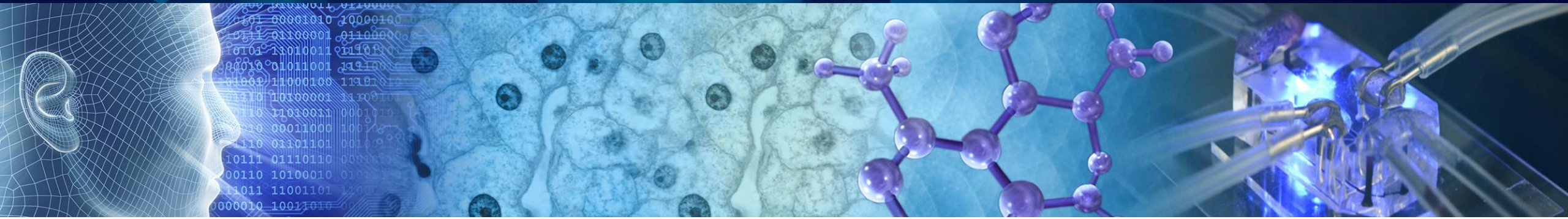
National Institute of
Environmental Health Sciences
Division of Translational Toxicology

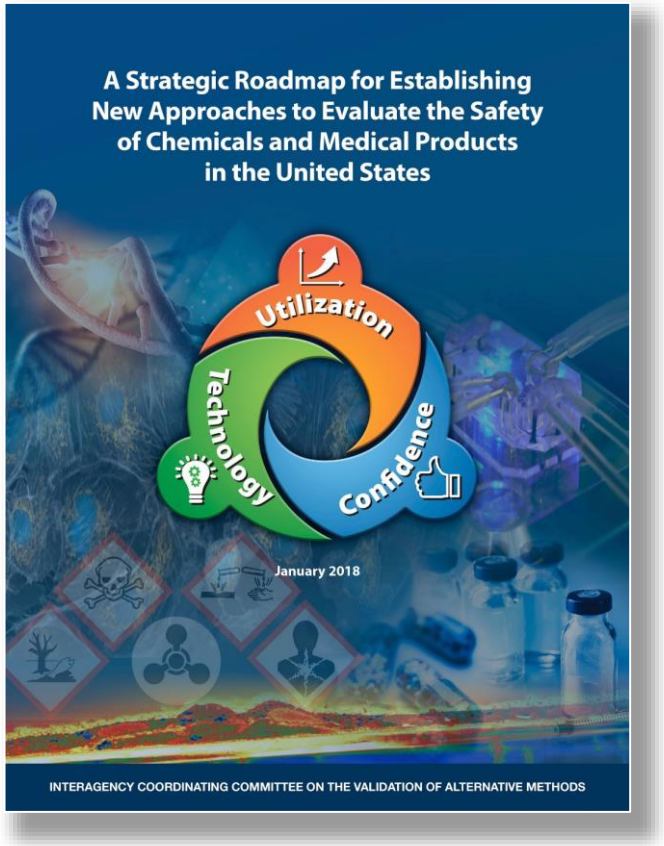


Flexible, Fit-For-Purpose NAMs Validation Across Federal Agencies

Nicole C. Kleinstreuer, PhD

NTP Interagency Center for the Evaluation of Alternative Toxicological Methods

"Advances in science and technology have not been effectively leveraged to predict adverse human health effects"

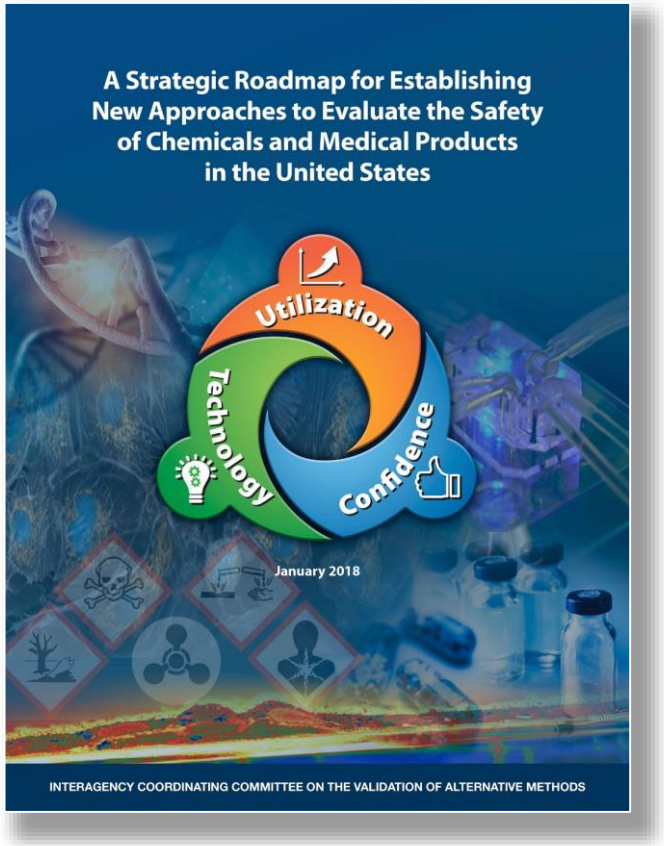


Help end-users guide the development of the new methods

Use efficient and flexible approaches to establish confidence in new methods

Encourage the adoption of new methods by federal Agencies and regulated industries

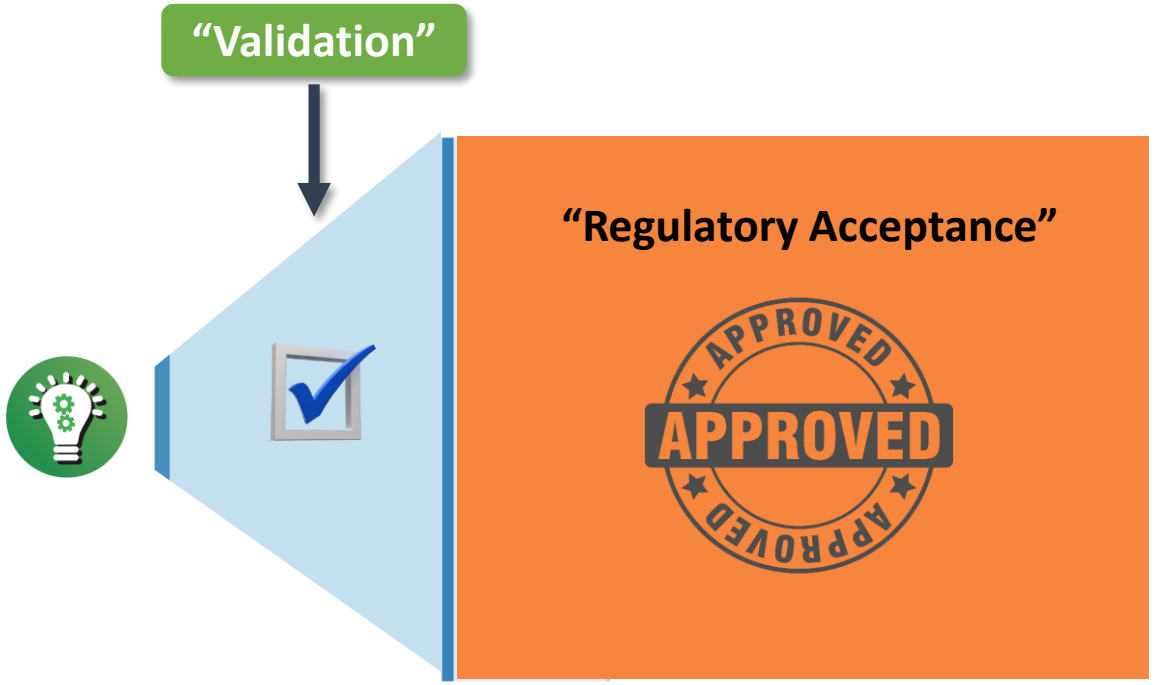
"Advances in science and technology have not been effectively leveraged to predict adverse human health effects"



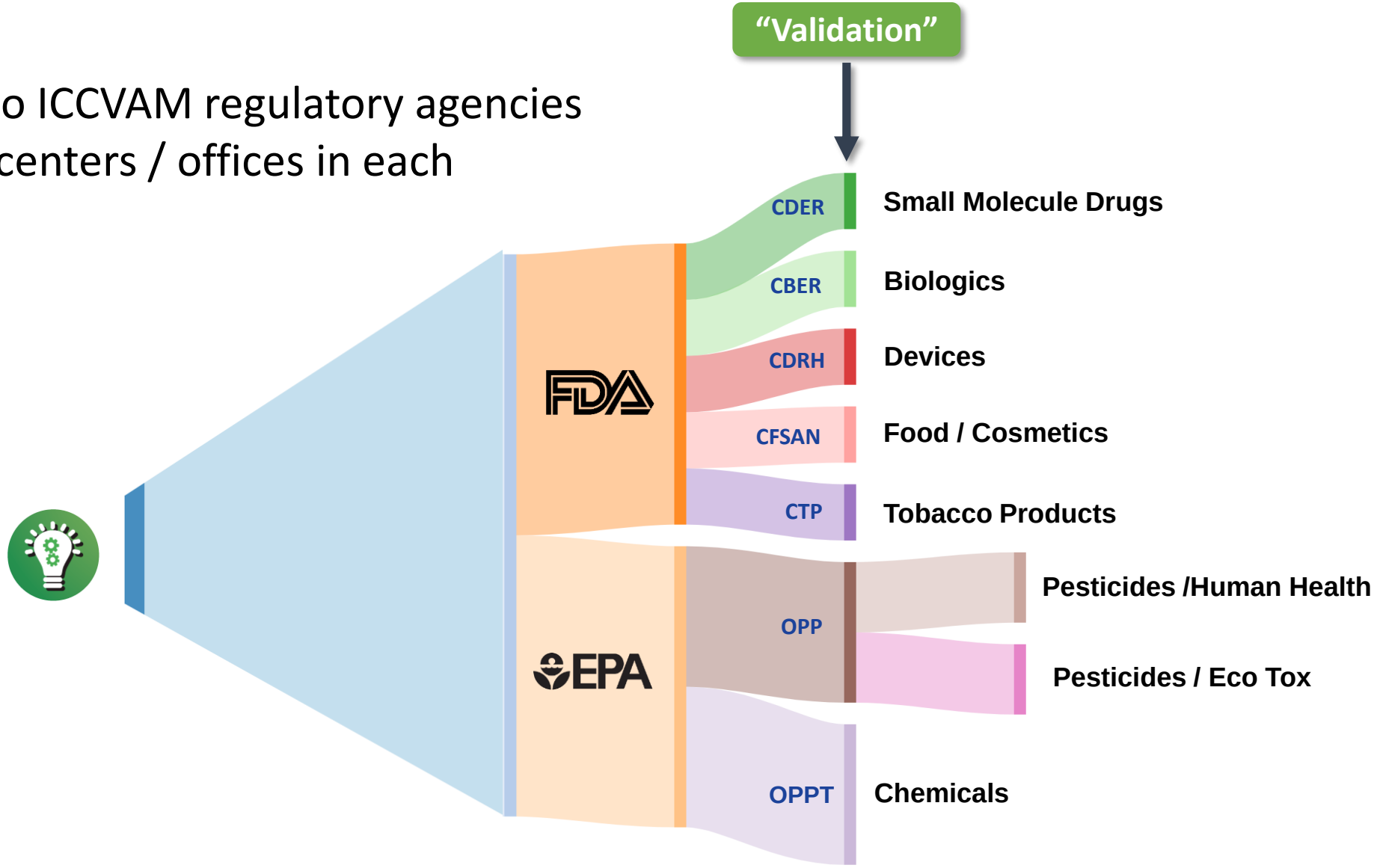
Help end-users [illegible] development of the new m[illegible]

Use eff[illegible]nt an[illegible]e ap[illegible]aches to establis[illegible]nce [illegible]new [illegible]thods

Encourage t[illegible] new methods by federal Agencies and regulated industries


"Validation"
"Regulatory Acceptance"
APPROVED
APPROVED
APPROVED

Example of two ICCVAM regulatory agencies with multiple centers / offices in each



From

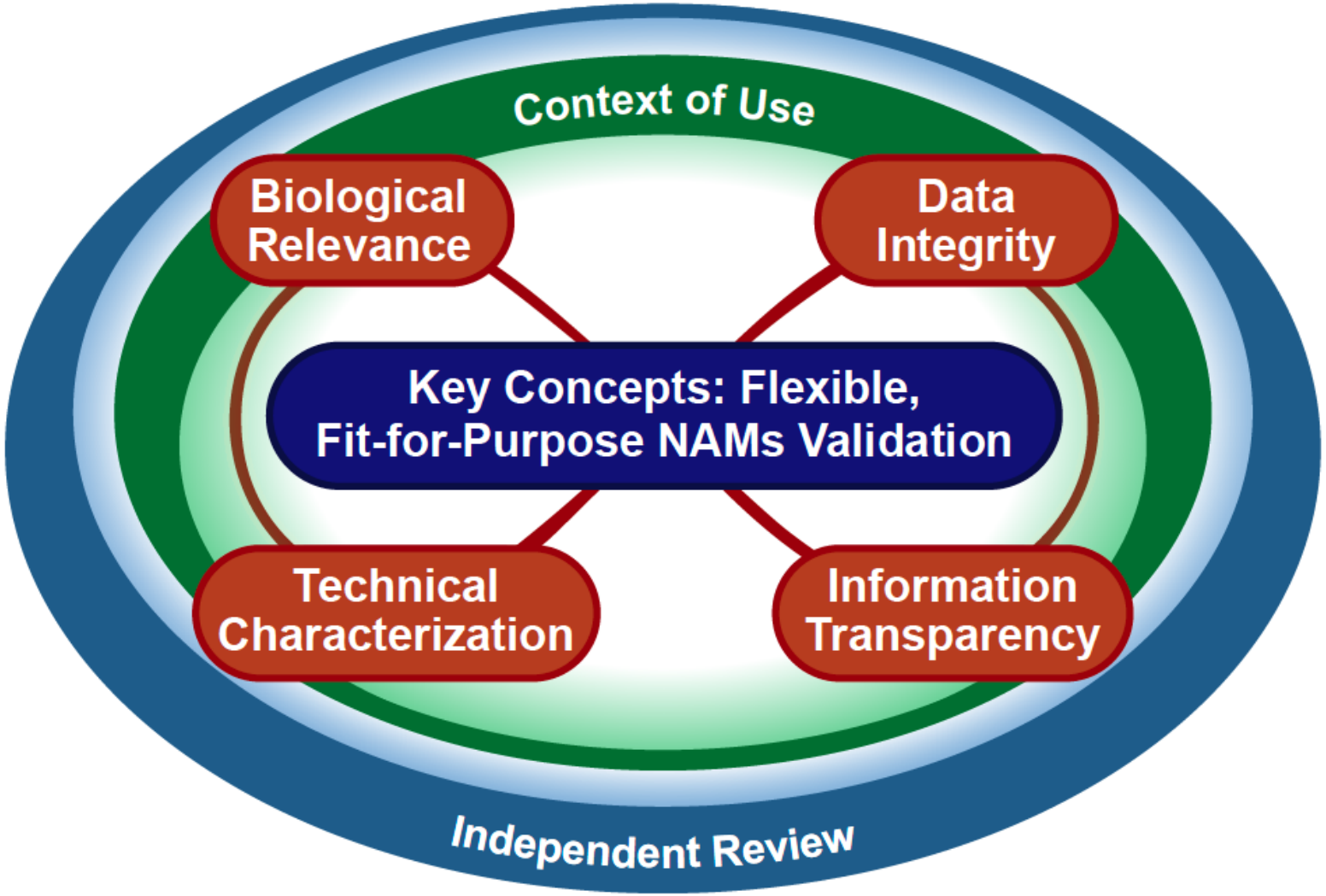
- Centralized ("VAMs")
- One Size Fits All
- Binary Status (Validated / Not)
- Stand Alone

TRANSITION

Towards

- Decentralized (End Users)
- Fit for Purpose
- Evolving Confidence
- Integrative

Updated ICCVAM Validation Guidance: Coming Soon!



National Institute of Environmental Health Sciences
Division of Translational Toxicology

Archives of Toxicology (2022) 96:2865–2879
https://doi.org/10.1007/s00204-022-03365-4

REVIEW ARTICLE

Check for updates

A framework for establishing scientific confidence in new approach methodologies

Anna J. van der Zalm[1] · João Barroso[2] · Patience Browne[3] · Warren Casey[4] · John Gordon[5] · Tala R. Henry[6] · Nicole C. Kleinstreuer[7] · Anna B. Lowit[6] · Monique Perron[8] · Amy J. Clippinger[1]

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

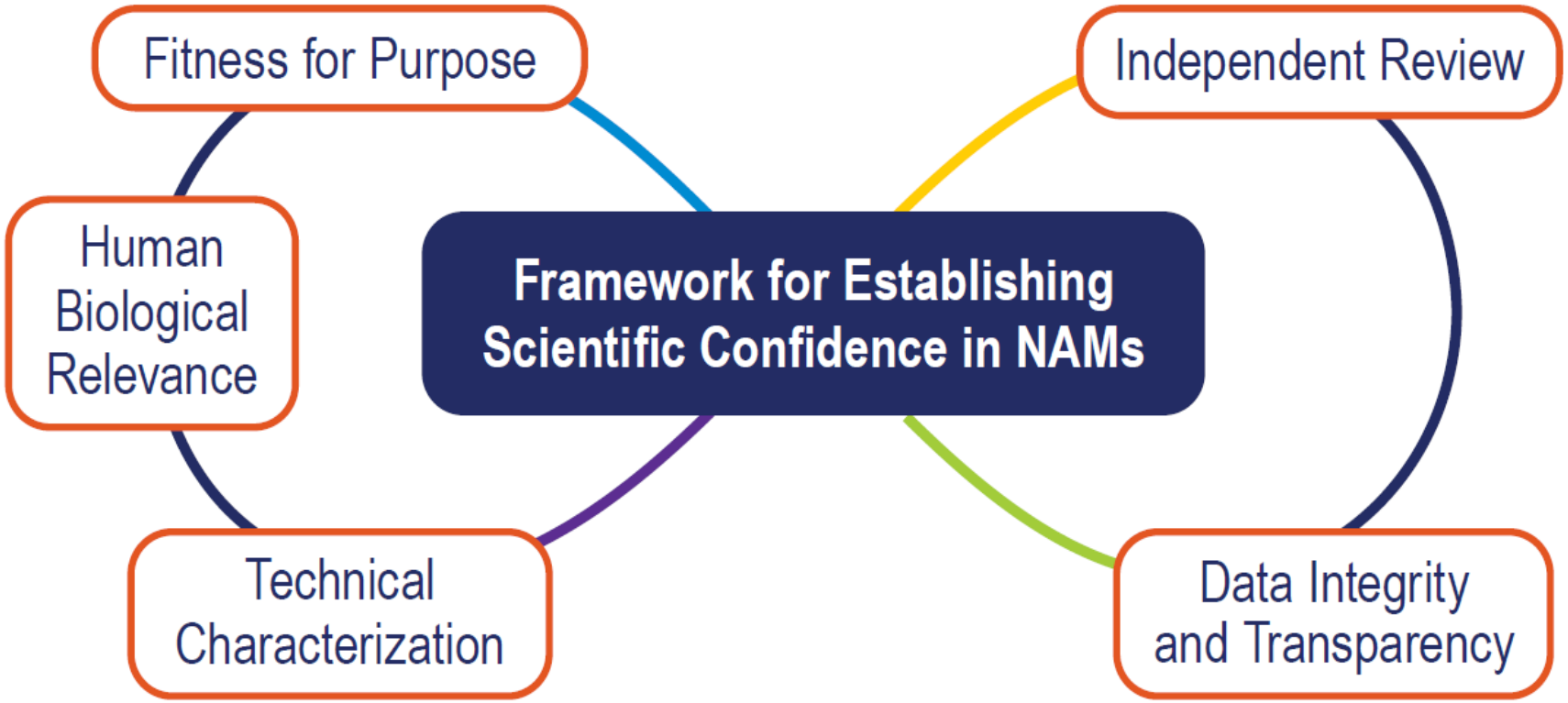
GUIDANCE DOCUMENT ON THE VALIDATION AND INTERNATIONAL ACCEPTANCE OF NEW OR UPDATED TEST METHODS FOR HAZARD ASSESSMENT

Arch Toxicol (2018) 92:611–617
https://doi.org/10.1007/s00204-017-2097-4

REGULATORY TOXICOLOGY

Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods

S. Casati[1] · K. Aschberger[1] · J. Barroso[1] · W. Casey[2] · I. Delgado[3] · T. S. Kim[4] · N. Kleinstreuer[2] · H. Kojima[5] · J. K. Lee[4] · A. Lowit[6] · H. K. Park[4] · M. J. Régimbald-Krnel[7] · J. Strickland[8] · M. Whelan[1] · Y. Yang[9] · Valérie Zuang[1]



van der Zalm et al. 2022 Arch Tox

Fitness for Purpose

Purpose = Context of Use

Fitness for Purpose

Which regulatory statutes are data from the NAM intended to comply with?

- U.S. TSCA
- EU REACH
- Other

How will the NAM be used?

- As a stand-alone assay
- As part of a defined approach
- As part of an integrated approach to testing and assessment or weight of evidence assessment

Is the information provided sufficient to address the regulatory endpoints of interest?

- Describe the relationship between the information measured by the NAM and the regulatory endpoints being addressed.
- Is the technical performance, including the level of uncertainty, acceptable?

What is the context in which the NAM is intended to be used?

- Preregulatory screening and prioritization
- Chemical grouping
- Hazard identification
- Quantitative risk assessment

van der Zalm et al. 2022 Arch Tox

(Human) Biological Relevance

- Similarities between the physiology of, or the biology measured by, the test system, and human biology
  - ✓ Consider human dosimetry modelling, cell types used, or the structure of the target organ/tissue
- Concordance with human responses

- Establishing biological relevance of a method can be used to benchmark its performance

Human Relevance

Prior GHS category	1	2A	2B	NC
1 (serious eye damage)	73%	16%	0%	10%
2A (irritant)	4%	33%	4%	59%
2B (mild irritant)	0%	4%	16%	80%
NC (non-irritant)	1%	4%	2%	94%

Adapted from Luechtefeld et al., ALTEX 33(2), 2016.

Consider strengths and limitations of all available methods with respect to:

- their relevance to human ocular anatomy
- the mechanisms of eye irritation/corrosion in humans

Clippinger et al. 2021 Cut Ocu Tox

Assessing approaches for eye corrosion/irritation potential

- The rabbit test should not be used as a reference method to demonstrate the validity of *in vitro/ex vivo* assays
- *In vitro/ex vivo* methods are as or more reliable and relevant than the rabbit test

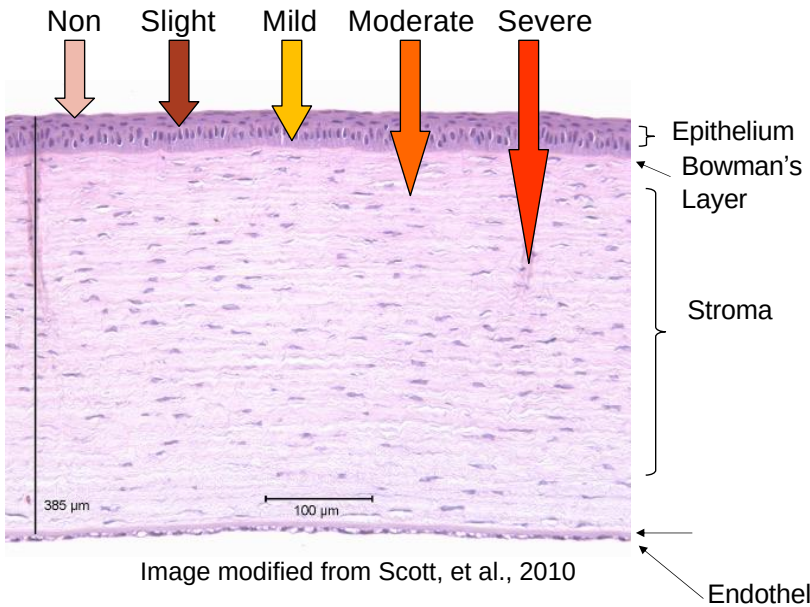
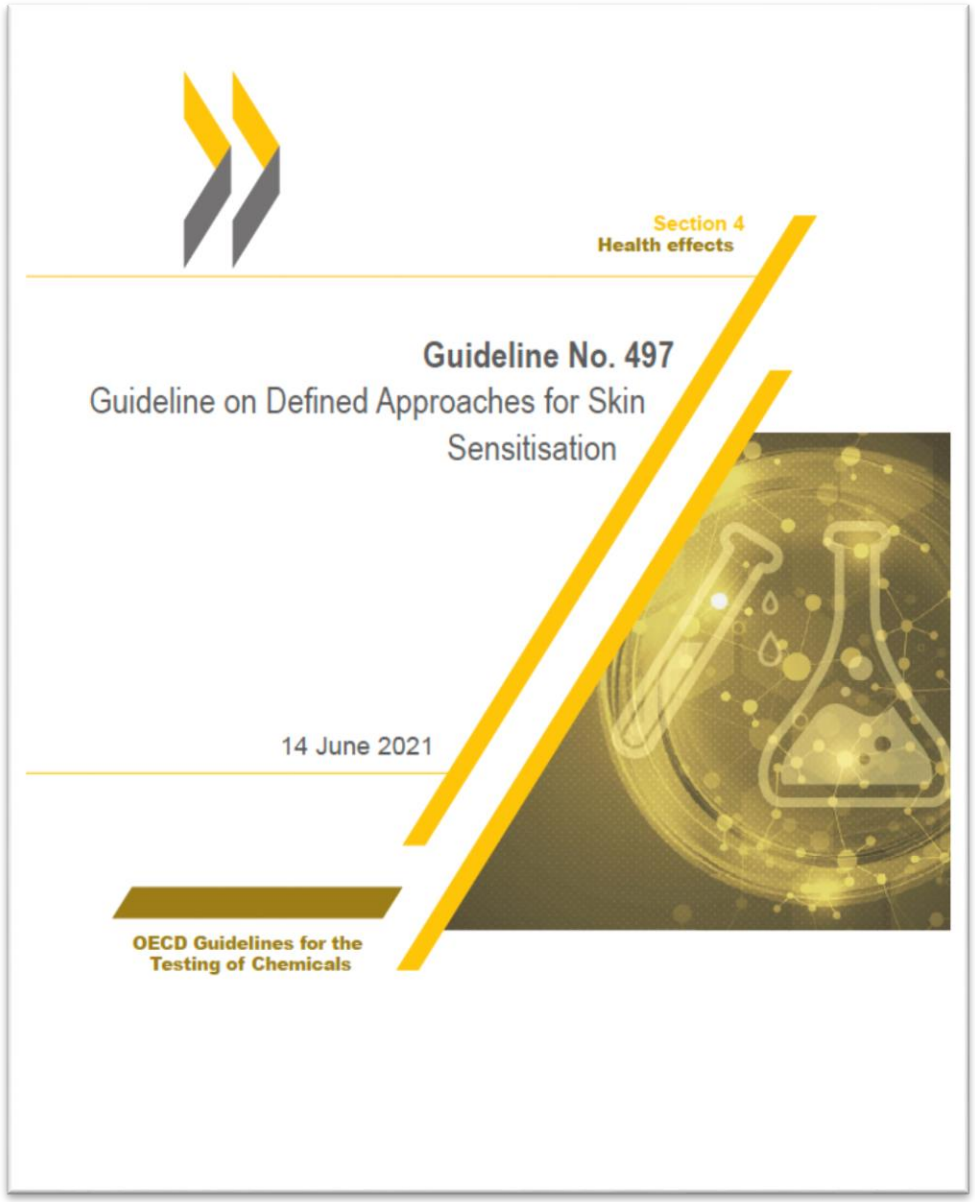
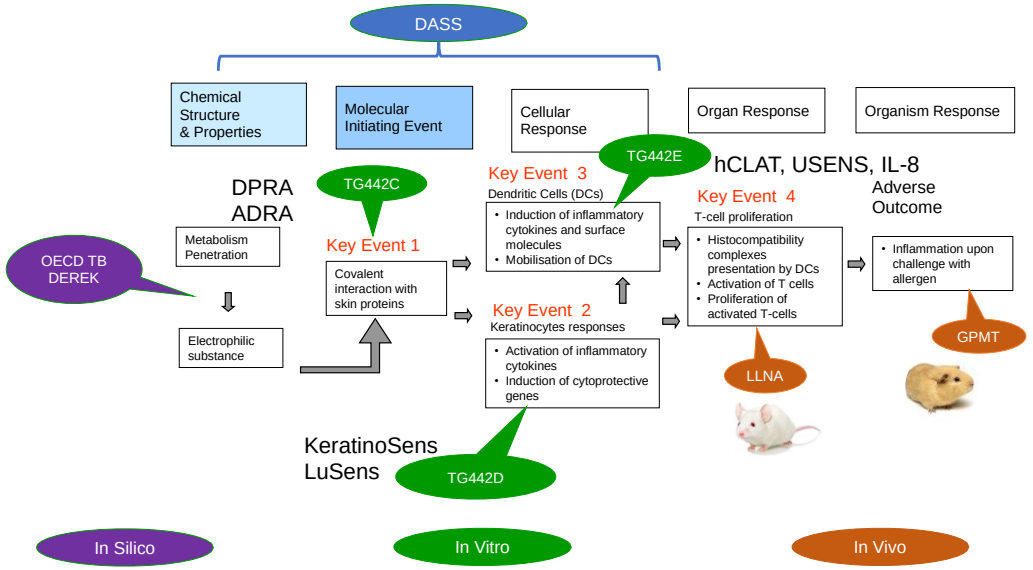


Image modified from Scott, et al., 2010

AOP-Anchoring





DA/Method	Information Sources	Capability (Hazard and/or Potency)	Hazard Performance vs. LLNA N~168	Hazard Performance vs. Human N~63	GHS Potency Performance vs. LLNA (Accuracy)	GHS Potency Performance vs. Human (Accuracy)
2o3 DA	DPRA, KeratinoSens™, h-CLAT	Hazard	84% BA, 82% Sens, 85% Spec	88% BA, 89% Sens, 88% Spec	-	-
ITSv1 DA	DPRA, h-CLAT, DEREK Nexus v6.1.0	Hazard, Potency (GHS)	81% BA, 92% Sens, 70% Spec	69% BA, 93% Sens, 44% Spec	70% NC, 71% 1B, 74% 1A	44% NC, 77% 1B, 65% 1A
ITSv2 DA	DPRA, h-CLAT, OECD QSAR Toolbox v4.5	Hazard, Potency (GHS)	80% BA, 93% Sens, 67% Spec	69% BA, 94% Sens, 44% Spec	67% NC, 72% 1B, 72% 1A	44% NC, 80% 1B, 67% 1A
LLNA (provided for comparison)	*in vivo*	Hazard, Potency	-	58% BA, 94% Sens, 22% Spec	-	25% NC, 74% 1B, 56% 1A

Technical Characterization

- Describe:
  - accuracy
  - intra-laboratory reproducibility
  - transferability
  - applicability domain
  - reference chemicals and controls
  - limits of detection and quantification

- Data reporting should allow for evaluation of the method, including:
  - protocol
  - equipment
  - computational models being used

- What is considered acceptable may depend on the method being evaluated and its intended use

Accuracy

- Traditional animal test methods should not be assumed to provide data relevant to human biology or mechanisms of toxicity and be the "right" answer to determine if another method is valid.
- When using benchmark animal data:
  - Relevance to predict human effects should also be considered, where possible (in the case of human health endpoints)
  - Variability of animal data should be characterised and considered when evaluating alternative approaches
- Instead, accuracy can be demonstrated by considering:
  - Consistency across methods/approaches
  - Ability to identify positive and negative reference chemicals
  - Greater emphasis on biological relevance and reproducibility

Toxicology and Applied Pharmacology
Volume 19, Issue 2, June 1971, Pages 276-360

Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests

Carrol S. Weil [a, b], Robert A. Scala [a, b]

Arch Toxicol (2017) 91:521–547
DOI 10.1007/s00204-016-1679-x

REVIEW ARTICLE

Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD)

João Barroso[1,2] · Uwe Pfannenbecker[3] · Els Adriaens[4] · Nathalie Alépée[5] · Magalie Cluzel[6] · Ann De Smedt[7] · Jalila Hibatallah[8] · Martina Klaric[1] · Jürsten R. Mewes[9] · Marion Millet[10] · Marie Templier[10] · Pauline McNamee[11]

Arch Toxicol (2014) 88:701–723
DOI 10.1007/s00204-013-1156-8

IN VITRO SYSTEMS

Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the in vivo endpoints under UN GHS/EU CLP for the development and evaluation of in vitro test methods

Els Adriaens · João Barroso · Chantra Eskes · Sebastian Hoffmann · Pauline McNamee · Nathalie Alépée · Sandrine Bessou-Touya · Ann De Smedt · Bart De Wever · Uwe Pfannenbecker · Magalie Tailhardat · Valérie Zuang

Analysis of Draize Eye Irritation Testing and its Prediction by Mining Publicly Available 2008-2014 REACH Data

Thomas Luechtefeld [1], Alexandra Maertens [1], Daniel P. Russo [2], Costanza Rovida [4], Hao Zhu [2,3] and Thomas Hartung [1,4]

Regulatory Toxicology and Pharmacology 122 (2021) 104920

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Analysis of variability in the rabbit skin irritation assay

John P. Rooney [a,*], Neepa Y. Choksi [a], Patricia Ceger [a], Amber B. Daniel [a], James Truax [a], David Allen [a], Nicole Kleinstreuer [b]

Toxicology in Vitro 34 (2016) 220–228

Contents lists available at ScienceDirect

Toxicology in Vitro

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

Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches

Coralie Dumont, João Barroso, Izabela Matys, Andrew Worth, Silvia Casati *

Concept Article

Uncertainties of Testing Methods: What Do We (Want to) Know About Carcinogenicity?

Martin Paparella [1], Annamaria Colacci [2] and Miriam N. Jacobs [3]

Toxicological Sciences

Evaluation of Variability Across Rat Acute Oral Systemic Toxicity Studies

Agnes L. Karmaus*, Kamel Mansouri†, Kimberly T. To*, Bevin Blake†,1, Jeremy Fitzpatrick‡,2, Judy Strickland*, Grace Patlewicz‡, David Allen*, Warren Casey†, and Nicole Kleinstreuer†

Review

A Section 508–conformant HTML version of this article is available at http://dx.doi.org/10.1289/ehp.1510183.

A Curated Database of Rodent Uterotrophic Bioactivity

Nicole C. Kleinstreuer,[1] Patricia C. Ceger,[1] David G. Allen,[1] Judy Strickland,[1] Xiaoqing Chang,[1] Jonathan T. Hamm,[1] and Warren M. Casey[2]

Reprod Toxicol. 2018 October ; 81: 259–271. doi:10.1016/j.reprotox.2018.08.016.

DEVELOPMENT OF A CURATED HERSHBERGER DATABASE

P Browne[a], NC Kleinstreuer[b], P Ceger[c], C Deisenroth[d], N Baker[e], K Markey[f], RS Thomas[d], RJ Judson[d], W Casey[b]

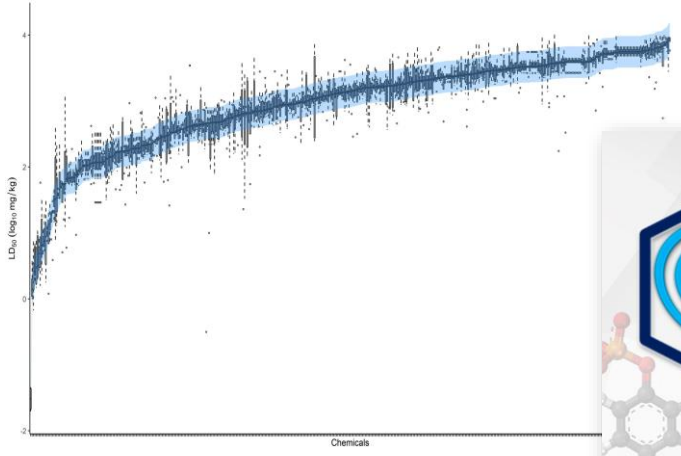
Published in final edited form as:
Comput Toxicol. 2020 August 1; 15(August 2020): 1–100126. doi:10.1016/j.comtox.2020.100126.

Variability in *in vivo* studies: Defining the upper limit of performance for predictions of systemic effect levels

Ly Ly Pham[1,2], Sean Watford[1,3], Prachi Pradeep[1,2], Matthew T. Martin[1,4], Russell Thomas[1], Richard Judson[1], R. Woodrow Setzer[1], Katie Paul Friedman[1]

Reference Data Variability

Data-driven Confidence Intervals for Model Evaluation/Predictions



Analyzing sources of variability in acute oral toxicity data & applying 95% confidence interval to predictions

	0	5	50	300	500	2000	5000 mg/kg
VT	0	0	1	1	1	1	1
NT	1	1	1	1	1	0	0
EPA	0	0	1	1	0	0	0
GHS	0	0	1	0	0	0	0
LD50	0	0	1 (160)	(-0.3) 1 (316)	(+0.3) 1 (613)	0	0
WoE	**1**	**1**	**5**	**4**	**3**	**1**	**1**

	Very Toxic		Non-Toxic		EPA		GHS	
	Train	Eval	Train	Eval	Train	Eval	Train	Eval
Sensitivity	0.87	0.70	0.88	0.67	0.81	0.62	0.80	0.58
Specificity	0.99	0.97	0.97	0.90	0.92	0.86	0.95	0.90
Balanced Accuracy	0.93	0.84	0.92	0.78	0.87	0.74	0.88	0.74
***In vivo* Balanced Accuracy**	0.81		0.89		0.82		0.79	

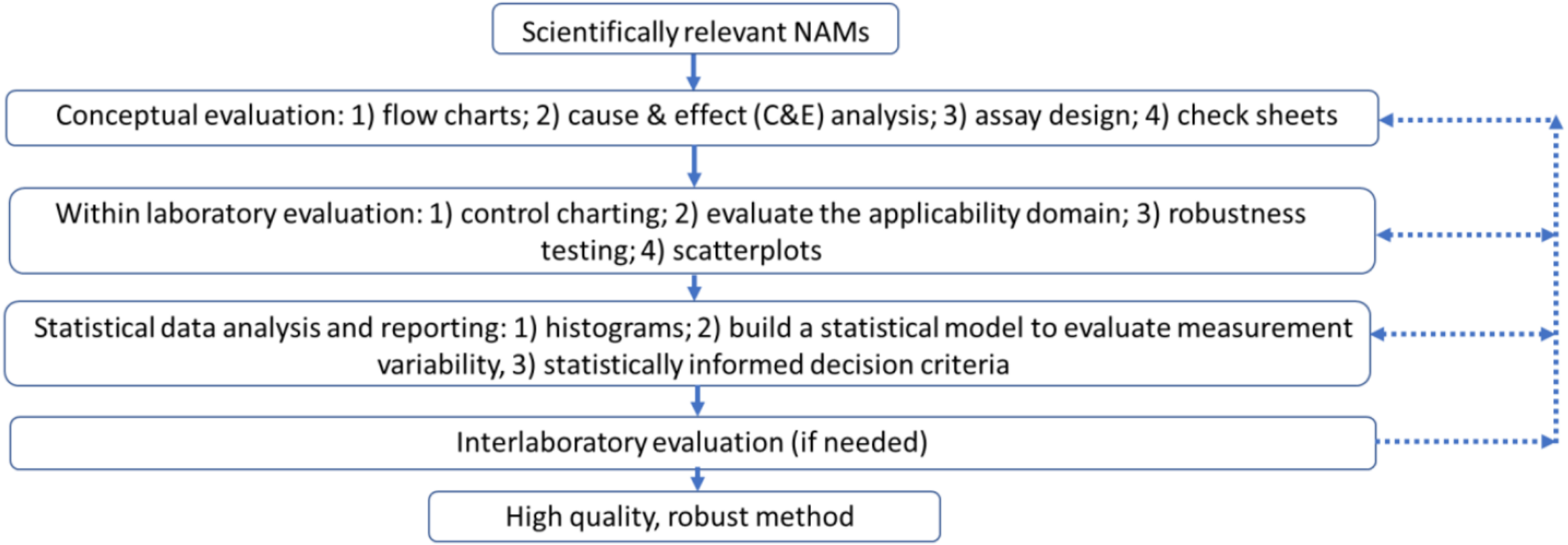
	LD50 values		LD50 values
	Train	Eval	***In Vivo***
R2	0.85	0.65	0.80
RMSE	0.30	0.49	0.42

CATMoS QSAR predictions perform just as well as replicate *in vivo* data at predicting oral acute toxicity outcome

Karmaus et al. Toxicol Sci. 2022; Mansouri et al. EHP 2021

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Technical Framework for Robust and Reliable NAMs



Petersen et al. 2022 ALTEX

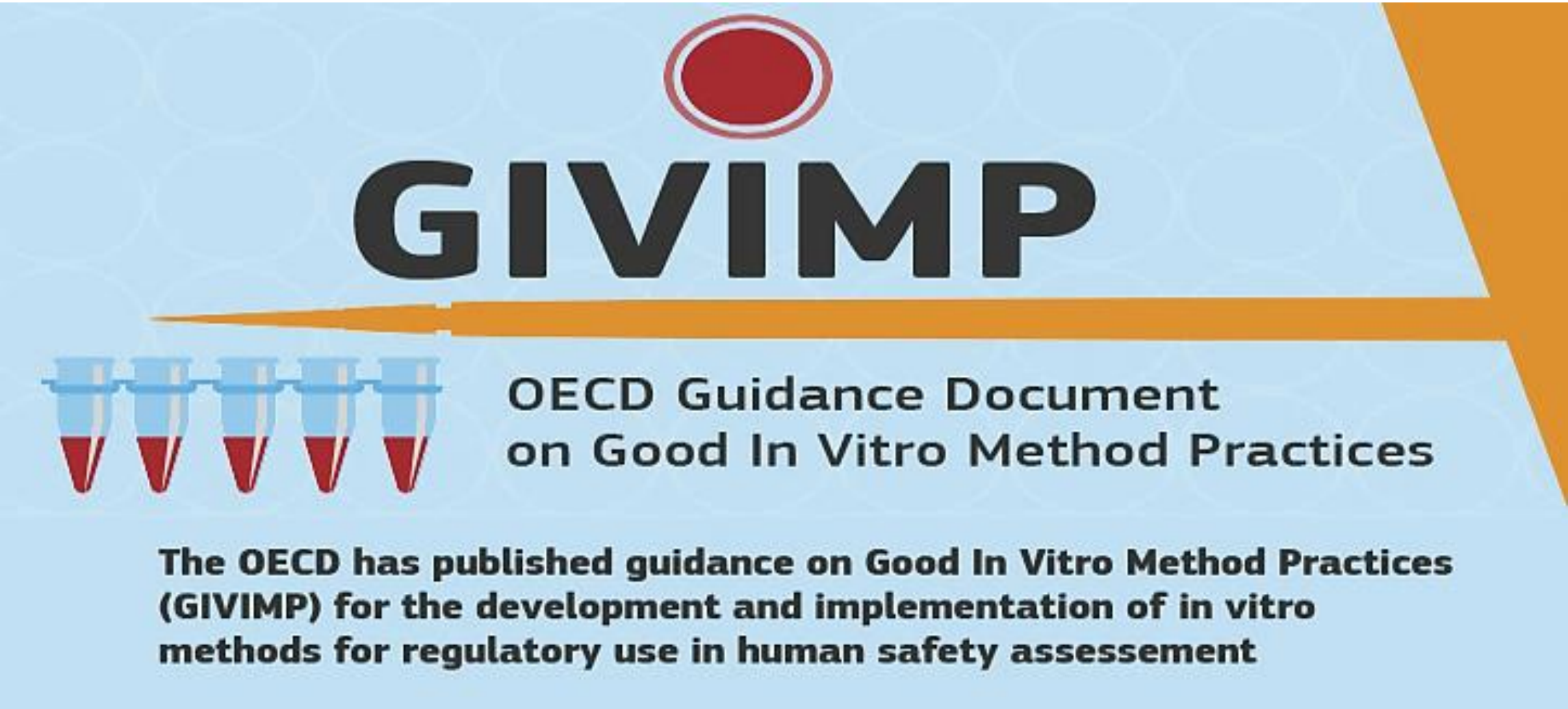
Data Integrity and Transparency

- Assess integrity and credibility of the raw data to the final report
- Communicate transparently and publicly
- Assess and describe the uncertainties and limitations
- Independently reproduce data
  - External implementation and training of the models
  - Processing of the raw data
  - Replicate predictions obtained in the validation study

Independent Review

- Important part of confidence building process
- Appropriate level of external review depends on the method and context of use
- Might include publication in peer-reviewed journal or review by an independent scientific advisory panel
- International adoption by OECD typically needs formal peer review
- Method developers may fund but should not manage peer review

Standards for Reliability and Data Integrity

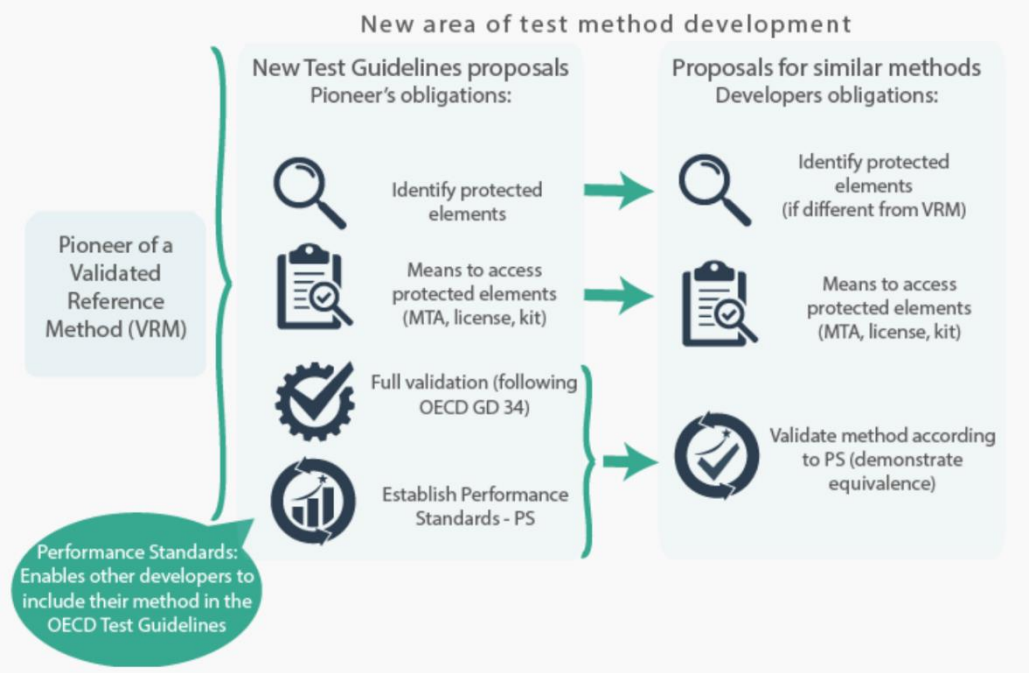


The OECD has published guidance on Good In Vitro Method Practices (GIVIMP) for the development and implementation of in vitro methods for regulatory use in human safety assessement

OECD Policy

Handling confidential information in candidate Test Guidelines

- Method developers encouraged to use other means than confidentiality to protect their intellectual property
- OECD will host confidential information on a protected webpage accessible to National Coordinators only during Test Guideline development
- Once the Test Guideline is adopted, this information will be made publicly available



Source: https://www.oecd.org/chemicalsafety/testing/intellectual-property-in-oecd-test-guidelines.htm

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Alternative Methods Tracking

TSAR – EURL ECVAM

European Commission | Log in | Search

TSAR - Tracking System for Alternative methods towards Regulatory acceptance

Genomic Allergen Rapid Detection test

Topic: Sensitisation

Test Method Number: TM2011-09 (EU)
Short Name of TM: GARD
Year received: 2011
Responsible Organisation: EURL ECVAM - European Union
Protocol(s)/SOP(s): GARDskin Assay Protocol
General Comments: Please note that the GARDskin Assay Protocol available in the link above is a revised version provided by the test method developer after completion of the ESAC peer review to address comments made by the ESAC.

Method Description

The Genomic Allergen Rapid Detection (GARD) is a transcriptomics-based *in vitro* assay proposed to assess the skin sensitisation potential/potency of chemicals.

GARD addresses the third key event of the skin sensitisation Adverse Outcome Pathway (activation of dendritic cells), step 5 (biochemical pathways related to skin sensitisation) and step 6 (immune recognition of chemical allergens and maturation of dendritic cells (DCs)).

The test method has two elements: the so-called GARDskin to assess skin sensitisation potential (first submission in October 2011) and GARDpotency to assess skin sensitisation potency (first submission in July 2018).

The test method is based on the nCounter system and measures the expression level of a panel of genes in the human myeloid cell line MUTZ-3 exposed to chemicals. In GARDskin, the expression of a panel of 200 genes (the GARD Prediction Signature, GPS) is used as input to a prediction model based...

[Read more]

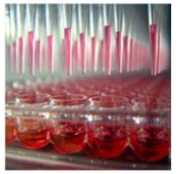
Track Approval Status

SUBMISSION Finalised | VALIDATION Finalised | PEER-REVIEW Finalised | n/a RECOMMENDATION Not applicable | REGULATORY ACCEPTANCE/STANDARDS Ongoing

Step Expand All

- Submission
- Validation
- Peer-review
- Recommendation
- Regulatory acceptance/Standards

NICEATM Website



Accepted Alternative Methods

NICEATM has compiled a list of alternative methods already accepted by U.S. agencies. Read more. Go »

Alternative Methods Accepted by US Agencies

SHARE THIS: 18
https://ntp.niehs.nih.gov/go/regaccept

The table below includes:

- Methods for chemical safety testing that are accepted by U.S. and international regulatory authorities as replacement, reduction, or refinement alternatives to required animal tests
- Guidances to support replacement, reduction, or refinement alternatives to animal use for required testing: these documents are recommendations that do not necessarily establish legally enforceable responsibilities

An overview of non-animal methods that have been proposed for regulatory safety or efficacy testing of chemicals or biological agents can be found in the Tracking System for Alternative Methods (TSAR) resource, provided by the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM). TSAR tracks progress of an alternative method from submission for validation through to its final adoption by inclusion into the regulatory framework.

NICEATM and ICCVAM interact with EURL ECVAM through the International Cooperation on Alternative Test Methods.

Show 25 entries | Search: Alternative Methods

Toxicity Area	Method	ICCVAM or ICCVAM Agency Contributions	Regulatory Acceptance/ Endorsement and Applicable Regulations
Acute Dermal Systemic	Acute dermal toxicity (includes provisions for waiving test and reducing	NICEATM and ICCVAM scientists participated in	• U.S.: Accepted via OECD Test Guideline 402 (1987, revised 2017)
...			
Biologics Testing	Serum neutralization test for potency testing of inactivated veterinary rabies vaccines (reduction and refinement of animal use)	ICCVAM workshop in 2011	• EU: Published in European Pharmacopoeia Monograph 0451 (2012)

Showing 1 to 25 of 128 entries

Previous 1 2 3 4 5 6 Next

In summary…

- Information about the model and data are publicly available to the extent possible and reviewed by independent third parties
- The purpose of the model is clearly identified
- The technical aspects of the model have been characterized, and the model captures key aspects of human biology or mechanisms of toxicity
- The model shows concordance with human data or across multiple methods

Increased Scientific Confidence

Confidence in a method should be determined with the species of interest (humans) in mind