ARIC Manuscript Proposal #1997

PC Reviewed: 9/11/12   Status: A   Priority: 2
SC Reviewed: _________   Status: _____   Priority: _____

1.a. Full Title:
Incidence of atrial fibrillation using the Centers for Medicare and Medicaid Services data

b. Abbreviated Title (Length 26 characters):
Atrial fibrillation incidence

2. Writing Group:
Writing group members: Alvaro Alonso, Aaron Folsom, Pamela Lutsey, Lin Chen, Sue Duval, Anna Kucharska-Newton, Lisa Wruck, Laura Loehr, Wayne Rosamond, Carla Sueta, Sally Stearns, Jessica Yeh and Ervin Fox

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

**First author: Lindsay Smith**
Address: West Bank Office Building
1300 South Second Street, Suite 300
Minneapolis, MN 55454

Phone: 612-624-5238   Fax: 612-624-0315
E-mail: smit4430@umn.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: Alvaro Alonso
Address: West Bank Office Building
1300 South Second Street, Suite 300
Minneapolis, MN 55454

Phone: 612-626-8597   Fax: 612-624-0315
E-mail: alonso@umn.edu
3. **Timeline:** This proposal is part of a dissertation thesis and work will begin in the fall of 2012 and should be completed by the fall of 2013.

4. **Rationale:**

   Increasingly, administrative data are being used for research purposes, including epidemiologic studies to identify patients with cardiovascular diseases (CVDs), including atrial fibrillation (AF) patients. The ability to efficiently and inexpensively access information on a large number of people makes administrative data an appealing data source for epidemiologic research. However, the usefulness of this approach varies by numerous factors, including the algorithm definition and the population studied. Generally, algorithms with high performance measures have been identified for major CVDs. A recent systematic review of algorithms used to identify atrial fibrillation (AF) patients in administrative data reported a median positive predictive value (PPV) of 89% (range: 70% - 96%) and a median sensitivity of 79% (range: 57% - 95%).

   Despite performance measures that indicate that administrative data could be a promising source for identifying AF patients, significant limitations exist in the data currently available as well as subsequent knowledge gaps about the appropriateness of this approach. Of the 16 individual studies included in the systematic review, only one examined the ability of administrative data to identify incident AF. In this single study, a sample of 125 hospital discharge summaries with a first ICD-9 code for AF and ECGs performed during that hospitalization were reviewed by a study physician to determine the validity of using hospital discharge codes; the PPV for AF was 89% and for incident AF was 62%. Additionally, no study has compared the incidence or prevalence rate of AF using only inpatient or only outpatient claims compared to using both in and outpatient claims; inpatient data were used in 10 studies, two used outpatient data and four used both but did not consider rates separately. An important limitation of some cohort studies, including ARIC and the Cardiovascular Health Study (CHS) cohorts, is reliance exclusively on inpatient claims to identify AF, which could result in under ascertainment of AF. Furthermore, the majority of studies have been performed in predominantly white populations. The validity of utilizing administrative data may vary by race/ethnicity as one study reported lower sensitivity for African Americans compared to whites using non-administrative data.

5. **Main Hypothesis/Study Questions:**

   We hypothesize that the incidence rate of AF will be higher based on CMS data (inpatient and outpatient claims data) compared to ARIC data (hospital discharge codes and study ECGs). Additionally, we hypothesize that the overall correspondence between ARIC and CMS will be high but differ by race, with lower concordance in African Americans. Finally, we hypothesize that age-specific prevalence will be similar in the two data sources.

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).

ARIC cohort participants, 65 years of age and older, who enrolled in fee-for-service Medicare will be eligible for inclusion. Participants must be enrolled in fee-for-service Medicare, both Parts A and B, because Medicare Advantage (HMO) plans are not required to submit claims for beneficiaries and those enrolled in only Part A are known to have incomplete claims data. Additionally, because we are interested in the incidence of AF, participants must be enrolled continuously in fee-for-service Medicare for at least two years without an AF diagnosis. AF following cardiac operative procedures is rather common and, therefore, patients who had heart revascularization (ICD-9 code 36.X) or other cardiac surgery involving heart valves or septa (ICD-9 code 35.X) during the index hospitalization will be excluded. CMS data on inpatient and outpatient services have been available for research since January 1, 1991; therefore, the earliest date of incident AF will be January 1, 1993. To be consistent with CMS data, on the basis of ARIC data, participants without ECG examination at baseline or with diagnosed AF or atrial flutter before January 1, 1993, as well as those without at least two years of continuous enrollment in fee-for-service Medicare prior to AF diagnosis will be excluded from analyses. For analyses involving age-specific prevalent AF, participants must be enrolled in fee-for-service Medicare for the entire calendar year for which AF prevalence will be calculated. Depending on the sample size shorter enrollment windows may be used.

Definition of atrial fibrillation

Using ARIC data, incident cases of AF will be ascertained through hospital discharge codes (ICD-9 code 427.3, 427.31 or 427.32; ICD-10 code I48), ECGs performed during three follow-up visits and death certificates (ICD-9 code 427.3, 427.31 or 427.32; ICD-10 code I48). The incidence date of AF will be defined as the date of first hospital discharge with an AF or atrial flutter (AFL) diagnosis, first ECG showing AF or death by AF (underlying cause of death), whichever occurs earlier.

Using CMS data, incident cases of AF will be ascertained from inpatient and outpatient claims data. The MedPAR file will be used to identify inpatient events. Outpatient events will be identified on the basis of ICD-9 and ICD-10 codes found in the Carrier file and Outpatient files. The primary definition of incident AF will be a diagnosis of AF on a single inpatient claim or two outpatient or carrier claims within 365 days (in any position on the claim: ICD-9 code 427.3, 427.31 or 427.32; ICD-10 code I48). The incidence date of AF will be defined as the discharge date for an inpatient claim or the date of the second qualifying outpatient or carrier claim, whichever occurs earlier. A minimum of two outpatient or carrier claims will be required to meet the definition of AF to minimize rule-out diagnoses and to improve the specificity of the algorithm. Secondary definitions of AF will include restriction to inpatient claims criteria and outpatient or carrier claims criteria.

Statistical analysis

Baseline characteristics (ARIC visit 1) of study participants, stratified by source of AF diagnosis (ARIC and/or CMS), will be presented. Age-, sex- and race-specific incidence rates of AF based on ARIC ascertainment of AF and CMS ascertainment of AF
will be calculated with Poisson regression; race-specific age- and sex-adjusted incidence rates also will be calculated with Poisson regression. Discrepancies between ARIC and CMS are expected in both directions for various reasons and therefore neither will be considered the gold standard for AF diagnosis. The concordance of incident AF events between ARIC and CMS will be assessed with a kappa (κ) statistic. The classification system proposed by Landis and Koch will be used to quantify the level of agreement.25 Due to inherent limitations of kappa, percent agreement, overall as well as positive and negative agreement, will be calculated to provide a more complete picture of concordance.26-28 Due to potentially less reliable CMS data prior to 1999 concordance of incident AF events between ARIC and CMS will be stratified by calendar year into two groups; concordance will be calculated for 1993 – 1999 and from 2000 – 2009. A descriptive analysis, restricted to participants with diagnosed AF in both ARIC and CMS, will be performed to determine the mean difference in incidence date. Subsequently, log-binomial regression will be used to determine predictors of earlier diagnosis from each data source. Additionally, log-binomial regression, restricted to participants with diagnosed AF in at least one data source, will be used to identify demographic and clinical factors associated with concordance. Age-, sex- and race-specific prevalence of AF by data source of diagnosis will be calculated with Poisson regression. Analyses will be repeated using the secondary definitions of AF based on CMS data.

**Limitations**

Claims data for ARIC participants who enroll in Medicare Advantage (HMO) are not uniformly available for research and thus participants who enroll in Medicare Advantage will be excluded from the analysis; the percent of ARIC participants enrolled in Medicare Advantage in 2005 was 42, 9, 31 and 3 in Forsyth County, the city of Jackson, MS, the suburbs of Minneapolis and Washington, County, respectively. Although exclusion of these participants limits the generalizability of the study findings, those findings will be applicable to the fee-for-service CMS population. Additionally, the cohort data can be used to assess differences in fee-for-service and Medicare Advantage enrollees. In ARIC, AF ascertainment relies primarily on hospital discharge codes and is not adjudicated. However, this method has been found to have acceptable validity.8 Finally, neither data source is considered the gold standard and subsequently high concordance between the two data sources will imply, but not prove, validity of these approaches to identify incident AF.

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?  ____ 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  
_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  
____ 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  
____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)?  

MS # 1528: Concordance of Heart Failure Codes Between ARIC and CMS 

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 (list number* 2008.12)  
___  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.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://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.


