A Simple and Portable Algorithm for Identifying Atrial Fibrillation in the Electronic Medical Record

Shaan Khurshid, MD[a], John Keaney, MD[b,c], Patrick T. Ellinor, MD, PhD[b,c], and Steven A. Lubitz, MD, MPH[b,c,*]

Atrial fibrillation (AF) is common and increases stroke risk and mortality. Many knowledge gaps remain with respect to practice patterns and outcomes. Electronic medical records (EMRs) may serve as powerful research tools if AF status can be properly ascertained. We sought to develop an algorithm for identifying subjects with and without AF in the EMR and compare it to previous methods. Using a hospital network EMR (n = 5,737,846), we randomly selected 8,200 subjects seen at a large academic medical center in January 2014 to derive and validate 7 AF classification schemas (4 cases and 3 controls) to construct a composite AF algorithm. In an independent sample of 172,138 subjects, we compared this algorithm against published AF classification methods. In total, we performed manual adjudication of AF in 700 subjects. Three AF schemas (AF1, AF2, and AF4) achieved positive predictive value (PPV) >0.9. Two control schemas achieved PPV >0.9 (control 1 and control 3). A combination algorithm AF1, AF2, and AF4 (PPV 88%; 8.2% classified) outperformed published classification methods including >1 outpatient *International Statistical Classification of Diseases, Ninth Revision* code or 1 outpatient code with an electrocardiogram demonstrating AF (PPV 82%; 5.9% classified), ≥1 inpatient *International Statistical Classification of Diseases, Ninth Revision* code or electrocardiogram demonstrating AF (PPV 88%; 6.1% classified), or the intersection of these (PPV 84%; 7.4% classified). When applied simultaneously, the case and control algorithms classified 98.4% of the cohort with zero disagreement. In conclusion, we derived a parsimonious and portable algorithm to identify subjects with and without AF with high sensitivity. If broadly applied, this algorithm can provide optimal power for EMR-based AF research. **(Am J Cardiol 2016;117:221−225)**

Atrial fibrillation (AF) is a prevalent arrhythmia of public health importance owing to its associated mortality, stroke risk, and economic costs.[1−4] Many knowledge gaps remain with respect to AF epidemiology, practice patterns, and resource utilization. To date, research addressing these gaps has largely used cohorts, registries, and claims-related databases. In contrast, electronic medical records (EMRs) are extensive repositories of clinical information that may serve as powerful tools for facilitating AF-related research if AF status can be properly and efficiently ascertained. Although several large registry[5,6] and cohort[7−9] studies have relied on EMRs to study AF, identification of AF was performed either solely or primarily using billing codes, which demonstrate modest and inconsistent accuracy.[7,8,10] Methods to identify AF in the EMR can likely be improved by supplementing billing codes with clinical information already available in most EMRs. We therefore sought to develop a simple and portable algorithm for identifying subjects with and without AF in the EMR. We then compared our algorithm to existing AF ascertainment algorithms previously used in cohort-related and claims-related studies.

Methods

We studied the Partners HealthCare EMR, which is used by 7 hospitals: Massachusetts General Hospital (MGH), Brigham and Women's Hospital, Faulkner Hospital, McLean Hospital, Newton-Wellesley Hospital, North Shore Medical Center, and Spaulding Rehabilitation Center. This repository represents a total of 5,737,846 subjects during the period of 1979 to 2015 (date of last assessment March 31, 2015). To facilitate a manual chart review, we limited the scope to records generated by encounters at MGH for this analysis.

The Research Patient Database Query Tool (RPDR) is a large database built entirely from data contained in the Partners HealthCare EMR. This tool allows researchers to query deidentified data in the EMR to generate data sets of interest. The RPDR matching tool also allows for the generation of age- and gender-matched controls for specific data sets. Once data sets of interest are built, one can request specific types of detailed EMR data on records contained within the data sets. For this analysis, we obtained detailed data regarding (1) medications, (2) cardiology tests, (3) procedure codes (current procedural terminology format), (4) diagnoses (*International Statistical Classification of Diseases, Ninth Revision* [ICD-9] codes), and (5) laboratory

[a]Department of Medicine, [b]Cardiovascular Research Center, and [c]Cardiac Arrhythmia Service, Department of Cardiology, Massachusetts General Hospital, Boston, Massachusetts. Manuscript received August 24, 2015; revised manuscript received and accepted October 24, 2015.

See page 225 for disclosure information.

*Corresponding author: Tel: (617) 643-7339; fax: (617) 726-3852.
E-mail address: slubitz@mgh.harvard.edu (S.A. Lubitz).

0002-9149/15/$ - see front matter © 2016 Elsevier Inc. All rights reserved.
www.ajconline.org
http://dx.doi.org/10.1016/j.amjcard.2015.10.031

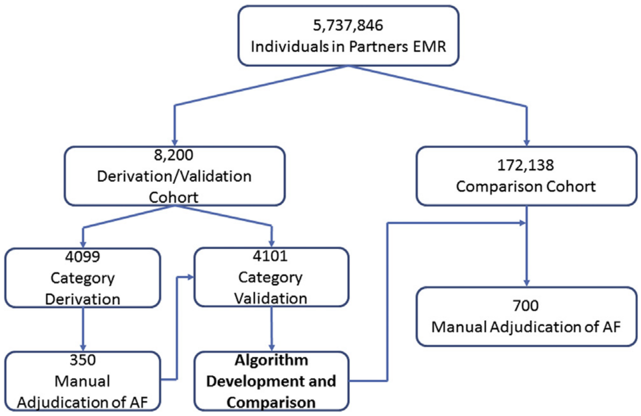

Figure 1. Patient flow through the study. AF classification schema were created a priori and internally validated in a cohort of 8,200 patients. High-performing schema were then combined and iteratively adjusted to develop an AF classification algorithm. This algorithm was compared with previously published methods in an independent sample of 172,138 subjects.

Table 1
Schema and composite algorithm definitions

A. Schemas	Definition
AF1	AF on ECG or AF/flutter ablation or cardioversion
AF2	not AF1 and ≥2 ICD-9 codes for AF
AF3	not AF1 and ≥5 ICD-9 codes for AF
AF4	not AF1 and 0 or 1 ICD-9 code for AF and on antiarrhythmic or anticoagulant or missing PR interval
Control 1	≥1 ECG and no ECG with AF and no ICD-9 code for AF and no AF/flutter ablation and no cardioversion
Control 2	not AF1 and not Control 1 and 1 ICD-9 code for AF and no antiarrhythmic and no anticoagulant
Control 3	not AF1 and not Control 1 and no ICD-9 code for AF and no antiarrhythmic and no anticoagulant
B. Composite Algorithms	**Definition**
Composite AF Algorithm	AF1 or AF2 or AF4
Modified AF Algorithm	AF1 or AF2
Comparator 1	>1 outpatient ICD9 code or 1 outpatient ICD9 code and ECG with AF
Comparator 2	≥1 inpatient ICD9 code or ECG with AF
Comparator 3	comparator 1 or comparator 2
Comparator 4	≥1 inpatient ICD9 code or ≥1 outpatient ICD9 code

Table 2
Schema derivation and validation

A. Derivation

Schema	No. in Sample (%)	No. Reviewed	PPV (95% CI)
AF1	508 (12%)	50	98 (94 - 100)
AF2	1,504 (37%)	50	96 (91 - 100)
AF3	1,316 (32%)	50	20 (09 − 31)
AF4	306 (7%)	50	92 (84 − 100)
Control 1	1,153 (28%)	50	96 (91 − 100)
Control 2	56 (1%)	50	34 (21 − 47)
Control 3	807 (20%)	50	98 (94 − 100)

B. Validation

Schema	No. in Sample (%) [Derivation Cohort]	No. in Sample (%) [Validation Cohort]
AF1	508 (12%)	485 (12%)
AF2	1,504 (37%)	1,541 (38%)
AF4	306 (7%)	283 (7%)
Control 1	1,153 (28%)	1,170 (29%)
Control 3	807 (20%)	798 (19%)
Unclassified	-	53 (1%)

Table 3
Comparison between performance of our derived algorithm and other accepted algorithms

Algorithm	Positive Predictive Value for AF (95% CI)	Proportion of Cohort Classified (%)
Composite AF algorithm (AF1 or AF2 or AF4 ≥1 inpatient ICD-9 code)*	88 (79 − 97)	8.2
Modified AF algorithm (AF1 or AF2)	92 (81 − 98)	7.2
Comparator 1[†] (>1 outpatient ICD9 code or 1 outpatient ICD9 code + ECG with AF)	82 (71 − 93)	5.9
Comparator 2[‡] (≥1 inpatient ICD9 code or ECG with AF)	88 (79 − 97)	6.1
Comparator 3 (comparator 1 or comparator 2)	84 (74 − 94)	7.4
Comparator 4 (≥1 inpatient ICD9 code or ≥1 outpatient ICD9 code)	84 (74 − 94)	8.2

* Missing PR interval criterion removed from AF4 (see text).
† Go et al, 1999[8].
‡ Alonso et al, 2009[7]; Patton et al, 2009[9].

tests. Data within these categories served as source data for classification schema.

To build an initial data set, we obtained a cohort of all patients seen at MGH during January 2014 with ≥1 ICD-9 code for AF or flutter (ICD-9 codes 427.31 and 427.32) at any time in their longitudinal medical record within the EMR (Figure 1). Using the RPDR matching tool, we identified age- and gender-matched individuals without any ICD-9 codes for AF. These cohorts were partitioned into derivation (n = 4,099) and validation (n = 4,101) sets.

A priori, we selected variables that we hypothesized would identify subjects with AF and that would be easily