



Published in final edited form as:

*Am J Cardiol.* 2016 January 15; 117(2): 221–225. doi:10.1016/j.amjcard.2015.10.031.

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

**Shaan Khurshid, MD<sup>a</sup>, John Keane, MD<sup>b,c</sup>, Patrick T. Ellinor, MD, PhD<sup>b,c</sup>, and Steven A. Lubitz, MD, MPH<sup>b,c</sup>**

<sup>a</sup>Department of Medicine, Massachusetts General Hospital, Boston, MA

<sup>b</sup>Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA

<sup>c</sup>Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA

### Abstract

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 (EMR) may serve as powerful research tools if AF status can be properly ascertained. We sought to develop an algorithm for identifying individuals 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 individuals seen at a large academic medical center in January 2014 to derive and validate seven AF classification schemas (4 case and 3 control) in order to construct a composite AF algorithm. In an independent sample of 172,138 individuals, we compared this algorithm against published AF classification methods. In total, we performed manual adjudication of AF in 700 individuals. Three AF schemas (*AF1*, *AF2*, and *AF4*) achieved 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 ICD9 code or 1 outpatient code with ECG demonstrating AF (PPV 82%; 5.9% classified), 1 inpatient ICD9 code or ECG 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 individuals with and without AF with high sensitivity. If broadly applied, this algorithm can provide optimal power for EMR-based AF research.

### Keywords

atrial fibrillation; electronic medical record; billing codes; database research

---

**Corresponding author:** Steven A. Lubitz, MD, MPH, Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, 55 Fruit Street, GRB 109, Boston, MA 02114; P:617-643-7339; F:617-726-3852; slubitz@mgh.harvard.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Disclosures**  
None

Atrial fibrillation (AF) is a prevalent arrhythmia of public health importance owing to its associated mortality, stroke risk, and economic costs.<sup>1–4</sup> Many knowledge gaps remain with respect to AF epidemiology, practice patterns, and resource utilization. To date, research addressing these gaps has largely utilized 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. Whereas several large registry<sup>5,6</sup> and cohort<sup>7–9</sup> 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.<sup>7,8,10</sup> 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 individuals with and without AF in the EMR. We then compared our algorithm to existing AF ascertainment algorithms previously utilized in cohort- and claims-related studies.

## METHODS

We studied the Partners HealthCare EMR, which is utilized 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 individuals over the period of 1979 to 2015 (date of last assessment 03/31/2015). 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 de-identified data in the EMR to generate datasets of interest. The RPDR matching tool also allows for the generation of age- and sex-matched controls for specific datasets. Once datasets of interest are built, one can request specific types of detailed EMR data on records contained within the datasets. For this analysis, we obtained detailed data regarding: 1) medications, 2) cardiology tests, 3) procedure codes (Current Procedural Terminology (CPT) format), 4) diagnoses (International Statistical Classification of Diseases, 9<sup>th</sup> revision (ICD-9) codes), and 5) laboratory tests. Data within these categories served as source data for classification schema.

To build an initial dataset, we obtained a cohort of all patients seen at MGH during January 2014 with 1 ICD-9 code for AF at any time in their longitudinal medical record within the EMR (Figure 1). We considered any individual with atrial flutter or fibrillation as having AF (ICD-9 codes 427.31 and 427.32). Using the RPDR matching tool, we identified age- and sex-matched controls without any ICD-9 codes for AF. These cohorts were partitioned into derivation (n=4099) and validation (n=4101) sets.

*A priori*, we selected variables that we hypothesized would identify individuals with AF and that would be easily ascertained in other datasets. These variables included AF presence on an ECG, number of ECGs with AF, electrical cardioversion, AF or atrial flutter ablation, antiarrhythmic use, warfarin use, novel oral anticoagulant use, missing PR interval on an

ECG, and number of ICD-9 codes for AF. We then compared pairwise agreement between each of these variables by measuring the Cohen's kappa coefficient between each variable. Agreement ranged from poor to modest between variables (-0.02 to 0.50), and therefore we assumed that each would potentially contribute independent information to AF classification. As such, all of the variables were retained in efforts to derive classification schemas.

From these nine preselected variables we derived seven *a priori* candidate classification schemas (4 for cases and 3 for controls). We applied these seven schemas to the derivation set. Each schema attempted to identify the presence or absence of AF using a combination of billing codes and clinical variables. We manually adjudicated AF in a random sample of patients chosen by each schema comprising a total of 350 individuals. Schemas meeting a pre-specified threshold with a positive predictive value (PPV) >0.9 for AF or its absence were retained and internally validated in the independent validation set. For each PPV, we report the 95% confidence interval calculated using an exact binomial method.

We then manually adjudicated AF within an independent random sample of 172,138 individuals treated at MGH in 2014 by combining AF case and control classification schemas into composite algorithms that maximized both the PPV and percent of individuals classified in the dataset. During the process, we iteratively adjusted the schemas to maximize performance, and reevaluated modified algorithms in independent samples after each iteration. In total, we manually adjudicated 700 cases. After identifying two composite algorithms that maximized both PPV and the percent of individuals classified for AF cases and controls, we applied them to the entire sample of 172,138 individuals to assess misclassification and no-classification rates.

After constructing our composite algorithms, we compared them to previously utilized methods of classifying AF. We applied our algorithms as well as two previously published comparator methods and two additional related methods.<sup>7-9</sup> To allow for uniform assessments, all comparator algorithms were modified to include atrial flutter. Comparison methods are outlined in Table 1. The primary metrics for assessing algorithm performance were PPV for the presence of AF (case algorithms) or for the absence of AF (control algorithms), and percentage of the total cohort classified by each algorithm. We then compared the concordance of ascertainment status between our algorithm and the best comparator with McNemar's test.

## RESULTS

Three AF case schemas achieved PPV >0.9 in the derivation set (Tables 1 and 2): *AF1* (AF on ECG, AF/flutter ablation, or cardioversion); *AF2* (not AF1, 2 ICD-9 codes for AF); and *AF4* (not AF1, 0 or 1 ICD-9 code for AF, on antiarrhythmic or anticoagulant or missing PR interval). Two control schemas achieved a PPV >0.9: *control 1* (1 ECG, no ECG with AF, no ICD-9 code for AF, no AF/flutter ablation, no cardioversion) and *control 3* (not AF1, not Control 1, no ICD-9 code for AF, no antiarrhythmic, no anticoagulant). Each of these algorithms performed similarly in the validation set (Table 2).

Schemas that achieved a PPV >0.9 (*AF1*, *AF2*, and *AF4*; *control 1* and *control 3*) were combined into composite algorithms for AF classification. In the course of this process, the missing PR interval criterion was eliminated from *AF4* after being found to be an unreliable indicator of AF on ECG during iterative adjustment, and stratification of ICD9 code by inpatient status was incorporated into each remaining schema.

When compared to previously published algorithms in an independent cohort (n = 172,138; 58% female, 78% white, median age 70 years), our composite case algorithm designated 14,116 individuals as having AF, corresponding to 8.2% capture of the cohort, while maintaining 88% PPV for AF (Table 3, Figure 2). A modified case algorithm comprised of features associated with the highest probabilities of AF (*AF1* and *AF2*) demonstrated a higher predictive value for AF as expected (PPV 92%) with a small reduction in the proportion of the sample classified (7.2%). As expected, the composite algorithm PPVs in the independent cohort were slightly lower than those obtained in the derivation/validation cohorts since in the latter individuals with AF were likely overrepresented by virtue of ascertainment. When compared to our primary algorithm, *comparator 1* and *comparator 3* captured less of the cohort with a lower PPV for AF. *Comparator 2* captured less of the cohort with a similar PPV for AF. *Comparator 4* captured a similar proportion of the cohort with a lower PPV for AF. Our control algorithm captured 90.2% of the cohort with a PPV of 98% for the absence of AF.

When our case and control algorithms were combined and applied to the entire sample (n = 172,138), 2,671 individuals (1.6%) remained unclassified due to indeterminate AF status. There was no overlap between classified cases and classified controls (Table 4). Since our algorithm and *comparator 4* identified similar proportions of individuals with AF in the dataset (8.2%) though with different PPVs (88% and 84%, respectively), we compared agreement between the two algorithms. Relative to *comparator 4*, our composite algorithm classified an additional 899 individuals as having AF, whereas 958 individuals classified as having AF by *comparator 4* were unclassified by our algorithm owing to an indeterminate probability of AF. Among the 169,467 individuals classified using our algorithm, ascertainment status differed significantly between our algorithm and *comparator 4* ( $P < 0.001$ ).

## DISCUSSION

Using a newly created AF ascertainment algorithm we were identified 14,116 individuals with AF and 155,351 without AF in a sample of 172,138 individuals from a large EMR. The AF algorithm achieved a higher PPV (88%) and generally classified a higher proportion of the total cohort (8.2%) than previously published methods. The primary algorithm successfully classified 98.4% of individuals with no disagreement between case and control classification. In sensitivity analyses, a modified algorithm restricted to features with the highest probability of AF achieved an even higher PPV (92%) with slightly lower classification performance (7.2%).

Our findings add to the growing body of evidence suggesting that billing code data alone are suboptimal for classifying disease. Even in the controlled setting of a prospective cohort

study, billing codes have shown low sensitivity for the diagnosis of AF.<sup>7</sup> Furthermore, a recent large EMR-based analysis of AF-related stroke revealed that only 72% of events identified by ICD-9 codes were true cases of stroke attributable to AF.<sup>10</sup> The incorporation of clinical parameters into our algorithm allowed for more discriminative and inclusive AF identification versus other commonly used billing code-based methods. As a result, use of our algorithm can maximize statistical power for large studies with a low misclassification rate.

Previous studies have investigated natural language processing (NLP), or use of algorithms designed to detect the presence of key phrases in free text notes, as a tool for identifying diseases such as AF.<sup>11</sup> NLP has been shown to improve classification performance modestly over clinical data in rheumatoid arthritis.<sup>12</sup> Unfortunately, the ability to search free text may not be available in all EMRs, and the programmatic resources required to apply NLP may be inaccessible to many investigators, thereby limiting the portability of AF classification algorithms based on NLP. Furthermore, NLP algorithm performance is likely to vary based on note structure and format. Whereas NLP may enable the proper classification of some diseases, the performance of our algorithm suggests that accurate and sensitive AF classification can be achieved using billing codes and clinical data alone.

Accurate classification of AF in the EMR will allow for improved characterization of the epidemiology, current practice patterns, and resource utilization associated with AF. As an example, broad application of our algorithm may allow clinical researchers to identify practice variation among different clinical and geographic settings and highlight successful management strategies in AF as well as areas for potential improvement. Specifically, in spite of strong evidence for its benefit,<sup>13</sup> anticoagulation in AF has been shown to be underutilized.<sup>5,14</sup> Analysis of data from the EMR may provide insight into specific predictors of suboptimal AF management and suggest strategies for closing treatment gaps. Furthermore, utilization of the EMR could facilitate targeted interventions that may benefit specific individuals.

The ability to accurately identify large cohorts of patients with and without AF using the EMR may also facilitate the evaluation of outcomes that may be too infrequent to study in prospective cohorts or trials. For example, utilization of the EMR may facilitate the efficient study of individuals with AF at low predicted risk of stroke, or patients with adverse effects of antiarrhythmic drug therapy that may only be evident in large numbers with long-term follow-up. Moreover, the large sample sizes generated using EMRs may be leveraged for genetic discovery, as has been performed successfully within eMERGE.<sup>11</sup> Importantly, our findings demonstrate that minor modifications can increase the predictive value of the algorithm with a small impact on the proportion of the sample classified, to enable investigators to balance the enrichment of AF and sample size required for a given application.

Our study has several potential limitations. Intrinsic to most EMR systems, data utilized in our analysis does not include encounters outside networked institutions. Nevertheless, we observed multiple instances where AF was diagnosed at an outside institution but later captured by new billing code or clinical data, suggesting that the ability to capture repeated

measures enables identification of patients diagnosed elsewhere. Second, although our algorithm performed favorably compared to previous methods, the final PPV for AF was only 90%. We believe that the proportion of AF cases captured by our algorithm will allow for identification of large enough AF cohorts to overcome imperfect accuracy. Third, although our algorithm classified 98% of individuals, it is not entirely exhaustive. Indeed, many individuals classified as cases by *comparator 4* were regarded as indeterminate in our algorithm, whereas additional cases were captured that had been regarded as controls by *comparator 4*. The reclassification and increased specificity observed using our algorithm may enhance studies requiring greater precision. Fourth, although our algorithm was designed to use basic clinical variables to maximize portability, it may be difficult to implement in rudimentary EMRs lacking procedure and medication data. Lastly, although our algorithm performed consistently on 3 separate datasets within the Partners HealthCare EMR, performance may differ outside large referral centers.

## Acknowledgments

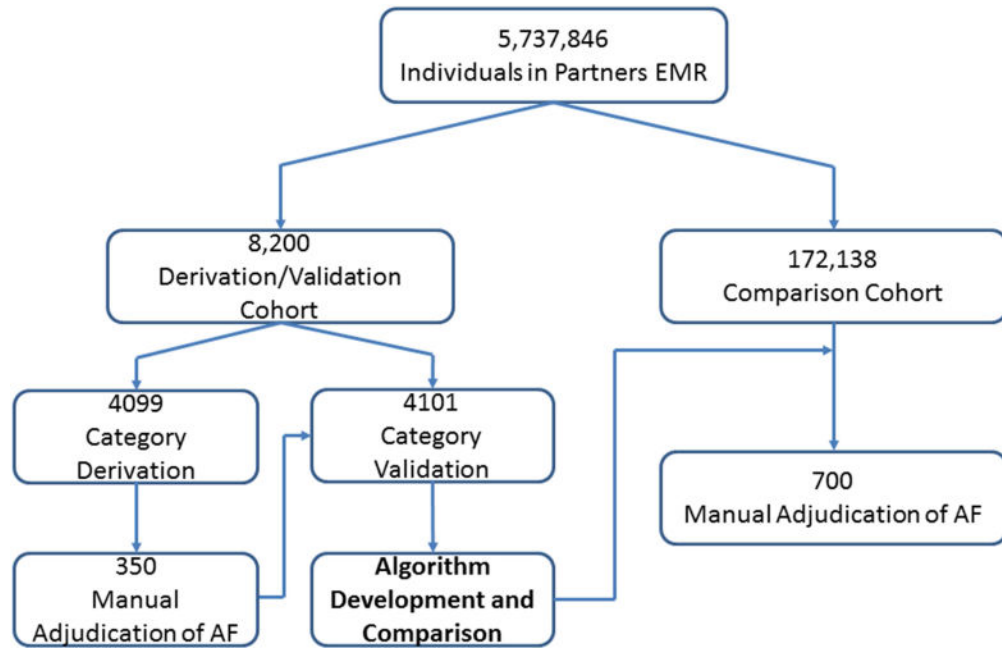
None

**Funding sources:** This work was supported by NIH grants K23HL114724 (Lubitz), 2R01HL092577, R01HL104156, K24HL105780 (Ellinor), a Doris Duke Charitable Foundation Clinical Scientist Development Award 2014105 (Lubitz), an American Heart Association Established Investigator Award 13EIA14220013 (Ellinor), and by support from the Fondation Leducq 14CVD01 (Ellinor).

## References

1. Corley SD, Epstein AE, DiMarco JP, Domanski MJ, Geller N, Greene HL, Josephson RA, Kellen JC, Klein RC, Krahn AD, Mickel M, Mitchell LB, Nelson JD, Rosenberg Y, Schron E, Shemanski L, Waldo AL, Wyse DG. Relationships between sinus rhythm, treatment, and survival in the Atrial Fibrillation Follow-Up Investigation of Rhythm Management (AFFIRM) Study. *Circulation*. 2004; 109:1509–1513. [PubMed: 15007003]
2. Stewart S, Hart CL, Hole DJ, McMurray JJ. A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study. *Am J Med*. 2002; 113:359–364. [PubMed: 12401529]
3. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation: a major contributor to stroke in the elderly. The Framingham Study. *Arch Intern Med*. 1987; 147:1561–1564. [PubMed: 3632164]
4. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, Newton-Cheh C, Lubitz SA, Magnani JW, Ellinor PT, Seshadri S, Wolf PA, Vasani RS, Benjamin EJ, Levy D. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. *Lancet*. 2015; 386:154–162. [PubMed: 25960110]
5. Wilke T, Groth A, Mueller S, Pfannkuche M, Verheyen F, Linder R, Maywald U, Kohlmann T, Feng YS, Breithardt G, Bauersachs R. Oral anticoagulation use by patients with atrial fibrillation in Germany. Adherence to guidelines, causes of anticoagulation under-use and its clinical outcomes, based on claimsdata of 183,448 patients. *Thromb Haemost*. 2012; 107:1053–1065. [PubMed: 22398417]
6. Friberg L, Rosenqvist M, Lip GY. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. *Eur Heart J*. 2012; 33:1500–1510. [PubMed: 22246443]
7. Alonso A, Agarwal SK, Soliman EZ, Ambrose M, Chamberlain AM, Prineas RJ, Folsom AR. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J*. 2009; 158:111–117. [PubMed: 19540400]
8. Go AS, Hylek EM, Borowsky LH, Phillips KA, Selby JV, Singer DE. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the anticoagulation and risk factors in atrial fibrillation (ATRIA) study. *Ann Intern Med*. 1999; 131:927–934. [PubMed: 10610643]

9. Patton KK, Ellinor PT, Heckbert SR, Christenson RH, DeFilippi C, Gottdiener JS, Kronmal RA. N-terminal pro-B-type natriuretic peptide is a major predictor of the development of atrial fibrillation: the Cardiovascular Health Study. *Circulation*. 2009; 120:1768–1774. [PubMed: 19841297]
10. Thigpen JL, Dillon C, Forster KB, Henault L, Quinn EK, Tripodis Y, Berger PB, Hylek EM, Limdi NA. Validity of international classification of disease codes to identify ischemic stroke and intracranial hemorrhage among individuals with associated diagnosis of atrial fibrillation. *Circ Cardiovasc Qual Outcomes*. 2015; 8:8–14. [PubMed: 25587093]
11. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, Basford MA, Brown-Gentry K, Balser JR, Masys DR, Haines JL, Roden DM. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010; 86:560–572. [PubMed: 20362271]
12. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, Szolovits P, Churchill S, Murphy S, Kohane I, Karlson EW, Plenge RM. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2010; 62:1120–1127. [PubMed: 20235204]
13. Stroke Prevention in Atrial Fibrillation Study. Final results. *Circulation*. 1991; 84:527–539. [PubMed: 1860198]
14. Waldo AL, Becker RC, Tapson VF, Colgan KJ. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. *J Am Coll Cardiol*. 2005; 46:1729–1736. [PubMed: 16256877]

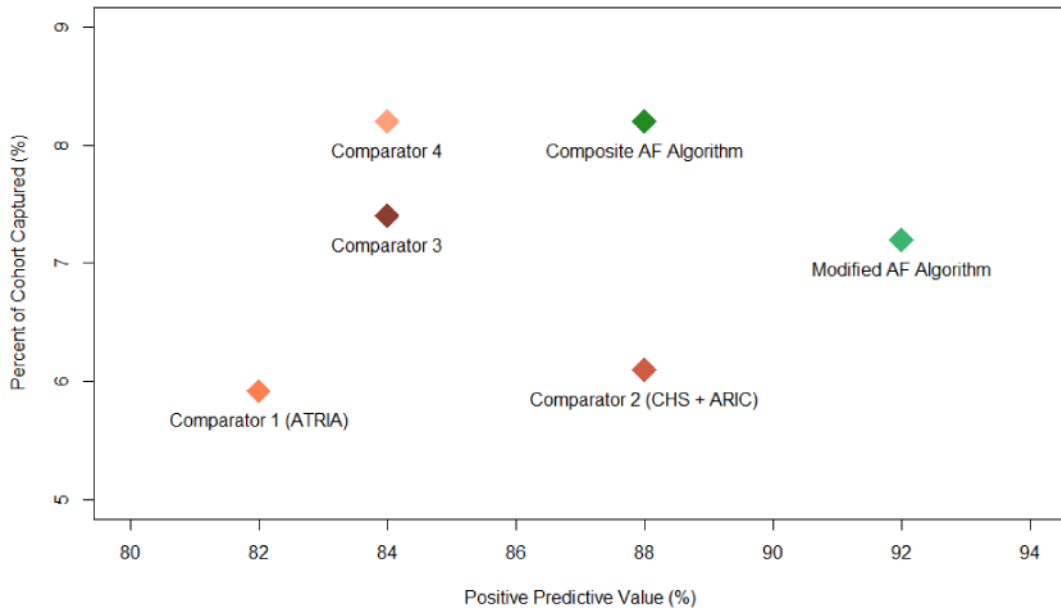


**Figure 1.**

**Patient Flow**

Patient flow through the study. AF classification categories were created *a priori* and internally validated in a cohort of 8,200 patients. High-performing categories were then combined and iteratively adjusted to develop an AF classification algorithm. This algorithm was compared to previously published methods in an independent sample of 172,138 individuals. AF – atrial fibrillation. EMR – electronic medical record.





**Figure 2.**

**Comparative Algorithm Performance**

Comparative algorithm performance on an independent comparison set. Our composite AF algorithms were compared against two previously published AF identification methods and two additional billing-code based methods in terms of both positive predictive value for AF and percent of the cohort captured. Definitions: *comparator 1* (>1 outpatient ICD9 code **or** 1 outpatient ICD9 code + 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), *composite AF Algorithm* (our primary AF algorithm), *modified AF algorithm* (our second AF algorithm). AF – atrial fibrillation. ARIC – Atherosclerosis Risk in Communities. ATRIA – Anticoagulation and Risk factors in Atrial Fibrillation. CHS – Cardiovascular Health Study.

**Table 1**

Schema and composite algorithm definitions.

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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Schema derivation and validation.

<b>A. Derivation</b>			
<b>Schema</b>	<b>No. in Sample (%)</b>	<b>No. Reviewed</b>	<b>PPV (95% CI)</b>
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>B. Validation</b>			
<b>Schema</b>	<b>No. in Sample (%) [Derivation Cohort]</b>	<b>No. in Sample (%) [Validation Cohort]</b>	
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%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**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)	88 (79 – 97)	8.2
Modified AF algorithm (AF1 or AF2)	92 (81 – 98)	7.2
Comparator 1 <sup>†</sup> (>1 outpatient ICD9 code or 1 outpatient ICD9 code + ECG with AF)	82 (71 – 93)	5.9
Comparator 2 <sup>‡</sup> ( 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

<sup>†</sup>Go et al., 1999<sup>8</sup>;

<sup>‡</sup>Alonso et al., 2009<sup>7</sup>; Patton et al. 2009<sup>9</sup>

**Table 4**

Case and control classification in 172,138 randomly sampled individuals.

Predicted case	Predicted control	
	Yes	No
Yes	0 (0)	14,116 (8.2)
No	155,351 (90.2)	2,671 (1.6)

Expressed as No. (%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript