTs were reported in 47 cases. The results showed the true positive HT concordance rate of 71% for rodent and nonrodent species, with nonrodents alone being predictive for 63% of HTs and rodents alone for 43%. The highest incidence of overall concordance was seen in hematological, gastrointestinal, and cardiovascular HTs, and the least was seen in cutaneous HT. Where animal models, in one or more species, identified concordant HT, 94% were first observed in studies of 1 month or less in duration. These survey results support the value of *in vivo* toxicology studies to predict for many significant HTs associated with pharmaceuticals and have helped to identify HT categories that may benefit from improved methods.

INTRODUCTION

0273-2300/00 $35.00

…data compiled from 150 compounds with 221 human toxicity events reported. The results showed the true positive human toxicity concordance rate of 71% for rodent and non-rodent species, with non-rodents alone being predictive for 63% of human toxicity and rodents alone for 43%.



Toxicology and Applied Pharmacology 334 (2017) 100–109

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/taap

Current nonclinical testing paradigm enables safe entry to First-In-Human clinical trials: The IQ consortium nonclinical to clinical translational database

Thomas M. Monticello[a,*], Thomas W. Jones[b], Donna M. Dambach[c], David M. Potter[d], Michael W. Bolt[e], Maggie Liu[f], Douglas A. Keller[g], Timothy K. Hart[h], Vivek J. Kadambi[i]

[a] *Comparative Biology and Safety Sciences, Amgen, Thousand Oaks, CA 91320, USA*
[b] *Eli Lilly and Company, Indianapolis, IN 46285, USA*
[c] *Safety Assessment, Genentech, South San Francisco, CA 94080, USA*
[d] *Drug Safety Research and Development, Pfizer, Groton, CT 06340, USA*
[e] *Drug Safety Research and Development, Pfizer, Cambridge, MA 02139, USA*
[f] *IQ Consortium, Washington, DC 20005, USA*
[g] *Preclinical Safety, Sanofi, Bridgewater, NJ 08807, USA*
[h] *GlaxoSmithKline, King of Prussia, PA 19406, USA*
[i] *Nonclinical Development Sciences, Blueprint Medicines, Cambridge, MA 02139, USA*

ARTICLE INFO

Keywords:
Animal testing
Translational
Nonclinical
Clinical
Safety
Concordance

ABSTRACT

The contribution of animal testing in drug
wide nonclinical to clinical translational
models translate to First-In-Human clini
contained animal toxicology data couple
clinical data were categorized by organ sy
positive, false positive, true negative, false
43% positive predictive value (PPV). The
verse effects, especially for gastrointestin
identified in both the rodent and nonrod
predictive value (NPV). The beagle dog
adverse effects. If no target organ toxicity
clinical studies can demonstrate great valu
stronger predictive performance measure
toxicity in animal studies strongly predict
regulatory paradigm of animal testing in a
alternate models.

1. Introduction

Major goals of nonclinical toxicology testing in drug development are to ensure human safety, aid in establishing a starting clinical dose, and identify potential organs of toxicity along with safety biomarkers that can be monitored in the clinic. The nonclinical data package required for a First-In-Human (FIH) trial includes animal toxicology studies that are designed to characterize potential toxic effects under the conditions of the supported clinical trial, as reported in the Investigational New Drug (IND) or Clinical Trial Application (CTA) and

* Corresponding author at: MS 29-2-A, 1 Amgen Center Drive, Thousand Oaks, CA 91320, USA.
E-mail address: tmontice@amgen.com (T.M. Monticello).

http://dx.doi.org/10.1016/j.taap.2017.09.006
Received 31 July 2017; Received in revised form 2 September 2017; Accepted 7 September 2017
Available online 08 September 2017
0041-008X/

…While nonclinical studies can demonstrate great value in the PPV for certain species and organ categories, the NPV was the stronger predictive performance measure across test species and target organs indicating that an absence of toxicity in animal studies strongly predicts a similar outcome in the clinic.

EPA United States Environmental Protection Agency

There are a Number of Other Studies That Have Also Looked at These Questions…

Concordance between sites of tumor deve… experimental animals for 111 agents that…

Review

A Curated Database of Rodent Ut…

Evaluation of Variability Across R… Systemic Toxicity Studies

Regulatory Toxicology and…

A big data approach to the concordance of the toxi… animals and humans

Potentials and limitations of nor… for predicting clinical adverse d… analysis of 142 approv…

Review

Development of a curated Hershberger database

Review Article

New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis†

Data Quality in Predictive Toxicolog… Carcinogenicity Experiments

Chemical Research in Toxicology

New Associations between Dru… Models and Humans Reveal N…

Negative Predictors of Carcin… Environmental Chemicals

Toxicology in V…

Analysis of the Local Lymph Node Assay (LLNA) … the prediction of skin sensitisation potential and … with non-animal approaches

Machine Learning of Toxicologi… Read-Across Structure Activity … Outperforming Animal Test Rep…

Concordance between Rats and Mice in E…

Species Sensitivities and Prediction of Tetratogenic Potential

by James L. Schardein,* Bernard A. Schwetz,† and Michael F. Kenel*

Many chemicals shown to be teratogenic in laboratory animals are not known to be teratogenic in humans. However, it remains to be determined if the unresponsiveness of humans is due to lessened sensitivity, to generally subteratogenic exposure levels, or to the lack of an appropriate means of identifying human teratogens. On the other hand, with the exception of the coumarin anticoagulant drugs, those agents well accepted as human teratogens have been shown to be teratogenic in one or more laboratory species. Yet, no single species has clearly distinguished itself as being more advantageous in the detection of human teratogens over any other. Among the species used for testing, the rat and mouse most successfully model the human reaction, but the rabbit is less likely than other species to give a false positive finding. Among species less commonly used for testing, primates offered a higher level of predictability than others. Regarding concordance of target malformations, the mouse and rat produced the greatest number of concordant defects, but they also were responsible for the most noncorrodant responses as well. Since no other species is clearly more predictive of the human response, it is concluded that safety decisions should be based on all reproductive and developmental toxicity data in light of the agent's known pharmacokinetic, metabolic and toxicologic parameters.

Introduction

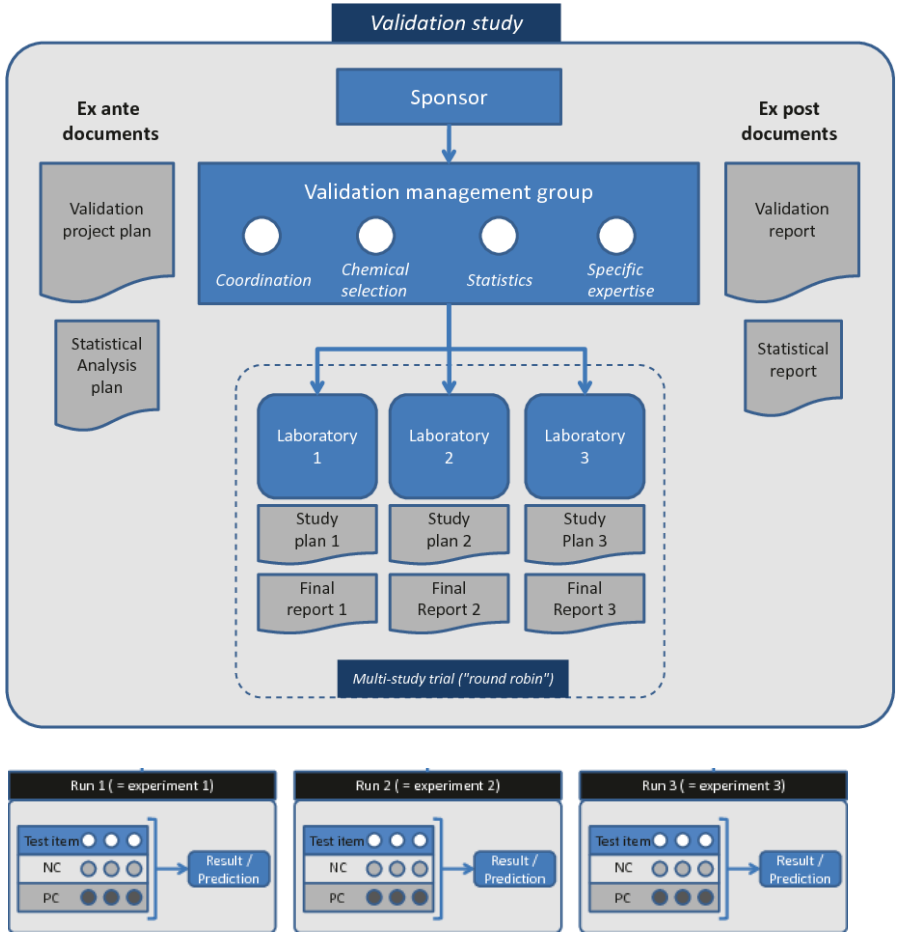
The extrapolation of animal data to the human is the foundation of safety evaluation of chemicals and drugs prior to human exposure. As will be shown, there is at present no perfectly suitable animal model from which to make these extrapolations in most cases, nor is it abundantly clear just what endpoints residing in the animal database are most important in this determination. In short, it is conceded that the predictive value of animal teratogenicity tests in extrapolating results into terms of human safety is imperfect. Consider the fact that of over 2800 chemicals now reported to have been assayed in animals for teratogenic potential, only about 14 chemicals or groups of chemicals have been shown to have this propensity in humans (1). Put another way, there are about 1000 or so chemicals that demonstrate some measure of teratogenicity in animals but for which we have no evidence at all that they share this property in humans. However, this discrepancy may only reflect that we cannot clearly recognize teratogenicity in humans. It should be realized from the onset that testing strategies of drugs for teratogenic potential on the one hand need to be clearly distinguished from testing of industrial or environmental chemicals on the other. Conditions of exposure, numbers of potential victims of exposure, benefit considerations, economics and other factors are radically different in the two cases and affect both the philosophy of testing and the acceptable risk (2).

Fortunate as we are that newly identifiable human teratogens have not wreaked havoc in the same manner as in the 10,000 thalidomide victims in 1961 to 1962, or in the 20,000 rubella-infected children in 1964, correlations of animal and human responses may not be providing us with the necessary answers to avoid potentially similar situations in the future. It is the apparent lack of association that precludes accurate prediction of teratogenic potential in extrapolating from animals to humans. We are really left at present with the dilemma of not being capable of selecting which animal species are the most predictive of the likely human response.

Two decades ago, with the thalidomide catastrophe, the rabbit emerged as the sensitive species with respect to this teratogen, highlighted in part because of the negative and/or inconclusive results produced in a large number of other laboratory species tested at the time.

*International Research and Development Corporation, Mattawan, MI 49071.
†National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.

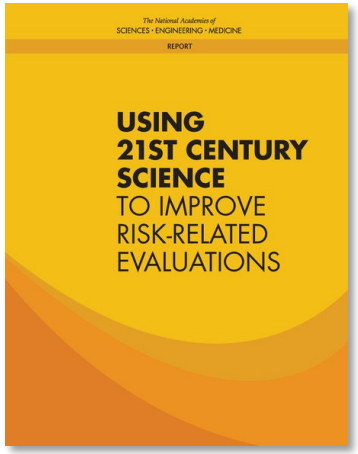
The Current Validation Process is Unsustainable and Not Necessary for All EPA Decision Contexts



Lambris and Paoletti, *Adv Exp Med and Biol*, 2016

- Traditional validation process in toxicology has focused on one-for-one replacement of a reference method using a ring trial design
  - Time frame (idealized): 3 years for trial, 2 years for review/publication (Altex 27:253, 2010)
  - Costs highly dependent on method being validated
- Challenges associated with the traditional approach:
  - Time and resource intensive
  - Accommodating rapidly changing technology and methods;
  - Lack of human data for most endpoints;
  - Many original reference methods were not themselves “validated”;
  - Limited reference chemicals for some endpoints/responses;
  - Difficulty in comparing overlapping, but not identical methods.
  - Expectation that individual methods that are part of integrated assessment and testing approaches (IATAs) are validated singly and in combination

ICCVAM Roadmap and National Academy Report Support Transition to Fit-for-Purpose Validation



NICEATM/ICCVAM Strategic Roadmap

- Promotes use of efficient, flexible practices to evaluate fitness for purpose and establish scientific confidence in new methods
- Emphasizes separating testing requirements from context of use
- Highlights need to establish confidence based on human biology and mechanistic relevance rather than comparisons with existing models

NAS Report

- Emphasizes importance in defining purpose and scope of the NAM
- Recognizes challenges in validating a NAM where there is no “gold standard” or against assays that have not themselves been validated
- Suggests establishing performance standards for data quality
- Recognizes that ring-trial design is not necessary for all purposes
- Emphasizes need for reporting standards and transparency

Recent EPA Documents Have Generally Adopted the Fit-for-Purpose Approach

EPA United States Environmental Protection Agency

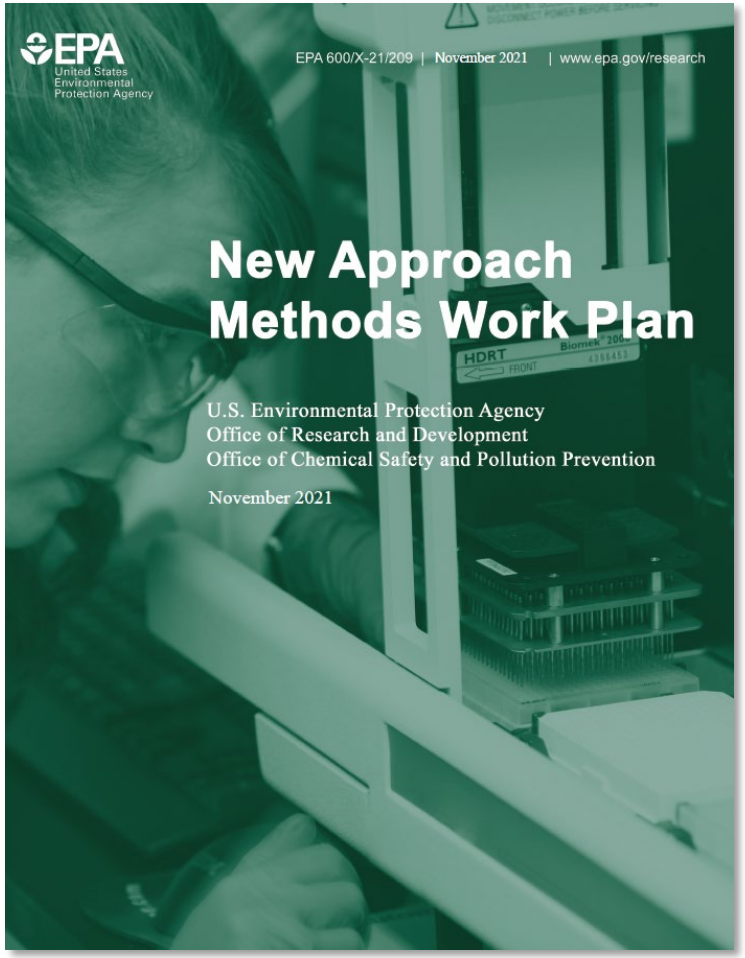
EPA Document# EPA-740-R1-8004
June 22, 2018
Office of Chemical Safety and Pollution Prevention

Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program

June 22, 2018

- Outlined general framework for establishing scientific confidence in NAMs for TSCA-related decisions
- Important elements in evaluating relevance
  - Fit for purpose and utilization
  - Considers both NAM and integrated NAMs
  - Flexibility for use in qualitative or quantitative predictions and as part of WOE
- Important elements in evaluating reliability
  - Evaluation of inter-laboratory reproducibility is not required for all NAMs and depends on context of use
  - Incorporates performance-based approach
- Criteria for establishing scientific confidence include:
  - Clear decision context
  - Mechanistically and/or biologically relevant where possible. If not, the chemical domain of applicability clearly defined
  - Reference chemicals adequately referenced
  - Reliability considered within the context of intended use and best practices
  - Transparently described and information available to the public (except TSCA CBI)
  - Characterize uncertainty
  - Evaluation and implementation by third parties must be possible (i.e., assays commercially available or protocols)
  - Independent scientific review

NAS Study in the Context of the EPA NAM Work Plan



- Five objectives to increase the scientific rigor and sophistication of Agency assessments using NAMs while reducing the reliance on vertebrate animals to test chemicals.
  - Evaluate Regulatory Flexibility
  - Develop Baselines and Metrics
  - Establish Scientific Confidence and Demonstrate Application
  - Develop NAMs to Address Information Gaps
  - Engage and Communicate with Stakeholders
- The Committee's Statement of Task help inform two strategies
  - *Characterize scientific quality and relevance of traditional toxicity tests – Due 2023*
  - *Develop a scientific confidence framework to evaluate the quality, reliability, and relevance of NAMs – Due 2024*
- Useful recommendations will..
  - Focus on evaluating the variability and relevance of the traditional *in vivo* mammalian toxicity testing models and not the variability and relevance of the NAMs.
  - Build on recommendations on validation and scientific confidence frameworks covered by previous NAS and OECD committees but tackle difficult issues of what to do when there is no standard test, the standard test is not relevant to the human response, or the standard test has not been benchmarked against human responses.
  - Accommodate the broad range of regulatory decisions the Agency needs to make

We Look Forward to the Committee's Input on These Difficult Issues…