ARIC Manuscript Proposal # 1595

PC Reviewed: 1/12/10              Status: A              Priority: 2
SC Reviewed: _________              Status: _____              Priority: ____

1.a. Full Title:

Historical Records as a Source of Early Life Socioeconomic Data: Feasibility and Data Quality

b. Abbreviated Title (Length 26 characters):

Ascertaining Early Life SES

2. Writing Group:

Writing group members:

Cynthia Owens, Kathryn Rose, Mehul Patel, Jay Kauffman, Heejung Bang, Gerardo Heiss, Kathryn O’Brien

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. [please confirm with your initials electronically or in writing]

First author: Cynthia Owens

Address:

Bank of America Center
137 E Franklin St
Chapel Hill, NC 27514

Phone: 919-966-1967    Fax: 919-966-9800
E-mail: cindy_owens@unc.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Kathryn Rose

Phone: 919-966-1967    Fax: 919-966-9800
E-mail: Kathryn_rose@unc.edu
3. **Timeline:**


4. **Rationale:**

As part of ARIC ancillary study 2003.07, we are assessing the feasibility and practicability of (1) obtaining birth certificates and census records for ARIC decedents and (2) using them as a source of information for early life socioeconomic status. One study of Alzheimer’s patients was able to obtain parental education from such records for a large number of participants (1) and in a pilot study based on a subset of ARIC Forsyth decedents also suggested that at least a birth certificate or a census record could be located for 85% of decedents and that parental (father’s) occupation data was generally complete and legible (2).

Obtaining this information is of interest for several reasons. In epidemiologic studies, childhood SES is typically ascertained by having adults recall their parents’ occupation when they (the adult respondents) were children (3 – 8). While problems related to recall bias (9) and (10) and the limited validity of proxy reports are well documented, neither these issues nor the potential for bias resulting from selective survival where early life SES is ascertained after a cohort’s inception, have been adequately addressed in the life course literature (11).

The ARIC study provides an excellent context in which to address these issues. In the ARIC ancillary Life Course SES, Social Context, and Cardiovascular Disease (LC-SES) study (AS 2003.07), early life SES was queried by phone during annual follow-up (2001-2002). At this time, data were obtained for approximately 12,700 of the original 15,792 cohort members. Most participants not providing early life SES at this annual follow-up survey were deceased. A primary goal of our current ancillary study (AS 2003.07) was to collect early life SES data from historical records on all decedents (including those who have died after participating in the LC-SES study) as of 2007 (N=3,444)

5. **Main Hypothesis/Study Questions:**

1. A major focus of this manuscript will be to describe the methods used to conduct our searches for the historical records described above. We will describe the strengths and limitations of all data sources (death certificates, the Social Security Death Index, birth certificates, 1930 US Census data, city directories and others) and discuss the challenges encountered with accessing data that is not publicly available, including rules imposed by state vital statistics offices and the current trends in data restriction. The primary purpose of this aspect of the manuscript will be to provide a blueprint for other life course researchers who may want to incorporate historical data as a primary data source or as a validation sample.

2. A second focus of the manuscript will be to describe the completeness of the data and to assess the comparability of the parental occupation data across three sources (birth
certificates, declassified census records, and city directories). Based on the results of our pilot study (3) we hypothesize that:

- Either a birth certificate or a census record will be obtained for most participants.
- Proportionately more records will be found for whites than blacks and for men than women.
- Parental occupation data will be available on most records located, though completeness of this data will vary by source.
- Concordance of parental occupation data across sources (birth certificates, census records, city directories) will be high.
- Concordance of place of birth/childhood residence will be high across data sources.

**6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).**

The historical parental occupation data has been collected over the past three years. Briefly, parent names were obtained from decedent death certificates and social security records. This information was used to conduct searches for birth certificates, declassified census records, and most recently, city directory listings. Access databases link the data across sources. After personal identifiers are removed, the data will be sent to CSCC for linkage with extant ARIC and LC-SES datasets.

The proportion of decedents for whom linkage with birth certificates, census records and city directory listings is made will be calculated overall and by socio-demographic characteristics (age/birth cohort, race, gender, adult SES). We will also assess the quality and completeness of the sociodemographic data contained in each record. Standard Chi-Square Tests will be used to evaluate variations in linkage with birth certificates and census records and in the completeness of data contained on these records by sociodemographic characteristics (age, race, gender, adult SES). Descriptive analyses will use simple frequency distributions to calculate percentage agreement. Chi Square analyses will be used to assess variations and associations by race and gender. Kappa statistics will be calculated to assess agreement of parental occupation (grouped into census occupational groups) on the birth, census, and city directory records.

**7.a. Will the data be used for non-CVD analysis in this manuscript?**

X Yes ______ No

This is a methodologic paper. However, papers based on this work will be used to address questions related to socioeconomic disparities in CVD.

**b. If Yes, is the author aware that the file ICTDER03 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

(This file ICTDER03 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)
8.a. Will the DNA data be used in this manuscript?
    ___ Yes  ___ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?  ___ Yes  ___ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php
    ___ X___ Yes  _______ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

We cannot find the ARIC MS # in the current listing on the ARIC website. However, it is an approved MS and has been published. See reference # 3 in the references section.

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?
    ___ X___ Yes  ___ No

11.b. If yes, is the proposal
    ___ X___ A. primarily the result of an ancillary study (list number* 2003.07)
    ___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

References


