ARIC Manuscript Proposal #2744

PC Reviewed: 4/12/16 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Regression Analysis of Multivariate Interval-Censored Data

b. Abbreviated Title (Length 26 characters): Interval-Censored Data

2. Writing Group:
Writing group members: Danyu Lin, Fei Gao, Donglin Zeng, David Couper

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

First author: **Danyu Lin**
Address: 3101E McGavran-Greenberg Hall, Chapel Hill, NC 27599

Phone: 919 843-5134 Fax: 919 966-3804
E-mail: lin@bios.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: **David Couper**
Address: 137 E. Franklin St. Suite 203, Campus Box 8030, Chapel Hill, NC 27599

Phone: 919-962-3229 Fax: 919-962-3265
E-mail: david_couper@unc.edu

3. Timeline: 2016-4 to 2017-12
4. Rationale:

Multivariate interval-censored data arise when each study subject can experience multiple events and the event times are not observed exactly but rather known only to lie in certain intervals. Such data are commonly encountered in epidemiological and clinical studies, including the ARIC study. These data allow characterization of the dependence of multiple events and evaluation of the effects of covariates on the multivariate outcome. The fact that event times are not exactly observed, together with their dependence, makes the analysis theoretically and

computationally challenging. In this work, we will develop valid and efficient statistical methods for regression analysis of multivariate interval censored data. The proposed methods will allow ARIC and other investigators to address specific scientific questions related to multivariate interval censored data in an unbiased and efficient manner.

5. Main Hypothesis/Study Questions:

We formulate the effects of covariates on multiple event times by a broad class of semiparametric regression models with random effects and consider maximum likelihood estimation under general interval censoring. The proposed estimators are consistent, asymptotically normal, and statistically efficient. We will evaluate the performance of the proposed methods through simulation studies. In addition, we will test our methods with the ARIC data, including the incidence of hypertension, diabetes, and peripheral arterial disease. Our purpose for using the ARIC data is to illustrate the new statistical methods, rather than addressing specific biological questions. The manuscript will be submitted to a statistical journal.

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).
Hypertension (HYPER), Diabetes (DIABTS): Visit 1-5
Peripheral arterial disease (PAD), Symptomatic Coronary Heart Disease, Prevalent Coronary Heart Disease: Visit 1-5
BMI, Drinker status, Cigarette smoking status, Family history of Stroke, Family history of CHD, Family history of diabetes, Visit Date: Visit 1-5
Age, Gender, Race, Education levels, Physical Activity (at work, sports during leisure time, during leisure time excluding sports), Systolic BP, Diastolic BP, Total calorie intake: Visit 1

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

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 ICTDER 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

_____ 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)?
None

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

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* _____)

___ 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/

12a. 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.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __✓__ No.