

# Yale SCHOOL OF PUBLIC HEALTH

## *Biostatistics*

### **A Framework for Automated and Reproducible Geomarker Curation and Computation at Scale**

**Cole Brokamp, PhD**  
**Associate Professor**  
**Department of Pediatrics**  
**Cincinnati Children's Hospital Medical Center and the**  
**University of Cincinnati College of Medicine**

**12:00 Noon Eastern time, Tuesday, October 25, 2022**  
**47 College Street, Room 106 A&B**

#### **ABSTRACT**

Environmental exposures and community characteristics (including air pollution, greenspace, crime, and indices of community deprivation) are powerful descriptors, predictors, and causes of health and health disparities. These data, termed "geomarkers", combined with the democratization of "big spatial data" and advances in geoinformatics, have paved the way for the expansion of precision medicine into precision public health. Automated linkage of these newly available data to new and existing residential addresses in health studies, however, is often difficult due to data heterogeneity with respect to spatiotemporal resolution and extent, annotation, storage, formats, retrieval methods, and computational strategies. This hinders data accessibility and interoperability. Because precise geolocation is considered protected health information (PHI), research regulations designed to protect the identities of study participants often present obstacles to sharing data and utilizing third party tools. To address these limitations, we have introduced DeGAUSS (\*Decentralized Geomarker Assessment for Multi-Site Studies\*), a curated and standardized library that researchers can use for efficient, automated, and reproducible linkage of geomarkers to their own data \*and\* a generalized framework to which exposure assessment scientists can contribute. DeGAUSS shines in applications where two or more organizations want to learn about their place-based data, but are not able to share individual geolocation data; for example a multi-site health study or a community health network of government and community institutions. Within community health, open and equitable data need to be disaggregated enough to illuminate opportunities for targeted actions, but must protect personal privacy. DeGAUSS can be used to share hyperlocal, disaggregated data among institutions to collectively learn in a decentralized manner how to best help residents thrive. DeGAUSS has proved to make geomarker data and methods more findable, accessible, interoperable, and reusable (FAIR) and we continue to seek to make our approach a widely-adopted, community-maintained, and sustainable resource to fuel the advancement of precision public health.