Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Combining Inpatient and Outpatient Data for Diagnosis of Non-Valvular Atrial Fibrillation Using Electronic Health Records: A Validation Study

This article was published in the following Dove Press journal:
Clinical Epidemiology

Orna Reges[1,2,*]
Hagay Weinberg[3,4,*]
Moshe Hoshen[1,5]
Philip Greenland[2]
Hana'a Rayyan-Assi[1]
Meytal Avgil Tsadok[1]
Asaf Bachrach[1]
Ran Balicer[1,6]
Morton Leibowitz[1]
Moti Haim[7,8]

[1]Clalit Research Institute, Clalit Health Services, Tel Aviv, Israel; [2]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; [3]Internal Medicine Department, Meir Medical Center, Kfar-Saba, Israel; [4]Department of Medicine, MidCentral District Health Board, Palmerston-North, New Zealand; [5]National Information Systems, Computational Authority, Ministry of Health, Jerusalem, Isarel; [6]Department of Epidemiology, Ben-Gurion University of the Negev, Beer-Sheva, Israel; [7]Department of Cardiology, Soroka University Medical Center, Beer Sheva, Israel; [8]Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel

*These authors contributed equally to this work

Purpose: Previous studies have demonstrated differences in atrial fibrillation (AF) detection based on data from hospital sources without data from outpatient sources. We investigated the detection of documented diagnoses of non-valvular AF in a large Israeli health-care organization using electronic health record data from multiple sources.

Patients and Methods: This was an open-chart validation study. Three distinct algorithms for identifying AF in electronic health records, differing in the source of their International Classification of Diseases, Ninth Revision code and use of the associated free text, were defined. Algorithm 1 incorporated inpatient data with outpatient data and the associated free text. Algorithm 2 incorporated inpatient and outpatient data regardless of the free text associated with AF diagnosis. Algorithm 3 used only inpatient data source. These algorithms were compared to a gold standard and their sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. To establish the gold standard (documentation of arrhythmia based on electrocardiography interpretation or a cardiologist's written diagnosis), 200 patients at highest risk for having non-valvular AF were randomly selected for open-chart validation by two physicians.

Results: The algorithm that included hospital settings, outpatient settings, and incorporated associated free text in the outpatient records had the optimal balance between all validation measures, with a high level of sensitivity (85.4%), specificity (95.0%), PPV (81.4%), and NPV (96.2%). The alternative algorithm that combined inpatient and outpatient data without free text also performed better than the algorithm that included only hospital data (82.9%, 95.0%, 81.0%, and 95.6%, compared to 70.7%, 96.9%, 85.3%, and 92.8%, sensitivity, specificity, PPV, and NPV, respectively).

Conclusion: In this study, involving a comprehensive data collection from inpatient and outpatient sources, incorporating outpatient data with inpatient data improved the diagnosis of non-valvular AF compared to inpatient data alone.

Keywords: atrial fibrillation, validation, electronic health records

Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia, associated with high rates of stroke, heart failure, and mortality.[1] Many current studies, aiming to identify patients with AF to study the epidemiology of the disease and improve quality of care, rely on diagnostic data from administrative sources or electronic health records (EHR).[2–4] Using EHR for this purpose has many challenges, such as misclassifications, documentation errors, and missing data.[5,6] Extraction and coding

Correspondence: Orna Reges
Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, 680 North Lake Shore Drive, Suite 1400, Chicago, IL 60611, USA
Email orna.reges@gmail.com

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/CLEP.S230677

of unstructured free text associated with diagnosis in the EHR, when available, may also be an asset or an added challenge. To our knowledge, there is no consistency in the way patients with AF are defined using EHR in various health settings.

Electrocardiographic data (ECG) are important for establishing a diagnosis of AF.[7,8] Unfortunately, many EHR databases, based on administrative data, do not have access to ECG tracings, which limits sensitivity for AF diagnosis. Previous studies have also demonstrated differences in AF populations based on accessing EHR data from hospital sources without access to data from outpatient sources.[9–25] In a review published in 2012,[4] sixteen algorithms for identification of AF from EHR were described. While the majority of the algorithms accessed data from hospital sources,[10–13,16–18,21,23,24] only four used both inpatient and outpatient data,[9,19,20,22] and two studies used only outpatient data.[14,15] Validation of different algorithms varied by the data sources, with lower sensitivity for algorithms based on only one source of information and higher positive predictive value (PPV) for algorithms based on hospital records only. Algorithms also varied by International Classification of Diseases, Ninth Revision (ICD-9) AF code. In addition, various studies used different gold standards and different populations with disparate levels of risk. Differences also occurred when using EHR data to define incidence as opposed to prevalence.

The present study aimed to investigate the detection of documented diagnoses of AF in a health system that had access to EHR data from inpatient as well as outpatient sources and from various administrative databases containing diagnostic information.

Methods

Study Design

This was an open-chart validation study of three distinct algorithms based on EHR data, each of which differed in the elements establishing the diagnosis, namely, data source and free text written by the physician while documenting the diagnosis. This study has been approved by Meir medical center review board and Clalit health services outpatients review board. No patient consent was required by the review boards as it is a retrospective study. Our research complies with the Declaration of Helsinki and the ethic of patient data confidentiality.

Data Source

In Israel, all citizens are registered for basic health care from any of the four payer-provider health funds. Payment amount for the health insurance is determined according to income, and every patient has the right to change provider with no explanation or any conditions. Clalit Health Services (Clalit), the largest integrated payer-provider health-care organization in Israel, provides total medical care, primary care, specialty care, and hospitalization coverage, to approximately 52% (over 4.2 million members) of the population. Membership turnover within Clalit is around 1% annually, facilitating long follow-up with the ability to identify many incident cases. Clalit has an extensive EHR data warehouse that contains data from multiple administrative and clinical sources, including inpatient and outpatient diagnoses (based on ICD-9 codes), billing codes, pharmacy data, medical procedures and chronic disease registries. In addition, in the outpatient setting, each diagnosis code is automatically associated with a supportive text, which may be entered by the physician. This supportive text becomes part of the retrievable diagnostic information and is available at Clalit's EHR data warehouse. By utilizing in-house automatic anonymous queries, all structured (ie, ICD-9 codes) and unstructured (ie, diagnosis associated free text) information are extracted and analyzed. These automatic queries enable the capture of a wide range of terminology, acronyms and abbreviations, and the use of this information to confirm individuals with or without a selected disease in the ICD-9 data set.

Process

Three separate algorithms for detection of documented diagnoses of AF in Clalit's EHR data warehouse were defined and compared to a gold standard, established by open-chart reviews by two physicians. Based on the most valid algorithm (CRI's AF algorithm), all Clalit members aged>40 years who had ≥2 years of continuous membership in Clalit and documentation of AF diagnosis through 31 December 2015 were identified.

Establishing the Gold Standard

Utilizing an in-house predictive model to identify members with AF, 200 patients at the 20% highest risk quartile and who had been treated in two (of 14) major Clalit medical centers, including both in-patient and clinic visits, were randomly selected for establishing the gold standard based on open-chart reviews (see Supplementary 1, e-Tables 1–3 for

Table 1 Source of Information for the Three AF Algorithms

Algorithm Data Source	1	2	3
Community: Supporting free text	✓		
Community: ICD-9 codes	✓	✓	
Hospital: ICD-9 codes	✓	✓	✓

Note: Check marks represent the specific source of information that was used for each algorithm.
Abbreviations: AF, atrial fibrillation; ICD-9, International Classification of Diseases, ninth revision.

more information). This method enabled the recognition of both positive and negative diagnoses.

Open-chart reviews of the 200 clinic records were carried out by two physicians, a cardiologist and an internist. An individual was considered as an AF patient if he/she had documentation of arrhythmia based on ECG interpretation or cardiologist's explicit and detailed written diagnoses. In case of disagreement between the two physicians, the chart was re-opened by the cardiologist to determine the final diagnosis. This review established the gold standard of identifying AF.

Table 2 Characteristics of Individuals with Atrial Fibrillation[a]

	Individuals Identified as Having AF Based on Open-Chart Review (n=41)	Clalit Members Aged >40 with Documented AF (n=174,188)
Age (years)		
40–44, n (%)	0	1555 (0.9)
45–54, n (%)	0	8957 (5.1)
55–64, n (%)	8 (4.0)	22,342 (12.8)
65–74, n (%)	76 (38.0)	43,679 (25.1)
75–84, n (%)	91 (45.5)	62,541 (35.9)
85+, n (%)	25 (12.5)	35,114 (20.2)
Mean (SD)	76.7 (6.5)	74.7 (11.5)
Median (IQR)	76 (72–80)	76 (68–83)
Sex, n (%)		
Male	131 (65.5)	82,671 (47.5)
Female	69 (34.5)	91,517 (52.5)
Comorbidity, n (%)		
Hypertension	41 (100)	134,627 (77.3)
Ischemic stroke	3 (7.3)	18,112 (10.4)
Ischemic heart disease	100 (50.0)	87,406 (50.2)
Diabetes	59 (29.5)	57,581 (33.1)

Note: [a]As of date of AF diagnosis.
Abbreviations: Clalit, Clalit Health Services; AF, atrial fibrillation; SD, standard deviation; IQR, interquartile range.

Table 3 Validity and Accuracy Measurements of the Three Different Algorithms Based on Different Data Sources

	Algorithm		
	1	2	3
Data source	Inpatient: ICD-9 codes Outpatient: ICD-9 codes and supporting free text	Inpatient: ICD-9 codes Outpatient: ICD-9 codes	Inpatient: ICD-9 codes
Sensitivity	85.4%	82.9%	70.7%
Specificity	95.0%	95.0%	96.9%
PPV	81.4%	81.0%	85.3%
NPV	96.2%	95.6%	92.8%

Abbreviations: PPV, positive predictive value; NPV, negative predictive value.

Defining Potential Algorithms for AF Detection in Clalit EHR Data Warehouse

Based on information in Clalit's EHR data warehouse, three separate algorithms, differing in the source of their ICD-9 code and use of the associated supportive text, were defined (Table 1). Algorithm 1 incorporated inpatient and outpatient data. Specifically, individuals were considered to have AF if they had at least one documentation of AF from hospital records based on ICD-9 codes (427.3x [427.3, 427.31 and 427.32]) or outpatient records, based on both ICD-9 codes and supportive text that was associated with the documented diagnosis. By using automatic anonymous query, the supportive text that is associated with the documented diagnosis in Clalit's data warehouse was automatically extracted and analyzed and the relevance for AF diagnosis was evaluated. This differs from the validation process which involved open-chart review of all the textual material. Phrases considered as relevant for AF diagnosis included "Atrial Fibrillation" or "Atrial Flutter" as a sole expression, or with additional information including: auricular, rapid, slow, paroxysmal, convert to sinus, chronic, permanent, persistent, Coumadin treated, status post, post cardioversion (D/C), recurrent, along with year of diagnosis. Phrases that ruled out a definitive diagnosis of AF included: suspected AF, AF ruled out, AF?, and tachycardia. Records with a relevant code and free text for AF diagnosis, as well as records with missing ICD-9 codes but with relevant text, were considered as an AF diagnosis. Records with a relevant code for AF but text that did not support a definitive diagnosis of AF were not considered as documentation of AF diagnosis.

Algorithm 2 incorporated inpatient and outpatient data regardless of the free text associated with AF diagnosis. Specifically, individuals were considered as AF patients if

they had at least one documentation of AF in hospital records or in the outpatient records – based on ICD-9 codes only without supportive free text. Records with missing ICD-9 codes but relevant text were not included in this Algorithm.

Algorithm 3 used only inpatient data source. Individuals were considered as having AF if they had at least one documentation of AF in hospital records based on ICD-9 codes.

These three algorithms were compared to the gold standard. The added effect of incorporating outpatient information to inpatient was evaluated, and the algorithm with the optimal balance between all validation measures (sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)) was selected as CRI's AF algorithm.

Identification of Individuals with AF Diagnosis in Clalit Data Warehouse

In order to provide Clalit's policy makers further valid evidence regarding AF to appropriately target interventions, all individuals aged >40 years who had ≥2 years of continuous membership in Clalit and documentation of AF diagnosis in Clalit's EHR through 31 December 2015 were identified based on CRI's AF algorithm. The date of the first documentation of AF was considered as the individual patient's index diagnosis date. The source of the first AF documentation was considered as an individual patient's index diagnosis source (inpatient or outpatient). Patients were classified as having a newly incident AF case if they had, after 2 years of continuous membership without an AF diagnosis, a first documentation of AF in either a hospital, as a primary or secondary diagnosis, or in the outpatient setting that included text suitable for new AF case. Patients who had a first documentation of AF in a hospital as a past diagnosis, as well as patients who had a first documentation of AF in the outpatient setting that included suitable text for a previous AF diagnosis, were classified as prevalent cases. Mortality during the 2 years after index diagnosis was evaluated. Recurrent documentations of AF diagnosis were evaluated only for patients who survived at least 2 years following the index diagnosis.

Baseline Measurements

Baseline measurements as of index diagnosis included age (years), sex (male, female), and diagnosis of ischemic heart disease, ischemic stroke, hypertension, and diabetes. Comorbidity variables were evaluated based on any documentation prior index date.

Statistical Analysis

Descriptive statistics (totals, proportions, means [standard deviation], and medians [interquartile range]) were used to present socio-demographic and clinical characteristics as of index diagnosis for individuals who were selected for open-chart validation and diagnosed with AF and for all individuals aged >40 years with documented AF in Clalit's EHR. For each algorithm, the following measurements were calculated: sensitivity (the proportion of actual patients with AF who were correctly identified as positive by the algorithm), specificity (the proportion of patients without AF who were correctly identified as negative by the algorithm), positive predictive value (PPV, proportion of patients identified as positive by the algorithm who actually had AF), and negative predictive value (NPV, proportion of individuals identified as not having AF by the algorithm who did not have AF).

Results

Open-Chart Validation

Of the 200 charts reviewed by two physicians, 41 (20.5%) were identified as having AF. Main characteristics of these individuals are described in Table 2.

A comparison between the three different AF algorithms with physician's definition (gold standard) is presented in Table 3. Algorithm 1, which defined AF diagnosis based on information from the hospital (using ICD-9 codes) and outpatient settings (using ICD-9 codes and supportive free text) provided the optimal balance between all validation measures, with a high level of sensitivity (85.4%), specificity (95.0%), PPV (81.4%), and NPV (96.2%). Algorithm 2, which was also based on information from the hospital and outpatient settings but did not include text validation of outpatient records, provided a lower level of sensitivity (82.9%), same level of specificity (95.0%), and similar levels of PPV (81.0%) and NPV (95.6%). Algorithm 3, which included ICD-9 codes from the hospital setting only, reached 85.3% PPV, but had the lowest level of sensitivity (70.7%). For policy planning, additional results were examined according to algorithm 1.

Documentation of AF Diagnosis in Clalit's EHR

Based on Algorithm 1, there were 174,188 individuals aged >40 years (mean, SD: 74.7, 11.15) who were identified with AF in the Clalit data warehouse (according to hospital and outpatient records as of 31 December 2015). Of these,

47.5% were males and 77.3%, 10.4%, 50.2%, and 33.1% had history of hypertension, ischemic stroke, ischemic heart disease, and diabetes, respectively. Compared to these patients, individuals who were identified as having AF based on open-chart review were more likely to be male and with history of hypertension (Table 2). For 95,980 patients, the source for first documentation of AF was from hospital records (as a sole source of information or with additional documentation in the outpatient records). For 78,208 patients, the index diagnosis source was from the outpatient records only. With respect to records from outpatient sources, 73,889 individuals were defined as having AF based on a relevant ICD-9 code together with text that was suitable for AF diagnosis, and 4319 patients were recognized as having AF based on supportive text only. There were 1590 patients with ICD-9 codes indicating AF but with text that ruled out a definitive diagnosis of AF. These individuals were not considered as having AF.

During the 2 years after index diagnosis, 47,835 patients with AF died. Of 126,353 patients who were labelled as having AF and survived 2 years after the index diagnosis, 108,803 (86%) had at least one recurrent documentation after the initial documentations, with more documentation in the outpatient setting [mean (standard deviation): 3.5 (5.4)] compared to in the hospital [mean (standard deviation): 1.8 (3.0)].

Incident versus Prevalent Cases

Of 174,188 individuals with a documentation of AF in Clalit's EHR, 171,222 were classified as newly incident AF cases; of these, 94,420 patients had first documentation of AF in a hospital as a primary or secondary diagnosis, and 76,802 had a first documentation of AF in the outpatient setting (that included text suitable for new AF diagnosis [atrial fibrillation, atrial flutter, rapid AF, slow AF, paroxysmal AF, AF converted to sinus, AF – post D/C]). The remaining 2966 patients diagnosed with AF were classified as prevalent cases; 1560 had a first documentation of AF in a hospital as a past diagnosis, and 1406 had a first documentation of AF in the outpatient setting (that included text suitable for old AF [chronic AF, permanent AF, persistent AF, Warfarin treated, status post (s/p) AF, and recurrent AF]).

Discussion

The present study utilizes an extensive hospital and outpatient EHR system and evaluated the validity of three algorithms differing on the sources utilized in establishing AF diagnosis and on use of free text that describe the written diagnosis. The major difference in the respective algorithms was the origin of diagnosis (hospital versus outpatient). The broadest of the three algorithms, which included both hospital and outpatient settings and incorporated associated free text in the outpatient records, was shown to be the most valid system. This algorithm was the most sensitive and with high level of PPV and NPV when assessed in an open-chart validation by a cardiologist and an internist, based on codes and text and review of archived ECGs and Holter reports when available.

The three algorithms yielded comparable specificities (≥95%) and NPV (≥93%), demonstrating that they all had good ability to rule out an AF diagnosis, and to correctly label AF status. They mainly differed with respect to sensitivity, with better performance when including outpatient sources. This is, however, a critical difference when the aim is to find candidates for an intervention. In 2012, Jensen et al[4] published a review of studies describing validated algorithms for identifying patients at risk for incident AF. Of 281 published articles, they identified 16 unique studies. Of these, 10 used only hospital data, two used only outpatient data, and only four used combined data sources. Most of these studies relied on ICD-9 coded data and only one study attempted to define incident cases. They demonstrated considerable variability in the sensitivity of the algorithms reported, ranging from 56.9% to 87.7%. The PPV yielded by the three algorithms, ranged from 81.4% (broadest algorithm) to 85.3% (inpatients only), lower than some previous studies.[26,27] Sundboll et al evaluated PPVs of cardiovascular diagnosis in the Danish National Patient Registry and reported PPV of 95% for atrial fibrillation or flutter. False information in the data warehouse may be a result of either misdiagnosis or a problem in the transfer of the diagnosis from the patient's chart into the data warehouse.

With advances in computer technologies, the capability of text analysis has recently been added as an adjunct in the construction of algorithms for defining cardiovascular conditions.[28] In the current study, the additional analysis of the free text fields that described AF diagnosis in the EHR yielded modest improvement in PPV and NPV (Table 3). This suggests that at present, diagnostic coding, which incorporates present and historical data from both outpatient and hospital settings without unstructured free text, may be sufficient to provide reliable AF case detection. This observation is quite consistent with that of Wang et al,[29] who examined seven different algorithms utilizing different elements and

found that combining outpatient and hospital diagnoses yielded an optimal model. The validation of these simpler algorithms is important as broad datasets may not be available to all potential users. In fact, in many insurance claims data sets and health-care sites, free text information and/or applicable automatic algorithms to extract and analyze free text are not available. Having said this, despite the modest contribution of textual analysis to the validity of AF definition, it is possible that free text analysis may have more significance for other disease entities.

Previous studies used populations with different levels of risk in order to facilitate the testing of potential algorithms. While some studies have been based on the general population, others have focused on high-risk AF patients, specifically, the elderly, hospitalized patients or those manifesting strokes, and individuals with heart disease, hypertension or other associated conditions. The validation dataset utilized for the chart reviews in the current study was defined as individuals at the highest risk by a five block stepwise logistic model. Despite the impact of utilizing such a select population, which may increase the PPV, this created an ideal sub-population with high numbers of patients with AF for open-chart validation. In addition, this approach provided the ability to estimate specificity and NPV which is often difficult in electronic data validation studies because of the challenge in identifying true negatives.

The creation and validation of the CRI's AF algorithm enabled Clalit's policy makers to identify all individuals (174,188) with a documentation of AF in Clalit's EHR. This is an important asset when considering newer and more costly interventions. The extremely low annual turnover in Clalit facilitated adequate follow up over a long period to identify new cases (171,222).

A major strength of the present study is the large stable population and the availability of free text in the outpatient records both of which combined to give more precise categorization between prevalent and incident AF. In addition, by choosing our validation sample from a population identified initially as high risk, we were able to measure both true positives and negatives.

This study also had some limitations. First, similar to most previous studies, the current study focused on validation of algorithms for identifying prevalent AF rather than incident AF. Despite this, the availability of free text associated with a written diagnosis enabled us to utilize free text associated with the outpatient records and by better identifying first events to distinguish true incidence and prevalence. Second, it was feasible to open and review only 200 charts. Therefore, the study population was selected from a list of individuals at highest risk. This method enabled the recognition of both positive and negative diagnoses. However, the performance metrics reported in this study represent the validation of the algorithms for a high-risk population and not for the general population. It is possible that different bounds could be achieved in broader patient populations. Third, individuals were selected for open-chart review only if they had been treated in two of 14 major Clalit medical centers. Although these medical centers include ambulatory clinic visits, the primary care clinics may be underrepresented. Therefore, this study focused on evaluating the added effect of incorporating outpatient information to inpatient and refrained from presenting an algorithm for the outpatient records only. Lastly, although Clalit is the largest health provider in Israel and the membership in one of the four health providers is not affected by the payment for the health insurance, the results of this study might not be representative for the entire population of atrial fibrillation patients in Israel.

Conclusion

In summary, the present study demonstrates the importance of combining inpatient and outpatient records for identification of individuals with AF in EHR data. Among individuals at highest risk for having non-valvular AF, incorporating outpatient records in the algorithm, in addition to inpatient records, may significantly contribute to increase sensitivity with a relatively small decrease in PPV.

Acknowledgments

We thank Sydney Krispin, MPH, of the Clalit Research Institute, for her editorial support.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study. *Circulation*. 2014;129(8):837–847. doi:10.1161/CIRCULATIONAHA.113.005119
2. Haim M, Hoshen M, Reges O, Rabi Y, Balicer R, Leibowitz M. Prospective national study of the prevalence, incidence, management and outcome of a large contemporary cohort of patients with incident non-valvular atrial fibrillation. *J Am Heart Assoc*. 2015;4(1):e001486. doi:10.1161/JAHA.114.001486

3. Tu K, Nieuwlaat R, Cheng SY, et al. Identifying patients with atrial fibrillation in administrative data. *Can J Cardiol*. 2016;32(12):1561–1565. doi:10.1016/j.cjca.2016.06.006
4. Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. Identifying atrial fibrillation from electronic medical data: a systematic review. *Pharmacoepidemiol Drug Saf*. 2012;21(1):141–147. doi:10.1002/pds.2317
5. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. *J Intern Med*. 2013;274(6):547–560. doi:10.1111/joim.12119
6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395–405. doi:10.1038/nrg3208
7. Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA*. 2003;290(20):2685–2692. doi:10.1001/jama.290.20.2685
8. Dewland TA, Olgin JE, Vittinghoff E, Marcus GM. Incident atrial fibrillation among Asians, Hispanics, Blacks, and Whites. *Circulation*. 2013;128(23):2470–2477. doi:10.1161/CIRCULATIONAHA.113.002449
9. Brophy MT, Snyder KE, Gaehde S, Ives C, Gagnon D, Fiore LD. Anticoagulant use for atrial fibrillation in the elderly. *J Am Geriatr Soc*. 2004;52(7):1151–1156. doi:10.1111/j.1532-5415.2004.52314.x
10. Yuan Z, Bowlin S, Einstadter D, Cebul RD, Conners AR Jr, Rimm AA. Atrial fibrillation as a risk factor for stroke: a retrospective cohort study of hospitalized medicare beneficiaries. *Am J Public Health*. 1998;88(3):395–400. doi:10.2105/AJPH.88.3.395
11. Hravnak M, Hoffman LA, Saul MI, et al. Atrial fibrillation: prevalence after minimally invasive direct and standard coronary artery bypass. *Ann Thorac Surg*. 2001;71(5):1491–1495. doi:10.1016/S0003-4975(01)02477-8
12. Shireman TI, Howard PA, Kresowik TF, Ellerbeck EF. Combined anticoagulant–antiplatelet use and major bleeding events in elderly atrial fibrillation patients. *Stroke*. 2004;35(10):2362–2367. doi:10.1161/01.STR.0000141933.75462.c2
13. Antani MR, Beyth RJ, Covinsky KE, et al. Failure to prescribe warfarin to patients with nonrheumatic atrial fibrillation. *J Gen Intern Med*. 1996;11(12):713–720. doi:10.1007/BF02598984
14. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *A J Med Qual*. 2004;19(5):201–206. doi:10.1177/106286060401900504
15. Go AS, Hylek EM, Phillips KA, et al. Implications of stroke risk criteria on the anticoagulation decision in nonvalvular atrial fibrillation. *Circulation*. 2000;102(1):11–13. doi:10.1161/01.CIR.102.1.11
16. Psaty BM, Manolio TA, Kuller LH, et al. Incidence of and risk factors for atrial fibrillation in older adults. *Circulation*. 1997;96(7):2455–2461. doi:10.1161/01.CIR.96.7.2455
17. Alonso A, Agarwal SK, Soliman EZ, et al. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J*. 2009;158(1):111–117. doi:10.1016/j.ahj.2009.05.010
18. Whittle J, Wickenheiser L, Venditti LN. Is warfarin underused in the treatment of elderly persons with atrial fibrillation? *Arch Intern Med*. 1997;157(4):441–445. doi:10.1001/archinte.1997.00440250097011
19. Glazer NL, Dublin S, Smith NL, et al. Newly detected atrial fibrillation and compliance with antithrombotic guidelines. *Arch Intern Med*. 2007;167(3):246–252. doi:10.1001/archinte.167.3.246
20. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *JAMA*. 2001;285(18):2370–2375. doi:10.1001/jama.285.18.2370
21. Shen AY-J, Yao JF, Brar SS, Jorgensen MB, Wang X, Chen W. Racial/ethnic differences in ischemic stroke rates and the efficacy of warfarin among patients with atrial fibrillation. *Stroke*. 2008;39(10):2736–2743. doi:10.1161/STROKEAHA.107.508580
22. Dublin S, French B, Glazer NL, et al. Risk of new-onset atrial fibrillation in relation to body mass index. *Arch Intern Med*. 2006;166(21):2322–2328. doi:10.1001/archinte.166.21.2322
23. Flaker GC, McGowan DJ, Boechler M, Fortune G, Gage B. Underutilization of antithrombotic therapy in elderly rural patients with atrial fibrillation. *Am Heart J*. 1999;137(2):307–312. doi:10.1053/hj.1999.v137.91403
24. Brass LM, Krumholz HM, Scinto JM, Radford M. Warfarin use among patients with atrial fibrillation. *Stroke*. 1997;28(12):2382–2389. doi:10.1161/01.STR.28.12.2382
25. Massicotte-Azarniouch D, Kuwornu JP, Carrero -J-J, et al. Incident atrial fibrillation and the risk of congestive heart failure, myocardial infarction, end-stage kidney disease, and mortality among patients with a decreased estimated GFR. *Am J Kidney Dis*. 2018;71(2):191–199. doi:10.1053/j.ajkd.2017.08.016
26. Sundboll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the Danish national patient registry: a validation study. *BMJ Open*. 2016;6(11):012832. doi:10.1136/bmjopen-2016-012832
27. Rix TA, Riahi S, Overvad K, et al. Validity of the diagnoses atrial fibrillation and atrial flutter in a Danish patient registry. *Scand Cardiovasc J*. 2012;46(3):149–153. doi:10.3109/14017431.2012.673728
28. Pakhomov SS, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: role of natural language processing of the medical record. *Am Heart J*. 2007;153(4):666–673. doi:10.1016/j.ahj.2006.12.022
29. Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. *J Am Med Inform Assoc*. 2017;24(2):339–344. doi:10.1093/jamia/ocw082