ARIC Manuscript Proposal #2350

PC Reviewed: 4/8/14  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: Use Whole Cohort Information to Improve the Efficiency of Multivariate Marginal Hazard Model for Case-Cohort Studies

b. Abbreviated Title (Length 26 characters): Case-cohort Studies

2. Writing Group:
   Writing group members: Hongtao Zhang, Jianwen Cai and David Couper

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

First author: Hongtao Zhang
Address: Department of Biostatistics, CB#7420
         University of North Carolina at Chapel Hill
         Chapel Hill, NC 27599-7420
         Phone: 919-316-0256  Fax: 919-316-0256
         E-mail: hongtaoz@live.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 27599-8030
            Phone: 919-962-3229  Fax: 919-962-3265
            E-mail: david_couper@unc.edu

3. Timeline: Work can begin as soon as approval is received.

4. Rationale: The case-cohort design is originally developed for large cohort studies, where the cost to assemble the exposure and covariates is huge. Such design is appealing especially when the event rate is low and the exposure measurements are expensive to obtain. The design effect needs to be taken into consideration when data
from case-cohort studies are analyzed. A key advantage of the case-cohort study design is its capacity to use the same subcohort for several diseases or for several subtypes of disease. In order to compare the effect of a risk factor on different types of diseases, times to different events need to be modeled simultaneously. Valid statistical methods that take the correlations among the outcomes from the same subject into account need to be developed. Recently, Kang and Cai (2009) proposed methods for fitting failure time data from case-cohort studies with multiple disease outcomes under marginal proportional hazards models. However, their method could potentially be improved by exploiting covariate information collected on the entire cohort. In our research, we will propose a doubly-weighted estimation procedure which utilizes the covariate information in the whole cohort. We will study the asymptotic properties of the proposed method. As part of the research, we will also apply our method to a real case-cohort study to illustrate its use. We are requesting data from the ARIC study for this purpose. These data will only be published as part of a methodological paper with acknowledgement of the ARIC study.

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

We want to use the proposed method to compare the effects of Lipoprotein-associated Phospholipase A2 and high-sensitivity C-reactive protein on incident coronary heart disease and on stroke using data collected under the case-cohort design while adjusting for possible confounding factors.

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 study design is a case-cohort study. We will consider two types of outcomes – incident CHD and incident stroke. We will use the same data set that was created for ARIC MS1621: Marginal Additive Hazards Model for Case-cohort Studies with Multiple Disease Outcomes. The primary predictors of interest are lipoprotein-associated Phospholipase A2 and high-sensitivity C-reactive protein measurements. Other covariates needed include age at visit 2, race, gender, smoking status, lipid measurements, blood pressure and diabetes status. A new statistical method is being developed by Hongtao Zhang as part of his Biostatistics PhD dissertation. This new method will attempt to use the covariate information in the whole cohort to improve the efficiency. Part of his method development includes using simulation studies to examine performance of the model. Then the new method will be applied to the ARIC data as an illustration.

7.a. **Will the data be used for non-CVD analysis in this manuscript?**  
_____ Yes  
__X__ 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  ___X__ 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.csc.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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?
_____ Yes  ___X__ 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