

## ARIC Manuscript Proposal # 932

PC Reviewed: 04/02/03

Status:   A  

Priority:   2  

SC Reviewed: 04/08/03

Status:   A  

Priority:   2  

**1.a. Full Title:**

Accuracy and Repeatability of Commercial Geocoding  
in the Life Course Socioeconomic Status (LCSES) Study

**b. Abbreviated Title (Length 26 characters):**

Geocoding in LCSES

**2. Writing Group (list individual with lead responsibility first):**

**Lead:** Eric A. Whitsel, MD MPH  
**Address:** University of North Carolina at Chapel Hill  
Departments of Medicine and Epidemiology  
Cardiovascular Disease Program  
Bank of America Center, Suite 306  
137 East Franklin Street  
Chapel Hill, NC 27514  
**Phone:** 919.966.3168  
**Fax:** 919.966.9800  
**E-mail:** ewhitsel@email.unc.edu

**Members:** Rose K, Wood J, Heiss G, Liao D, Smith R

**3. Timeline:** Initial data analyses will be completed in preparation for an April 1, 2003 American College of Epidemiology (ACE) abstract submission deadline. Remaining analyses will be completed prior to the September 7-9 ACE meeting. The associated manuscript will be submitted thereafter.

**4. Rationale:** As emphasized by Krieger (2001), commercial geocoding of residential addresses is not entirely accurate. Addresses can be assigned, for example, the wrong latitude, longitude or block group. Moreover, the repeatability of this process has not been adequately explored to date. Further exploration of these issues is warranted because unreliable geocoding would bias e.g. estimates of the modifying effects of neighborhood-level characteristics on inverse associations between individual-level socioeconomic exposures and cardiovascular disease outcomes. Based on these observations and published recommendations, we have therefore decided to systematically evaluate commercial geocoding methods using data from the Life Course Socioeconomic Status (LCSES) study and the U.S. Environmental Protection Agency (EPA) Air Quality System (AQS) database.

**Main Hypotheses:**

[1] Commercial geocoding methods are accurate.

[2] Commercial geocoding methods are repeatable.

**6. Data (variables, time window, source, inclusions/exclusions):**

[1] Commercial geocoding methods are accurate. The Office of Air Quality Planning and Standards will extract from the AQS database (1997-2002) an electronic census of all spatially characterized addresses located in or adjacent to counties housing the majority of ARIC participants. We will submit the addresses—which are associated with fixed-site, ambient criteria pollutant (PM<sub>10</sub>; NO<sub>2</sub>; SO<sub>2</sub>; CO; O<sub>3</sub>) monitors operated by the U.S. EPA Aerometric Information Retrieval System—to a commercial geocoder (CG) for assignment of longitudes and latitudes. We will convert the assigned longitudes and latitudes in decimal degrees to radians and estimate their accuracy in three ways: [1] as an average distance (m) between the CG and AQS coordinate pairs determined using the Haversine spherical earth formula, [2] as an average bearing (°) from the CG to AQS coordinate pairs using angle trigonometry, and [3] as an error rate in block group assignment. Sample size permitting, we also will investigate the influence of address characteristics (e.g. Residential vs. Commercial vs. Industrial vs. Agricultural) on these estimates.

[2] Commercial geocoding methods are repeatable. Under the legally binding and IRB-approved terms of our institutional contract with the commercial geocoder, we will confidentially submit (using a secure data transfer protocol) an identical list of ARIC participant addresses on two occasions separated in time by approximately nine months for assignment of longitudes and latitudes. We will convert the assigned longitudes and latitudes in decimal degrees to radians and estimate their repeatability (before and after log transformation) in several ways: [1] as an intra-class correlation coefficient obtained from a balanced, random-effects, one-way analysis of variance and [2] as one or more of the measures described above. We also will investigate the influence of address characteristics (e.g. Urban vs. Suburban vs. Rural) on these estimates.

**7.a. Will the data be used for non-CVD analysis in this manuscript?  Yes  No**

Although this study will use ancillary study data collected in LCSES for non-CVD analyses (to determine the accuracy and repeatability of geocoding), it will do so to justify use of geocodes in contextual analyses of CVD outcomes. Therefore, we do not intend to use the file ICTDER02 to exclude participants with a value of RES\_OTH='CVD Research', as suggested below.

**7b. If yes, is the author aware that the file ICTDER02 must be used to exclude persons with a value RES\_OTH="CVD Research" for non-DNA analysis, and for DNA analysis RES\_DNA="CVD Research" would be used?  Yes  No  NA (This file ICTDER02 has been distributed to ARIC Pis, and contains the response to consent updates related to stored sample use for research.)**

**8.a. Will the DNA data be used in this manuscript?  Yes  No**

**8b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER02 must be used to exclude those with value RES\_DNA = "No use/storage DNA"?  Yes  No  NA**