# Quantitative Non-Targeted Analysis for Risk-Based Prioritization of Emerging Contaminants

Jon R. Sobus, Ph.D.,
Physical Scientist, Center for Computational Toxicology and Exposure

# Why Do "We" Need Measurement Data?

- **Measurement data needed to ensure chemical safety**
  - Characterize risk
  - Regulate use & disposal
  - Manage human & ecological exposures
  - Ensure compliance under legal statutes

**Toxic Substances Control Act (TSCA) Compliance Monitoring**

To protect
federal, sta
with statute
import), pr
chemical su
substances

**Safe Drinking Water Act (SDWA) Compliance Monitoring**

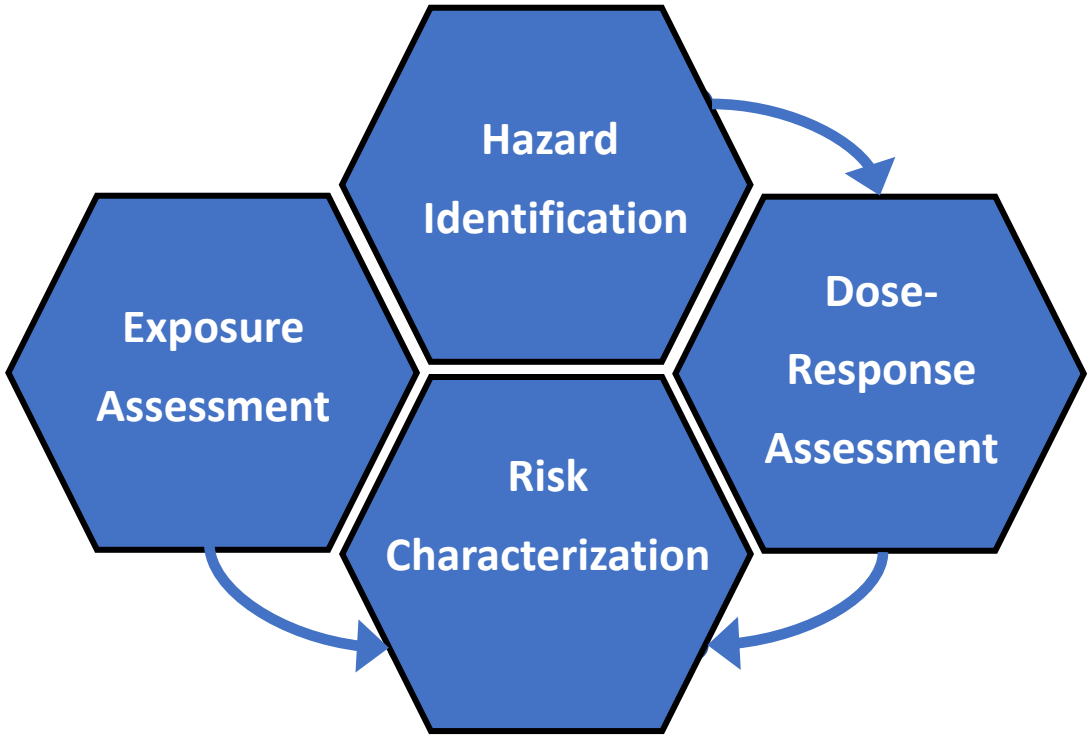
Providing safe drin
states, tribes, publi
certified laboratori
water samples coll
the tribes monitor
Water Act regulato

**Federal Insecticide, Fungicide and Rodenticide Act Compliance Monitoring**

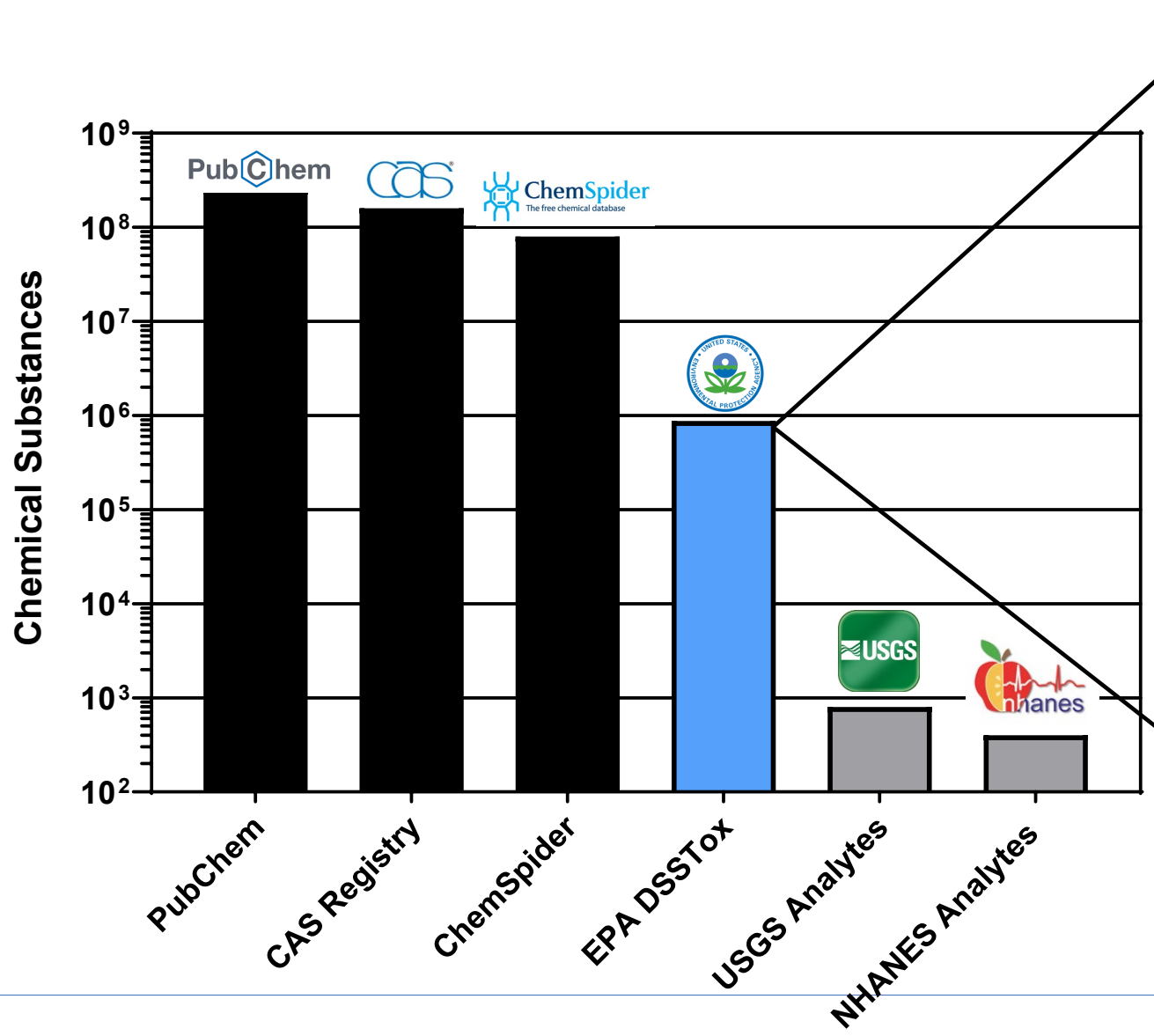
The Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) gives EPA the authority to regulate the registration, distribution, sale and use of pesticides. FIFRA applies to all types of pesticides, including:

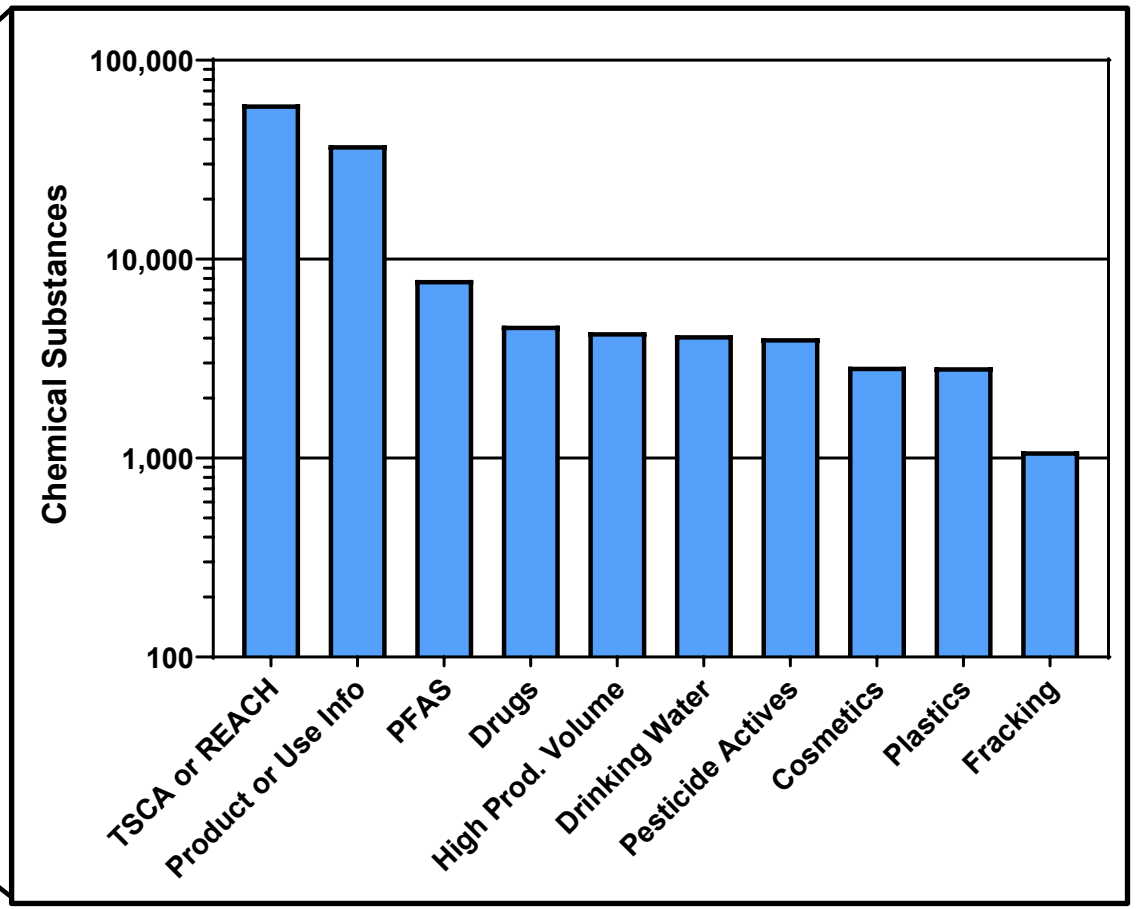
Resources and Guidance Documents

**Chemical Monitoring Needs**



# Data Disparity: Have vs. Need




# Challenges

- High-quality monitoring data are unavailable for most chemicals
- Measurement data traditionally generated using “targeted” methods
- Targeted analytical methods:
  - Require *a priori* knowledge of chemicals of interest
  - Produce data for few selected analytes (10s-100s)
  - Require standards for method development & compound quantitation
  - Are blind to emerging contaminants
  - Can’t keep pace with the needs of 21$^{st}$ century chemical safety evaluations

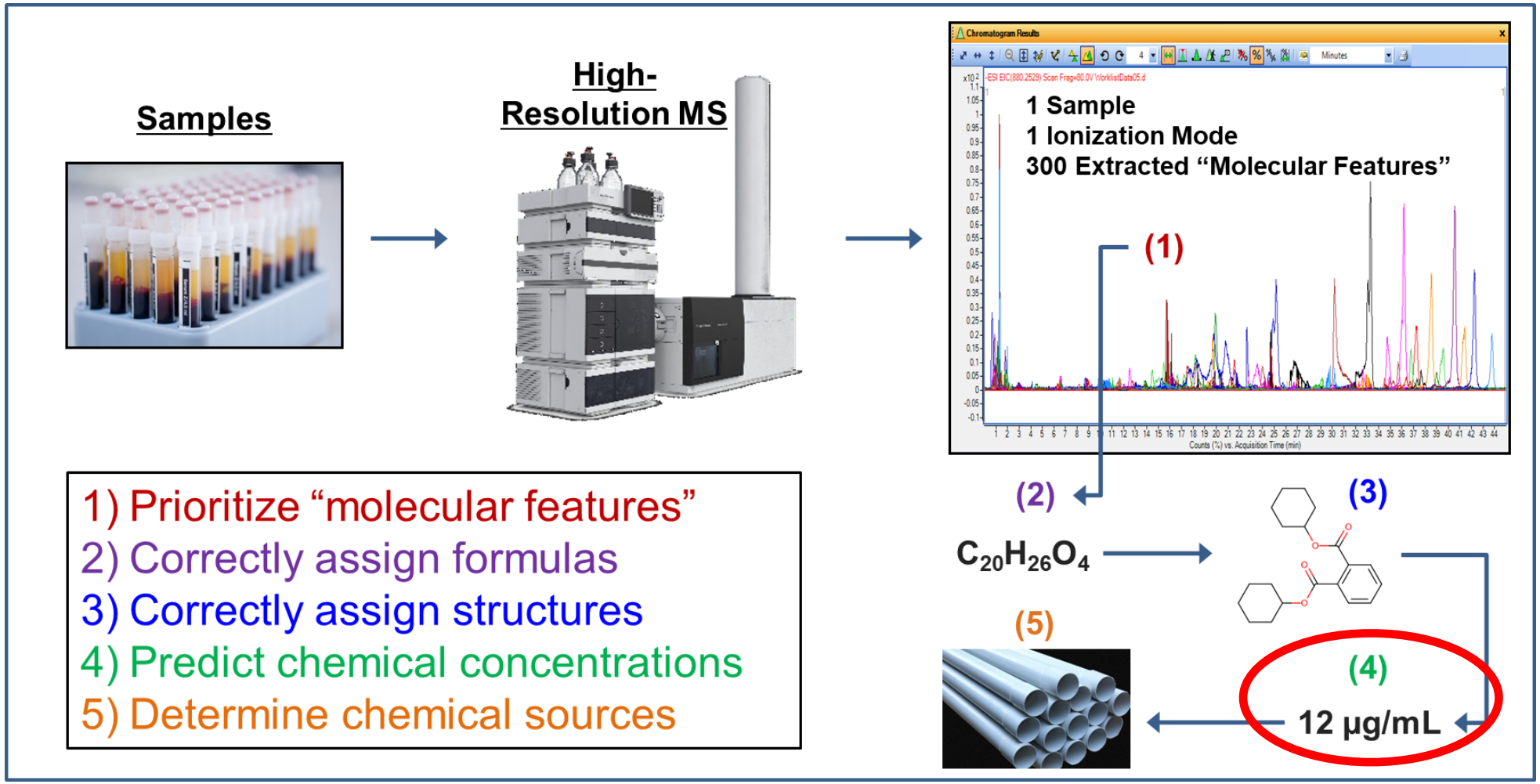
# What's So Great About NTA?

Rapidly screen for "knowns"

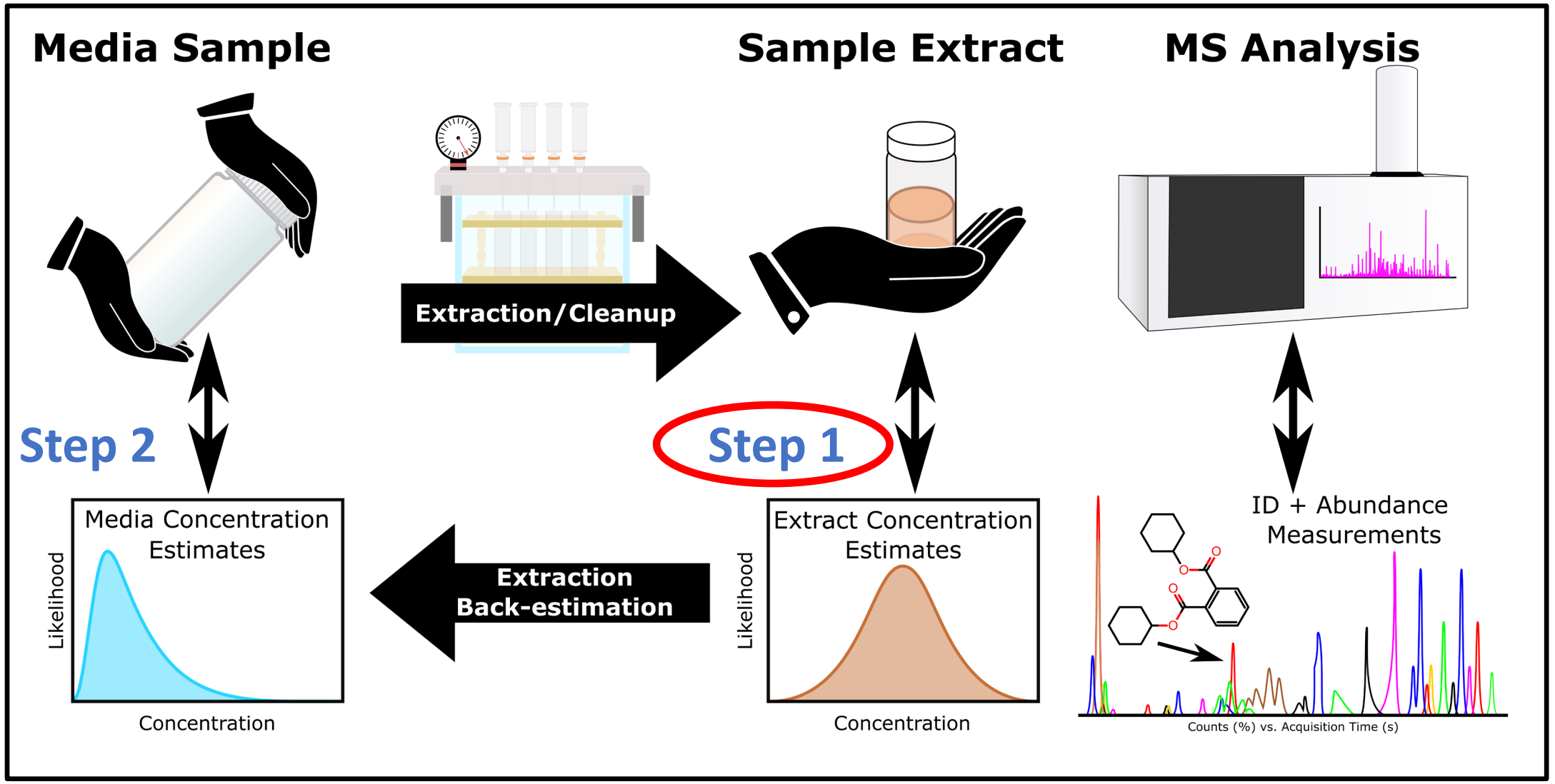
Discover "unknowns"

Uncover historical exposures

Generate source fingerprints…

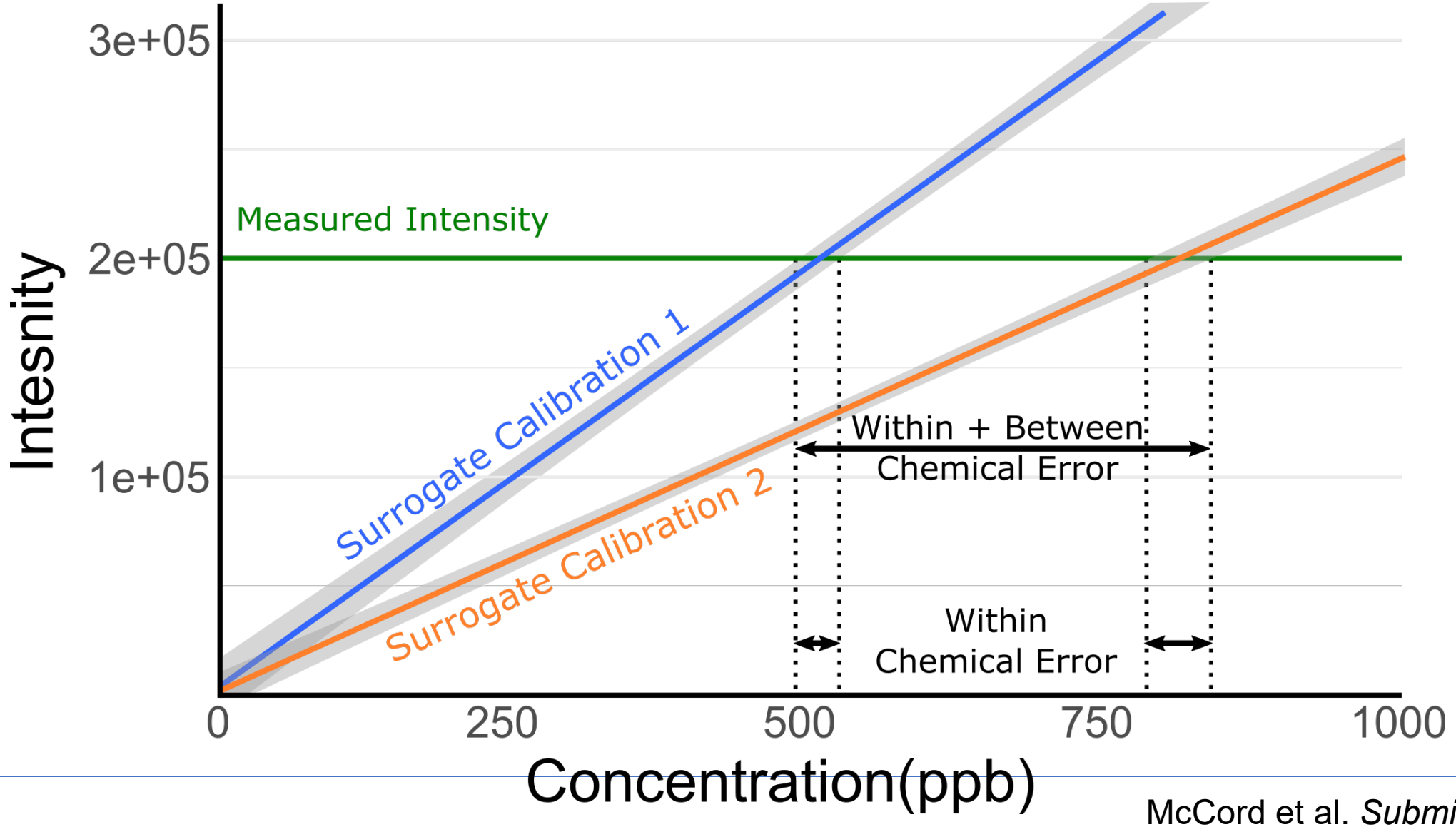


# Quantitative NTA (qNTA) Workflow




McCord et al. *Submitted*

# Step 1 Uses "Surrogate Calibration"



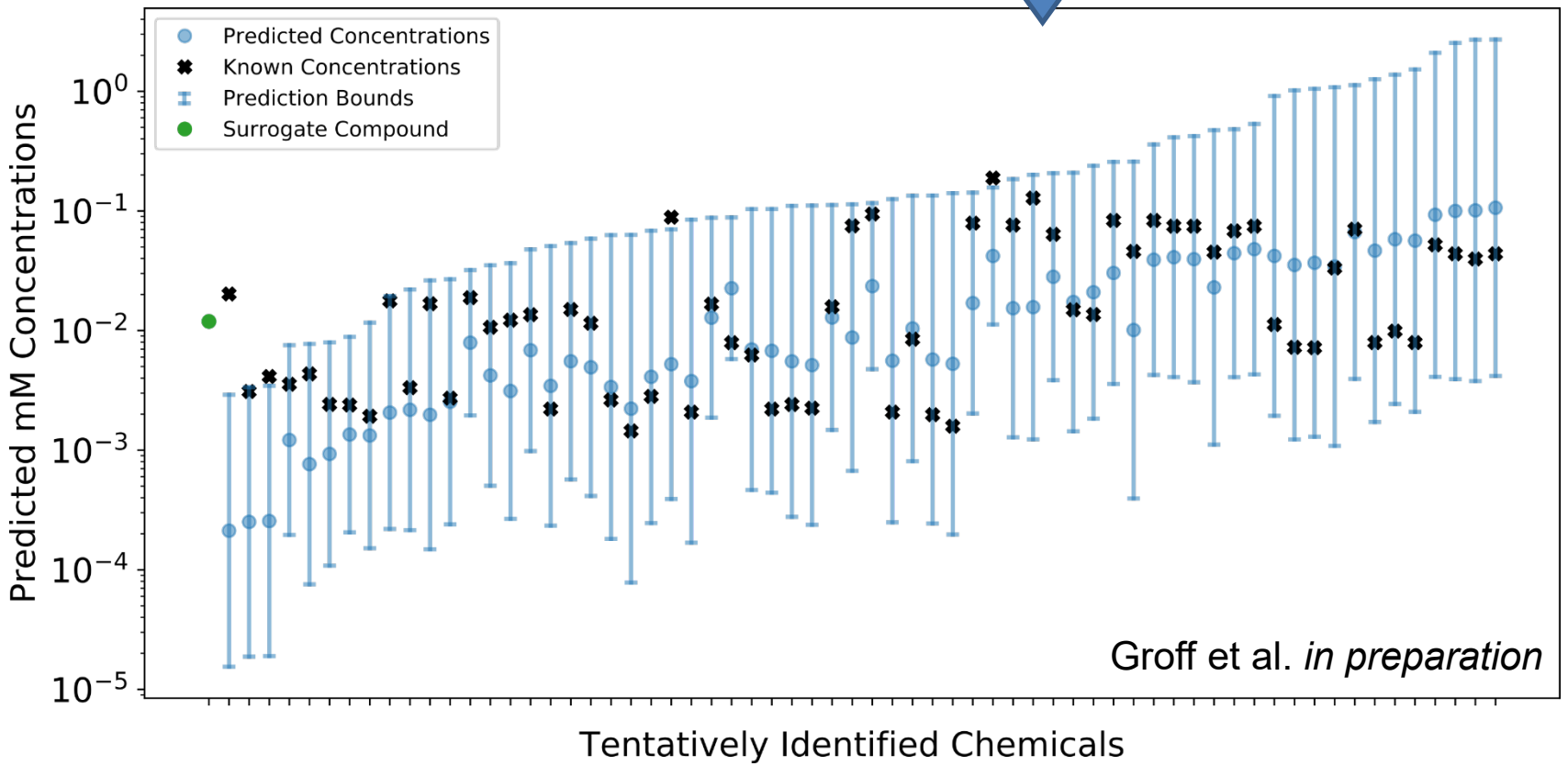

McCord et al. *Submitted*

# Considerations for Surrogate Calibration

- Multiple methods for choosing a surrogate calibrant
  - Single surrogate (i.e., an “average” responder)
  - Structurally similar surrogate
    - Nearest neighbor (e.g., based on elution time)
    - Within chemical class
    - Based on calculated similarity (e.g., Tanimoto index)
    - Based on known parent/metabolite relationship
  - Model-predicted value (e.g., based on expected ionization efficiency)

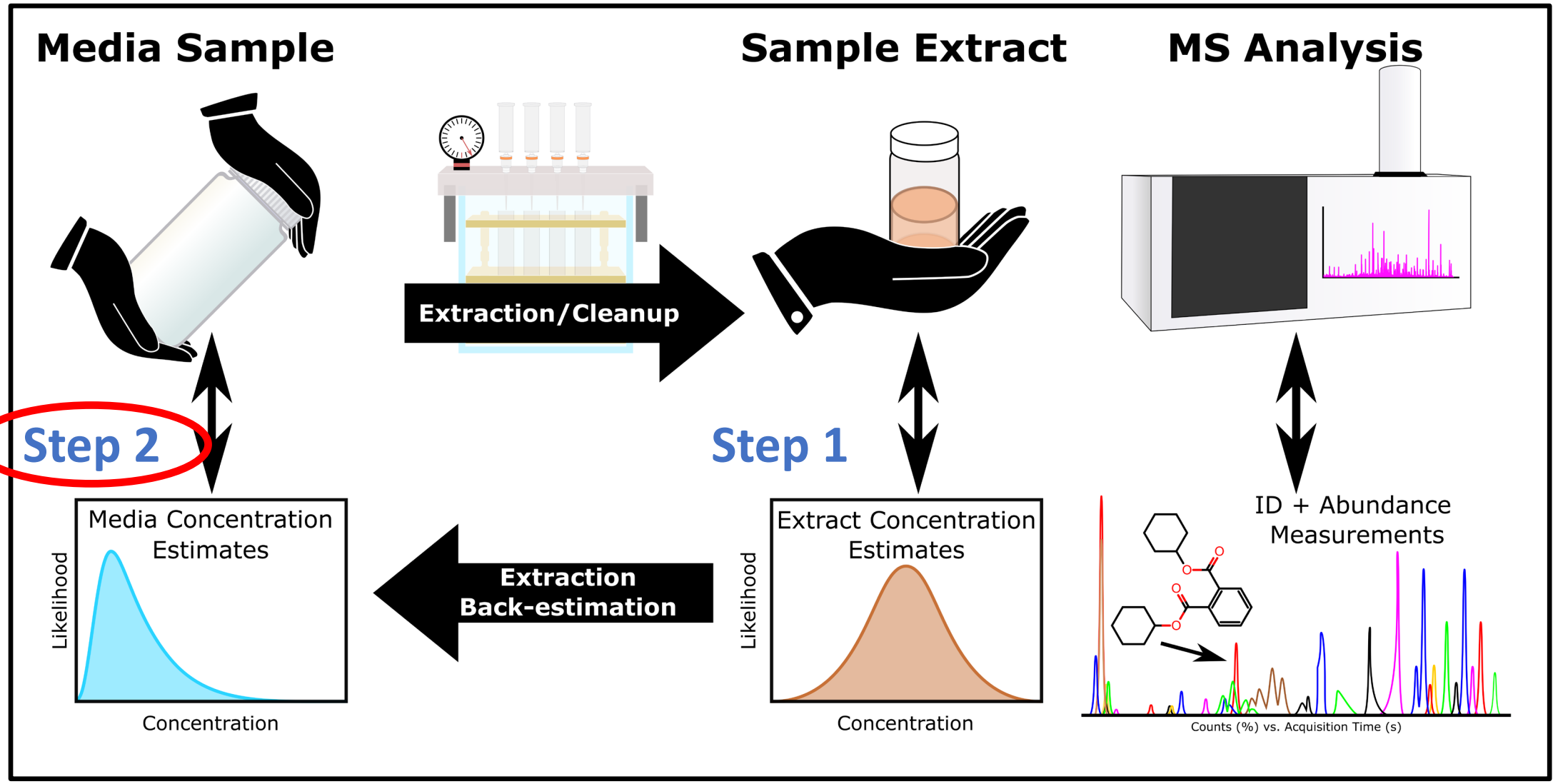
- Prediction error within and between chemicals
  - Affected by sample & batch correction techniques
  - Affected by surrogate selection techniques
  - Consider all error when estimating confidence intervals for individual predictions

# qNTA Proof-of-Concept


- Analysis of Brita filter extracts via GC-HRMS.
- Single surrogate selected and applied to all identified analytes
- Concentration estimates can be above or below true value.
- Prediction intervals used to bound concentration estimates.
- 95% prediction intervals shown; Can use 99%, 99.9%, etc.
- Tentatively identified compounds ranked by upper-bound estimates.
- **Upper-bound estimates compared to level-of-interest to set priorities.**

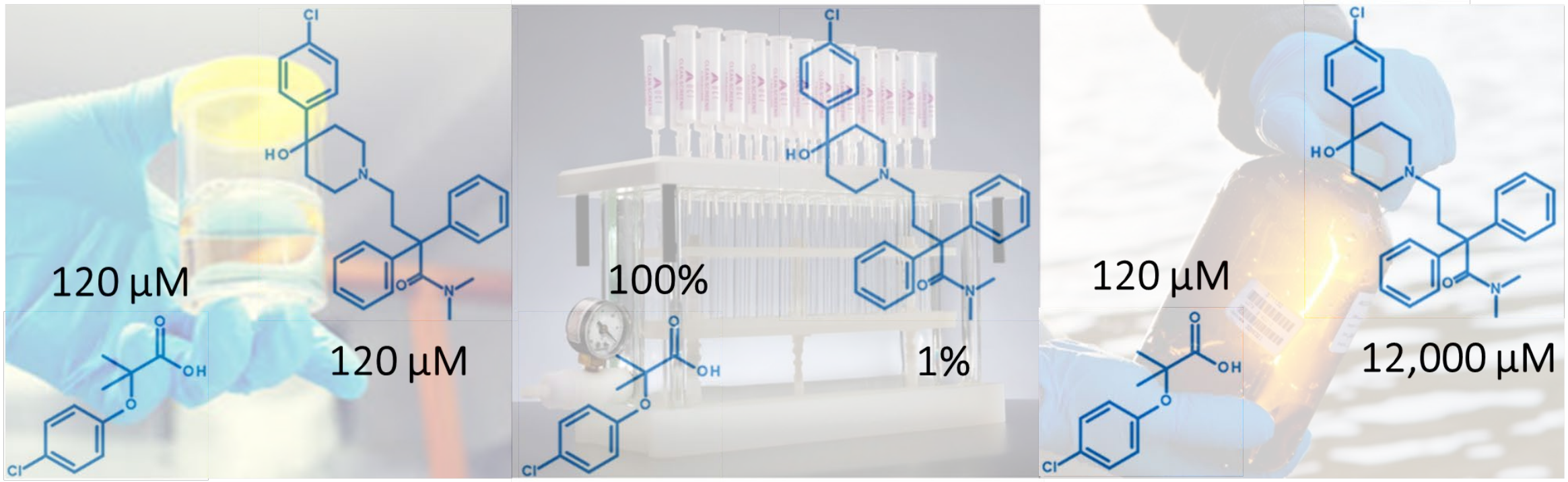
# Quantitative NTA (qNTA) Workflow




McCord et al. *Submitted*

# Why is “Recovery” A Critical Parameter?

Max. Percent Recovery = 100% → known lower bound on media conc.
Min. Percent Recovery = ?% → no upper bound on media conc.




Sample Extract Estimates → Percent Recovery from Media → Media Estimates

# No Existing Models for Predicting Recovery

- General rule of thumb:
  - At least 10× more data points than explanatory variables

- Type of media: 10s
- Conditions of media: 10s
- Extraction solvents: 10s
- Extraction conditions: 10s
- Clean-up procedures: 10s
- Interactions terms (e.g., media × condition × solvent…): ???

- >100,000 possible recovery scenarios → >1,000,000 required data points

- So we can't bound it, and we can't predict it. Now what???

# Defining "Margin of Recovery" (MoR)

## Traditional "Recovery" Definition:

$$\frac{\text{Amount in Sample Extract}}{\text{Amount in Sampled Media}} \times 100 = \%\text{Recovery} \Rightarrow \frac{80\ \mu g}{100\ \mu g} \times 100 = 80\%\ \text{Recovery}$$

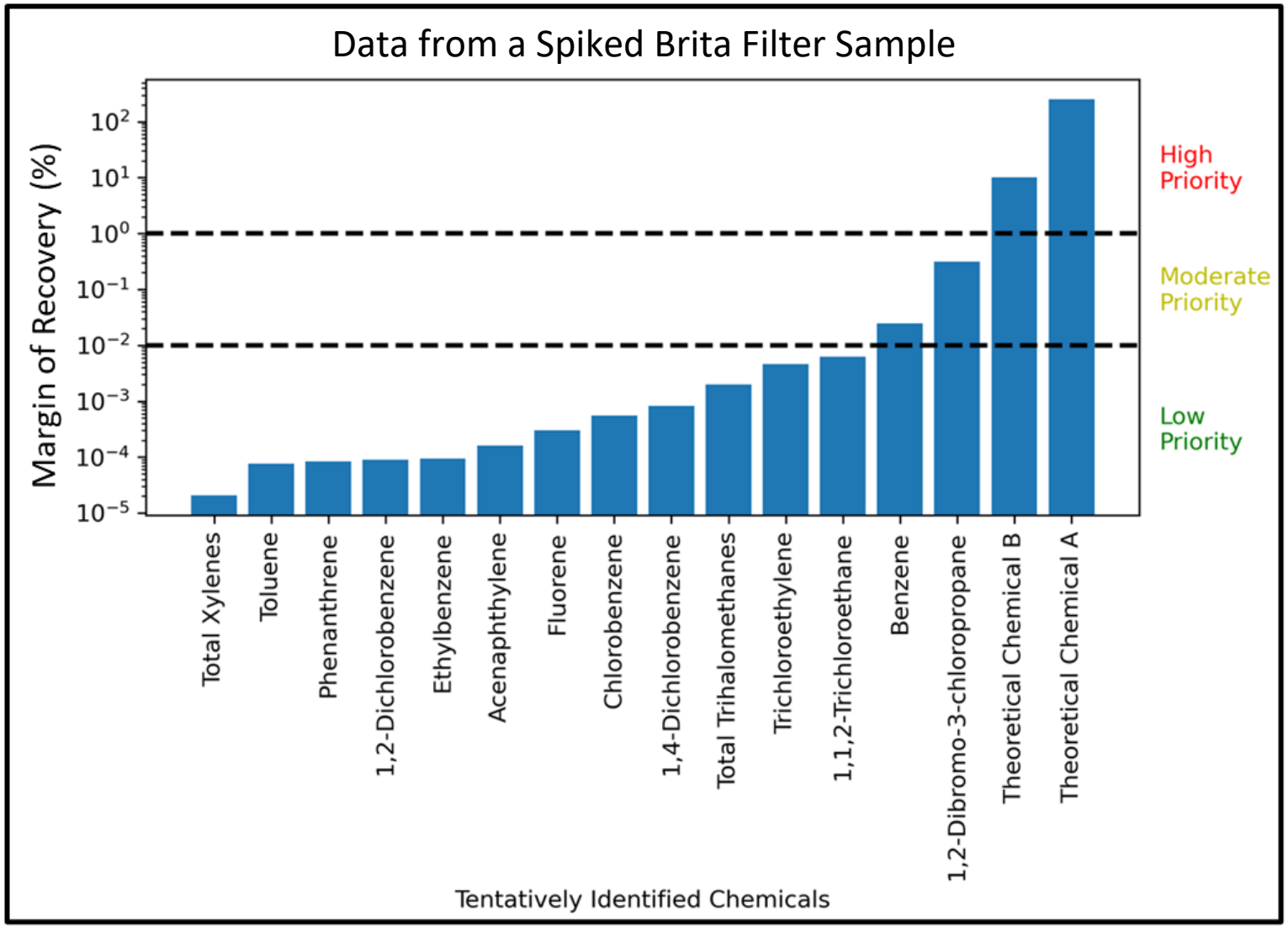
## "Margin of Recovery" (MoR) Definition:

$$\frac{\text{Upper Bound qNTA Estimate (amount in sample extract)}}{\text{Level of Concern (amount in sampled media)}} \times 100 = \%\text{MoR}$$

## *Important interpretation:*

***- What recovery is needed for the qNTA estimate to match the level-of-concern?***
***- Is that calculated recovery plausible enough to warrant further targeted analysis?***

# Example Risk-Based Prioritization

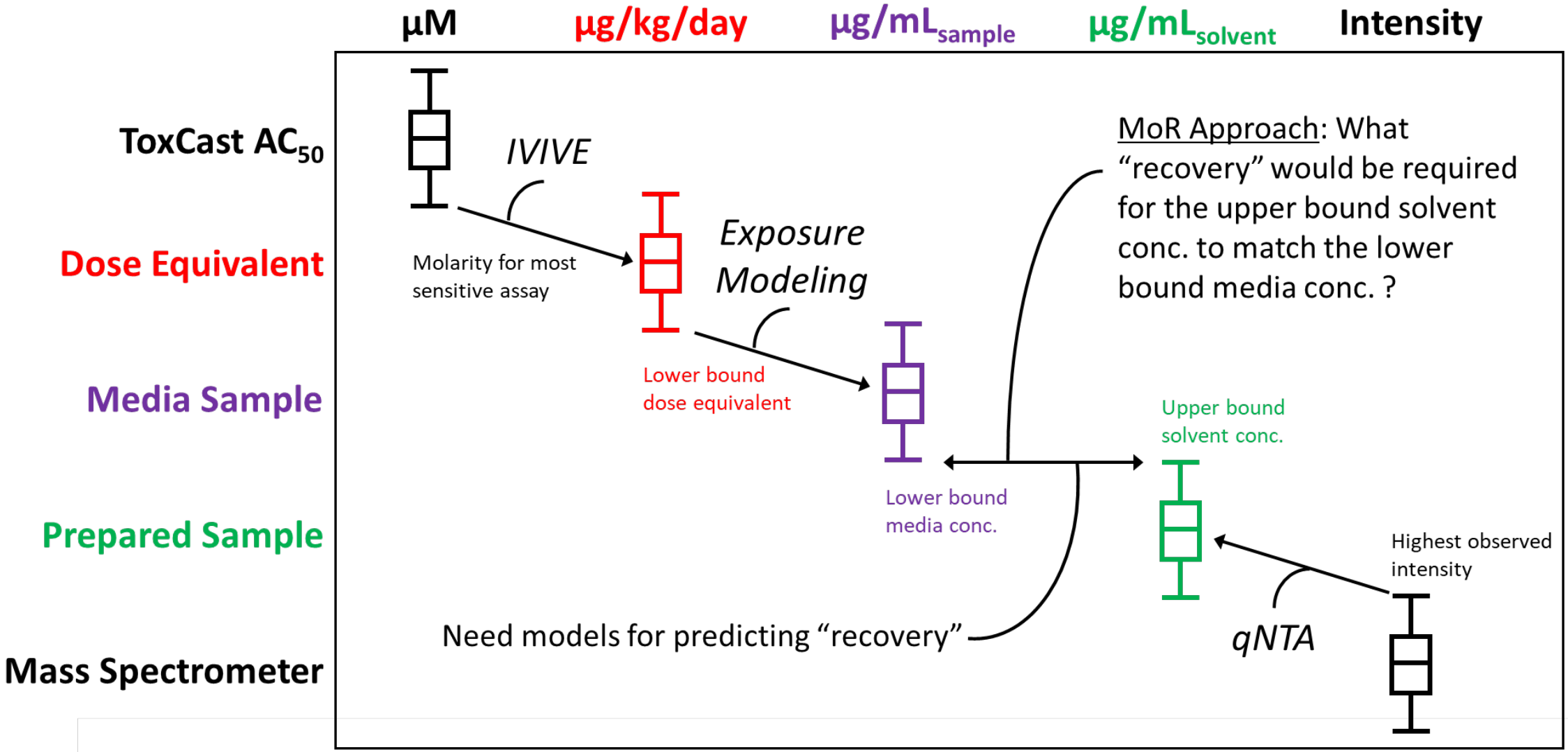



For "**High Priority**" Chemicals, a 1-100% experimental recovery would be needed for the upper bound qNTA estimate to match a drinking water concentration associated with bioactivity/toxicity.

Recoveries < 1% and > 0.01% are considered somewhat unlikely. Chemicals in this range are therefore considered "**Moderate Priority**"

Recoveries < 0.01% are considered highly unlikely. Chemicals in this range are therefore considered "**Low Priority**"

# Conceptual Model for Rapid Risk Evaluation



# The Future of NTA and Chemical Risk Assessment

- The number of labs performing NTA will increase dramatically!

- We're expecting a wealth of NTA data for known (but data-poor) chemicals
  - These data cannot be interpreted using traditional performance metrics
    - **How will risk assessors use new NTA data to support decisions?**

- We're expecting a steady stream of NTA data for newly discovered chemicals
  - Chemical standards won't be readily available (via purchase or synthesis)
    - **How will risk assessors rapidly evaluate the safety of these CECs?**

- ORD efforts will enable translation of NTA data to support Agency decisions

# INTRODUCING... BP4NTA

BENCHMARKING AND PUBLICATIONS FOR NON-TARGETED ANALYSIS

## WHAT IS BP4NTA?

**A consensus organization of >100 members from ~55 worldwide, multi-sector institutions.**

BP4NTA was formed in 2018 to address challenges in non-targeted analysis (NTA) studies using mass spectrometry. We hope to accelerate the broad acceptance and use of HRMS studies in scientific and regulatory communities.

## BP4NTA PRODUCTS

*Visit www.nontargetedanalysis.org for more information.*

**1. Collated resources for new NTA researchers traversing the learning curve**
Access detailed, NTA reference content at
www.nontargetedanalysis.org/reference-content/

**2. The NTA Study Reporting Tool (SRT)**
Access a peer-reviewed tool that aids study design & review of NTA manuscripts at www.nontargetedanalysis.org/srt/

**3. A glossary of commonly used NTA terms, concepts, and performance metrics**
Access the list of NTA definitions at
www.nontargetedanalysis.org/glossary/

## UPCOMING EVENTS

**SETAC SciCon4 2021: Nov. 14-17**

- **Poster:** Non-Targeted Analysis Study Reporting Tool: A New Framework to Improve Reproducibility and Transparency (4.08.13)
- **Session:** Non-Targeted Analysis: Approaches Toward Identification of Chemical Contaminants (4.08.01- 4.08.25)
- **Session:** Non-Target Analysis: Prioritization of Organic Contaminants for Monitoring and Toxicological Studies (4.09.01- 4.09.20)

**SETAC Focused Topic Meeting: May 22-26, 2022 "Non-Target Analysis for Environmental Assessment"**

Follow us on Twitter!
@BP4NTA

TO LEARN MORE ABOUT BP4NTA, VISIT
WWW.NONTARGETEDANALYSIS.ORG

# THE STUDY REPORTING TOOL (SRT)

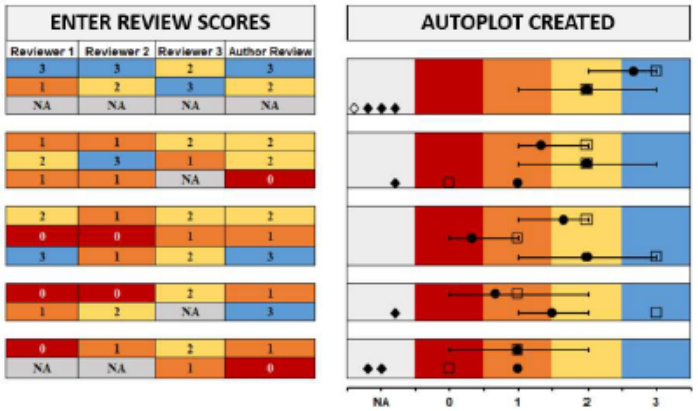
## A FRAMEWORK FOR CONSISTENT PEER REVIEW

NTA Study Chronology

| Section | Category | Sub-Category | Example Information to Report | Score | Rationale |
|---|---|---|---|---|---|
| Methods | Study Design | Objectives & Scope | • 3-4 examples of representative information to report for each of the 13 sub-categories.<br>• Not an exhaustive list – intended to guide researcher/reviewer and relies on expertise/discretion. | NA, 0, 1, 2, 3 | Space for reviewer to explain assigned score (i.e., typical peer review rationale) |
| | | Sample Info &Preparation | | | |
| | | QC Spikes & Samples | | | |
| | Data Acquisition | Analytical Sequence | | | |
| | | Chromatography | | | |
| | | Mass Spectrometry | | | |
| | Data Processing & Analysis | Data Processing | | | |
| | | Statistical &Chemometric Analysis | | | |
| | | Annotation &Identification | | | |
| Results | Data Outputs | Statistical &Chemometric Outputs | | | |
| | | ID & Confidence Levels | | | |
| | QA/QC Metrics | Data Acquisition QA/QC | | | |
| | | Data Processing & Analysis QA/QC | | | |

**Enables rigorous evaluation of reporting quality in NTA studies**

## READILY VISUALIZE REVIEWER FEEDBACK




**Excel auto-plot function provides quick visual representation of peer-reviewer and self-assigned scores**

**Download the SRT:**
www.nontargetedanalysis.com/srt/

**Evaluation of the SRT:**
10.1021/acs.analchem.1c02621

# Contributing Researchers

This work was supported, in part, by ORD's Pathfinder Innovation Program (PIP) and an ORD EMVL award

* = ORISE/ORAU


**EPA ORD**

Hussein Al-Ghoul*
Alex Chao
Jacqueline Bangma*
Matthew Boyce*
Kathie Dionisio
**Louis Groff***
Jarod Grossman*
Chris Grulke
Kristin Isaacs
Sarah Laughlin*
**Hannah Liberatore**
Charles Lowe
Kamel Mansouri*
Aurelie Marcotte*
**James McCord**
Andrew McEachran*
Kelsey Miller
Jeff Minucci

**EPA ORD (cont.)**

**Seth Newton**
Grace Patlewicz
Allison Phillips
Katherine Phillips
Tom Purucker
Ann Richard
Randolph Singh*
Jon Sobus
Mark Strynar
Elin Ulrich
Ariel Wallace
John Wambaugh
Antony Williams

**GDIT**

Ilya Balabin
Tom Transue
Tommy Cathey

# Questions?

sobus.jon@epa.gov

*The views expressed in this presentation are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.*