Unlocking the Mysteries of Mixed Exposures: Targeted Learning for Robust Discovery and Causal Inference in Epidemiology

In the realm of epidemiological research, particularly when dealing with high-dimensional data involving mixed exposures such as air pollution, pesticides, pharmaceuticals, or nutrition, researchers confront the complex task of understanding intricate interactions and identifying vulnerable subpopulations. These challenges, coupled with the need to determine safe exposure limits, are pivotal for effective regulation. Traditional methods often fall short in addressing the multitude of possible interactions and lack the subtlety needed to uncover complex relationships inherent in mixed exposure studies.

Enter Targeted Learning (TL) – a dynamic and evolving field that blends machine learning, causal inference, and semiparametric statistical theory to provide credible causal answers. TL stands at the forefront, bridging the gap between discovery (unearthing exposure relationships including interactions and effect modifiers) and precise estimation (quantifying effects considering these variables). This workshop is designed to offer a deep dive into TL's potent data-adaptive methodologies and its comprehensive R package suites within the TLverse ecosystem.

Structured around real epidemiologic data on mixed exposures, the workshop combines lectures, practical worksheets, and hands-on programming exercises. Participants will emerge with not only a deeper understanding of these advanced strategies but also practical, transferable skills.

Workshop Outline:

1. **Study Design Exercise with NHANES Data:**

We commence with a participatory study design exercise using the National Health and Nutrition Examination Survey (NHANES) dataset, previously employed in evaluating statistical methods for exposure mixtures. This exercise will focus on the relationship between persistent organic pollutants (POPs) exposure and leukocyte telomere length (LTL), accounting for various covariates. Students will break into groups and discuss how they would investigate this data.

2. **Introduction to Data-Adaptive Parameters:**

The concept of data-adaptive parameters, crucial for tackling unknown interactions and modifications in mixed exposures, will be introduced by Mark van der Laan and Alan Hubbard. This segment includes exploring data-splitting, using one part to discover and another to estimate a data-adaptive target parameter.

3. **Efficient Estimation and Super Learning:**

Alejando Schuler delves into efficient estimation, target parameter definition, and the statistical framework for data-adaptive parameters. Rachel Philips will provide a focused session on Super Learning, integrating both cross-validation and advanced learning strategies.

4. **Software and Estimation Strategies:**

A thorough exploration of software tools and estimation strategies begins with identifying policy-relevant targets in mixed exposures. We discuss how to define and measure interactions in identified exposures using flexible estimators. David McCoy will lead the focus on the SuperNOVA package, which adeptly identifies interactions, effect modifications, and mediation, employing stochastic interventions for outcome analysis.

5. **Practical Application and Method Comparison:**

Participants will apply the SuperNOVA method on the NHANES data in group settings, juxtaposing it with other mixture methods to understand underlying assumptions. This hands-on session emphasizes real-world application and interpretation.

6. **Closing Discussion on Robustness and Consistency in Research:**

The workshop concludes with a discussion on the robustness of data-adaptive findings and the importance of a structured statistical approach in enhancing consistency in research literature. Closing remarks by David McCoy and Mark van der Laan will encapsulate key insights.

Join Us:

Embark on an enlightening journey in advanced epidemiologic research tailored for high-dimensional, mixed exposure data. This workshop is ideal for individuals with intermediate statistical knowledge and R programming proficiency.

Course Teachers

David McCoy
Mark van der Laan
Alan Hubbard
Alejandro Shuler
Rachael Phillips
Ivana Malenica