ARIC Manuscript Proposal #1971

1.a. Full Title: Time to event methods where the event is crossing a threshold level for a biomarker

b. Abbreviated Title (Length 26 characters): Time until cross threshold

2. Writing Group:
   Writing group members: Noorie Hyun, Donglin Zeng, David Couper, Jim Pankow, others welcome

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

First author: Noorie Hyun
   Address: Department of Biostatistics, CB#7420
   University of North Carolina at Chapel Hill
   Chapel Hill, NC 27599-7420
   Phone: 919-260-5717
   Fax: NA
   E-mail: nrhyun@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: Department of Biostatistics
   University of North Carolina at Chapel Hill
   137 E. Franklin St., Suite 203
   Chapel Hill, NC 27599-7420
   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. Expect to be able to submit to a journal within 1 year.
4. **Rationale:**

Disease can be diagnosed in various ways, such as medical symptoms, signs, and/or biomarkers. The use of biomarkers has revolutionized disease diagnosis because screening or diagnostic tests using biomarkers are often cheaper and easier to measure than true endpoints.

For some diseases or conditions, it is the presence or absence of a particular biomarker that is used in the diagnosis. For instance, HIV infection is diagnosed on the basis of the presence or absence of antibodies to HIV. For others, the biomarker is always present and it is the level of the biomarker that is used to make the diagnosis. For such biomarkers, there is often an accepted threshold used as a cut-off, with values exceeding the threshold being used to diagnose (or even to define) the condition. For example, diabetes is often defined in terms of a fasting plasma glucose threshold of 7:0 mmol/l (126 mg/dl).

There are several limitations to the use of such thresholds that are usually ignored in practice. The threshold is generally regarded as a fixed constant that is appropriate for everyone. However, two people with the same level of a biomarker may differ in terms of their other symptoms of the disease of interest. Although a threshold is typically evidence-based and set by a panel of experts, the choice of the precise value is at least somewhat arbitrary. Also, every assay has some inherent variability, so if the assay is run twice on a sample from an individual the results may not be identical. Further, even though there may be a smooth underlying trend in an individual’s value of the biomarker, there is likely to be short-term intra-individual variability, resulting in variation around that underlying trend. Assay variability and within-person effects complicate determination of whether an individual’s biomarker has exceeded the threshold. Some biases may be more likely in a particular direction, such as with “white-coat hypertension”, whereas others are essentially random. In clinical practice, ad hoc approaches that are used to take into account biomarker variability include taking two or more measurements over a period of time.

In longitudinal studies, using a threshold-defined diagnosis raises additional issues. For many investigations of potential associations between an exposure and an incident disease outcome, time-to-event (“survival analysis”) methods are used. For some events, such as a myocardial infarction, the exact date of the event can usually be determined accurately. For threshold-defined events, what is usually known is the date of the last study visit at which an individual’s biomarker value was below the threshold and the first study visit at which it was above the threshold. Assuming for the moment that we know the person’s “true” value (without error) at each of these times, the underlying smooth trend function must have crossed the threshold at some (unknown) time in the interval. Analysts sometimes use as event time the visit at which a value above the threshold is first recorded. This approach can lead to invalid inferences (Lindsey and Ryan, 1998). In the absence of any information other than the time of the last event-free visit and the first visit at which the condition is noted, interval censoring methods are appropriate. When the condition is defined in terms of a biomarker exceeding a threshold, more information is available – the value of the biomarker at each visit – and this information could be used in analyses.

Reference:

5. Main Hypothesis/Study Questions:

We propose a method for investigating associations between exposures and current status data that relaxes all the assumptions mentioned above. We assume that the outcome is defined in terms of a biomarker but that the biomarker may be measured with error or there may be short-term variability within each individual and that the appropriate threshold for defining the condition of interest may vary from person to person. The proposed method involves a semi-parametric likelihood-based approach based on the Cox model.

We will develop the model and associated inference procedures (estimation of parameters and their variances). We will investigate theoretical properties of the model and use simulation to test its performance under known conditions. We will then use real data from the ARIC Study to illustrate the method.

The new statistical model is being developed by Ms Hyun as part of her Biostatistics PhD dissertation.

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).

Analyses will use either all ARIC participants providing appropriate consent or the case-cohort subset in the Inflammatory Precursors of Diabetes ancillary study (AS# 1995.09). The outcome is time to incident diabetes, with time of onset being determined using the new methodology being developed here. Predictors may include traditional predictors of diabetes, such as BMI and/or the novel markers investigated in AS# 1995.09. The analyses will be used as examples of the use of the new methodology rather than to discover or replicate new associations.

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 ICTDER03 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  ___ 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  
  ___ Yes  ______ No  

Since this is a methodology paper and includes no novel risk factors, it does not  
overlap with ongoing research.

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)?  

Related manuscripts are those using time of onset of diabetes defined using interpolation,  
as calculated by the Coordinating Center.  The first such manuscript was:

MS#853: Duncan BB, Ballantyne C, Chambless LE, Couper D, Folsom AR, Heiss G,  
Hoogeveen R, Pankow J, Schmidt MI. Low-grade systemic inflammation and the  

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

11.b. If yes, is the proposal  
  ___ A. primarily the result of an ancillary study (list number* _________)  
  ___ X ___ B. primarily based on ARIC data with ancillary data playing a minor  
  role (usually control variables; list number(s)* _1995.09_ _________  
  _________)  

*ancillary studies are listed by number at  
http://www.csc.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.