ARIC Manuscript Proposal # 3142

PC Reviewed: 03/20/18  Status: _____  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: A proportional hazards model for the subdistribution in generalized case-cohort designs

b. Abbreviated Title (Length 26 characters): Generalized case-cohort designs

2. Writing Group:
Writing group members:
Yayun Xu, Division of Biostatistics, Medical College of Wisconsin
Soyoung Kim, Division of Biostatistics, Medical College of Wisconsin
Kwang Woo Ahn, Division of Biostatistics, Medical College of Wisconsin
Mei-Jie Zhang, Division of Biostatistics, Medical College of Wisconsin

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

First author:  Yayun Xu
Address:  8701 W Watertown Plank Rd, Milwaukee, WI 53226

Phone:  (414) 955-7424  Fax:
E-mail:  yayunxu@mcw.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 East Franklin Street, Suite 203, CB #8030, Chapel Hill, NC 27599

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

3. Timeline: plan to submit the paper by 12/30/2018

4. Rationale: In large cohort studies, it can be expensive for obtaining expensive covariate information on all members in the entire cohort. In order to reduce cost, the case-cohort study design was proposed by Prentice (1986). In such design, a random sample from the full cohort,
namely subcohort, is selected via simple random sampling, then all subjects having events of interest outside this subcohort are sampled. In many cohort studies, the number of cases can be large, because the event is relatively common, the cohort size is large, or the follow-up duration is long. Under the situation, it is not feasible to collect expensive covariate information on all cases. Instead of collecting exposure information from all cases, a generalized case-cohort design was proposed where only a fraction of the nonsubcohort cases were sampled for exposure assessment.

Many methods have been proposed for analyzing generalized case-cohort data under the proportional hazards model with survival outcomes. However, statistical methods for generalized case-cohort data with competing risks outcomes have not been studied. Competing risk often occurs in the analysis of survival data. A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest. For example, time to diabetes and time to death are competing risks. If subjects die before having the event of diabetes, time to diabetes cannot be observed. In this research, we propose a subdistribution hazards model to analyze competing risks data for generalized case-cohort studies.

5. Main Hypothesis/Study Questions:

How can we analyze the competing risks data under the generalized case-cohort study?

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).

We would like to apply the proposed subdistribution hazards model for generalized case-cohort designs to the ARIC data. In the paper of Duncan et al (2003), they examined the association of low-grade systemic inflammation with diabetes and the data was based on a generalized case-cohort design from the ARIC study. They fitted weighted proportional hazards model for diabetes and inflammation (IL-6, CRP, Oromucoid, Sialic acid) after adjusting age, center, sex, ethnicity, parental history of diabetes, hypertension, BMI, waist-to-hip ratio, fasting glucose, and fasting insulin.

To apply our method to ARIC data, we need three more variables including time to death, death indicator, and cause of deaths in addition to a generalized case-cohort data from the ARIC study. Then diabetes and death are competing risks.

Study design: Using a generalized case cohort data with competing risks (diabetes and death) from ARIC study, we will fit the proposed subdistribution hazards model for a generalized case-cohort design to examine the association between competing risks outcomes (diabetes/death) and inflation marker.

Inclusion/exclusion: We will use same inclusion/exclusion criteria with Duncan et al. (2003). They excluded 2,018 participants with prevalent diabetes, 95 members of minority ethnic groups
with small numbers, 853 not returning to any follow-up visit, 26 having no valid diabetes determination at follow-ups, 7 with restrictions on stored plasma use, 12 with missing baseline anthropometrics, and 2,506 participants in previous ARIC case-control studies involving cardiovascular disease for whom stored plasma was either previously exhausted or held in reserve. This resulted in a final sample of 10,275 individuals.

**Outcome:**

a) time to death
b) time to diabetes defined as diabetes on the basis of 1) a reported physician diagnosis, 2) use of antidiabetes medications, 3) a fasting ($\geq 8$ h) glucose $\geq 7.0$ mmol/l, or 4) a nonfasting glucose of $\geq 11.1$ mmol/l.

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

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”? ____ X____ 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](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)?


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

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.cscn.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.cscn.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.