



Statistical Approaches for Assessing Health Effects for Environmental Chemical Mixtures in Epidemiology Studies

Location: NIEHS Main Campus, Rodbell ABC; Research Triangle Park, North Carolina

*Shuttles will be available to all workshop participants. For participants staying at the Comfort Suites, please walk to the Wingate Hotel to take the shuttle to NIEHS. Shuttles will depart as they are filled starting at the time listed.

If you would like to attend Monday's dinner at Mez, please sign up at the **Registration Table by 10:30 a.m.

Agenda

Day 1, Monday, July 13, 2015

7:30 a.m. Shuttles depart Wingate Hotel for NIEHS*

8:30 a.m. – 8:40 a.m. Welcome/Logistics about Workshop – Danielle Carlin, Ph.D. (NIEHS)

Moderator: Dr. Danielle Carlin (NIEHS)

8:40 a.m. – 8:45 a.m. Opening remarks – Gwen Collman, Ph.D. (NIEHS)

8:45 a.m. – 9:05 a.m. Classic single chemical approach in epidemiology studies – Joseph Braun, Ph.D. (Brown University)

9:05 a.m. – 9:20 a.m. Setting the Stage and Charge for Workshop – Russ Hauser, M.D. (Harvard University)

Presentation of Statistical Approaches of Simulated Data Sets #1 and #2

Simulated Data Set #1

Moderator: Cynthia Rider, Ph.D. (NIEHS)

9:20 a.m. – 9:40 a.m. Thomas Webster, Ph.D. (Boston University) – Development of the Simulated Data Set #1 – What was the right answer?

9:40 a.m. – 9:55 a.m. Birgit Claus Henn, Sc.D. (Boston University) – Bayesian Kernel Machine Regression (BKMR)

9:55 a.m. – 10:10 a.m. Katrina Waters, Ph.D. (Pacific Northwest National Laboratory) – Variable Selection and Multivariate Adaptive Spline Assessments

10:10 a.m. – 10:30 a.m. Break (Posters should be put up during break)

10:30 a.m. – 10:45 a.m. Sung Kyun Park, Sc.D. (University of Michigan – Ann Arbor) – Shrinkage methods including Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, and Least Angle Regression (LARS)

10:45 a.m. – 11:00 a.m. Chris Gennings, Ph.D. (Mt Sinai Hospital) – Weighted Quantile Sum (WQS) Regression

Moderator: Thomas Webster, Ph.D. (Boston University)

11:00 a.m. – noon Group/Panel Discussion

Panel Discussion Speakers:

- Brent Coull, Ph.D. (Harvard University) – BKMR
- Katrina Waters, Ph.D. (Pacific Northwest National Laboratory) – Variable Selection and Multivariate Adaptive Spline Assessments
- Sung Kyun Park, Sc.D. (University of Michigan – Ann Arbor) – Shrinkage methods including LASSO, Elastic Net, and LARS
- Chris Gennings, Ph.D. (Mt Sinai Hospital) – WQS Regression
- Veronica Vieira, Ph.D. (University of California, Irvine) – Exposure Surface Smoothing (ESS)
- Shuo Chen, Ph.D. (University of California, Irvine) – LARS
- Ghassan Hamra, Ph.D. (Drexel University) – Bayesian Estimation of Weighted Sum

Noon – 1:00 p.m. Lunch (on your own at NIEHS cafeteria)

Simulated Data Set #2

Moderator: Kyla Taylor (NIEHS)

1:00 p.m. – 1:20 p.m. Chris Gennings, Ph.D. (Mt Sinai Hospital) – Development of the Simulated Data Set #2 – What was the right answer?

1:20 p.m. – 1:35 p.m. Emily Mitchell, Ph.D. (NICHD) – Principal Component Analysis (PCA)

1:35 p.m. – 1:50 p.m. Jenna Krall, Ph.D. (Emory University) – Classification and Regression Trees (CART)/PCA

1:50 p.m. – 2:05 p.m. Thomas Webster, Ph.D. (Boston University) – ESS

2:05 p.m. – 2:20 p.m. Alexander Keil, Ph.D. (University of North Carolina – Chapel Hill) – Regression, Bayesian LASSO

2:20 p.m. – 2:35 p.m. David Dunson, Ph.D. (Duke University) – Bayesian Additive Regression Tree (BART)

2:35 p.m. – 2:45 p.m. Break

Moderator: Chris Gennings, Ph.D. (Mt Sinai Hospital)

2:45 p.m. – 3:45 p.m. Group/Panel Discussion

Panel Discussion Speakers:

- Emily Mitchell, Ph.D. (NICHD) – PCA
- Jenna Krall, Ph.D. (Emory University) – CART/PCA
- Thomas Webster, Ph.D. (Boston University) – ESS
- Alexander Keil, Ph.D. (University of North Carolina – Chapel Hill) – Regression, Bayesian LASSO
- David Dunson, Ph.D. (Duke University) – BART
- Shuo Chen, Ph.D. (University of Maryland) – Shrinkage, LARS
- Sung Kyun Park, Sc.D. (University of Michigan – Ann Arbor) – Shrinkage; LASSO, LARS
- David Wheeler, Ph.D. (Virginia Commonwealth University) – WQS Regression (with corrected analysis)

Overall Discussion of Simulated Data Sets 1 & 2

Moderators: Thomas Webster, Ph.D. (Boston University) and Chris Gennings, Ph.D. (Mt Sinai Hospital)

3:45 p.m. – 4:15 p.m. Overall Discussion of Simulated Data Sets 1 & 2

4:15 p.m. – 5:30 p.m. Poster Session (posters to stay up both on Day 1 & 2)

5:30 p.m. Shuttles depart NIEHS for Wingate hotel*

6:45 p.m. Dinner (Mez; dinner for purchase only)**

Day 2, Tuesday, July 14, 2015

7:30 a.m. Shuttles depart Wingate Hotel for NIEHS*

8:30 a.m. – 8:35 a.m. Welcome Day 2 / Logistics – Caroline Dilworth, Ph.D. (NIEHS)

Moderator: Caroline Dilworth, Ph.D. (NIEHS)

8:35 a.m. – 8:40 a.m. Opening Remarks – William Suk, Ph.D. (NIEHS)

8:40 a.m. – 9:00 a.m. Introduction and Overview of Real–World Epi Data Set – Joseph Braun, Ph.D. (Brown University)

Presentations of Analyses of Real World Data Set

Exposure–Response Surface Estimation Strategies

Moderator: Kimberly Gray, Ph.D. (NIEHS)

9:00 a.m. – 9:15 a.m. Birgit Claus Henn, Sc.D. (Boston University) – Bayesian Kernel Machine Regression (BKMR)

9:15 a.m. – 9:30 a.m. Thomas Webster, Ph.D. (Boston University) – Exposure Surface Smoothing (ESS)

9:30 a.m. – 9:45 a.m. David Dunson, Ph.D. (Duke University) – Bayesian Additive Regression Tree (BART)

Moderators: Kimberly Gray, Ph.D. (NIEHS) and Joseph Braun, Ph.D. (Brown University)

9:45 a.m. – 10:15a.m. Panel Discussion

Panel Discussion Speakers:

- Birgit Claus Henn, Sc.D. (Boston University) – BKMR
- Brent Coull, Ph.D. (Harvard University) – BKMR
- Thomas Webster, Ph.D. (Boston University) – ESS
- David Dunson, Ph.D. (Duke University) – BART
- Amy Herring, Sc.D. (University of North Carolina – Chapel Hill) – BART

10:15 a.m. – 10:35 a.m. Break

Principal Component Analysis, Classification, and Prediction Strategies

Moderator: Bonnie Joubert, Ph.D. (NIEHS)

10:35 a.m. – 10:50 a.m. Emily Mitchell, Ph.D. (NICHD) – Principal Component Analysis (PCA)

10:50 a.m. – 11:05 a.m. Sophia Banton (Emory University) – Principal Component Analysis (PCA)

11:05 a.m. – 11:20 a.m. Katrina Waters, Ph.D. (Pacific Northwest National Laboratory) – Variable Selection and Multivariate Adaptive Spline Assessments

Moderators: Bonnie Joubert, Ph.D. (NIEHS) and Russ Hauser, M.D. (Harvard University)

11:20 a.m. – 11:50 a.m. Panel Discussion

Panel Discussion Speakers:

- Emily Mitchell, Ph.D. (NICHD) – PCA
- Sophia Banton (Emory University) – PCA
- Katrina Waters, Ph.D. (Pacific Northwest National Laboratory) – Variable Selection and Multivariate Adaptive Spline Assessments
- Anne Starling, Ph.D. (University of Colorado) – Random Forest
- John Molitor, Ph.D. (Oregon State University) – Bayesian Profile Regression

11:50 a.m. – 12:50 p.m. Lunch (on your own at NIEHS cafeteria)

Variable Selection Strategies

Moderator: Caroline Dilworth, Ph.D. (NIEHS)

12:50 p.m. – 1:05 p.m. Sung Kyun Park, Sc.D. (University of Michigan – Ann Arbor) – Shrinkage methods including Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, and Least Angle Regression (LARS)

1:05 p.m. – 1:20 p.m. Chris Gennings, Ph.D. (Mt Sinai Hospital) – Weighted Quantile Sum (WQS) Regression

1:20 p.m. – 1:35 p.m. Ghassan Hamra, Ph.D. (Drexel University) – Bayesian Estimation of Weighted Sum

Moderators: Caroline Dilworth, Ph.D. (NIEHS) and Richard Kwok, Ph.D. (NIEHS)

1:35 p.m. – 2:05 p.m. Panel Discussion

Panel Discussion Speakers:

- Sung Kyun Park, Sc.D. (University of Michigan – Ann Arbor) – Shrinkage; LASSO, LARS
- Chris Gennings, Ph.D. (Mt Sinai Hospital) – WQS Regression
- Ghassan Hamra, Ph.D. (Drexel University) – Bayesian Estimation of Weighted Sum
- David Wheeler, Ph.D. (Virginia Commonwealth University) – WQS Regression (with corrected analysis)
- Sandra Taylor, Ph.D. (University of California – Davis) – Bootstrap techniques
- Harrison Quick, Ph.D. (CDC) – Quantile-Specific Regression Coefficients
- Sarah Kreidler, Ph.D. (Neptune Incorporated) – Bayesian Networks

2:05 p.m. – 2:25 p.m. Break

Concluding Session: Overall General/Large Group Discussion

Moderators: Russ Hauser, M.D. (Harvard University) and Joseph Braun, Ph.D. (Brown University)

2:25 p.m. – 3:45 p.m. Overall General/Large Group Discussion

3:45 p.m. – 4:00 p.m. Closing Remarks/Paper/Next steps – Danielle Carlin, Ph.D. (NIEHS) and Caroline Dilworth, Ph.D. (NIEHS)

4:00 p.m. Poster removal

4:30 p.m. Shuttles depart NIEHS for RDU airport or Wingate hotel*

Day 3, Wednesday, July 15, 2015

8:00 a.m. – 10:00 a.m. Discussion of workshop and manuscript (Workshop Planning Committee Members)
Location: USEPA–RTP Cafeteria

Questions

Simulated Data Sets #1 and #2:

The presentations (i.e., speakers) should answer the following questions for their analyses of Simulated Data Sets #1 and #2:

- Which exposures contributed to the outcome? Are there any that did not? What metric did you use to make this assessment? (Qualitative)
- How much did the exposures contribute to the outcome? (Quantitative)
- Was there evidence of interaction among the chemicals? Do you see evidence that the association between the exposure(s) and outcome was modified by any of the covariates? Be explicit with your definition of interaction and/or effect measure modification (toxicologists, epidemiologists and biostatisticians tend to think about this quite differently).
- What was the effect of joint/cumulative exposure to the mixture? (Qualitative and/or Quantitative)
- If you didn't get the correct answer for this data set, can you speculate why this might have occurred?

Questions for panel discussions of Simulated Data Sets #1 and #2:

- Why did some approaches work well and others not as well for a given data set?
 - What are the pros and cons of the different approaches presented by the speakers?
 - If your approach was similar to an approach used by others, what did you do differently and why? Do these variations in approach help explain differences in results?
- How easily can the results from the different approaches be interpreted?
 - Are additional complexities worth it?
 - Can results be used by decision makers?
 - What are the potential pitfalls for drawing conclusions on the impacts of exposure?
- Are there any other notable aspects of different approaches that are important to consider (e.g., availability of existing software packages for analysis)?
- Any other advice from what we have learned?

Questions for Overall Discussion of Simulated Data Sets #1 and #2:

- Did the same approaches work for both Simulated Data Set #1 and #2?
- If not, are there differences between the data sets that might provide clues as to when a certain approach may perform better than another and vice versa?

Real World Data Set

The presentations (i.e., speakers) should answer the following questions for their analyses of the Real World Data Set:

- Which exposures contributed to the outcome? Are there any that did not? What metric did you use to make this assessment? (Qualitative)
- How much did the exposures contribute to the outcome? (Quantitative)
- Was there evidence of interaction among the chemicals? Do you see evidence that the association between the exposure(s) and outcome was modified by any of the covariates? Be explicit with your definition of interaction and/or effect measure modification (toxicologists, epidemiologists and biostatisticians tend to think about this quite differently).
- What was the effect of joint/cumulative exposure to the mixture? (Qualitative and/or Quantitative)

Questions for panel discussions of the Real World Data Set:

- Why did some approaches work well and others not as well for this data set?
 - What are the pros and cons of the different approaches presented by the speakers?
 - If your approach was similar to an approach used by others, what did you do differently and why? Do these variations in approach help explain differences in results?
- How easily can the results from the different approaches be interpreted?
 - Are additional complexities worth it?
 - Can results be used by decision makers?
 - What are the potential pitfalls for drawing conclusions on the impacts of exposure?
- Are there any other notable aspects of different approaches that are important to consider (e.g., availability of existing software packages for analysis)?

Questions for General/Large Group Discussion:

- Is there a preference for specific methods across the synthetic and real-world data sets? Is there a preference for a specific method for certain settings, studies or research questions?
 - Is there a limitation to the number of chemicals that we can study with all the approaches discussed at this workshop?
 - Are we limited by the sample size and/or richness of exposure assessment performed in epidemiologic studies?
- With respect to next steps:
 - What are the gaps/needs in developing or utilization of statistical approaches in epidemiological studies and policy associated with complex exposures?
 - Do we need to develop new approaches or improve upon those we already have? Please be specific.
 - What role should funding agencies/researchers play in recommending approaches and helping in further development and application of approaches?
 - How should we publicize findings from this workshop?