ARIC Manuscript Proposal #2115

PC Reviewed: 4/9/13	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Sensitivity Analyses with Shared-Parameter Models for studying Cognitive Change in the presence of potentially Informative Dropout – the Atherosclerosis Risk in Communities (ARIC) Neurocognitive Study

b. Abbreviated Title (Length 26 characters): SPM NCS

2. Writing Group:
(Alphabetical) Karen Bandeen-Roche, Andrea L.C. Schneider, Josef Coresh, Jennifer Deal, Rebecca Gottesman, Michael Griswold, Thomas Mosley, Andreea Rawlings, A. Richey Sharrett, Daniel Scharfstein, Lisa Wruck

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

First author: **Michael Griswold**
Address: Center of Biostatistics
University of Mississippi Medical Center
2500 North State Street
Jackson, MS 39216
Phone: (601) 984-4933 Fax: NA
E-mail: mgriswold@umc.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: **Richey Sharrett**
Address: 615 North Wolfe Street, Room W6009
Baltimore, MD 21205
Phone: (443) 287-6178 Fax: (410) 955-0863
E-mail: rsharret@jhsph.edu

3. Timeline:
Manuscript will be completed in 3 months.

4. Rationale:
Analysis of longitudinal studies are often predicated on strong, unverifiable assumptions about missing data and primary conclusions drawn from such studies often critically depend on these assumptions. Sensitivity analyses can be used to examine the robustness of the results to a set of plausible alternative assumptions and thereby add an important verification component to the results obtained. For example, standard generalized linear mixed models (GLMM) assume that the mechanism giving rise to missing outcome values is dependent on observed data, but is independent of data that is unobserved, i.e. data is missing at random (MAR) in the framework of Little and Rubin[1]. The MAR assumption does not hold, for example, if the outcome changes of primary interest are more dramatic among those participants who have dropped out compared to

those (often healthier) participants remaining in the study; MAR estimates are then biased. Critically, assumptions about missing data are untestable.

In order to minimize and evaluate the effect of missing data in longitudinal studies, therefore, (1) additional data are needed (i.e., other information is available) and/or (2) mechanisms are needed by which to evaluate assumptions about missingness.

In epidemiologic studies of longitudinal cognitive outcomes, some design mechanisms exist for capturing cognitive data from participants who may refuse a study visit, such as telephone[2] or proxy[3] assessments, and can help increase participation. However, such alternative assessments are generally simpler, surrogate measures and are often still subject to informative selection pressures related to the outcomes under study. Therefore, in longitudinal studies of cognitive function, statistical approaches that flexibly incorporate additional data and assist with understanding possible impacts of missing data assumptions on study inferences are necessary.

Shared parameter models (SPMs), also known as joint models, simultaneously model both longitudinal outcomes as well as time to dropout, linking the two models through a set of "shared parameters" which are most often specified through random effects. We will focus on SPMs, which have been less well-described in the literature than alternative methods to examine missingness assumptions (e.g., selection models, pattern-mixture models, etc). SPM's derive primary estimates for both longitudinal outcomes and attrition from a simultaneous (joint) model but additionally allow for the specification of a sensitivity parameter for attrition effects. This makes these models ideally suited for sensitivity analysis in which model estimates are based on a variety of plausible missingness assumptions that can be easily compared.

Here we propose a manuscript with the goal of translating the value of shared parameter models for primary results and sensitivity analyses to a clinical and/or epidemiologic audience. We will illustrate differences in estimates from models incorporating several MNAR assumptions and their related MAR formulation, as well as juxtapose these results with an alternative MAR approach using inverse probability of attrition weights (as reported in our companion piece, ms# 1982). We will use education as the primary exposure of interest. As a translational piece, we will emphasize the research questions addressed by SPMs, and on the interpretation of the model estimates. Programming code for common statistical software packages (e.g., SAS, Stata) will be presented.

This paper will broadly highlight the methodological challenge of missing cognitive data in longitudinal epidemiologic studies, and specifically focus on the utility of shared parameter models to calculate estimates and evaluate assumptions about missing data. We will highlight the utility of SPMs as a family of models for primary estimation and sensitivity analysis.

5. Main Hypothesis/Study Questions:

- We will optimize models of cognitive decline accounting for attrition using shared parameter models.
- We will develop a model using covariates similar to those used in ARIC ms# 1982, to facilitate comparison with an alternative inverse weighting approach.
- We hypothesize that, compared with a model which does not consider informative attrition, the SPM estimates will show
    - faster rates of cognitive decline regardless of education level
    - altered associations of education with decline only if dropout effects are differential by education strata

- When both methods are specified under similar assumptions (i.e. MAR), we expect that estimated relationships between education and cognitive decline will be similar between an SPM and an inverse probability of attrition weighting approach.
- The SPM model will have utility for both estimating primary results and examining the sensitivity of those primary results, additionally facilitating comparisons with alternative model results such as those reported in our companion piece, ms# 1982.

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 analysis will follow closely that by Gottesman et al. (MP 1982**)** in order to juxtapose results from a model of inverse probability of attrition weighting with results from a shared parameter model.

Study design: Prospective observational study of male and female ARIC participants seen at Visit 2 (1990-92) (mean age at visit 2: 57.5 ± 5.7 years; 24.2% black race; 55.7% female).

Exclusion criteria:
The study population is all ARIC participants examined at visit 2, after excluding those that had no cognitive test scores or were missing years of education, or values for the needed covariates. For comparison with ms# 1982, the same population (n=14,069) will be considered eligible.

Outcome: Cognitive change from visit 2 (1990-92) to the current visit (2011-13). Cognitive function will be measured using three neurocognitive tests. The Delayed Word Recall Test (DWRT)[5] is a test of verbal learning and memory. Participants are asked to learn ten common nouns and recall those nouns after a filled interval of 5 minutes. The Digit Symbol Substitution Test (DSST)[6] is a common test of executive attention. in which participants are asked to translate numbers to symbols using a key. The score is the total number of numbers correctly translated to symbols within 90-seconds. Adult performance is influenced by motor persistence, sustained attention, response speed, and visuo-motor coordination[7]. The Word Fluency Test (WFT)[8] is a test of verbal fluency that consists of three 1-minute word-naming trials. Participants are asked to name as many words as possible (excluding) that begin with the letters "F", "A" and "S". The word fluency score is the total number of words generated across the three trials.

The primary outcome for this analysis will be a global cognitive score created from all three neurocognitive tests. Tests will be standardized to z-scores (observed test score – mean test score/ standard deviation of test score at Visit 2), then averaged and standardized (global score – global mean / standard deviation of global score at Visit 2).

The DWRT, DSST, and WFT were offered to all ARIC participants at three time points: Visit 2 (1990-92), Visit 4 (1996-98), and Visit 5 (2011-present). These tests were offered to a subset of participants at two additional timepoints: Visit 3 (1993-95) and in 2004-06 as part of the Brain MRI substudy. The primary analysis will be limited to V2, V4 and V5; data from V3 and the Brain MRI study will be used in sensitivity analysis.

Exposure:
Education was assessed as years of schooling at ARIC visit 1 (1987-89). The primary analysis will examine education as a categorical variable: less than a high school degree, high school degree, GED or vocational school, or greater than a high school degree.

Statistical analysis:
We will stratify all analyses by race. The primary analysis will adjust for age (years) and sex. Covariates such as smoking, hypertension, and other comorbid conditions have not significantly added to previous models of education and cognitive change in ARIC NCS (Gottesman et al.); we will confirm those findings in this analysis.

We will juxtapose the results of various SPM formulations with those from both a standard GLMM (MAR) model, as well as those from an inverse probability of weighting for attrition (IPAW) model. Compared to IPAW, we expect lower residual variance with the SPM, permitting more sensitive detection of exposure effects on the rate of cognitive change. Although perhaps not part of the present paper, we realize that a comparison between alternative models to handle informative missingness (IPAW and SPM) may be necessary for any future exposures studied for which the attrition bias may have different effects.

References

1. Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* 2nd ed. Hoboken, N.J.: Wiley; 2002.
2. Brandt J, Spencer M, Folstein MF. The telephone interview for cognitive status. *Neuropsychiatry, neuropsychology, and behavioral neurology*. 1988;1:111-117.
3. Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia. *Psychol Med*. 1991;21:785-790.
4. Arnold AM, Newman AB, Dermond N, Haan M, Fitzpatrick A. Using telephone and informant assessments to estimate missing Modified Mini-Mental State Exam scores and rates of cognitive decline. The cardiovascular health study. *Neuroepidemiology*. 2009;33:55-65.
5. Knopman DS, Ryberg S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol*. 1989;46:141-145.
6. Wechsler D. *Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation; 1981.
7. Schear JM, Sato SD. Effects of visual acuity and visual motor speed and dexterity on cognitive test performance. *Arch Clin Neuropsychol*. 1989;4:25-32.
8. Benton AL, Hamsher K, Sivan AB. *Multilingual Aphasia Examination.* 3rd ed. Iowa City: AJA; 1994.

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? N/A
____ 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"? N/A
____ 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)?

MP1967. Adjusting for Measurement Error in Baseline Measures of Cognitive Function: The ARIC Neurocognitive Study. Wruck L et al.

MP1982. Estimation of cognitive change from repeat measures in observational studies; associations with education: the ARIC NCS. Gottesman R. et al.

MP2033. Cognitive domains in elderly ARIC blacks and whites. Rawlings et al.

MP2041. The effect of selection bias on the relationship between cardiovascular risk factors and mortality. Banack et al. (Doctoral student at McGill University. Lisa Wruck is senior author.)

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 N/A

___ 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.