1.a. **Full Title**: Data mining for risk factors of age-related maculopathy

**b. Abbreviated Title (Length 26 characters)**: Age-related Maculopathy

2. **Writing Group (list individual with lead responsibility first):**

   - **Lead**: Hong C Lai  
     - **Address**: Department of Ophthalmology  
       - Johns Hopkins Medical Institutions  
       - 600 N Wolfe Street, Wilmer B29  
       - Baltimore, MD 21205  
       - **Phone**: 410-955-2216  
       - **Fax**: 410-955-2217  
       - **E-mail**: hlai1@jhmi.edu

   - **Writing group members**: Tien Wong, David Couper, John Kempen, Joe Coresh

3. **Timeline**: Manuscript preparation is expected to be completed in two years.

4. **Rationale:**

According to the National Eye Institute (www.nei.nih.gov/news/pressreleases/032002.htm), more Americans than ever are facing the threat of blindness from age-related eye diseases. ARM is one of the most common causes of blindness and vision impairment in the United States, with more than 1.6 million Americans older than age 60 being affected. These numbers are expected to double over the next 30 years as the baby boomer generation ages. Estimates for the prevalence of AMD range from 5.8% in a national population-based survey among American adults older than 45 years to 8.8% among residents of Framingham, Massachusetts, over the age of 52 years. Among residents of Framingham between 75 and 85 years old, estimates increase to 28%. These rates have been estimated to be as high as 41.5% in England and 14.3% in New Zealand, thus demonstrating that AMD is one of the most commonly occurring conditions among older adults in countries with large Caucasian populations. The prevalence of ARM is thought to be increasing in the United States and elsewhere, in part because of the increasing longevity of the population.

ARM is the leading irreversible cause of blindness in the elderly. Despite extensive research, much remains to be elucidated regarding its etiology. Identification of the risk factors is an important step in developing prevention and treatment strategies for ARM. Modification of relevant risk factors for ARM may prove to be much more effective, from the perspective of public health, than treatment of established disease already causing progressive visual...
degeneration and impaired quality of life. Also, identification of risk factors may suggest primary and secondary prevention strategies.

Although age has been the most consistent risk factor for ARM, it has been demonstrated that age is just one of those factors, which influence the occurrence of ARM. Some other factors including gender, hypertension, alcohol consumption, and cigarette smoking have been shown to associate with ARM. These observations are supported by a theory that darker pigmentation may be protective against the development of aging changes in the retina. Several studies have suggested that ARM is associated with other cardiovascular diseases, plasma fibrinogen levels, nutritional and antioxidant status, and a family history of ARM.

Epidemiological studies have been inconsistent, and occasionally contradictory, in their identification of risk factors for ARM. This inconsistency probably has arisen from differences in study designs, in the populations studied, and in the methods of data analysis, and we now know of only one clear-cut modifiable risk factor, namely, cigarette smoking (and possibly hypertension). Further information is needed to corroborate or refute the associations observed in the early studies. Also, the strength and consistency of the association between age and ARM make it a critical factor in the study of other risk factors. Furthermore, the interactions between age and other risk factors have not been fully investigated.

For example, when the relationship between vitamin E and myocardial infarction (MI) was investigated, the interaction between age (V1AGEZ1) and vitamin E (VITE) could not be ignored: the relationship seemed to be unimportant (the slope was flat) when age was 50, however, when age was 60, a higher Vitamin E level was associated with a lower risk of MI based on a multiple additive regression trees analysis.
Dr. Klein R and his colleagues have published two papers on age-related maculopathy with the use of the ARIC data. However, these papers were focused on prevalence of ARM and the relationship between ARM and cognitive function.

5. Main Hypothesis/Study Questions:

We propose to employ newly developed data mining techniques, including multivariate adaptive regression splines (MARS), classification and regression trees (CART), multiple additive regression trees (MART), and patient rule induction method (PRIM) to perform a secondary analysis on the data derived from the ARIC study, evaluating characteristics of approximately 600 ARM cases and 12,000 control individuals free from ARM at the time of retinal evaluation.

Since epidemiological studies have been inconsistent, and occasionally contradictory, in their identification of risk factors for ARM, data mining approaches will be adopted for this proposed study. Data mining is to search a large number of factors from the ARIC study, seeking statistical associations, interactions, and hidden patterns and structures. Data mining approaches will be used to perform exploratory data analysis to find novel relationships, and statistical inference to see if the relationships are due to chance alone. Thus, we do not propose any hypotheses for this study. We would like to perform knowledge discovery from the ARIC data and compare the data mining approaches with more conventional ones.

This analysis is not meant to exclude other more focused hypothesis driven manuscripts on the same outcome which may yield different inferences.

6. Data (variables, time window, source, inclusions/exclusions):

The following standard ARM variables are needed:

id: id number
pigabnor: any pigmentary abnormality
anypig: any pigmentary abnormality 2
softdru: soft drusen soft drusen
armearly: early ARM summary variable for early ARM
anyarm: any ARM (early and late)
armlate: late ARM summary variable for late ARM
rpedepig: RPE depigmentation
hyperpig: increased pigmentation

The exclusion criteria will be, as usual (ie):

<table>
<thead>
<tr>
<th></th>
<th>V3</th>
<th>V2</th>
<th>V3</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not in V3</td>
<td>2905</td>
<td>18.3954</td>
<td>2905</td>
<td>12887</td>
</tr>
<tr>
<td>Not in V2</td>
<td>211</td>
<td>1.3361</td>
<td>3116</td>
<td>12676</td>
</tr>
<tr>
<td>Race not B or W</td>
<td>38</td>
<td>0.2406</td>
<td>3154</td>
<td>12638</td>
</tr>
<tr>
<td>Race B in M or W</td>
<td>42</td>
<td>0.2660</td>
<td>3196</td>
<td>12596</td>
</tr>
<tr>
<td>No retinal photo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retinal photo not gradable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 ICTDER02 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  x No
(This file ICTDER02 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 ICTDER02 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

___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. 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.

References

The committee appreciates the value of exploring a new methodology. However, the current write up of the hypotheses contradicts previous publications which show distinct differences between age related maculopathy (ARM) and atherosclerosis risk factors. Therefore, we decided to defer the proposal pending submission of a revised proposal. We foresee two approaches to making the hypotheses better. One approach is to specify a list of specific risk factors which you think may predict ARM directly or through interactions which have not been explored in previous papers. The other approach is to say that this is a data mining exercise, a true fishing expedition, without prior hypotheses, and that its value is in discovering signals which will need replication and comparison of the new methodology to more traditional approaches. We anticipate that you will be able to make these revisions relatively quickly. To avoid delays, this deferral will be treated as a conditional approval where the condition is submission of a revised proposal to the satisfaction of two members of the committee.