Air Toxics Data Analysis

Madeleine Strum
EPA/OAQPS
Air Quality Assessment Division

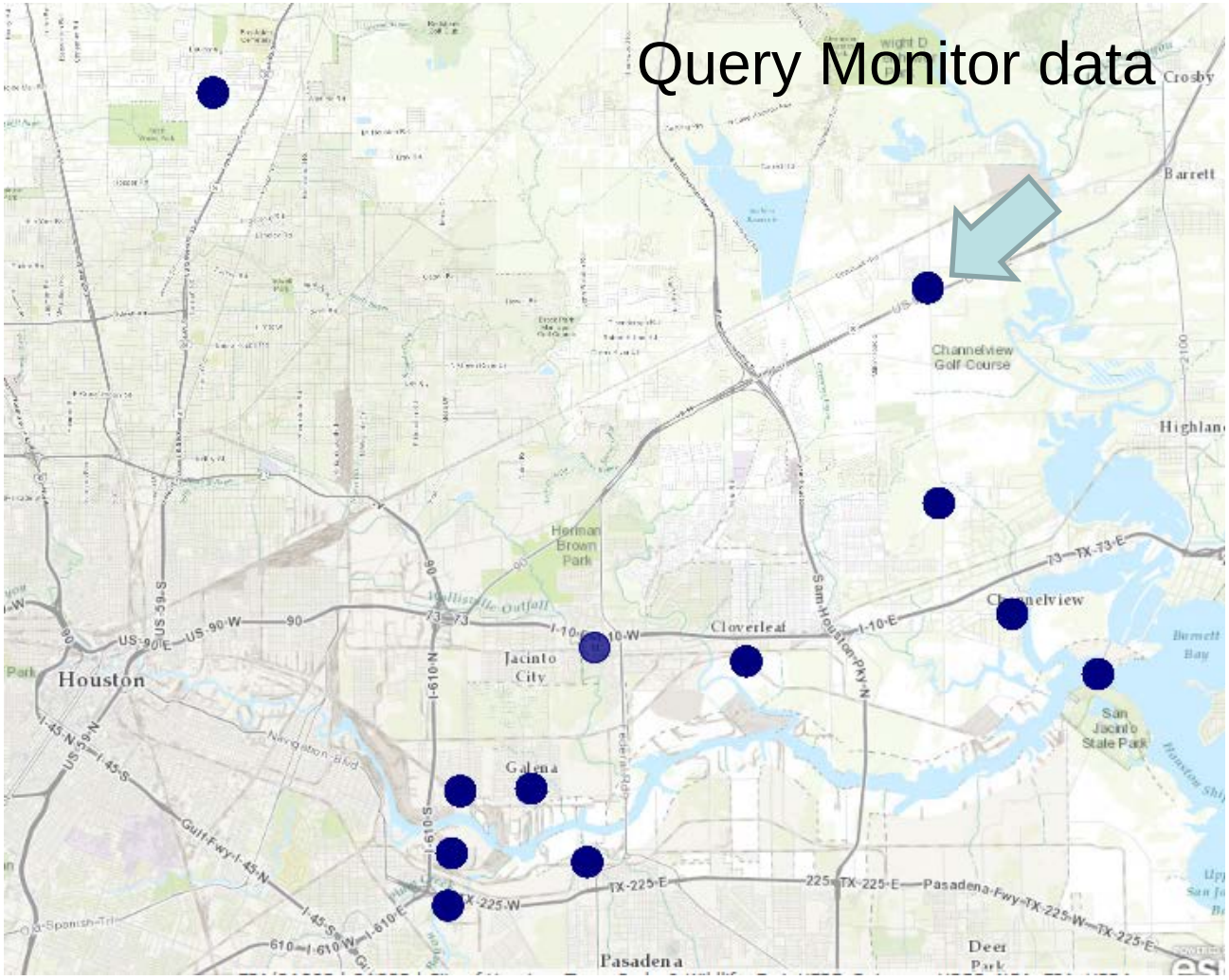
National Monitoring Conference
Annual Meeting
August 2016

Uses of Ambient Data of Air Toxics

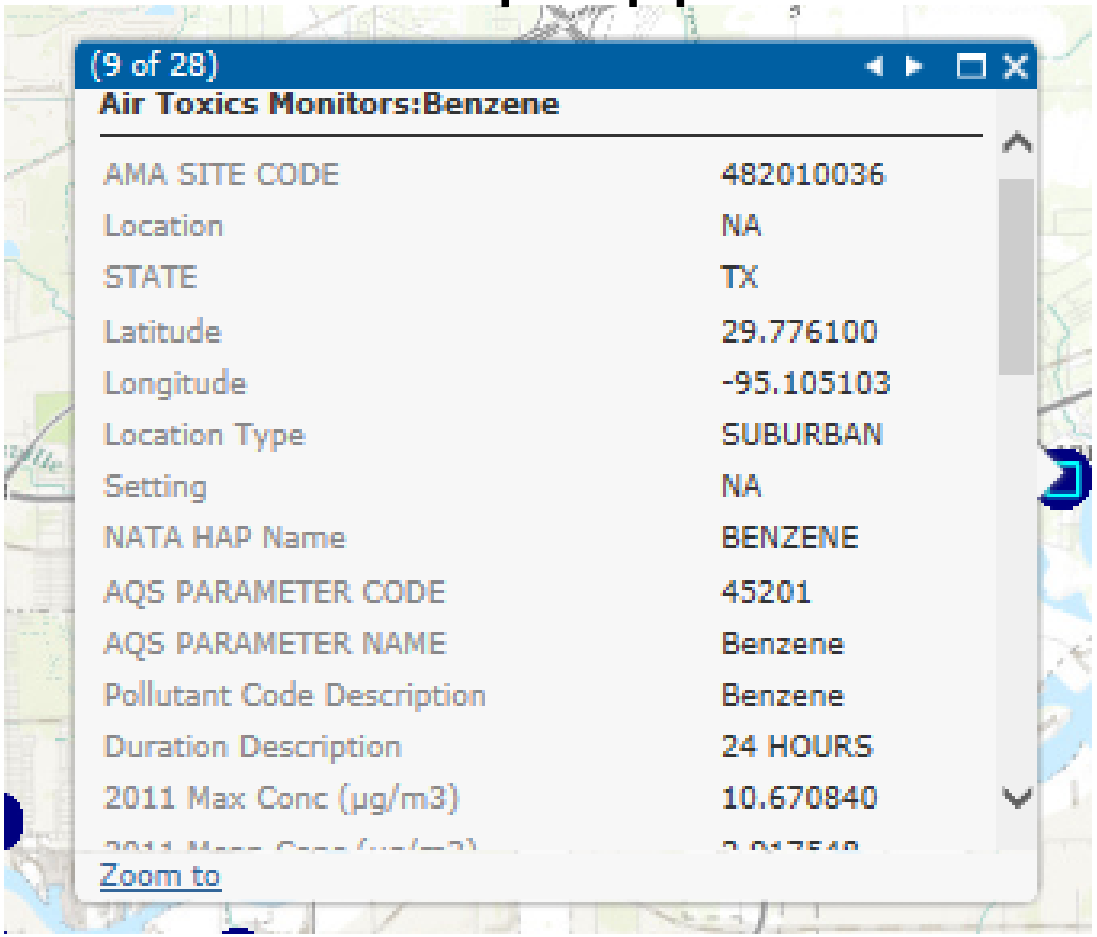
- Confirm risk model results
  - Assess risk predicted by model in key locations (i.e., schools, communities)
- Model Evaluation
- Trends
  - Tract progress of air toxic reduction efforts for key pollutants (i.e., benzene, formaldehyde)
- Local assessments
  - Near roadway, fenceline monitoring downwind from industrial facilities

http://www.epa.gov/national-air-toxics-assessment

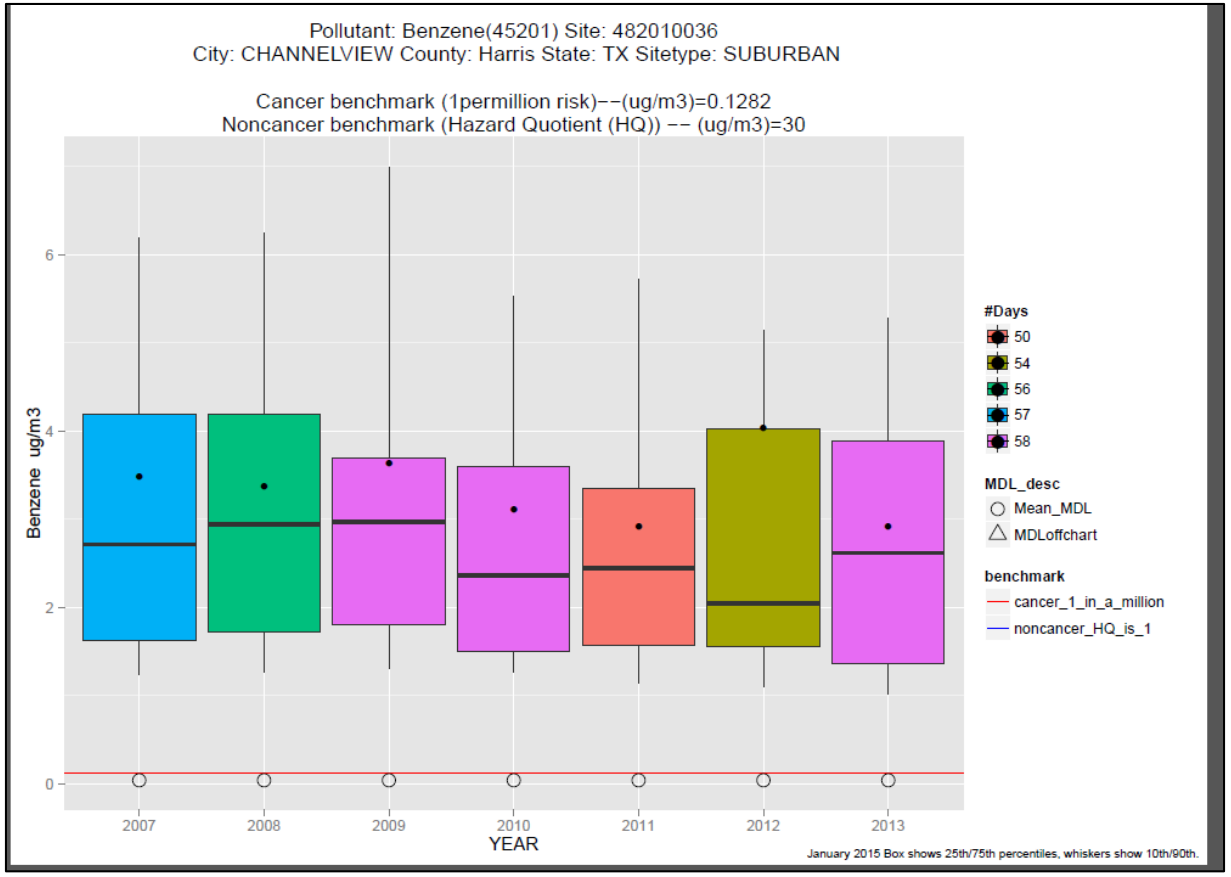
Monitoring Data Included in the NATA Map App



Monitoring Data Included in the NATA Map App



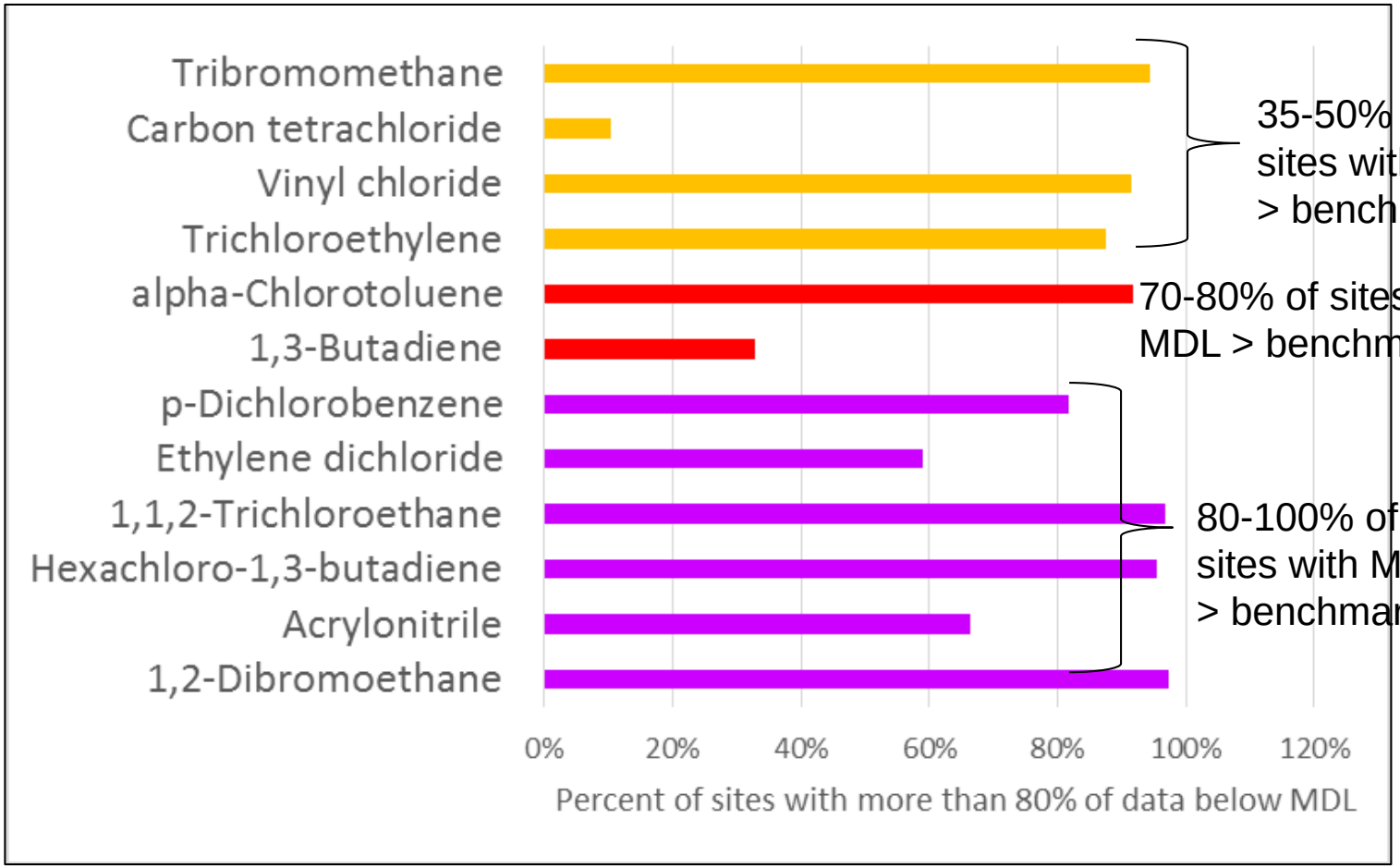
Monitoring Data Included in the NATA Map App



Air Toxics Challenges

- What is level of completeness to require to creating annual average/statistics: (11 measurements vs 12 per quarter)
- Differences in measurements across POCS (monitors) at same monitoring site
  - Average across all POCS or pick a POC?
- Treatment of data below MDL and ND
  - cancer benchmarks below MDL

Pollutants where MDL > 1 in 1-million cancer benchmark



Data analysis challenges: data below MDL and non detects

- MDL - statistical construct, estimate of the concentration at which there is 99% confidence that the analyte, when positively identified, is present

- ND – analyte not identified based on set criteria such as signal to noise
  - Does not mean the concentration is zero
  - We see NDs where remote concentration (background) is nonzero

Don't see these distinguished in literature on handling data below detect

What does below MDL/ND mean: not there? vs can't measure it?

	REMOTE (ug/m3)	Range of Site Minimum MDL 2014	Sites with NDs In 2014
Carbon tetrachloride	0.547	0.006 to 0.53	6 sties
Chloroform	0.058	0.01 to 0.4	12 sites
Benzene	0.116	0.006 to 0.27	5 sites
Methyl chloroform	0.06	0.01 to 0.46	7 sites
Methyl bromide	0.0294	0.02 to 0.4	16 sites
Dichloromethane (methylene chloride)	0.146	0.01 to 7	6 sites
Trichloroethylene	0.0041	0.01 -0.46	3 sites
Tetrachloroethylene	0.0131	0.01-0.6	10 sites

ND versus below MDL

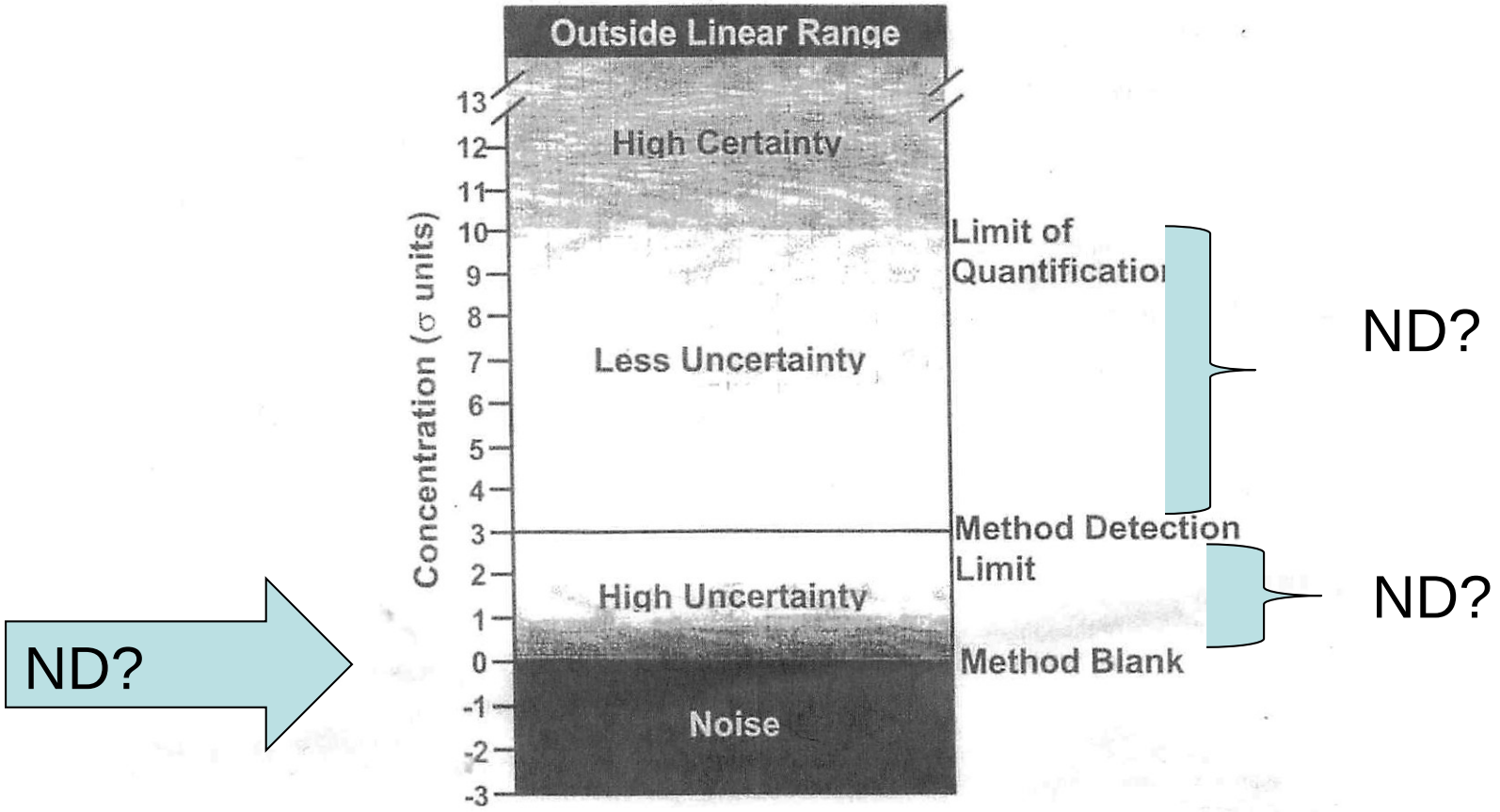


Figure 2. Analyte Concentration vs. Reporting Certainty. (The y axis represents signal strength, in units of the standard deviation (σ) used to determine the MDL.) Adapted from Keith, 1991 by Johnson, 2001.

What to do with data below MDL/Nondetect

- When using censored data: Helsel (Chemosphere, 2006) – substituting is fabricating. Use statistical techniques.

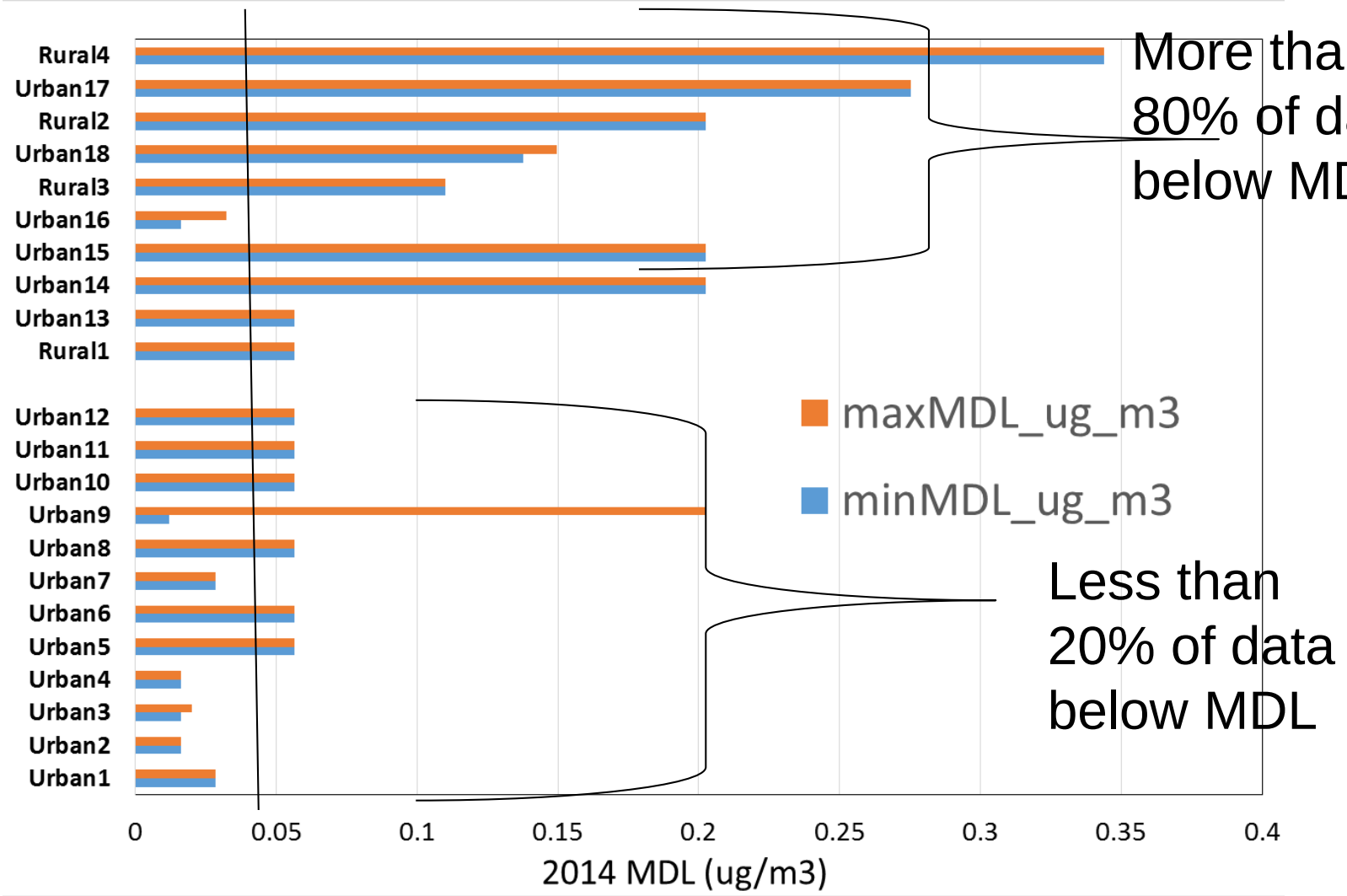
- When reporting the data: many scientists including EPA's Science Advisory Board; Analytical Methods Committee, Royal Society of Chemistry: Report the result found, with statement of its uncertainty

Sensitivities to different treatment at individual sites

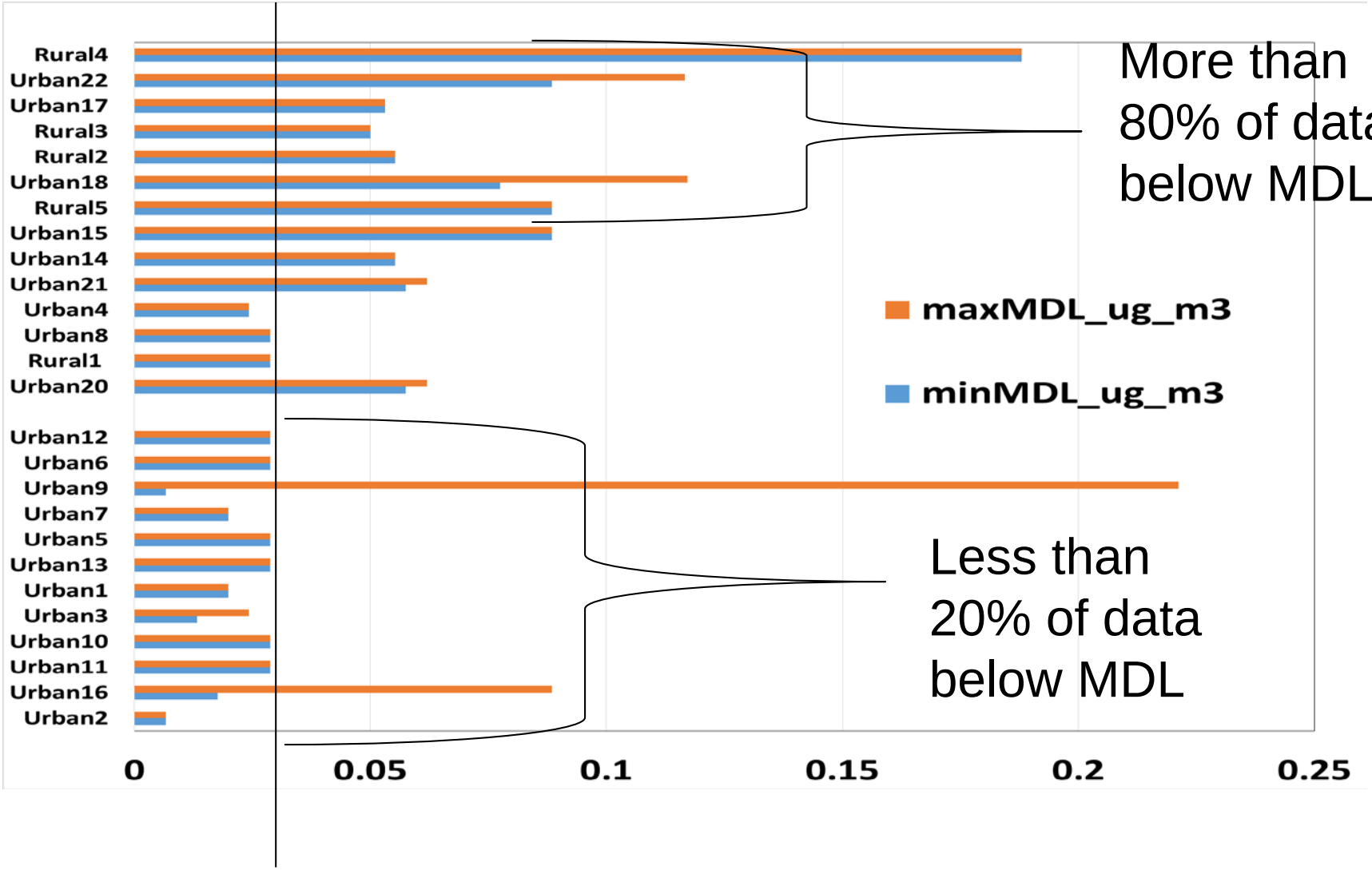
- Ethylene dichloride
- 1,3 butadiene

First, we look at MDLs for across the NATTS sites

Ethylene Dichloride MDLs – 2014



1,3 Butadiene MDLs – 2014



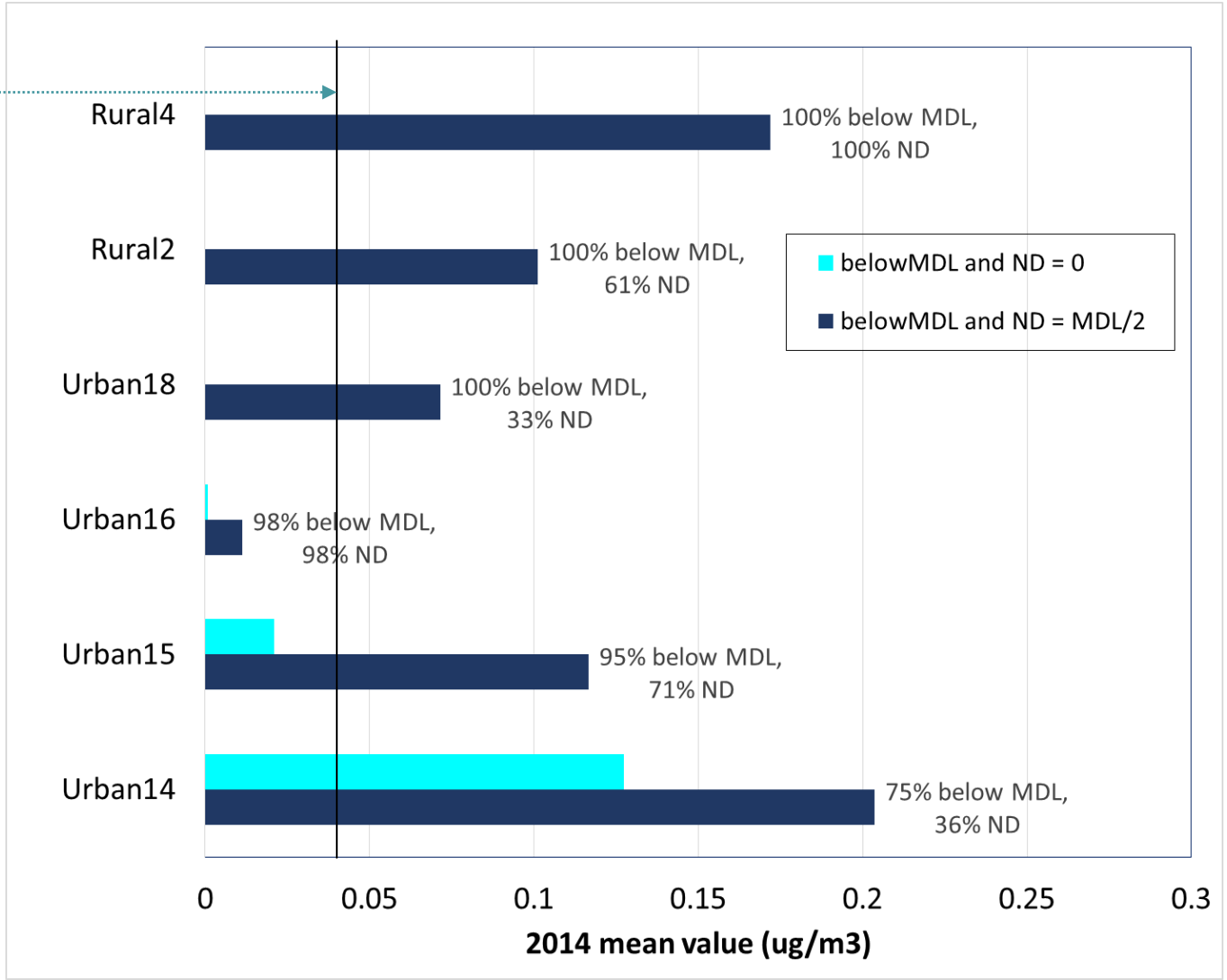
Can treat below MDL and ND many different ways

- Substitution approaches
  - Data below MDL = ½ MDL
  - Data below MDL = 0
  - Use data as-is. No change to data below MDL; ND=0
  - Data below MDL=1/2 MDL; ND=0
- Statistical Approaches
  - Use Regression on Order statistics (ROS) for ND
  - Use Kaplan-Meir (KM) for ND
  - Use ROS for all data below MDL
  - Use K-M for all data below MDL

Ethylene dichloride– sensitivity to substitutions

Treat <MDL; ND as the same

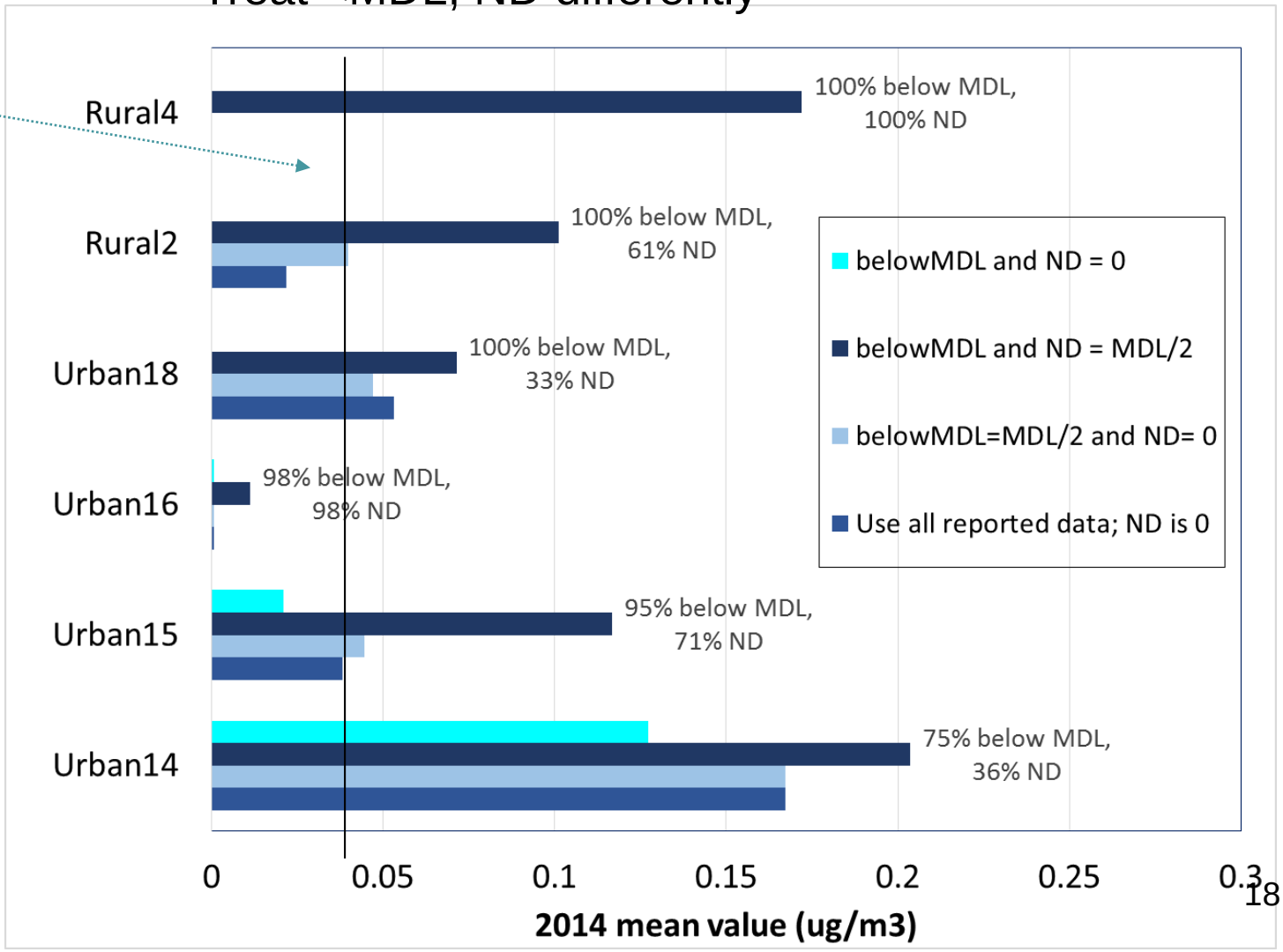
Cancer benchmark



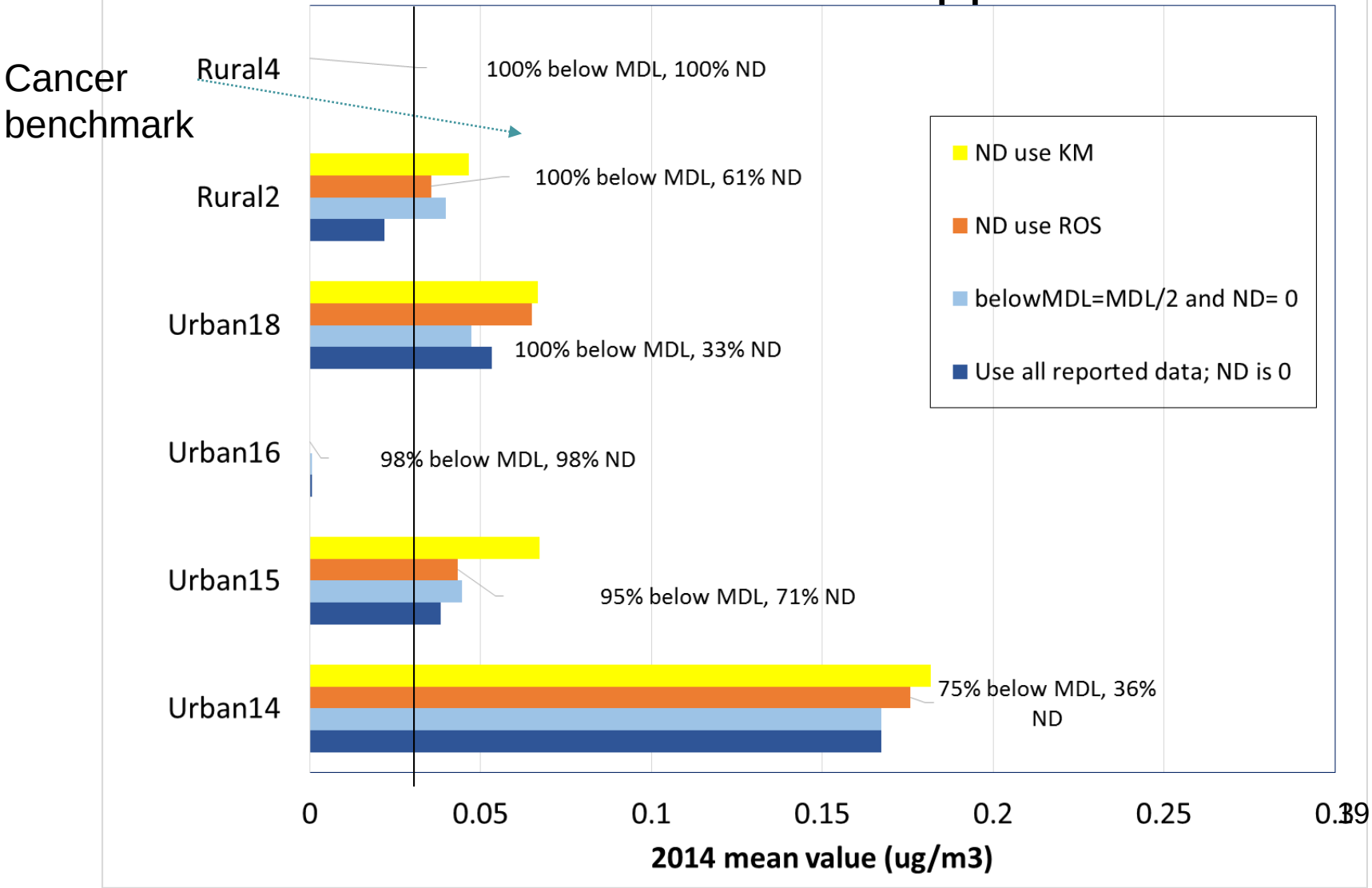
Ethylene dichloride– sensitivity to substitutions

Treat <MDL; ND differently

Cancer benchmark



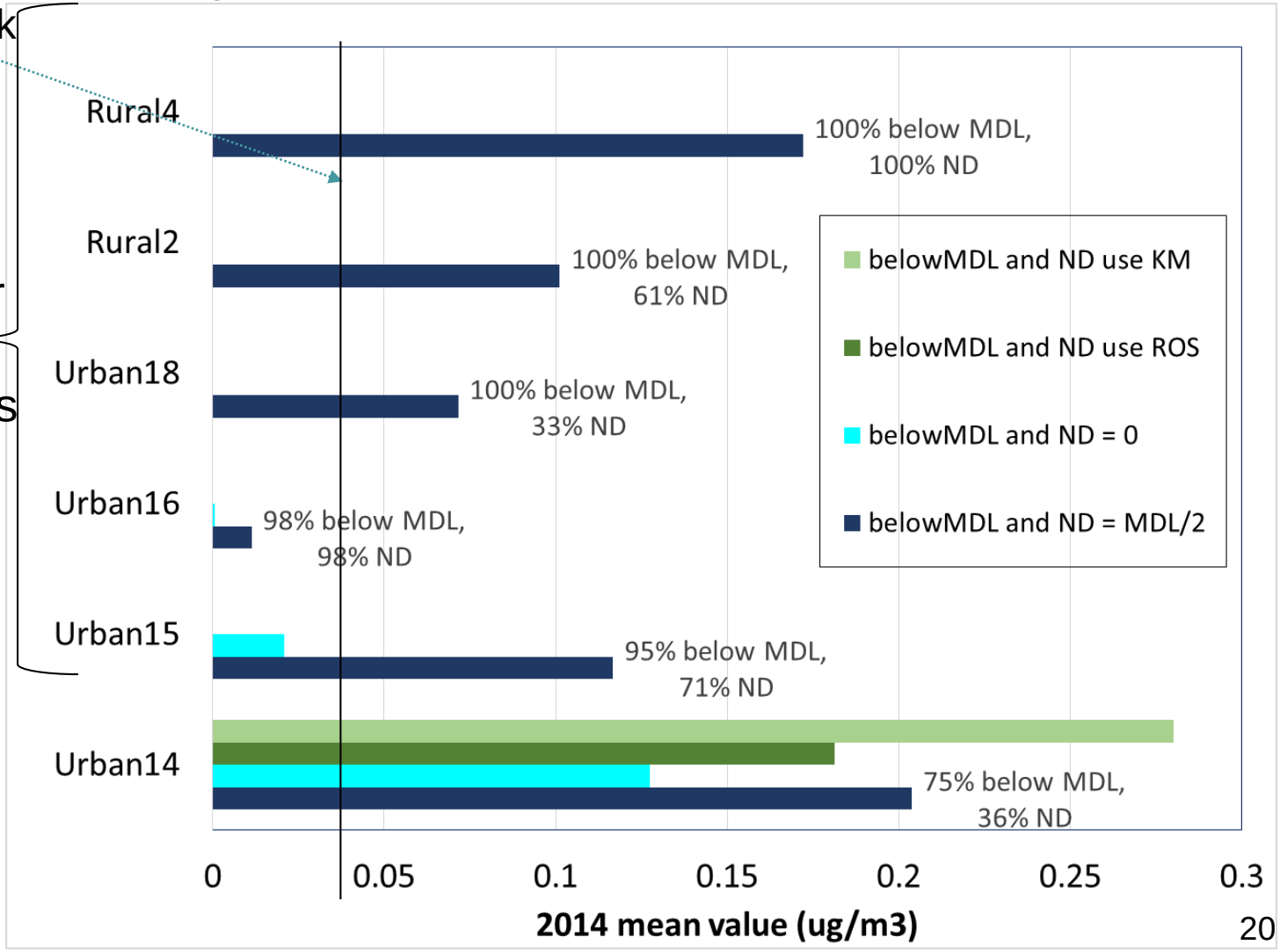
Ethylene dichloride– sensitivity to substitutions/statistical approaches



Ethylene dichloride– sensitivities

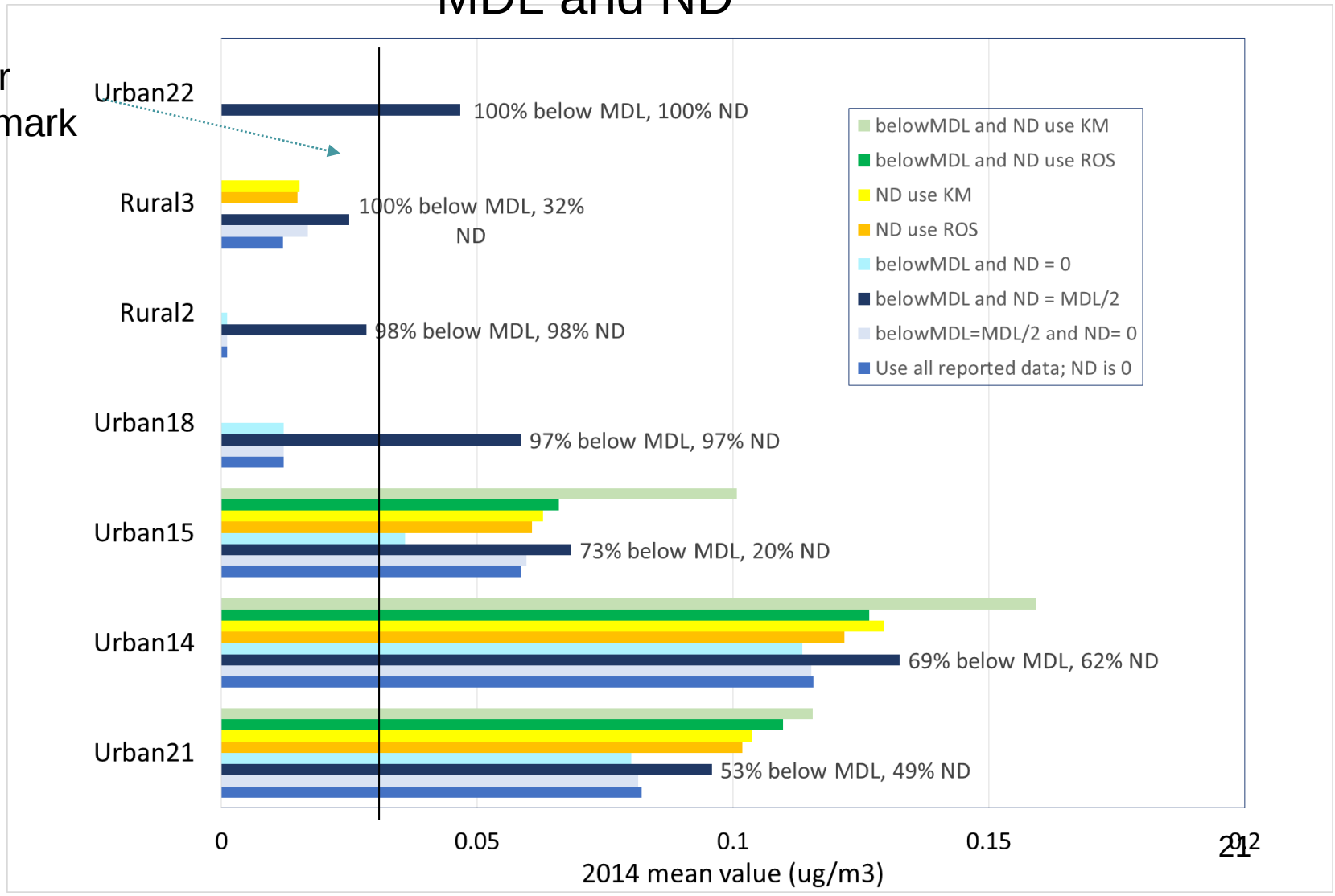
Cancer benchmark

No data for statistical approaches



1,3 butadiene – sensitivity to approaches for below MDL and ND

Cancer benchmark

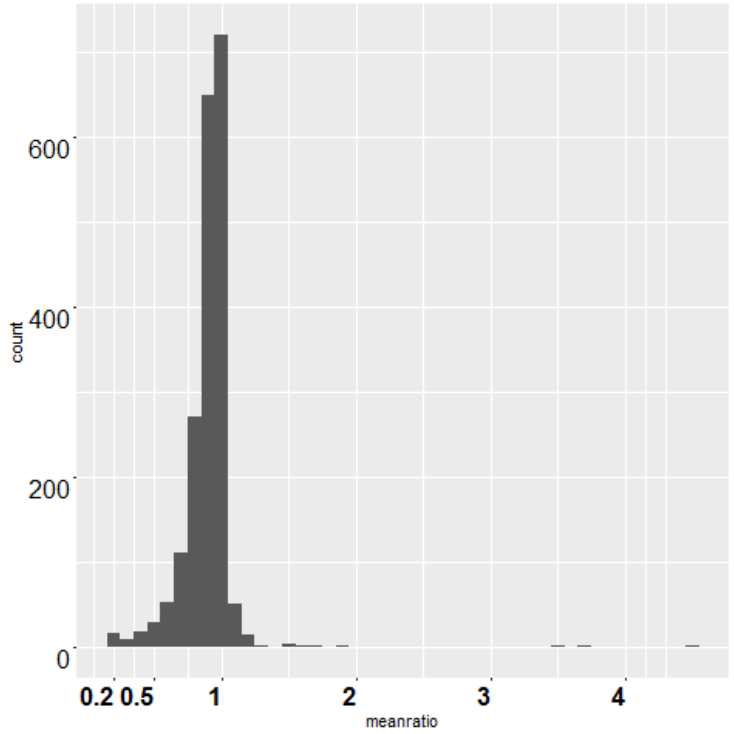


Sensitivities to different treatments - NATTS wide, pollutant wide – 2005 and later

RATIO OF MEAN:

(½ MDL; ND= 0) / (ND=ROS)

- Looked at 2005 and later
- sites with between 20% and 80% data below MDL

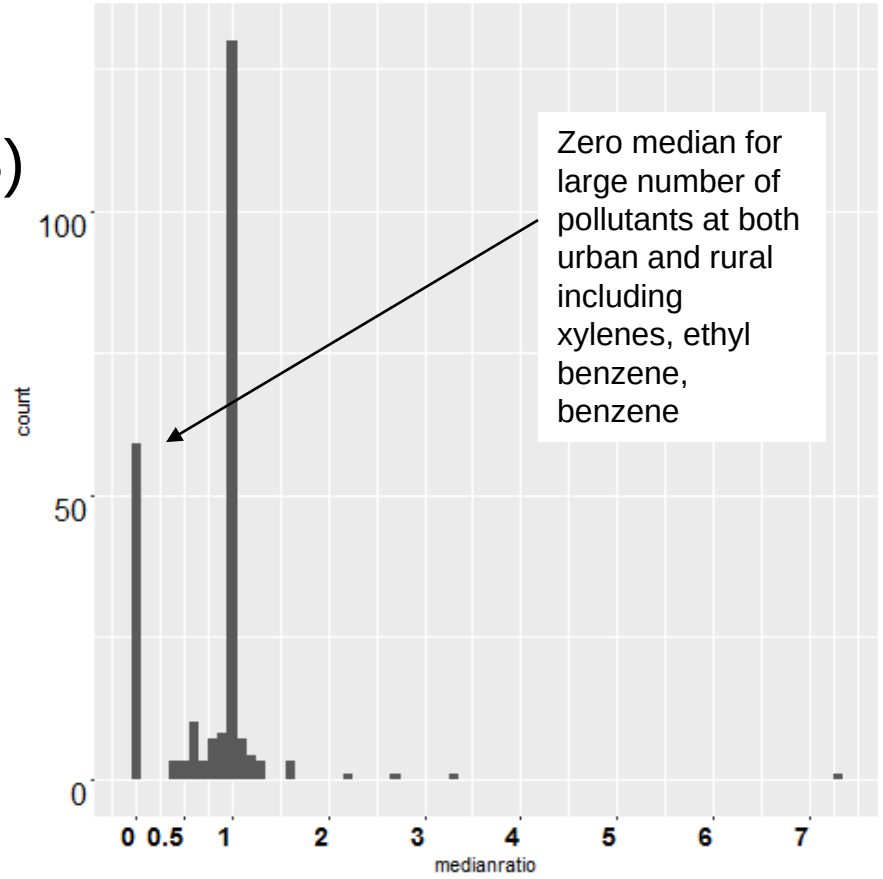


Sensitivities to different treatments - NATTS wide, pollutant wide – 2014

RATIO OF MEDIAN:

(½ MDL; ND= 0) / (ND=ROS)

- sites with between 20% and 80% data below MDL



Analysis of data below MDL

- Just beginning –lots of questions- e.g.,
  - distinguish between ND and MDL?
  - What is the percent of data below MDL above which the site should not be used for model evaluation or trends?
- Deciding what to do with values below MDL should not be data-reporting dependent; it should be done at the analysis stage not the reporting/recording stage
- Once values are removed or replaced, we cannot get them back
  - Can utilize values to learn about uncertainty/distribution of data below MDL if you have the actual values
  - these data could help us determine best practices for data analysis
- Substituting a value (0, MDL/2, MDL) will bias the result
- Foster consistency