Toxicogenomic Dose Response Analysis to Inform Risk Assessments

Lyle D. Burgoon, Ph.D.

Leader, Bioinformatics and Computational Toxicology Group

Environmental Laboratory

AOPXplorer

U.S. ARMY

US Army Corps of Engineers®

ERDC

Engineer Research and Development Center

Approach Overview

- Preprocessing
  - Log2 transform
  - Quantile normalization
- Hypothesis-testing vs Screening
  - Screening for differentially expressed genes
    - Analyze only probes with at least 1.5x up/down regulation (normal space)
    - Bayesian Region of Practical Equivalence and 95% Highest Density Interval Analysis
  - Hypothesis-testing
    - Identify probes associated with genes associated with adverse outcome pathway networks of interest
    - Bayesian Region of Practical Equivalence and 95% Highest Density Interval Analysis
- Point of Departure determination
  - Monotonic dose-response
  - GRAVEE: Good Risk Assessment Values for Environmental Exposures
  - https://github.com/datasciburgoon/gravee
- Overlay data onto Adverse Outcome Pathway Networks
  - AOPXplorer: http://apps.cytoscape.org/apps/aopxplorer

Bayesian Analysis to Identify Differentially Expressed Probes/Genes

Bayesian Analysis Basics

- Not concerned with “significance”
  - We don’t deal with p-values
  - Aside: p-values tell you how well your data fit a particular statistical model – that’s it

- Bayesian statistics are focused on **probability**
  - What is the probability of some event?
  - What is the probability the data fit a model?
  - What is the probability a chemical changes the expression of a gene?

Bayesian Statistics Without Equations

- Prior probability
  - Reflects our knowledge of events
    - Probability that a chemical causes Gene X to change
    - Sometimes we don't have prior knowledge
      - We use uninformative prior probabilities or conjugate priors
      - More weight is given to the data in this case

- Likelihood
  - The data we observed

- Posterior probability
  - Proportional to the Prior probability times the Likelihood
  - Posterior probability is the probability of an event given conditions using any available prior knowledge (if it exists) and the likelihood
    - The probability that a gene changes in expression given a chemical exposure at a particular dose and time

Bayesian Statistics Without Equations

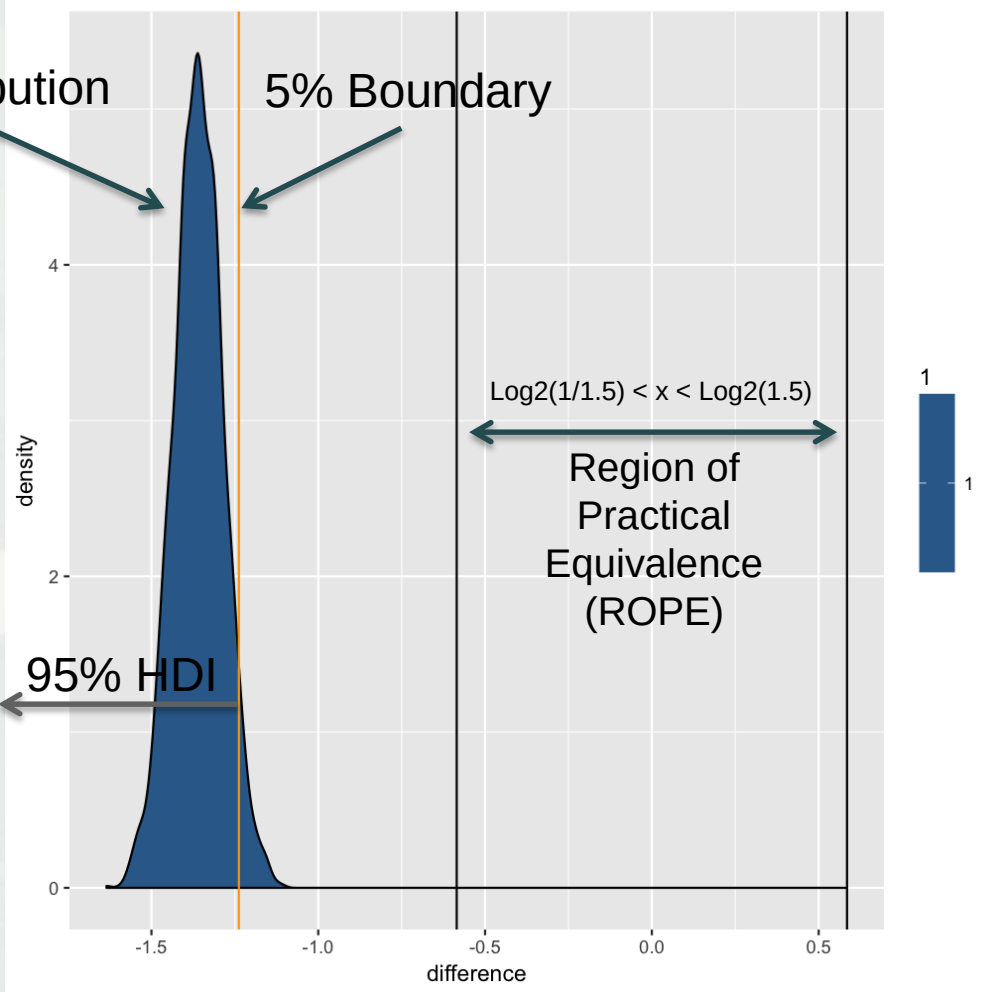
- Posterior probability
  - Usually we get a distribution called Posterior Distribution
  - The distribution represents our uncertainty
  - Central tends (median) is typically used to represent the most likely value
  - Posterior interval represents the range where the posterior probability is likely to exist
    - 95% Posterior Interval represents the region where we have 95% probability where the posterior probability is

Bayesian Statistics Without Equations

- 95% Highest Density Interval
  - The values of the distribution from 5% and above
  - This represents the 95% most likely values

- We typically couple this with decision rules…

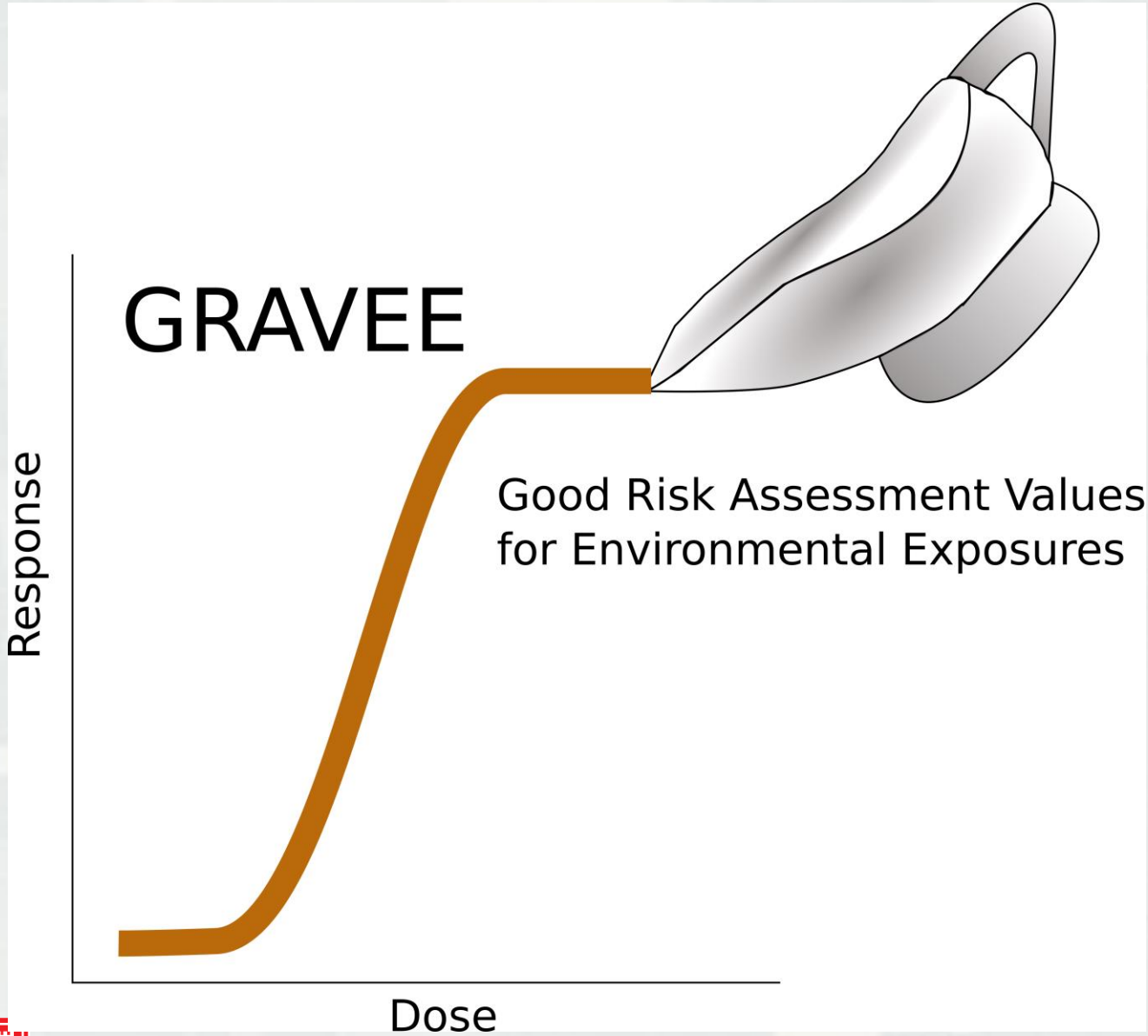
Bayesian 95% HDI Analysis



Acyl-coA Synthetase

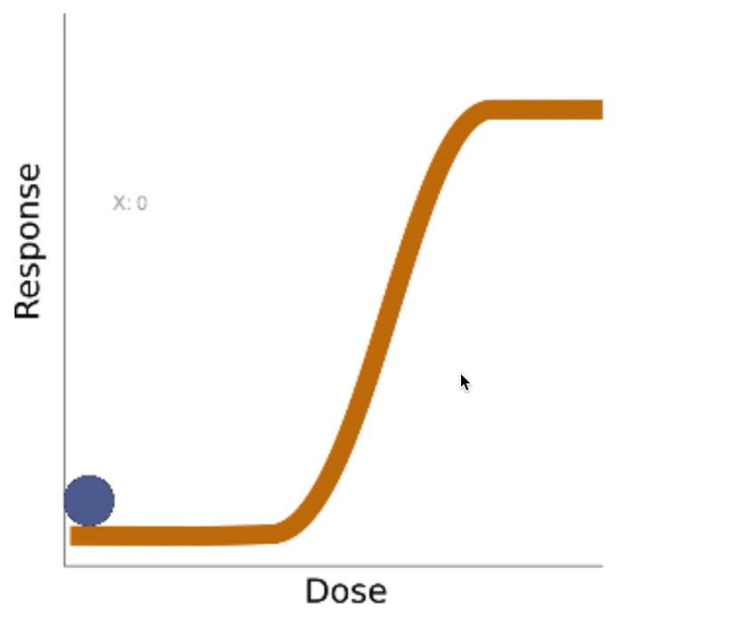
Significant Difference:

All of the 95% HDI is outside the ROPE



How Can We Avoid Needing a Benchmark Response??

- Interpolate curve data
  - Spline-based metaregression
- Menger Curvature
  - Measures the curvature of a curve
  - We identify the point of the maximal Menger Curvature in the interpolated curve
  - This point is the POD



How Do We Get The Uncertainty

- Bootstrap!
  - Sample with replacement across the dose response dataset to create a lot (let's say 1,000) dose response curves
- So now we have a lot (let's say 1,000) dose response curves
- And now on each of these models we:
  - Interpolate curve data using spline-based metaregression
  - Find POD using Menger Curvature
- So now we have a lot (let's say 1,000) of PODs
  - You can do a lot with a lot (let's say 1,000) of PODs

Parametric vs Nonparametric Modeling

- Parametric
  - We “know” the general mathematical family the data follow
  - Start with predefined mathematical models
    - Hill Model (a variant of a sigmoidal model)
    - Exponential models
    - Polynomial
  - Example
    - Benchmark Dose Software

- Nonparametric
  - Does not start with a predefined mathematical model
  - Lets the data speak for themselves
  - Example
    - LOESS or LOWESS commonly used in microarray normalization
    - GRAVEE (Good Risk Assessment Values for Environmental Exposures)

Uncertainty Around Our POD

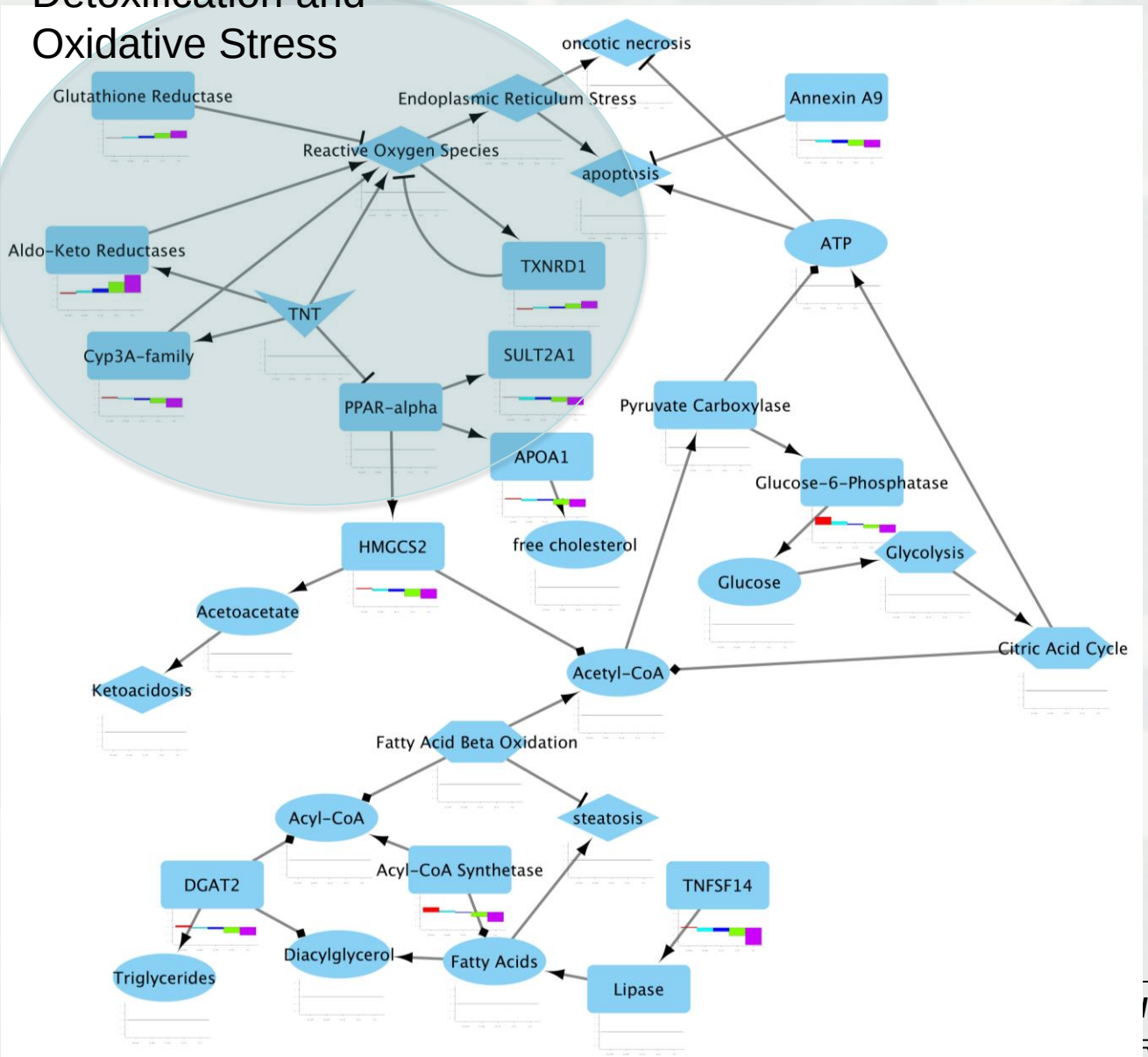
GRAVEE POD	TNT, in vitro (ug/mL)
POD (5%)	0.60
POD (50%, median, most likely POD)	3.00
POD (95%)	4.80

Uncertainty can be propagated through IVIVE and RfD/RfC calculations

Revisiting 2,4,6-Trinitrotoluene…A Case Study

Placing Gene Expression in Biological Context

Detoxification and Oxidative Stress



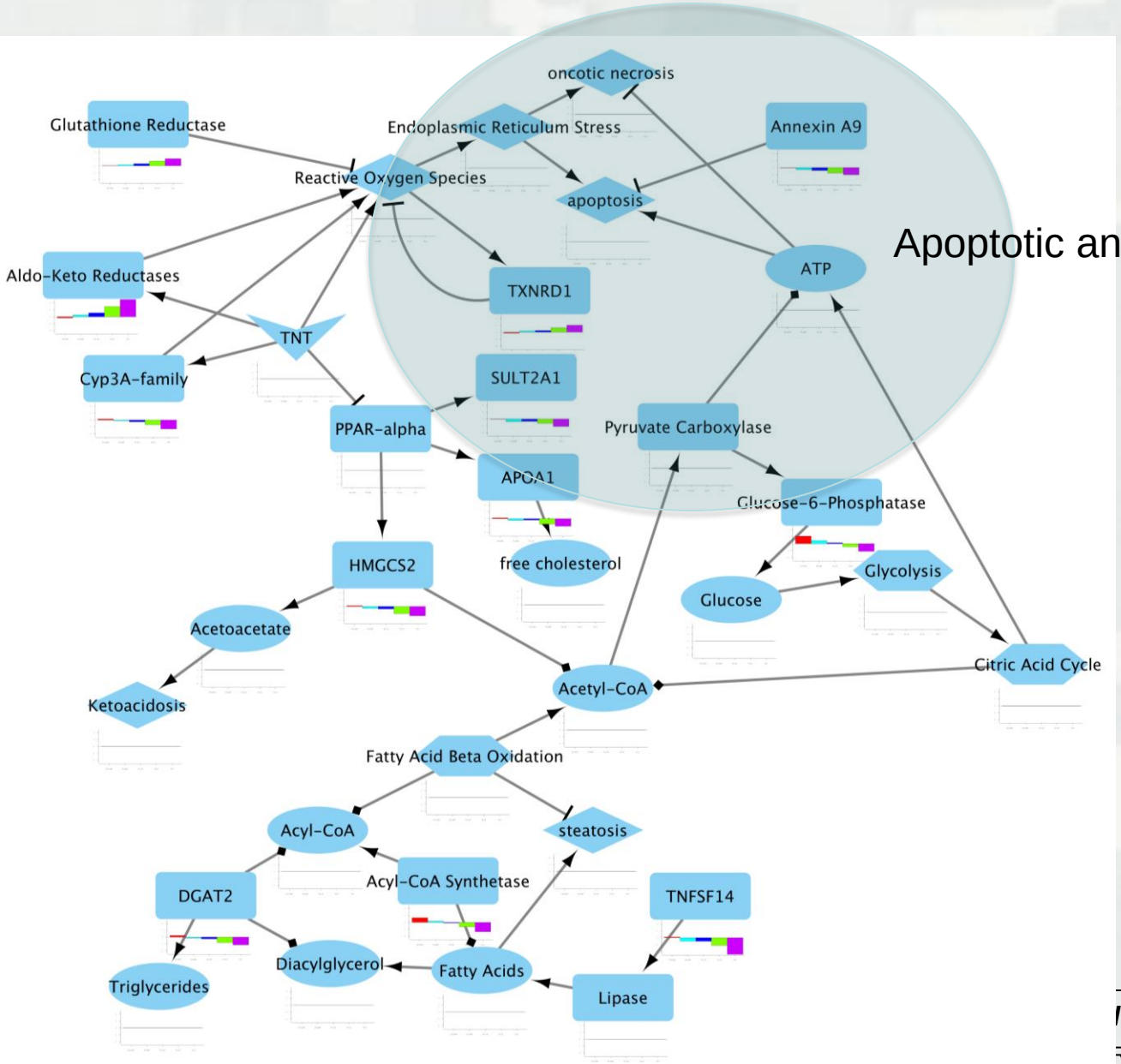
AOPN:
Steatosis and
Primary Metabolism

ERDC

lutions for a safer, better world

RIBUTION IS UNLIMITED.

Placing Gene Expression in Biological Context



AOPN:
Steatosis and
Primary Metabolism

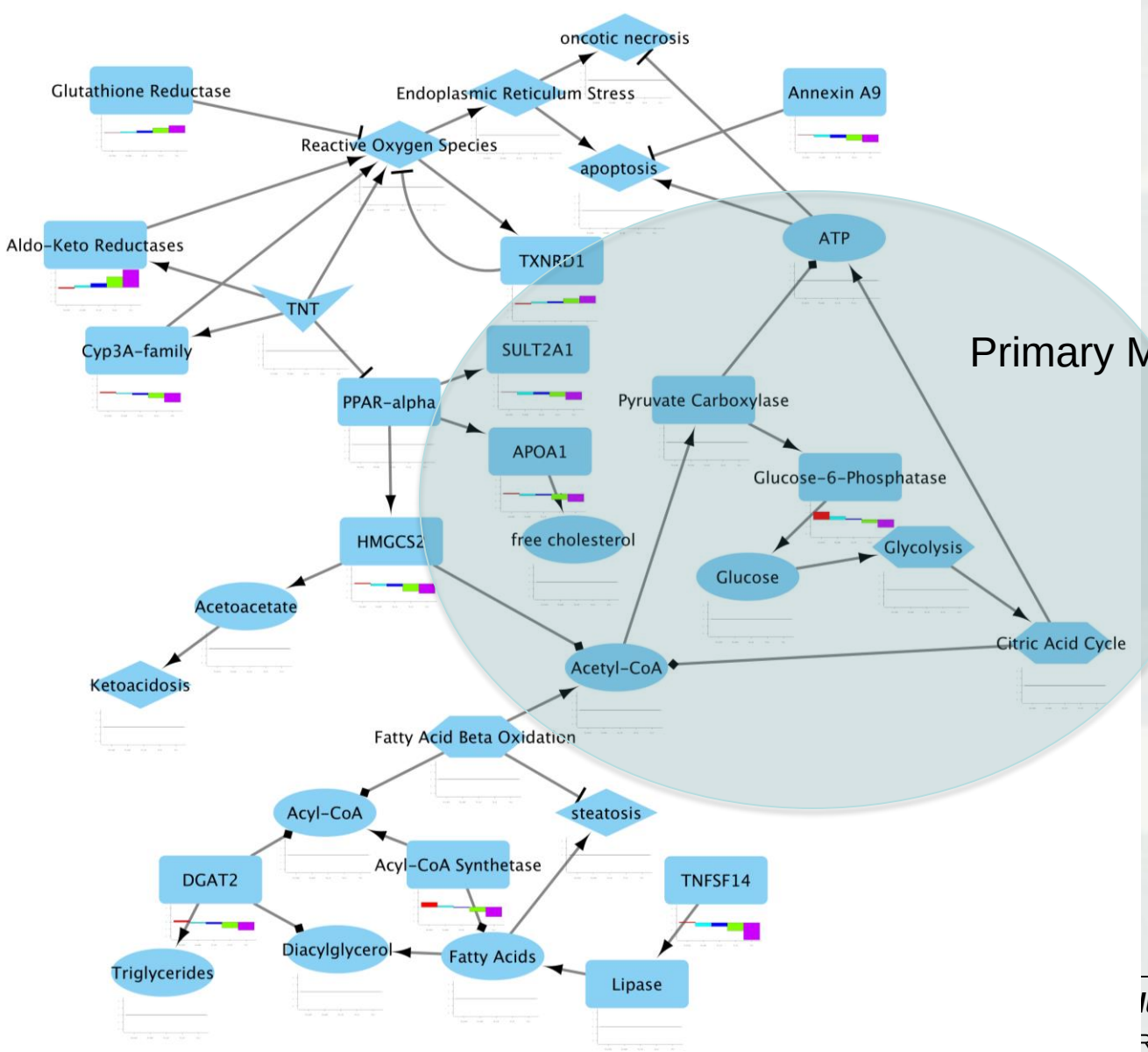
Apoptotic and Oncotic Necrosis

ERDC

lutions for a safer, better world

RIBUTION IS UNLIMITED.

Placing Gene Expression in Biological Context



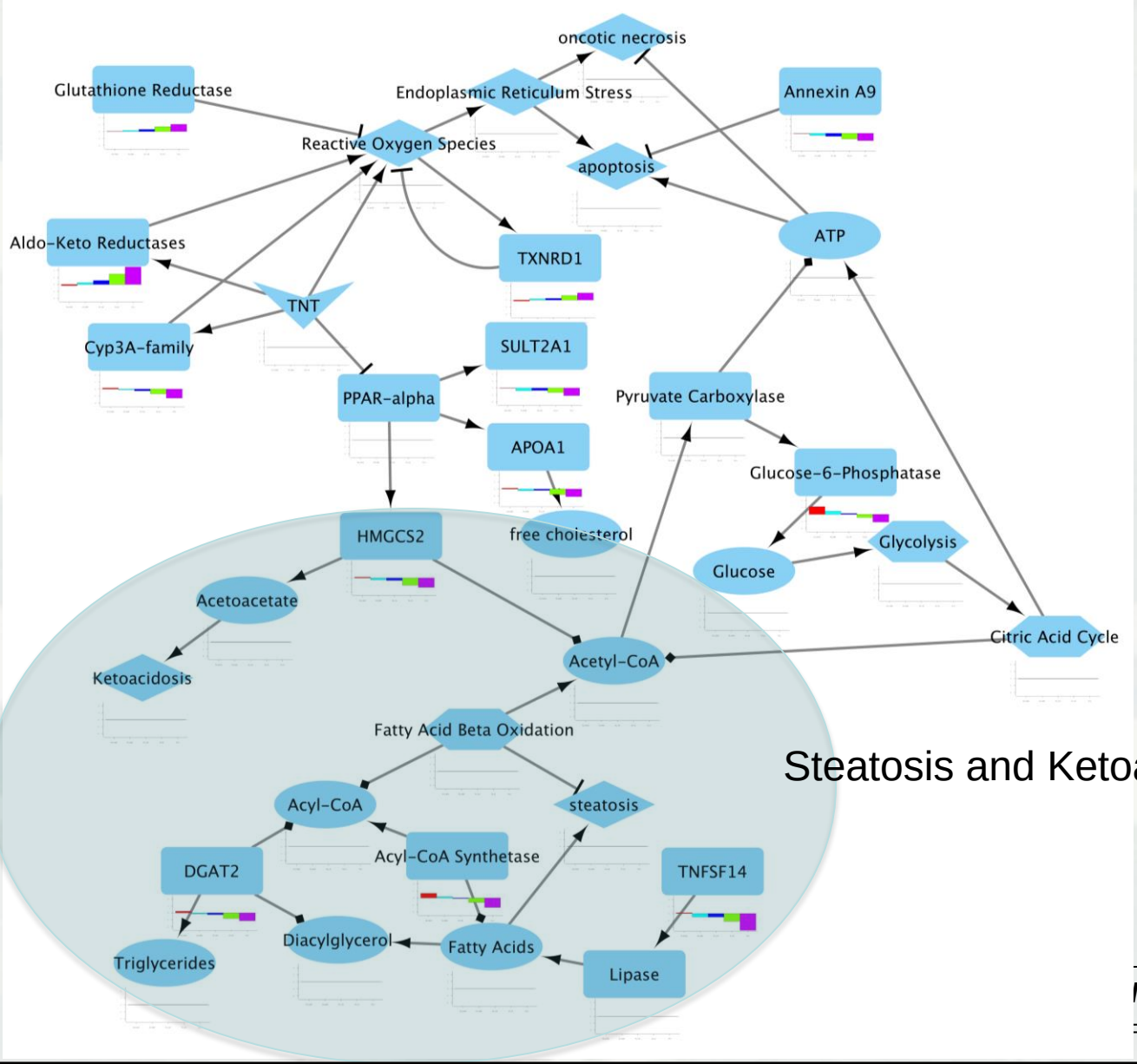
AOPN:
Steatosis and
Primary Metabolism

ERDC

lutions for a safer, better world

RIBUTION IS UNLIMITED.

Placing Gene Expression in Biological Context



AOPN:
Steatosis and
Primary Metabolism

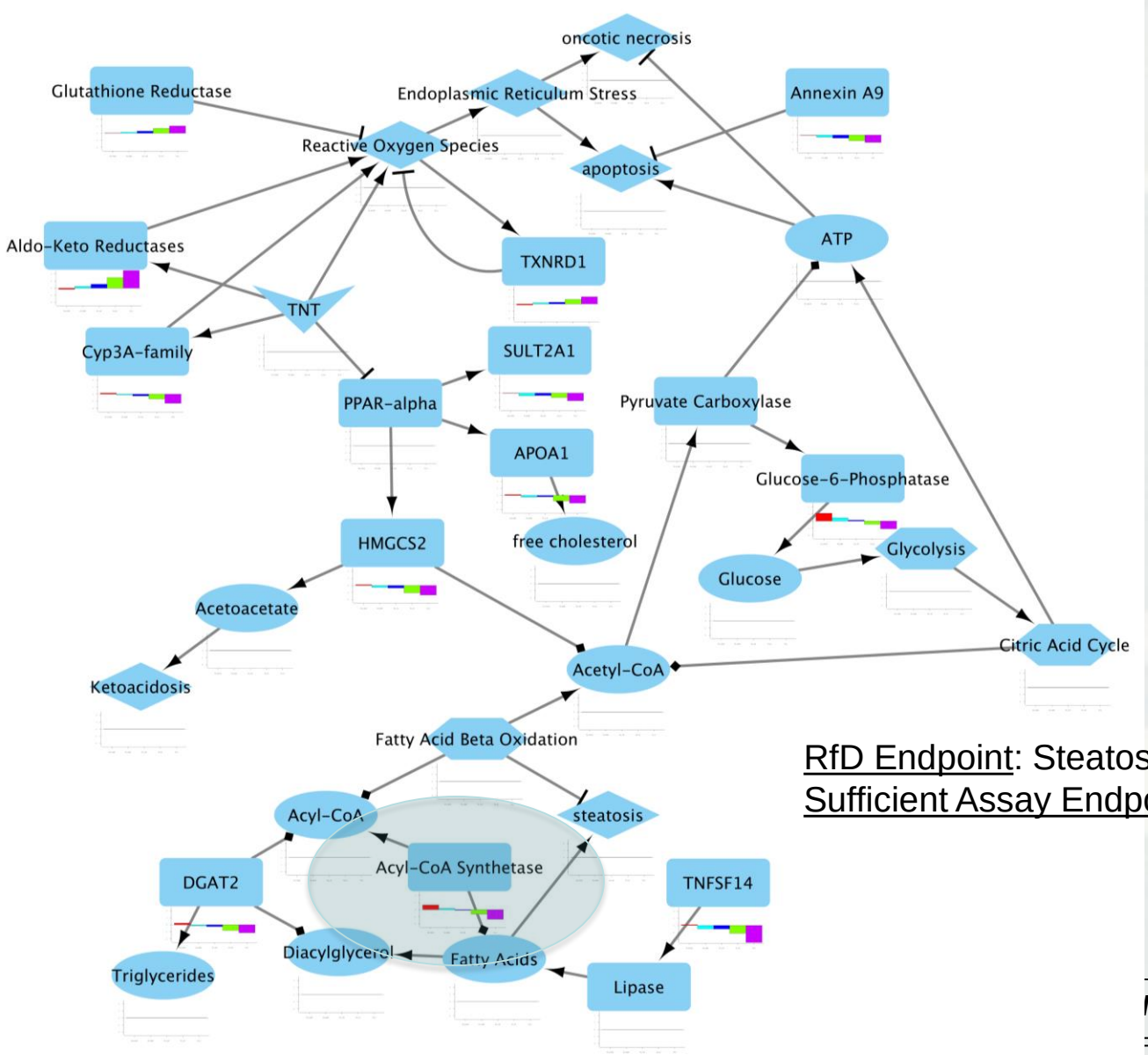
Steatosis and Ketoacidosis

ERDC

lutions for a safer, better world

RIBUTION IS UNLIMITED.

Placing Gene Expression in Biological Context



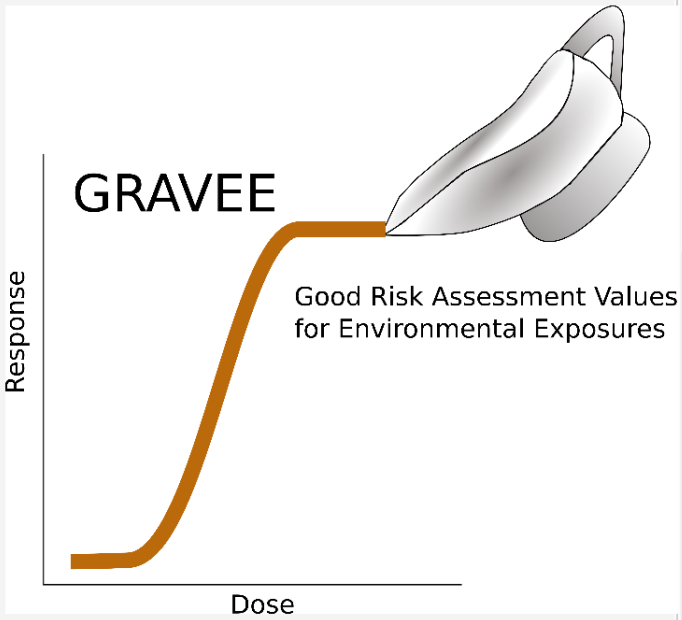
AOPN:
Steatosis and
Primary Metabolism

RfD Endpoint: Steatosis
Sufficient Assay Endpoint: Acyl-CoA Synthetase

ERDC

lutions for a safer, better world

RIBUTION IS UNLIMITED.



Point of Departure Results

GRAVEE calculates the distribution of predicted PODs using bootstrapping.

POD-5%: 5% of the distribution is below this value.

POD-50%: This is the median of the POD distribution (the most likely POD value given the data)

POD-95%: 95% of the distribution is below this value.

Identifier	POD-5%	POD-50%	POD-95%
ACSM2B	0.6	3.1	4.8

Get your PODs in 2 easy steps!

1) Choose your input data file

Data File tnt_dr_table.txt

2) Press Go!

Go!

Input File Specs:

Tab-delimited text file. Should look something like this:

Assay/Gene Name >> Dose 1 >> Dose 2 >> Dose 3 >> Dose 4 >> Dose 5
Assay/Gene Name >> Dose 1 >> Dose 2 >> Dose 3 >> Dose 4 >> Dose 5
Assay/Gene Name >> Dose 1 >> Dose 2 >> Dose 3 >> Dose 4 >> Dose 5

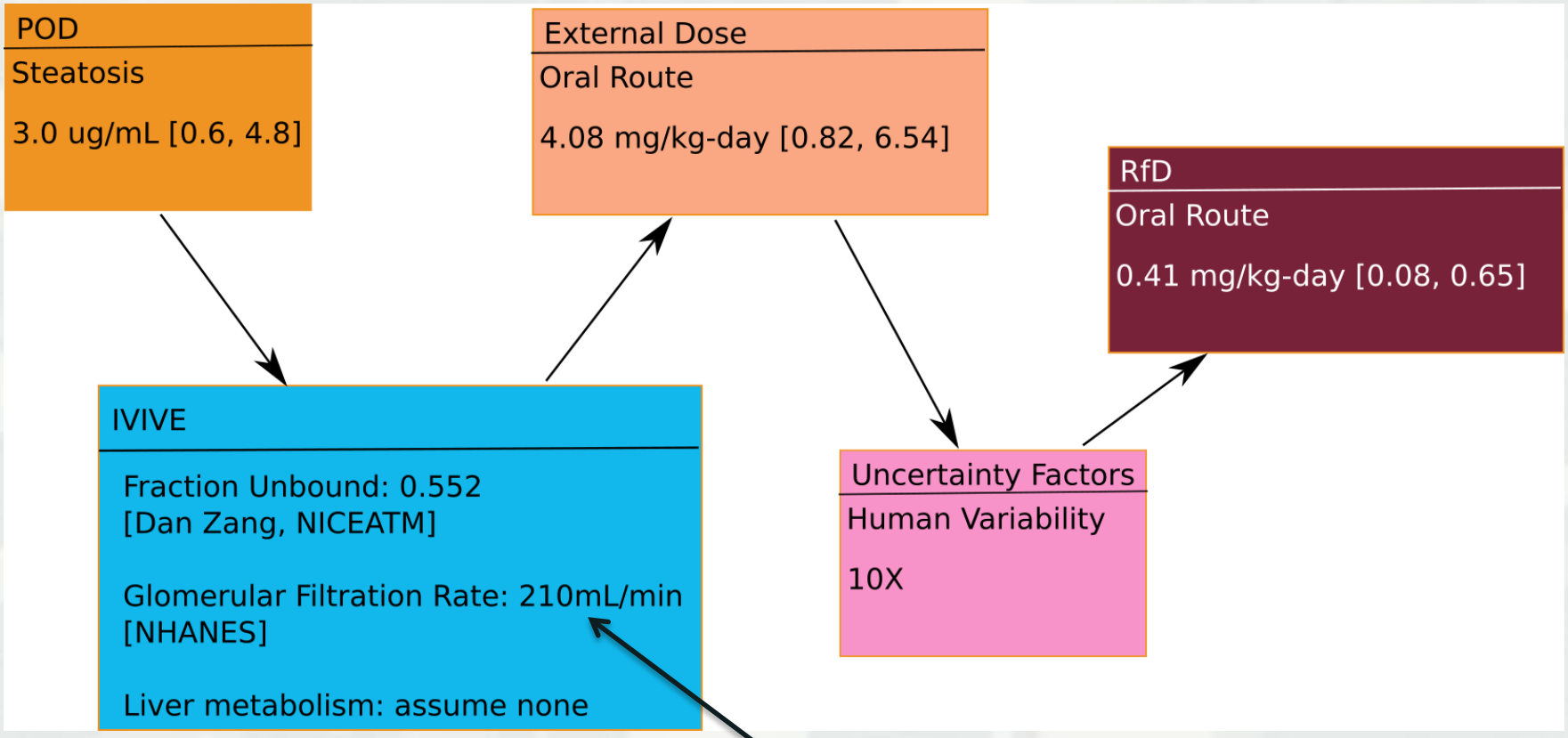
Please see the example_data.txt file for an example (you can open and save these in Excel)

Version 1.0, 1 September 2017

Contact: Lyle D. Burgoon, Ph.D. (lyle.d.burgoon@usace.army.mil)

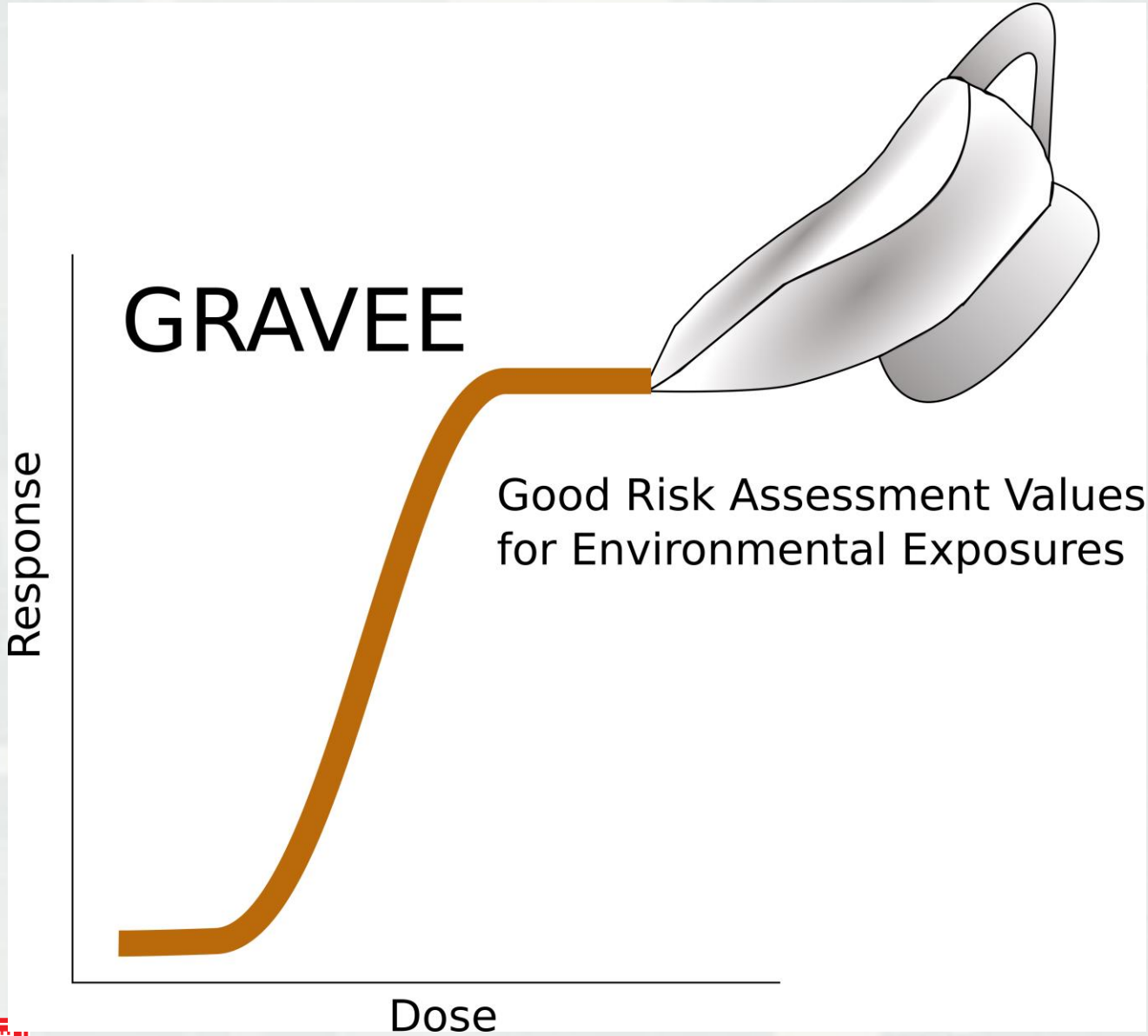
U.S.ARMY

TNT RfD Synthesis for Steatosis



In the future we will consider uncertainty at GFR

http://apps.cytoscape.org/apps/aopxplorer



Acknowledgements

- US Army Military Materials in the Environment
  - NexGen Predictive (Eco)toxicity ToolBox Focus Area
    - Big Data Driven Hazard Screening Task

- US Army ERDC
  - Dr. Ed Perkins
  - Dr. Natalia Vinas
  - Dr. Keri Donohue
  - Anne Mayo
  - Mitchell Wilbanks

- ILS (supporting NICEATM)
  - Dr. Shannon Bell
  - Dr. Dan Zang