ARIC Manuscript Proposal # 3070

PC Reviewed: 11/14/17 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Use of Electronic Health Records in Community Surveillance

b. Abbreviated Title (Length 26 characters):

2. Writing Group: Anna Kucharska-Newton, Brittany Bogle, Carlton Moore, Stephanie Haas, Eric Whitsel, Wayne Rosamond, Matthew Loop, Elsayed Soliman, Aaron Folsom, Gerardo Heiss.

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

First author: **Anna Kucharska-Newton**
Address: University of North Carolina at Chapel Hill,
137 E. Franklin St, Suite 306
Chapel Hill, NC 27514

Phone: 919 966-4564 Fax: 919 966 9800
E-mail: anna_newton@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: **Gerardo Heiss**
Address: University of North Carolina at Chapel Hill,
137 E. Franklin St, Suite 306
Chapel Hill, NC 27514

Phone: 919 962 3253 Fax: 919 966 9800
E-mail: gerardo_heiss@unc.edu

3. Timeline:
Preliminary study data will be prepared for presentation at AHA Epi/Lifestyle meeting, March 2018. Initial manuscript preparation is aimed for summer 2018.

4. Rationale:

Electronic health records hold the promise of "big data" that can help translate epidemiological findings into clinical practice[1]. They further provide an attractive opportunity for conducting health research at a fraction of the current costs[2]. Increasingly greater expenditures associated with patient recruitment, data collection, as well as data storage and processing have contributed to recent considerable reductions in the funding for existing observational cohort studies which for many years have been the mainstay of epidemiological research. Studies such as the Atherosclerosis Risk in Communities (ARIC) study[3,4] which provide valuable ongoing information concerning population-level trends in disease burden by conducting surveillance of acute cardiovascular disease events, can be enhanced considerably with the use of electronic health records as the source of information on disease occurrence.

5. Main Hypothesis/Study Questions:

Our goal is to develop and evaluate EHR-based CHD and HF case finding algorithms, equivalent to those attained in concurrent abstractor-based ARIC Community Surveillance. To achieve this goal, we propose the following study aims:

Aim 1: Validate the optimal combination of EHR-based and observer-based data collection to achieve prespecified levels of data completeness and agreement with CHD and HF event classification vs. ARIC community surveillance of these health events.

Aim 2: Evaluate the gains in cost/time efficiency and scale of community-based coverage in the surveillance of CHD and HF events, by integrating EHR-based case finding and classification supported by partial observer-based data collection.

Aim 3: Contrast the attack rates of CHD and HF estimated by ARIC Surveillance to those estimated from EHR-based, observer supported case finding and classification for one or more ARIC study sites.

Aim 4: Examine the challenges and opportunities in the portability of the procedures developed in the ARIC pilot sites to other EHR systems, through interoperable replication and validation studies.

Aim 5: Examine the potential for EHR-based community surveillance of conditions such as atrial fibrillation, stroke, COPD, and PAD.

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).

Data retrieval:

For this work, which will be conducted as part of the ARIC Study contractual agreement with the National Heart Lung and Blood Institute, we will engage six hospitals in the ARIC study area by requesting numeric and text digital data from medical records that meet established eligibility criteria for investigation. Analyses will be limited to community surveillance events for the year 2015. A structured data request will be submitted to each hospital requesting retrieval of data in standardized formats.

Data processing

Analytical approaches to structured data elements and natural language processing algorithms will be tested on records generated from EHR from one of the ARIC surveillance network hospitals (hospital #11) and applied to the remaining five hospitals.

Analytical approaches and data security
All data received from participating hospitals will be de-identified using commercial de-identification software to remove patient and clinician names. Medical record numbers will be replaced by occurrence IDs using cross-walk files provided by the ARIC Coordinating Center. Data will be received from the hospitals by the ARIC Coordinating Center. Analyses will be conducted using the secure environment of the ARIC Coordinating Center (CC) and the secure data workspace of the Department of Epidemiology at UNC-Chapel Hill.

We will use existing hospital records that have been manually extracted according to existing ARIC abstraction forms, as a "gold" standard for the derivation and validation of EHR-based algorithms for terms contributing to classification algorithms and in the final disease classification. Analyses will focus initially on terms abstracted as part of the Hospital Record Abstraction (HRA) and Heart Failure Abstraction (HFA) forms and the classification of coronary heart disease and heart failure events.

We will estimate the percent agreement, kappa coefficients for the agreement for continuous variables, and the sensitivity, specificity, positive predictive value, and negative predictive value for the EHR-based findings and ARIC abstraction.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ 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? ____ 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 ____ 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.cscc.unc.edu/ARIC/search.php

_____ 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 2014.24

11.b. If yes, is the proposal

__x_ A. primarily the result of an ancillary study (list number* _2014.24_____)
___ 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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes ____ No.

1. Toh S, Platt R. Is Size the Next Big Thing in Epidemiology? *Epidemiology*. 2013;24(3):349-351 310.1097/EDE.1090b1013e31828ac31865e.
2. Khoury MJ, Lam TK, Ioannidis JPA, et al. Transforming Epidemiology for 21st Century Medicine and Public Health. *Cancer Epidemiology Biomarkers & Prevention.* 2013;22(4):508-516.
3. White AD, Folsom AR, Chambless LE, et al. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *Journal of clinical epidemiology.* 1996;49(2):223-233.
4. The ARIC Investigators. The Atherosclerosis Risk in Community (ARIC) Study. Design and Objectives. *Am. J. Epidemiol.* 1989;129(4):687-702.