What do we need and what do we expect from a new approach methods-based point of departure?

Katie Paul Friedman, PhD

Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency

Presentation to the IIVS/PETA Webinar Series

March 12, 2025

What do we need and what can we expect from a NAM-based point-of-departure, or POD_{NAM}?

Needs

- Rapidly address data-poor chemicals
- Inform priority for additional information
- Alternative to 90d repeated dose study
- Flexible "toolbox"
- Protection

Expectations

- Use of existing data to benchmark protection and prediction
- NAMs cannot predict traditional animal study PODs with less error than traditional animal study PODs replicate themselves
- Reproducibility

Outline of presentation

- **What we need from a POD_{NAM}**
  - Understand how traditional POD are used
  - Identify a "toolbox"
  - Evaluate protection and predictivity of POD_{NAM}
- **What we expect from a POD_{NAM}**
  - Prediction of traditional animal-based POD will not be perfect
  - Traditional PODs have uncertainty and variability also
  - Target toxicities may be "flagged" but definitive evaluation is separate from POD_{NAM}
- **Conclusions**

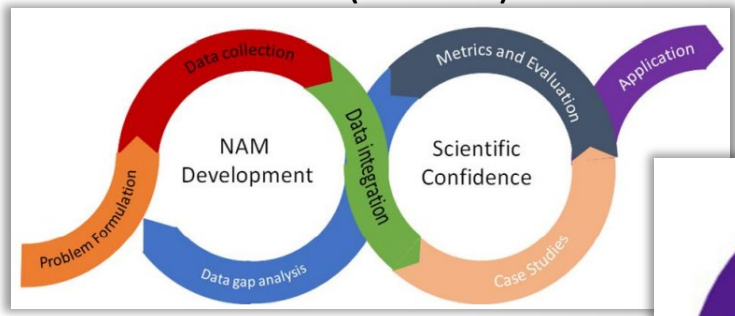
What do we need from a new-approach methods-based POD (POD_{NAM})?

Frameworks indicate comparison of POD_{NAM} with traditional studies to build scientific confidence

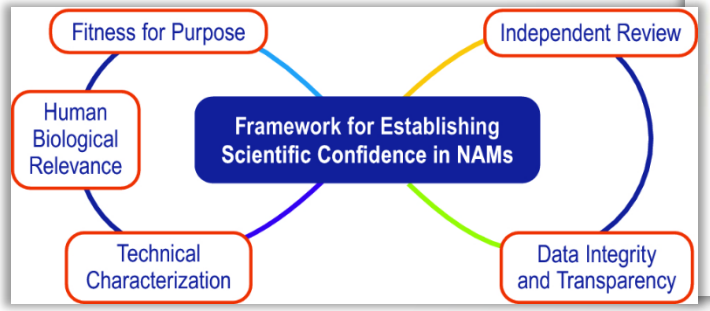
- In Section 4(h) in the Lautenberg amendment to Toxic Substances Control Act:
  - *"…Administrator shall reduce and replace, to the extent practicable and scientifically justified…the use of vertebrate animals in the testing of chemical substances or mixtures…"*
  - New approach methods (NAMs) need to provide *"information of equivalent or better scientific quality and relevance…"* than the traditional animal models

- Multiple frameworks suggest scientific confidence may depend in part on characterization of NAM performance in comparison to traditional animal study performance.
- Improving confidence by understanding how well pre-clinical data relates to human clinical data may be another benchmark for "equivalent of better scientific quality and relevance"

US EPA NAMs WorkPlan (2020-2021)



Parish et al. (2020). 10.1016/j.yrtph.2020.104592



van der Zalm et al. (2022). 10.1007/s00204-022-03365-4

Traditional animal-based PODs for systemic toxicity are used for protection, regardless of study type

- Regulatory decisions often made on basis of body/organ weight changes interpreted as adverse
  - Not interpreted as **predictive** of a similar % body/organ weight decrease in humans
  - Safety factors included to account for uncertainties

- Animal data have been used in a **protective** manner
  - Mechanism of action generally not included for weight changes or histopathology
  - Even in tests for specific toxicity types, nonspecific endpoints or effects may be used to support selection of a POD that is protective

Source of PODs	% studies where non-specific endpoints/effects used for POD
498 chemicals, US EPA IRIS Non-cancer	33% Body wt
608 chemicals, non-cancer Chiu et al. 2018	29% Body wt + NT organ wt
55 chemicals EOGRTs ECHA 2023	24% Body wt + NT organ wt
331 chemicals MGR, US EPA ToxRefDB	73% Body wt + NT organ wt
839 chemicals DEV, US EPA ToxRefDB	55% Body wt + NT organ wt

From text, Browne P et al. (2024). https://doi.org/10.1016/j.yrtph.2024.105579

NAM-based assessment of systemic toxicity parallels current practice

- Some tests are designed to provide systemic toxicity information
- Systemic toxicity may be observed in studies designed to be used to measure specific toxicities

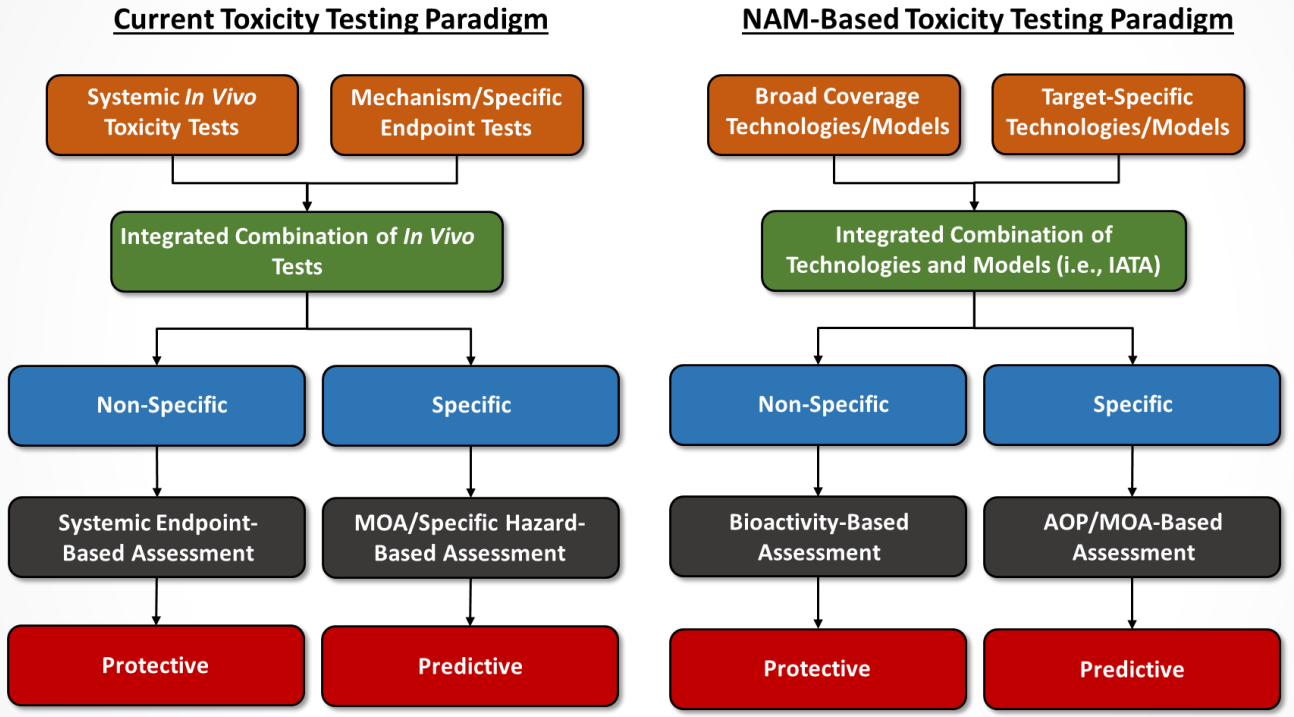
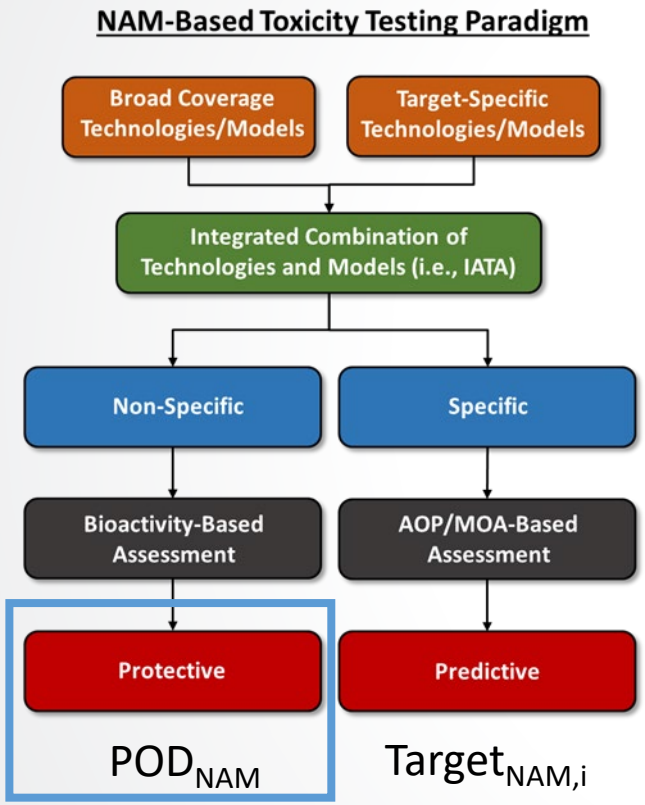


Figure 2, Browne P, Paul Friedman K, Boekelheide K, Thomas RS. (2024). https://doi.org/10.1016/j.yrtph.2024.105579

- NAMs can detect non-specific effects (systemic toxicity)
- NAMs can detect specific effects (defined apical endpoint)

Future safety assessments are anticipated to use batteries of protective NAMs or predictive NAMs, leading to a combined approach with both non-specific and specific NAMs as appropriate to address different regulatory and health protection goals

For systemic toxicity, we need a NAM-based point-of-departure (POD_{NAM}) estimate that…



Is <u>protective</u> of non-specific effects

- developed based on a battery of assays covering many biological targets and processes
- informed by multiple technologies while maintaining resource efficiency

Though the mechanism, mode of action, or type of toxicity may not always be understood (or may have been mischaracterized) using these *in vivo* approaches, the data support the conclusion that the tested chemicals alter biology of living organisms and setting limits for allowable exposures based on these data are <u>generally protective of human health</u>.

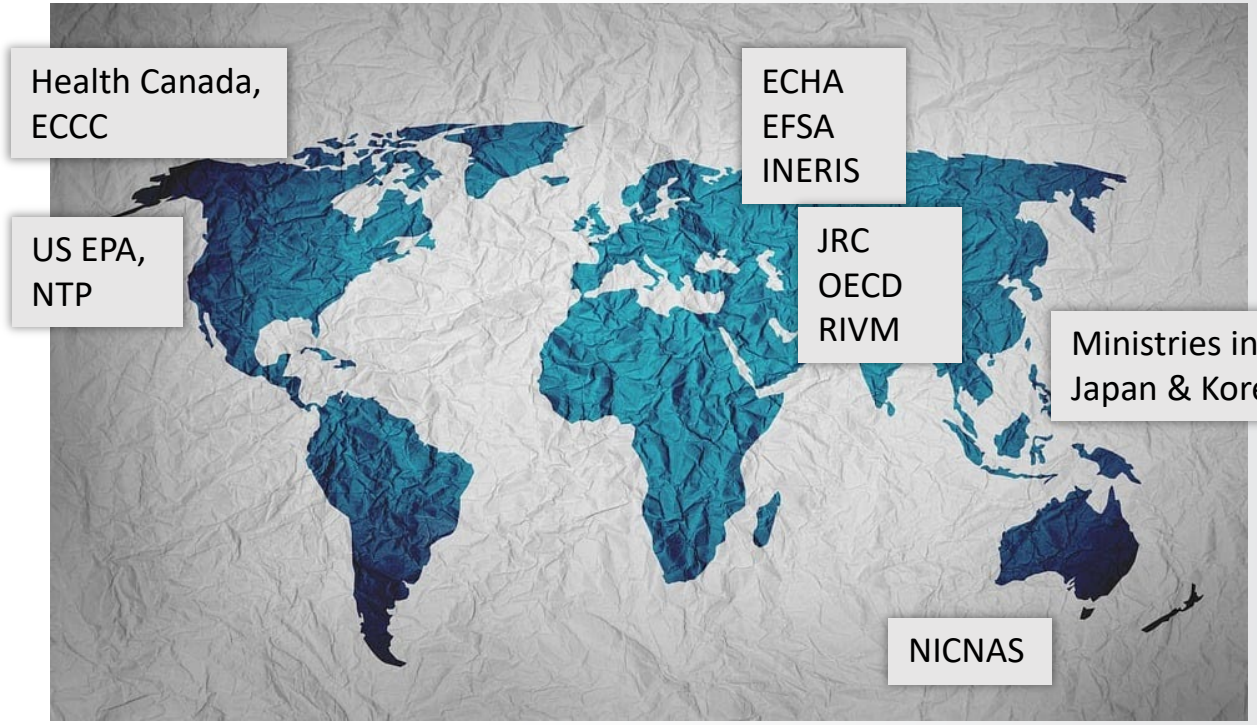
Practical derivation of POD_{NAM} and why a flexible design is needed

Work within the Accelerating the Pace of Chemical Risk Assessment consortium

Goals of the Accelerating the Pace of Chemical Risk Assessment (APCRA) initiative

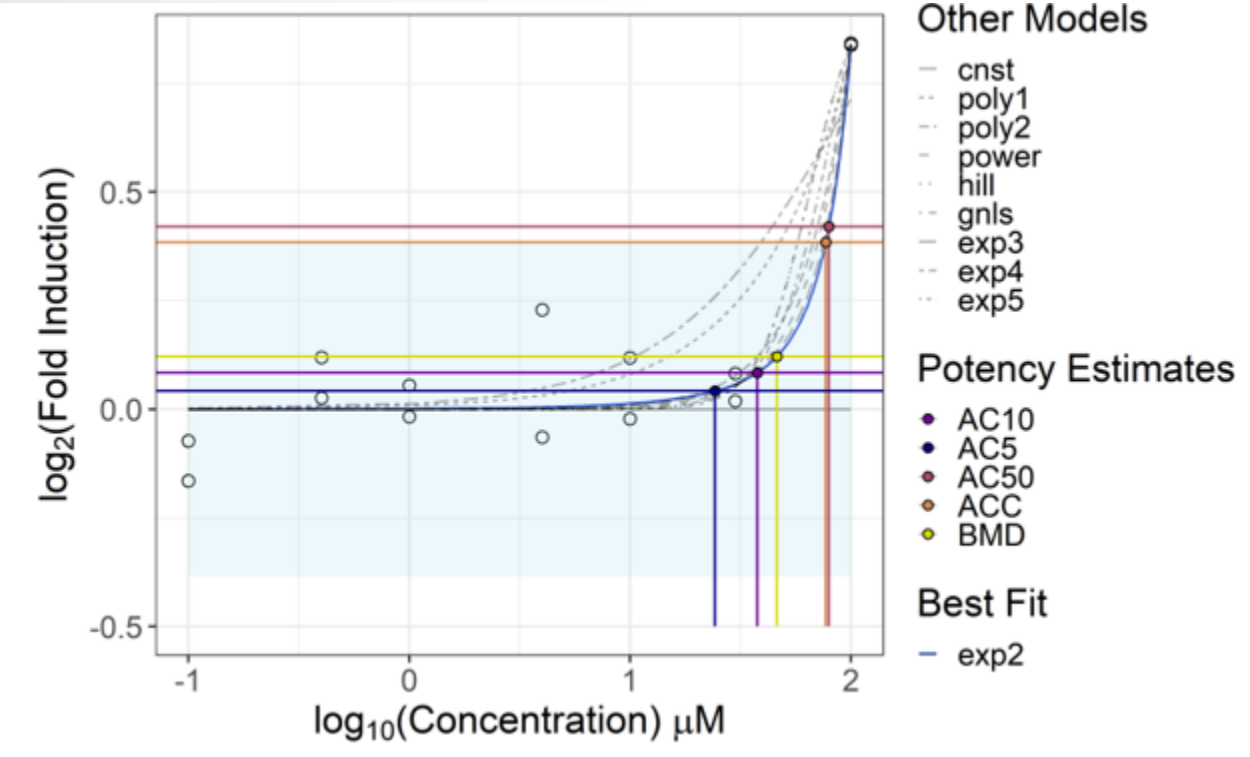
APCRA
ACCELERATING THE PACE OF CHEMICAL RISK ASSESSMENT

- To bring together international regulators to discuss progress and barriers in applying new approach methods (NAMs) to prioritization, screening, and quantitative risk assessment applications
- To formulate and execute collaborative case studies to advance this primary objective



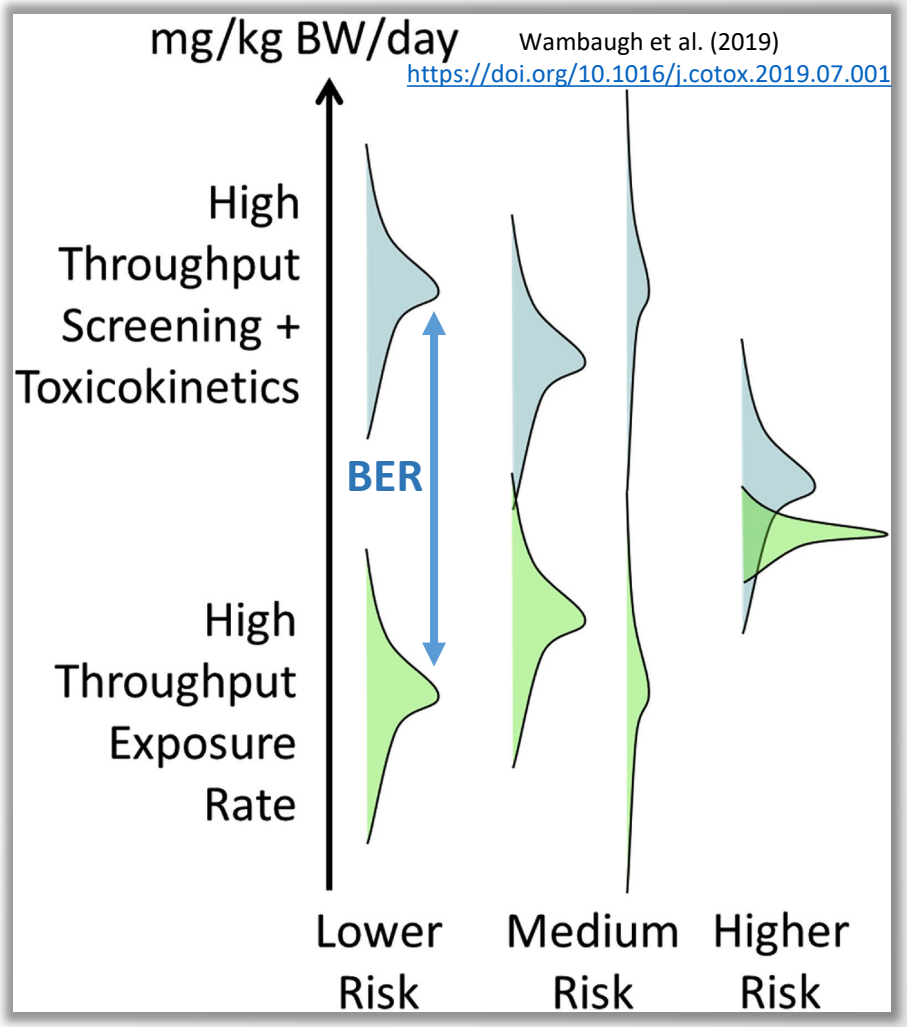
Defining POD and BER

A point-of-departure (POD) describes a point on a concentration (or dose) response curve where the activity moves away from the background and can be a first basis for setting health-protective limits



Feshuk et al (2023)
https://doi.org/10.3389/ftox.2023.1275980

Bioactivity:exposure ratio: quantitative difference between bioactive dose and possible exposure dose



POD_{NAM} toolbox must be adaptable and evolve with new science

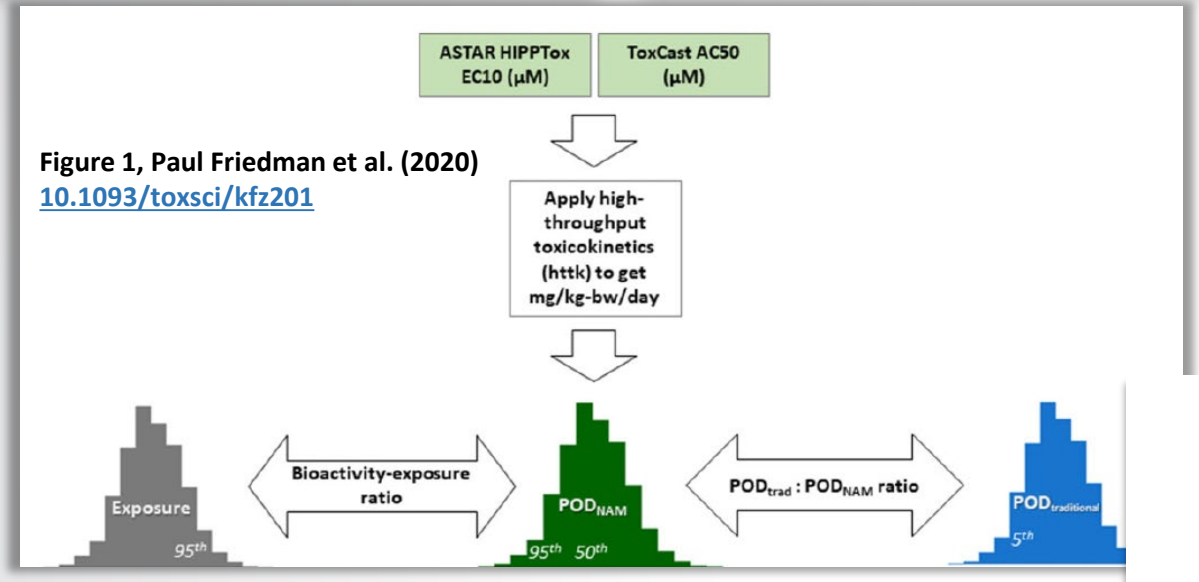


Figure 1, Paul Friedman et al. (2020)
10.1093/toxsci/kfz201

In the retrospective case study with 448 "data-rich" chemicals, we developed a POD_{NAM} based on all AC50s from ToxCast and a high-throughput phenotypic profiling platform.

In the prospective case study with 200 chemicals including more "data-poor" chemicals, we developed a POD_{NAM} based on a specified battery of **broad and targeted assays** and increased the complexity of the in vitro-to-in vivo extrapolation approach. We had to generate the data for many chemicals.

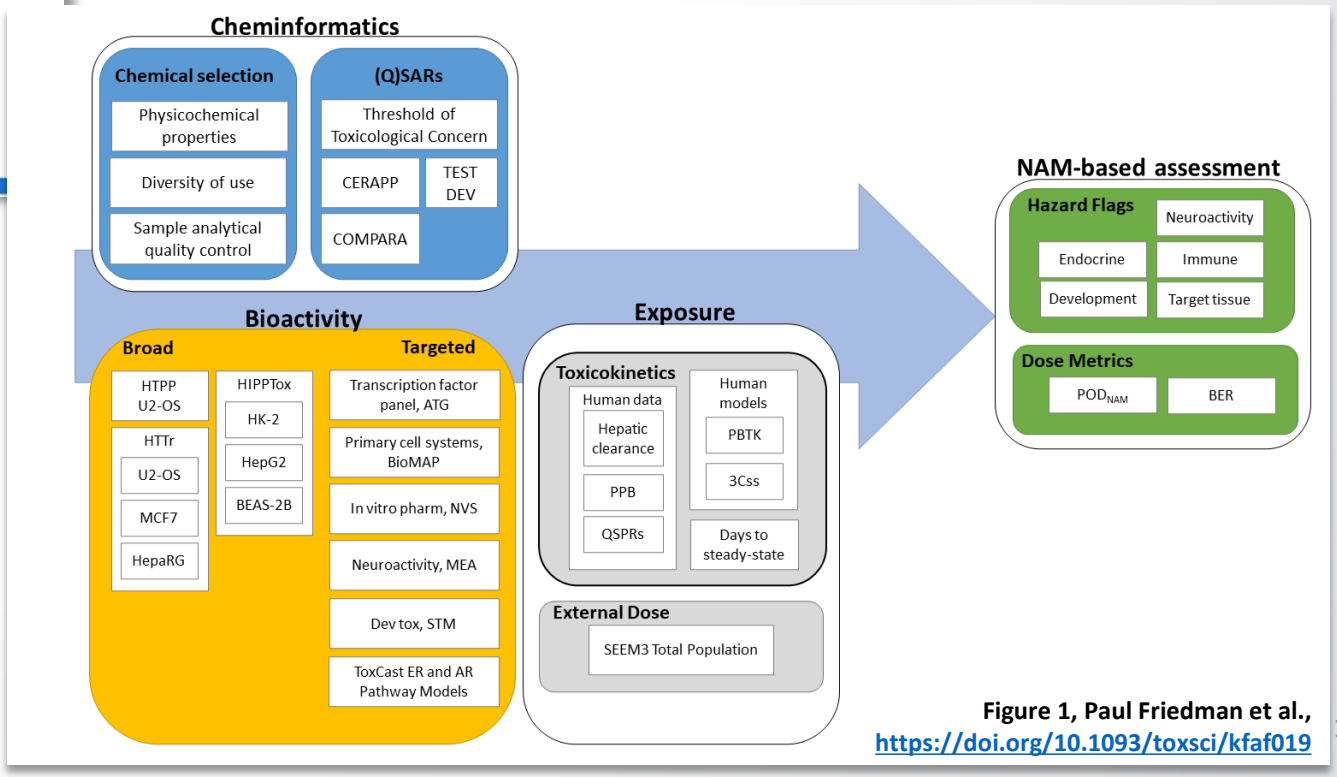


Figure 1, Paul Friedman et al.,
https://doi.org/10.1093/toxsci/kfaf019

Many technical decisions to define an overall POD_{NAM}

Toxicodynamic NAMs

Toxicokinetic NAMs

Evaluation of POD_{NAM}

- Which assays are needed: broad profiling, targeted, both
  - How to demonstrate biological coverage as an alternative to a test like the 90-day subchronic test?
  - How to define quantitative performance metrics?
  - Which cell lines may be needed and/or how many?
- *In vitro* to *in vivo* extrapolation decisions
  - Are some chemicals in or out of domain for high-throughput toxicokinetic modeling?
  - Which toxicokinetic models to consider for the specific context?
  - Should population variability be included in this step or in a separate step, similar to consideration in current chemical assessment practice?
- How to evaluate the POD_{NAM}?
  - How concordant are POD_{NAM} to current traditional (POD_{trad}) from animal studies?
    - How to summarize the POD_{trad} for a large number of chemicals that may or may not have a regulatory POD?

We evaluated myriad options for the assays included

Which assays?	Summary AED50	POD_{trad} percentile	Predictive POD_{NAM}?			Protective POD ratio?						
			RMSE	R2	RMSD	Count	# Greater than 0	# Within ± 2	# Greater than -2	% Greater than 0	% Within ± 2	% Greater than -2
All	min AED50	5th	1.264	0.149	2.226	158	140	90	155	88.6	57	98.1
	min AED50	10th	1.201	0.151	2.326	158	140	83	155	88.6	52.5	98.1
	min AED50	15th	1.138	0.154	2.433	158	143	81	156	90.5	51.3	98.7
	min AED50	20th	1.091	0.154	2.529	158	146	77	158	92.4	48.7	100
	min AED50	25th	1.039	0.149	2.625	158	148	69	158	93.7	43.7	100
	min AED50	30th	0.998	0.153	2.689	158	150	66	158	94.9	41.8	100
	med AED50	5th	1.278	0.13	1.782	158	84	134	146	53.2	84.8	92.4
	med AED50	10th	1.207	0.142	1.803	158	94	135	148	59.5	85.4	93.7
	med AED50	15th	1.143	0.147	1.853	158	104	134	150	65.8	84.8	94.9
	med AED50	20th	1.089	0.157	1.9	158	108	133	152	68.4	84.2	96.2
	med AED50	25th	1.035	0.156	1.962	158	110	131	153	69.6	82.9	96.8
	med AED50	30th	0.994	0.16	2.006	158	112	133	156	70.9	84.2	98.7
Broad profiling	min AED50	5th	1.317	0.076	1.617	158	45	117	121	28.5	74.1	76.6
	min AED50	10th	1.248	0.083	1.622	158	49	122	126	31	77.2	79.7
	min AED50	15th	1.18	0.09	1.648	158	56	127	131	35.4	80.4	82.9
	min AED50	20th	1.124	0.102	1.679	158	62	131	136	39.2	82.9	86.1
	min AED50	25th	1.063	0.11	1.718	158	67	135	140	42.4	85.4	88.6
	min AED50	30th	1.02	0.116	1.753	158	70	135	140	44.3	85.4	88.6
Targeted	min AED50	5th	1.244	0.175	2.165	145	84	118	138	53.2	74.7	87.3
	min AED50	10th	1.184	0.174	2.263	145	91	116	139	57.6	73.4	88
	min AED50	15th	1.124	0.175	2.369	145	102	116	140	64.6	73.4	88.6
	min AED50	20th	1.079	0.174	2.463	145	106	116	141	67.1	73.4	89.2
	min AED50	25th	1.028	0.167	2.558	145	109	114	142	69	72.2	89.9
	min AED50	30th	0.988	0.17	2.622	145	112	114	143	70.9	72.2	90.5
Broad profiling	med AED50	5th	1.326	0.062	1.672	158	68	128	139	43	81	88
	med AED50	10th	1.255	0.073	1.577	158	77	130	142	48.7	82.3	89.9

- **How "predictive" are the POD_{NAM} for POD_{trad}?**
- **How "protective" are the POD_{NAM} for POD_{trad}?**

Using minimum AED50 values results in nearly perfect protection but less concordance (higher RMSE and RMSD values) than using median AED50 values

There are numerous scenarios that result in quantitatively similar outcomes

Using the median of broad profiling assays alone resulted in similar performance to a battery with broad and targeted assays

Table 2, Paul Friedman et al. (2025)

Decisions in the POD_{NAM} derivation can be tuned for protectiveness

Retrospective Case Study (2020)

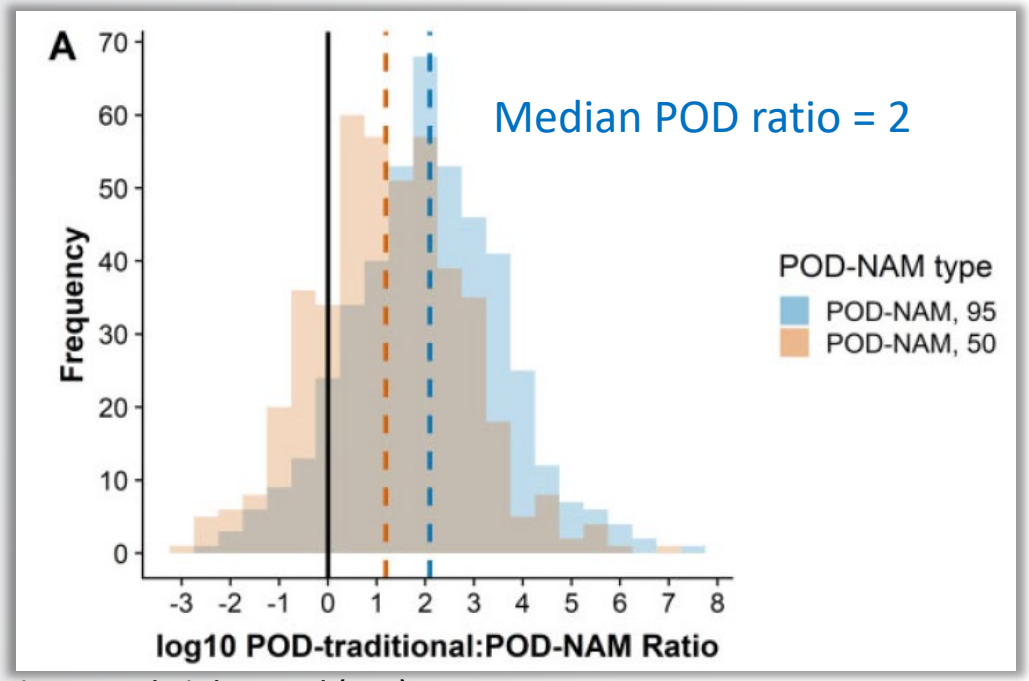


Figure 7, Paul Friedman *et al.* (2020)

- Most toxicokinetically-sensitive individual ($POD_{NAM,95}$) [vs. median individual, $POD_{NAM,50}$]
- 5th%ile of ToxCast AC50 values
- 5th%ile of ToxVal POD values

Prospective Case Study (In Review)

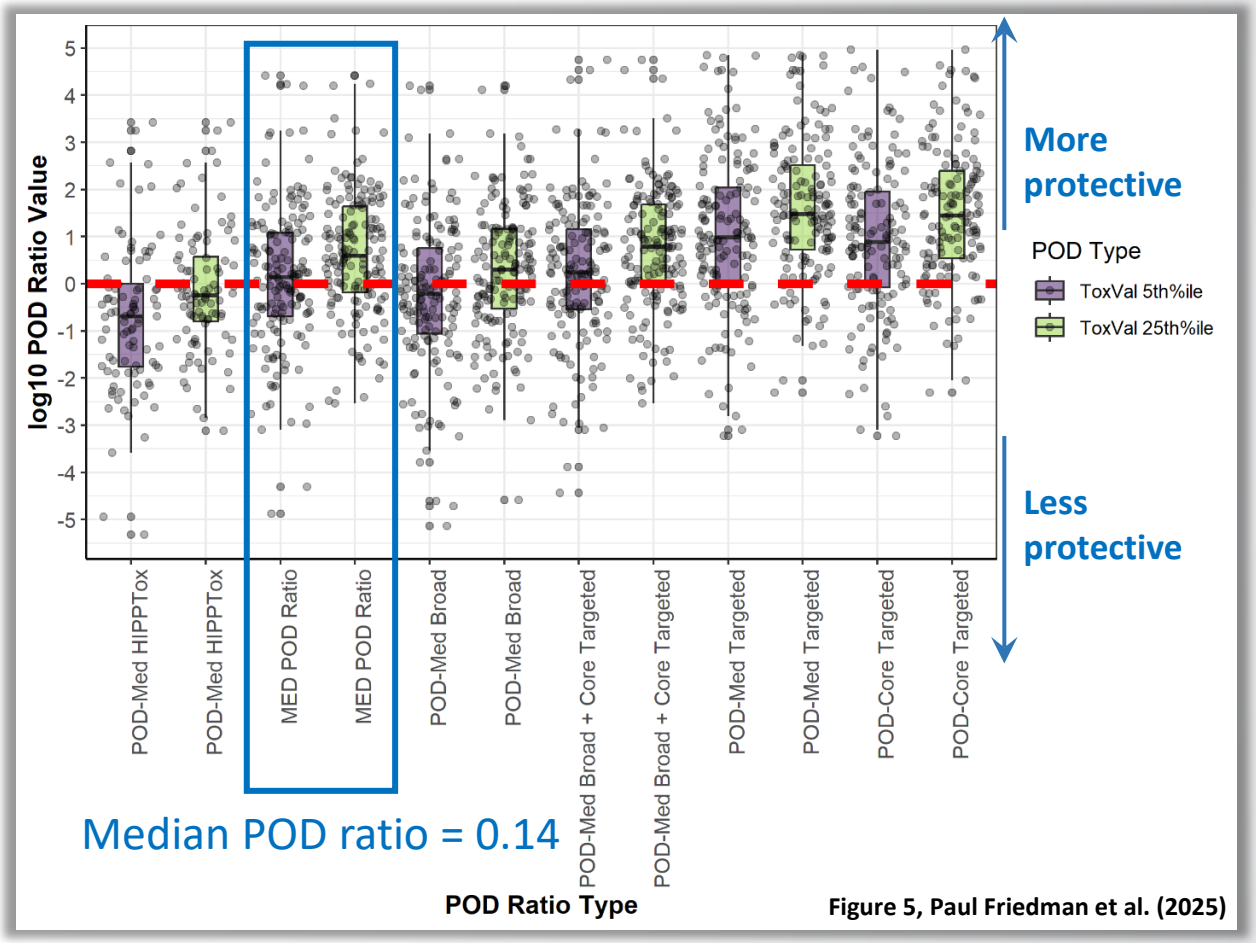


Figure 5, Paul Friedman et al. (2025)

- Median individual, PBTK model if available
- Median of minimum *in vitro* POD per assay, constrained to battery
- 5th %ile of ToxVal values [vs. 25th%ile of ToxVal values]

Using a conservative estimate of an animal-based POD suggests POD_{NAM} can already be protective

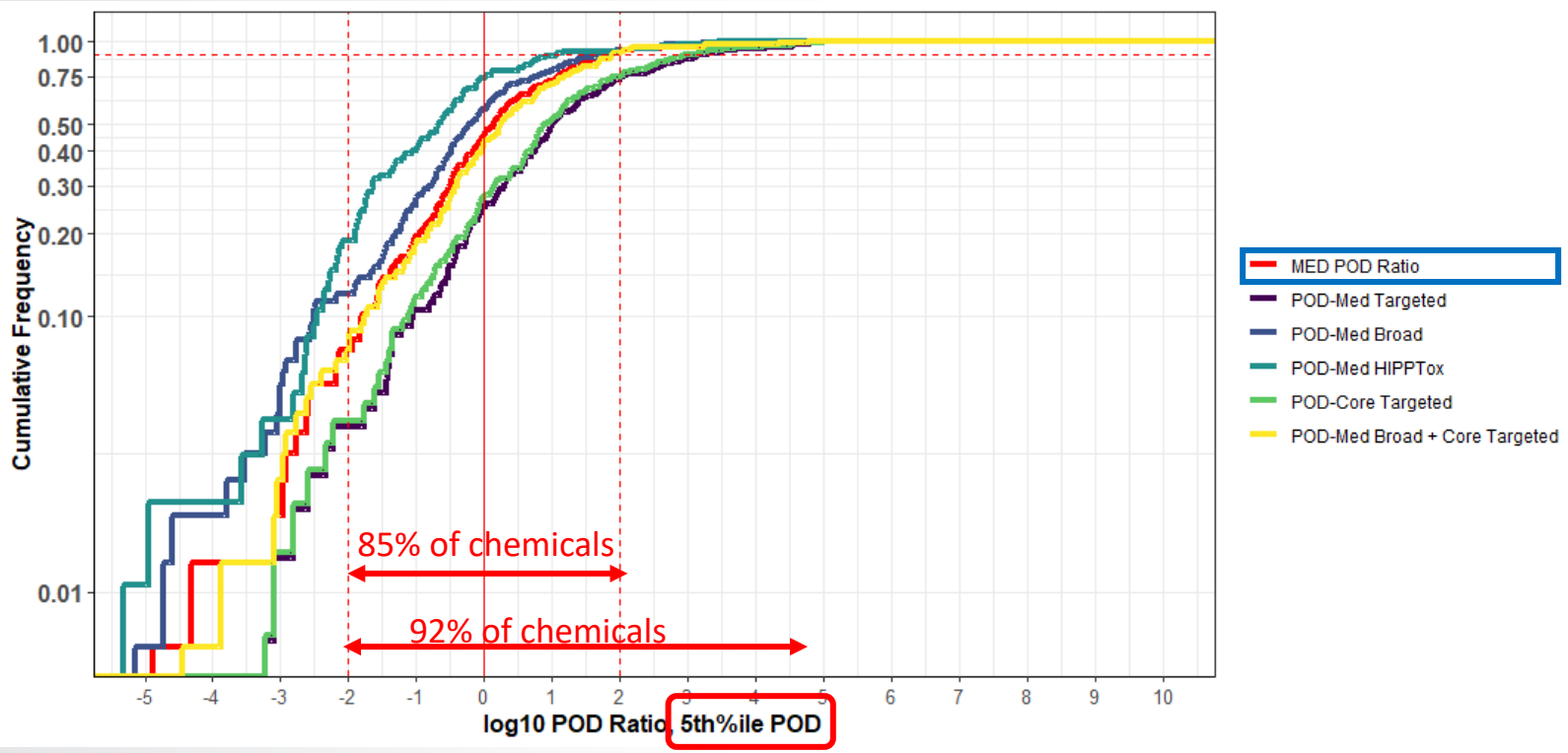


Figure 10, Paul Friedman et al. (2025)

- 12 of 158 chemicals have a log POD ratio < -2 (POD_{NAM} not protective)
- 12 of 158 chemicals have a log POD ratio > 2 (POD_{NAM} very protective)
- "Protective" (POD ratio > -2) for 92% of chemicals

Despite the limitations on the size and chemical diversity of this case study, it is notable that approximately 85% of chemicals (134/158 with a calculable POD ratio) for which POD_{NAM} and $POD_{traditional}$ were available were within ± 2 log10-mg/kg/day of each other.

If we redefine our estimate of an animal-based POD, protectiveness shifts slightly

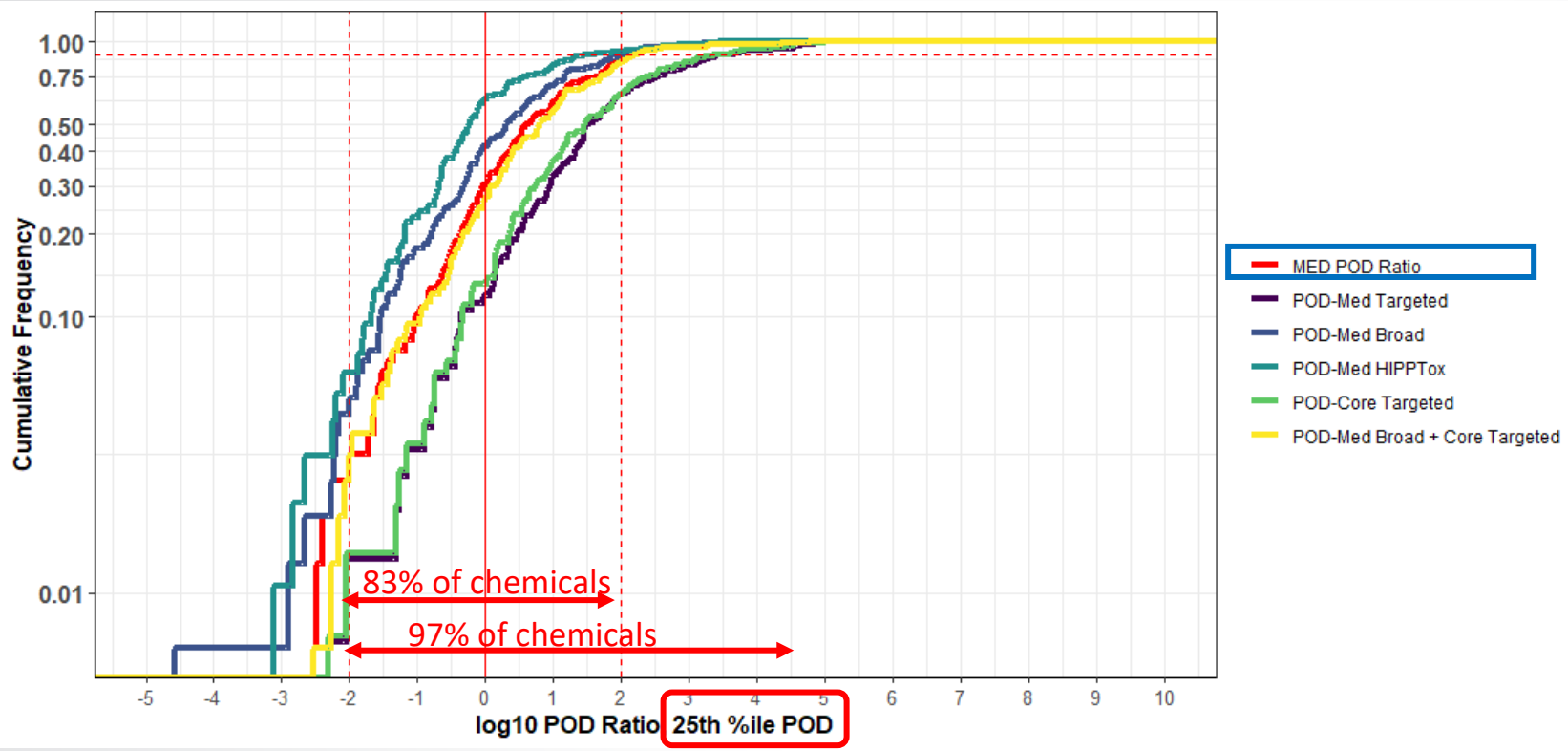


Figure 10, Paul Friedman et al. (2025)

- 5 of 158 chemicals have a log POD ratio < -2 (POD_{NAM} not protective)
- 22 of 158 chemicals have a log POD ratio > 2 (POD_{NAM} very protective)
- "Protective" (POD ratio > -2) for 97% of chemicals

What about potential hazards of interest as a complement to POD_{NAM} like what would be obtained from a 90-day repeat dose toxicity test?

Hazard flag for developmental and reproductive toxicity

- Here, focused on chemicals with BER < 4 (10,000)
- The conceptual goal of these hazard flags was to provide preliminary information, similar to a repeated dose or subchronic study, on the types of target toxicities that might be of interest for the chemical. In part, the hazard flags helped to illustrate the biological learnings from the NAM data generated for this case study. However, these hazard flags represent a conceptual experiment that has not undergone a performance evaluation.
- Bioactivity in *in silico* (0.5) and *in vitro* (1) NAMs can be used to indicate putative endocrine and/or developmental hazard.

DEV = Stemina positive
DEV-S = Stemina positive that is selective
DEV-TEST = TEST model prediction > 0.7
ER/AR = combined *in silico* and *in vitro* indicator

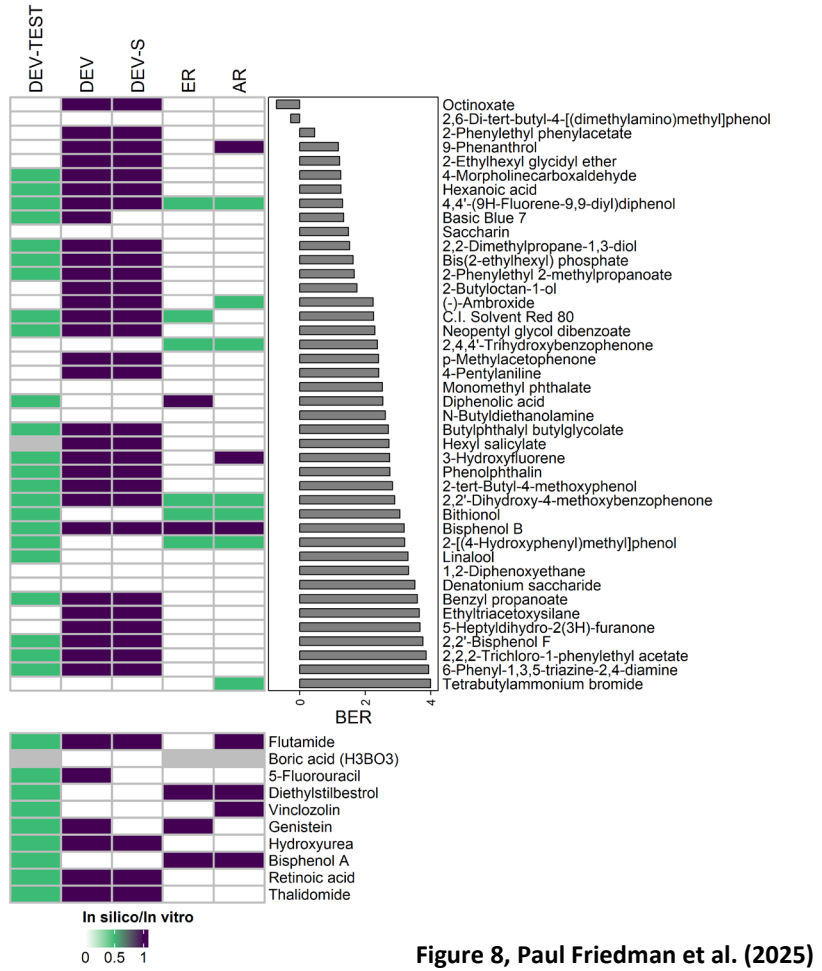


Figure 8, Paul Friedman et al. (2025)

Hazard flags for target cell types can indicate some next directions for further consideration

- Bioactivity in models of organ-based toxicity can be used as hazard flags and can be reviewed by potency
- Limitations in translating in vitro NAMs for neurotoxicity to in vivo effects (lack of BBB in current IVIVE approach)
- Further refinement to hazard flags likely needed as NAMs become available or become fully validated

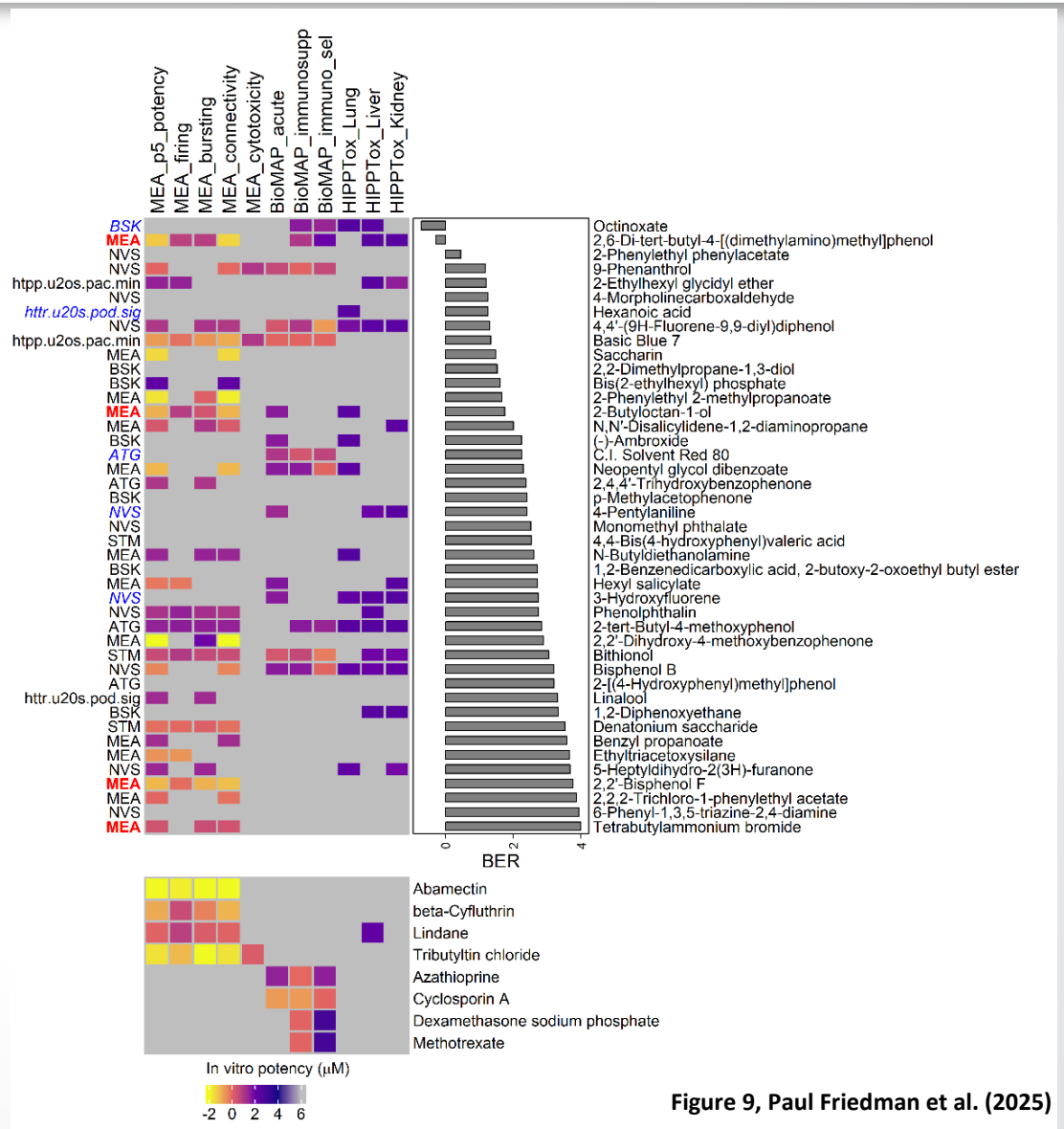


Figure 9, Paul Friedman et al. (2025)

What should we expect from a POD_{NAM} ?

POD_{NAM} estimates may not be highly predictive of animal-based PODs, but there is more to the story

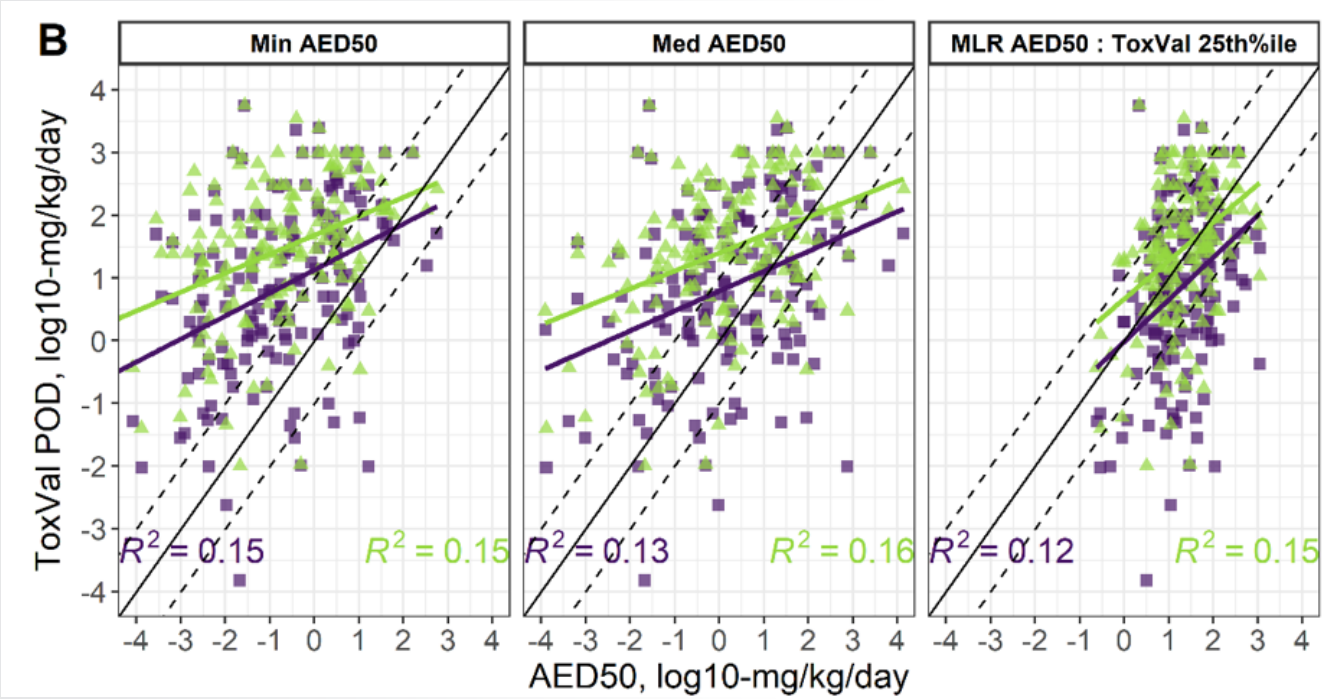


Figure 4, Paul Friedman et al. (2025)

Regardless of how we summarize the *in vitro* and *in vivo* data for comparison, we tend to see: **RMSE ~ 1 to 1.2 log10-mg/kg/day** and **R^2 < 0.2** when we look at large numbers of chemicals

- Animal data we compare to has variance and uncertainty which limits our predictivity
- *In vitro* to *in vivo* extrapolation has variance and uncertainty which limits our predictivity
- *In vitro* bioactivity data, how it is curve-fit, and then summarized, has variance and uncertainty

POD_{NAM} could provide an empirical POD indication for data-poor substances, alongside POD_{TTC} (as we explored previously), POD_{QSAR} for repeat dose toxicity, and other structure-based predictions or alerts.

We should not expect perfect prediction of *in vivo* animal-based study level POD values

- Animal reference POD values have inherent variability, likely on the order of ± 0.5 log10-mg/kg/day based on variance in replicate study PODs

- The % variance explained by detailed meta-data of these *in vivo* studies approaches 55-73% of the variability in replicate study-level PODs for a given chemical

	Total Variance (log_{10}-mg/kg/day)^2	Unexplained Variance (MSE) (log_{10}-mg/kg/day)^2	RMSE (log_{10}-mg/kg/day)	% explained variance	Minimum prediction interval (log_{10}-mg/kg/day)
Range	0.744 - 1.013	0.2 - 0.395	0.448 - 0.629	54.9 - 73.3	± 0.878 - ± 1.23
Median (MAD)	0.825 (0.065)	0.301 (0.068)	0.549 (0.061)	66.1 (4.89)	± 1.07 (0.12)
Mean (SD)	0.838 (0.070)	0.300 (0.055)	0.545 (0.050)	65.3 (4.86)	± 1.07 (0.098)

Based on tables from Pham LL, Watford S, Pradeep P, Martin MT, Thomas RS, Judson RS, Setzer RW, Paul Friedman K. 2020. 10.1016/j.comtox.2020.100126

The minimum prediction interval for an animal systemic effect level is likely ± 1 log10-mg/kg/day, based on the variance in replicate animal systemic effect levels alone

Meaning: we should not expect POD_{NAM} to predict POD_{trad} with less than an order of magnitude error

Table 3
Comparison of performance of the current model with previous publications.

Study	Reference	Number of chemicals	RMSE (log_{10}-mg/kg/day)	R^2
Current	Current	3592	0.70	0.57
Mumtaz et al.	[16]	234	0.41	0.84
Hisaki et al.	[17,18]	421	0.53, 0.56, 0.51	–
Toropova et al.	[19]	218	0.51–0.63	0.61–0.67
Veselinovic et al.	[20]	341	0.46–0.76	0.49–0.70
Novotarskyi et al.	[22]	1,854	1.12 ± 0.08	0.31
Truong et al.	[24]	1247	0.69	0.43

Pradeep P, Paul Friedman K, Judson RS. (2020). 10.1016/j.comtox.2020.100139

Organ level findings are not necessarily predictive across species, but these information can be used for protection

$$\%\ Concordance = \frac{chemical\ with\ positive\ finding\ in\ all\ studies + chemicals\ with\ negative\ finding\ in\ all\ studies}{total\ chemicals\ tested}$$

- Qualitative reproducibility of organ-level effect observations in repeat dose studies of adult animals was 33-88%, depending on grouping.
- Organs associated with more negative chemicals (stomach, thyroid, adrenal) had higher rates of concordance.
- Within-species concordance tended to be greater than within-study concordance.

We should not expect higher rates of concordance of NAMs for prediction of target organ than replicate animal studies can achieve

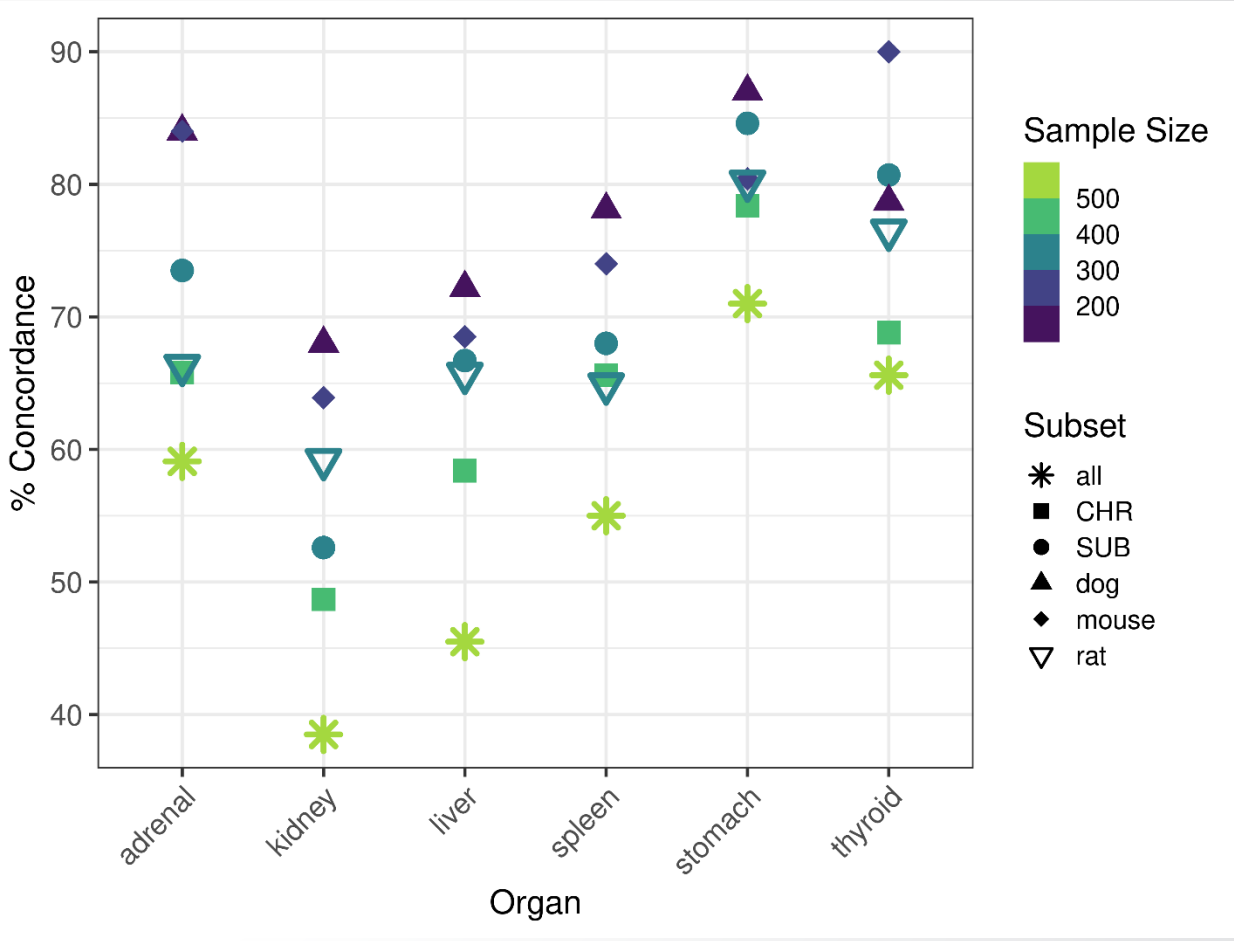


Figure 2, Paul Friedman et al. (2023). 10.1016/j.comtox.2023.100287

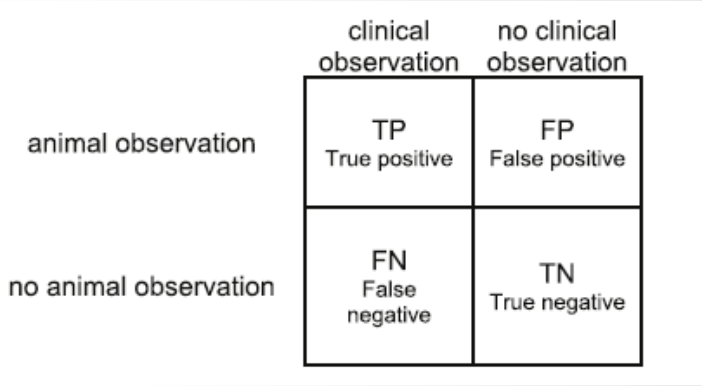
What else can we do to inform expectations of POD_{NAM}?

Studies evaluating qualitative human concordance report high negative predictive value

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]

Monticello *et al*. 2017; https://doi.org/10.1016/j.taap.2017.09.006



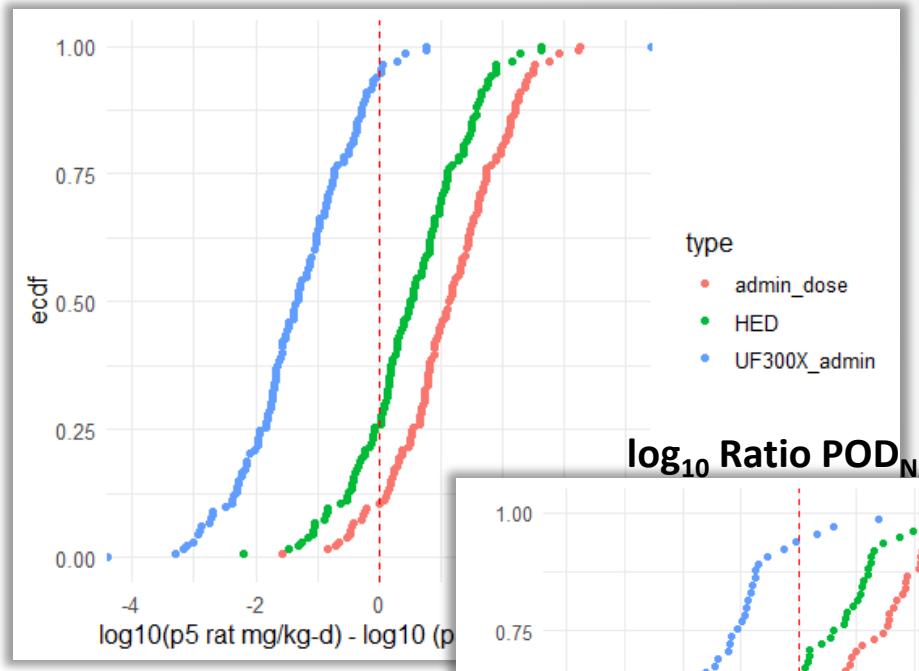
- Nonclinical and clinical trial data have been analyzed from perspective of *qualitative* concordance (i.e., presence/absence of effects)
- Monticello et al. (2017) found low PPV (~30%) but high NPV (~86%)
- Suggests protection over prediction for preclinical to clinical comparisons for drugs that proceed to market

Measure	Equation	Interpretation
Positive predictive value (PPV)	TP/(TP+FP)*100	% positive nonclinical effects with positive clinical effects
Negative predictive value (NPV)	TN/(TN+FN)*100	% negative nonclinical effects with negative clinical effects

POD_{NAM} for pharmaceuticals – work in progress

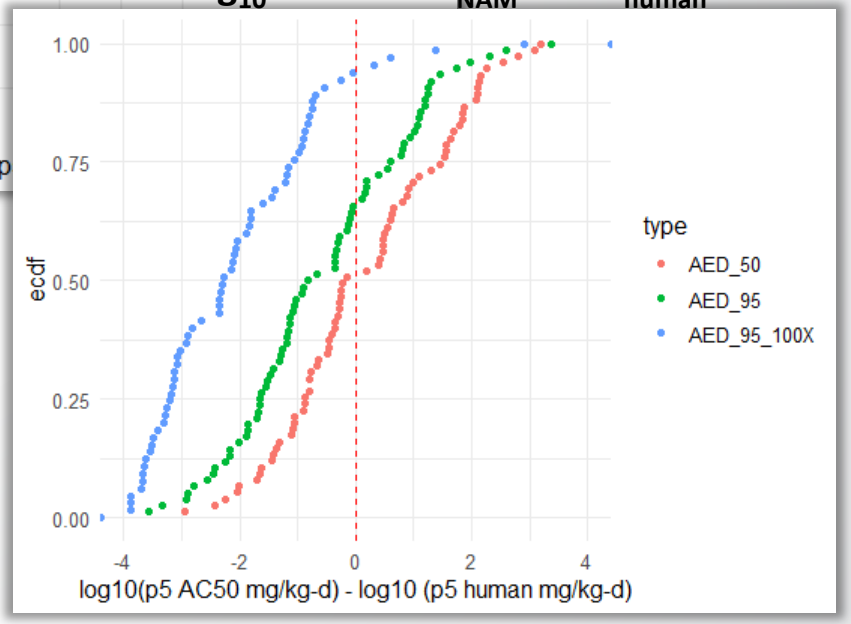
- Traditional animal-based toxicology provides protection, especially for systemic toxicity or non-specific findings lacking mechanistic information.
- Comparison of preclinical to clinical outcomes suggests that animal studies provide protection.
  - POD_{rat} was protective over 90% of the time using a 300x adjustment factor.
- We have developed examples of POD_{NAM} toolboxes.
  - POD_{NAM} was protective over 90% of the time using a 100x adjustment factor.

log_{10} Ratio POD_{rat} : POD_{human}



Similar rate of protection for POD_{human} using POD_{rodent} and 300x as using POD_{NAM} and 100x

log_{10} Ratio POD_{NAM} : POD_{human}



Preliminary results from Weitekamp et al., APCRA

Do not cite or quote

Conclusions

Performance expectations for NAMs (1/2)

- **What we need: start with protection**
  - Neither *in vivo* nor *in vitro* approaches for developing protective POD for systemic toxicity necessarily indicate the mechanism, mode of action, or type of toxicity
  - Using threshold doses for either *in vivo* and *in vitro* biological perturbation can be generally protective of human health
  - Evolving assay "toolbox" that combines broad and/or targeted NAMs can provide enough biological coverage to develop a protective POD_{NAM} for data-poor chemicals
  - Expectations of protection are also consistent with observations from aggregate analyses of outcomes in preclinical and clinical studies

Performance expectations for NAMs (2/2)

- **What we expect: limited linear relationships and variance in POD estimates**
  - Some amount of uncertainty is expected for POD_{NAM} as uncertainty and variability are present in the data we use for comparison and the NAMs themselves
  - The median difference between POD_{NAM} and $POD_{traditional}$ approaches 0, but may be larger for some chemicals
    - Current methods suggest we can typically approximate $POD_{traditional}$ within ± 2 log10-mg/kg/day for 85% of chemicals in large datasets
  - Experience with POD_{NAM} suggests that additional context may increase POD_{NAM} utility
    - *Other estimates of POD, such as $POD_{traditional}$ or POD_{QSAR}, that might enable a consensus POD*

Thank you to my APCRA and EPA colleagues, especially:

EPA

Katie Paul Friedman
Antony Williams
John Wambaugh
Josh Harrill
Richard Judson
Tim Shafer
Rusty Thomas
Chelsea Weitekamp
Alison Harrill
Many previous coauthors on variability work

Health Canada

Matthew Gagne
Alexandra Long
Tara Barton-Maclaren

ECHA

Tomasz Sobanski
Ulla Simanainen
Mounir Bouhifd
Mike Rasenberg

A*STAR

Lit-Hsin Loo
Jia-Ying Joey Lee

JRC

Maurice Whelan