ARIC Manuscript Proposal #2686

PC Reviewed: 1/12/16	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: A Novel Method for Power Analysis and Sample Size Determination in Metabolic Phenotyping

b. Abbreviated Title (Length 26 characters): **Power analysis methods**

2. Writing Group:
Benjamin Blaise, Gonçalo Correia, Adrienne Tin, J. Hunter Young, Jeremy Nicholson, Elaine Holmes, Timothy Ebbels

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __BJB__ __GC__

First author:

Name: Benjamin Blaise
Email: b.blaise@imperial.ac.uk
Address: Sir Alexander Fleming Building, South Kensington Campus

Name: Gonçalo Correia
Email: gd2212@imperial.ac.uk
Address: Room 365, Sir Alexander Fleming Building, South Kensington Campus

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: **J. Hunter Young, MD**
Address: Johns Hopkins Bloomberg School of Public Health
Department of Medicine, The Johns Hopkins University
The Welch Center for Epidemiology and Clinical Research
2024 E. Monument Street
Baltimore, MD 21205
Phone: 410-502-5808
E-mail: jhyoung@jhmi.edu

3. Timeline:

All the work on method development has already been performed. Manuscript writing is also close to conclusion. The method will be applied to the ^{1}H NMR data acquired at Imperial College London from the ARIC study, to showcase application of the

methodology to urine metabolic phenotypic data from a human cohort and exemplify how it could be used to perform power analysis for large cohort studies.

4. Rationale:

Power analysis is very important in study design and interpretation, particularly when study costs and exposure of human subjects to previously uncharacterized and potentially harmful or less optimal procedures are an issue. Traditional methodologies for power analysis are usually associated with univariate hypothesis testing and do not fully account for major characteristics of *omics* data, including metabolic phenotyping, mainly the multivariate nature and the dense between variable correlation networks present in the measurements. This complicates sample size determination and power analysis for metabolic phenotypic analysis.

5. Main Hypothesis/Study Questions:

The main aim of this work is to develop a flexible (compatible with multiple statistical analysis methods and analytical chemistry platforms) computational framework capable of providing meaningful estimates of power and guide metabolic phenotyping study design.

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).

This is a computational method development study. However, it uses data from multiple human and metabolic phenotypic studies to show its application.

<u>The ARIC study data included:</u>
Spot urine samples taken at visit 4 from 1861 individuals without prevalent diabetes or end-stage renal disease

<u>Brief method description:</u>
The methodology can be decomposed in 4 aspects:

1) Obtain statistical models of preexisting data (pilot study data for e.g.), incorporating information from each variable marginal distribution and their joint distributions, from which we can obtain new samples with characteristics matching the original dataset.
2) Introduction of a simulated effect size with a varying magnitude.
3) Repeated analysis of the generated datasets, with the desired statistical method, to estimate distributions of true and false positives and negatives.
4) Visualization and analysis of these ratios and quantities derived from them (for e.g., power, false discovery rate) on a variable by variable basis or globally to infer overall power of the study.

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?
__ 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 __X__ 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)?

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

11.b. If yes, is the proposal
_X_ A. primarily the result of an ancillary study (2009.02)
__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2015.12**)**

*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.

References