**Applications of Causal-Inference Concepts and Machine-learning Methods to Investigate Cancer Clusters**

John Maddox (University of Tennessee, Knoxville, TN 37659)

Ashley Rice (Oak Ridge National Laboratory, Oak Ridge, TN 37830)

An enduring public health goal is that of discovering environmental exposures that may explain cancer clusters. The Standardized Incidence Ratio (SIR), a ratio of observed and expected cancer incidence within a designated population, is commonly used to identify these localized regions of increased cancer incidence. However, the SIR is statistically flawed, because it assigns causality to suspected environmental factors within the area of interest independent of exposure. A similar but improved cluster-identifying tool, called the causal SIR (cSIR), has been proposed as a solution to the SIR-related problem. The cSIR accounts for exposure to chemicals of concern by establishing community exposure *before* identifying the cancer cluster. Thus, subsequent causal links between greater cancer incidence and exposures are statistically valid. Our research involved using public data to search for causal relationships between dioxin exposure and a lung- and brain-cancer cluster in Davis County, Utah. We implemented this search by calculating the cSIR metric for exposed census tracts. Before evaluating cancer incidence, we performed cosine-similarity matching on socioeconomic, demographic, and health-confounding variables to ensure similarity between populations in exposed and non-exposed census tracts. Then, we tested multiple methods to resolve granularity differences between a larger county-level cancer incidence and the smaller census tract-level covariate dataset including areal interpolation schemes, multiple-imputation schemes, and machine learning. Gradient boosting regressor machine learning models performed the best and were used to calculate the cSIR scores. Our analysis of cancer incidence between dioxin-exposed and the matched non-exposed census tracts using cSIR supported/did not support the hypothesis of a causal relationship between brain/lung/brain and lung cancer and dioxin exposure. However, the possibility of a causal link between dioxin exposure and increased cancer incidence needs to be confirmed. Overall, community-reported cancer clusters are relatively common, and if an elevated incidence is identified mathematically, future investigations should use cSIR to identify environmental causes.