ARIC Manuscript Proposal #2839

PC Reviewed: 09/13/16 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Multiple imputation of missing data in nested case-control and case-cohort studies

b. Abbreviated Title (Length 26 characters): MI and study design

2. Writing Group:
Writing group members: Angela Wood, Ruth Keogh, Shaun Seaman, Jonathan Bartlett

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

First author: **Dr Angela Wood**
Address: Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN

Phone: 01223 748652 Fax:
E-mail: amw79@medschl.cam.ac.uk

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: ARIC Study Coordinating Center PI, Clinical Professor and Deputy Director, Collaborative Studies Coordinating Center, UNC Dept. of Biostatistics, 137 E. Franklin St., Ste. 203, CB# 8030, Chapel Hill, NC 27514

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

3. Timeline: 1 year

4. Rationale:

Approaches to perform multiple imputation of missing data need to be adapted when applied to nested case-control and case-cohort studies.

5. Main Hypothesis/Study Questions:

We will outline extensions of (1) multiple imputation by chained equations and (2) the substantive model compatible imputation approach for imputation of missing values in nested case-control and case-cohort studies. Our methods will be assessed using simulation studies and illustrated using example data from ARIC. Example code in Stata and R will be provided in supplementary materials.

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).

Here we describe the proposed analysis on the ARIC study, to be used as an illustration of our methods.

We will assume the model of interest is a Cox proportional hazards model, with time to first CVD event as the outcome of interest, and (Framingham) covariates of interest to include: sex, age, HDL, total cholesterol, smoking status, SBP and history of diabetes. We will first explore the extent of missing values, and if not substantive for illustrating our methods we will induce further missing data under an assumed Missing At Random assumption.

From the full ARIC cohort, we will then create (i) a nested case-control study by matching on age and sex each case to 2 controls and (ii) an unmatched case-cohort study using a subcohort of 5% of individuals.

The two proposed multiple imputation approaches will then be applied to the full cohort, the nested case-control study and the case-cohort study. We will compare the model regression coefficients and their standard errors along with the 10-year baseline survival estimate. Similar results from the different study designs and MI methods will illustrate how our methodological extensions work in this study (note – we have already performed more rigorous simulation studies).

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

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)?**

    Other methodological papers which have used the ARIC study as an example include "Using repeated measurements in risk prediction models" proposal 1958, led by colleague Michael Sweeting.

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