

## BIostatistics AND DATA MANAGEMENT CORE ABSTRACT

**EPA Grant #:** RD83479701      **EPA Project Officer:** Mel Peffers/Sherri Hunt

**Title:** Biostatistics and Data Management Core

**Investigators:** Bhramar Mukherjee, Cathie Spino, Brisa Sanchez and Lu Wang

**Institution:** The University of Michigan, Ann Arbor, MI

**Project Period:** 12/1/2010 – 11/30/2015      **Core Costs:** \$721,515

**RFA:** Clean Air Research Centers      **Research Category:** Air Quality

**Description:** The Biostatistics and Data Management Core (BDMC) is an integral part of all three Projects of the Great Lakes Center for Integrative Environmental Research (GLACIER).

**Objectives/Hypothesis:** The objectives of the GLACIER BDMC are to provide guidance for the statistical design and analysis of studies and data management services that allow for the integration of the data into a single platform that facilitates timely analysis for GLACIER investigators and projects. Specific aims for this Core are (1) to assist GLACIER investigators with statistical and data aspects of their research by providing expertise in the design, conduct and analysis of studies in Projects 1-3; (2) to establish a database for each project on a secure computerized system; and (3) to implement a website to allow data import and export in a secured, controlled environment with a user-friendly interface.

**Approach:** Aim 1. The core biostatisticians were involved in designing the preliminary analytic plans in the proposals, but as the research progresses the team will adapt to the analytic needs of the projects and will collaborate on new modeling and methodological issues that may arise. This will be accomplished by reviewing data periodically during the conduct of the study to assess distributional and other model assumptions that will inform the development of models and methods. Beyond the standard analytical and design needs, the BDMC will focus on certain methodological challenges in the context of modeling mixtures of multipollutants that are closely related to Projects 1-3, such as (1) the issue of fitting random effects threshold models, (2) the choice of variable selection and use of functional data analytic techniques with high frequency correlated data and multiple pollutants, and (3) identification of exposure windows of vulnerability. Aim 2. Separate Oracle® database will be developed for each study project. Data structures will be flexible to allow for the storage of the study data in multiple ways, specific to each study design. SAS data files, codebooks and study reports will be available on the web, allowing access to the researchers and the analysts. Aim 3. A network-only (internal) web site for the GLACIER team will be created and maintained. Access to privileged information on the web site will be granted using password authentication. Any information that is transferred between the web server and remote sites will be encrypted to ensure security.

**Expected Results:** The success of the GLACIER BDMC will be the timely completion of analyses owing to the integration of biostatisticians and clinical study professionals in all aspects of the studies --- from design of studies and data collection instruments, quality control measures and monitoring during the conduct of the studies, and development of analysis plans with close collaboration with clinical investigators. Thus, the publication of study findings in a timely manner, with application of innovative and rigorous statistical methods, is the ultimate expected result of this Core.

**Supplemental Keywords:** biostatistics, data management, modeling of multipollutant mixtures, high dimensional correlated data