REVIEW

Sensitivity, specificity, positive and negative predictive values of identifying atrial fibrillation using administrative data: a systematic review and meta-analysis

Ren Jie Robert Yao[1]
Jason G Andrade[1]
Marc W Deyell[1]
Heather Jackson[2]
Finlay A McAlister[3]
Nathaniel M Hawkins[1]

[1]Division of Cardiology, University of British Columbia, Vancouver, Canada; [2]Cardiac Services British Columbia, Vancouver, Canada; [3]Division of General Internal Medicine, University of Alberta, Edmonton, Canada

Introduction: Atrial fibrillation (AF) is the commonest arrhythmia and a major cause of stroke and health care utilization. Researchers and administrators use electronic health data to assess disease burden, quality and variance in care, value of interventions and prognosis. We performed a systematic review and meta-analysis to assess the validity of AF case definitions in administrative databases.

Methods: Medline was searched from 2000 to 2018. Extracted information included sensitivity, specificity, positive and negative predictive values (PPV and NPV) for various AF case definitions. Estimates were pooled using random-effects models due to significant heterogeneity between studies.

Results: We identified 24 studies, including 21 from North America or Scandinavia. Hospital, ambulatory and mixed data sources were assessed in 10, 4 and 10 studies, respectively. Nine different AF case definitions were evaluated, most based on ICD-9 or 10 codes. Twenty-two studies assessed case definitions in patients diagnosed with AF and thus could generate PPV alone. Half the studies sampled unrestricted populations including a mix of those with and without AF to assess sensitivity. Only 13 studies included ECG confirmation as a gold standard. The pooled random effects estimates were: sensitivity 80% (95% CI 72–86%); specificity 98% (96–99%); PPV 88% (82–94%); NPV 97% (94–99%). Only 3 studies reported all accuracy parameters and included rhythm monitoring in the gold standard definition.

Conclusion: Relatively few studies examined sensitivity, and fewer still included rhythm monitoring in the gold standard comparison. Administrative data may fail to identify a significant proportion of patients with AF. This, in turn, may bias estimates of quality of care and prognosis.

Keywords: atrial fibrillation, registries, validation studies, accuracy, sensitivity, specificity

Introduction

Atrial fibrillation (AF) increases risk of stroke, heart failure and death, and is one of the few cardiac conditions whose prevalence continues to rise.[1,2] Most developed health systems collect reasons for hospital and ambulatory encounters for administration, service planning, quality improvement and reimbursement. Health services researchers use these administrative electronic databases to monitor the burden of disease, quality of care, and ascertain exposures or outcomes. The accuracy of AF

Correspondence: Nathaniel M Hawkins
University of British Columbia, BC Centre for Improved Cardiovascular Health, St. Paul's Hospital, 1081 Burrard Street, Vancouver, BC V6Z 1Y6, Canada
Tel +1 604 875 4111
Fax +1 604 875 5504
Email nat.hawkins@ubc.ca

753

identification is central to these applications. Sensitivity and specificity, though theoretically independent, typically trade-off and are inversely related.[3] The "optimal" approach to identifying AF depends on the purpose. High sensitivity more completely captures a population, improves generalizability and is important when defining AF as an exposure. By contrast, high specificity ensures persons identified truly have AF and is central to adjudicating treatment uptake, which appears inappropriately low if patients with sinus rhythm are misclassified as having AF.[4]

Conceptually, the AF patient journey involves ambulatory and acute contacts dissociated in time and space, between which information flows by varying amounts and rates. Interrogating data sources over short time intervals or single environments may miss infrequent encounters. A previous systematic review examined the accuracy of AF detection, but was limited to ICD-9 codes only, non-contemporary electronic sources, North American cohorts and narrative synthesis without consideration of the impact of different health care settings (indeed the focus was largely on hospitalization data).[5] We, therefore, undertook a systematic review to address these evidence gaps.

Methods
Participants, outcomes and study designs
Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were followed (Table S1). We examined the accuracy of AF case definitions in electronic administrative health data, namely sensitivity (SN), specificity (SP), positive and negative predictive values (PPV and NPV). Inpatient, outpatient and mixed populations were included. All study designs were accepted. The study protocol was not published.

Search strategy and data collection
MEDLINE was searched from January 2000 to February 2018, limited to adult humans and English language, excluding case studies, reviews and conference abstracts. Search terms were determined by literature review and database query. The search strategy combined Medical Subject Headings (MeSH) terms and keywords in title and abstract to define three groups: atrial fibrillation (including atrial flutter (AFL) if not differentiated); administrative and electronic medical databases; and studies examining accuracy of AF identification within these records (Table S2). The search returned 1007 unique

records. Manual bibliography searches identified an additional 31 publications (Figure S1). Titles and abstracts were screened for inclusion, and 302 full-text articles reviewed. Studies fulfilling the participant, outcomes and study design criteria were included. Variables of interest were decided a priori, expanded iteratively after pilot and collected in Microsoft Excel. The following information was extracted: bibliographic details, sample size, population characteristics, inclusion and exclusion criteria, codes and algorithms, AF confirmation gold standard and accuracy parameter outcomes.

Data synthesis
Weighted averages of sensitivity, specificity, positive and negative predictive values were calculated using the DerSimonian-Laird random effects model.[6] Forest plots of estimates with 95% confidence intervals (CI) were generated. Publication bias was assessed through visual inspection of funnel plots and the Begg-Mazumdar rank correlation test for asymmetry.[7,8] Heterogeneity was tested with visual forest plot inspection, Cochrane Q, I^2 and Tau^2 statistics.[9] Estimates with significant heterogeneity (I^2>90%) were examined manually and formally for moderating effects including country, publication year and reference standard, none of which were significant. The leave-one-out method was used to determine if the results were sensitive to the inclusion of extreme values from specific studies.[10]

Results
Study characteristics
Twenty-four studies were identified (Table 1). Most originated from countries with established administrative databases that are often interrogated by health services researchers, including 10 from the United States, 3 from Canada and 8 from Sweden or Denmark. The populations were heterogeneous, including general unselected, stroke and post-operative cohorts. Hospital, ambulatory and mixed data sources were assessed in 10, 4 and 10 studies, respectively. Only 3 studies outside Scandinavia examined mixed populations.[4,11,12] One Canadian study included administrative data from emergency departments separate from hospitalizations.[11]

Coding and case definition algorithms
Most reports investigated International Classification of Diseases codes: ICD8 (427.93 AF, 427.94 AFL), ICD9 (4

Table 1 Study characteristics

Validated	Prevalent/incident	Time period defining incident	Year	Country	Population sampled	AF (%)	Data source	Exclusion
Mixed data								
Frost 07[28]	Prevalent	–	80–02	Denmark	Stroke + AF	–	National patient registry	Stroke, valvular
Norberg 13[29]	Prevalent	–	04–10	Sweden	General	3.0	National patient register	–
Baturova 14[15]	Prevalent	–	01–11	Sweden	Stroke	28.2	National patient register	–
Tu 16[11]	Prevalent	–	11	Canada	General	2.6	Administrative databases	–
Navar-Boggan 15[4]	Prevalent	–	11–12	US	In/outpatient	–	Duke University Health System	AFL alone
Sung 16[12]	Prevalent	–	08–10	Taiwan	Stroke	10.1	Taiwan National Health Insurance	–
Sundboll 16[30]	Incident	–	10–12	Denmark	General	–	National patient registry	–
Rix 12[31]	Incident	Start records	93–09	Denmark	General	6.1	National patient registry	–
Frost 05[32]	Incident	Start records	93–99	Denmark	General	1.2	National patient registry	Prior cancer
Smith 10[33]	Incident	Not specified	91–05	Sweden	General	1.3	National hospital/death registry	–
Ambulatory data								
Brophy 04[34]	Prevalent	–	97–01	US	General	–	Veteran Affairs clinics	Valvular disease
Borzecki 04[13]	Prevalent	–	98–99	US	Hypertension	6	Veteran Affairs clinics 10 sites	–
Go 01/00[16,18]	Prevalent	–	96–97	US	AF	1.0	HMO Kaiser Permanente	Prior AF ECG
Ruigomez 05/02[14,35]	Incident	Start records	96	UK	AF	0.3	General Practice Research Database	–
Hospital data								
Quon 04[36]	Prevalent	–	96–97	Canada	General	9.4	Discharge Abstract Database	–
Kokotailo 05[20]	Prevalent	–	00–01	Canada	Stroke	–	Discharge Abstract Database	–
Thigpen 15[19]	Prevalent	–	06–10	US	Stroke + AF	–	Administrative databases	Prosthetic valve
Shen 08[37]	Prevalent	–	95–00	US	General	–	HMO Kaiser Permanente	Valvular disease
Shireman 04[38]	Prevalent	–	98–99	US	AF warfarin	–	Medicare discharges, 750/state	–
Chotchaisuwatana[22]	Prevalent	–	07–08	Thailand	AF	–	Thailand community hospitals	–
Munkholm 15[39]	Incident	Unspecified	11–13	Denmark	Cardiac surgery	36.4	Western Denmark Heart Registry	–
Walkey 11[40]	Incident	Unspecified	07–08	US	Sepsis	5.9	California Inpatient Database	–
Alonso 09[17]	Incident	Start study	87–04	US	General	7.0	ARIC study	–
Hravnak 01[41]	Incident	Unspecified	96–98	US	CABG	31.9	Medical Archive Pittsburgh	–

Abbreviations: AF, atrial fibrillation; AFL, atrial flutter; ARIC, Atherosclerosis Risk in Communities study; CABG, coronary artery bypass grafting; ECG, electrocardiogram; HMO, health maintenance organization; UK, United Kingdom; US, United States.

digit code 427.3 and more explicit 427.31 AF and 427.32 AFL), ICD10 (I48). Overall, 9 different combinations of codes were studied (Table 2). The impact of coding position (primary versus secondary diagnosis in hospitalization data) was never examined. Four studies compared the accuracy of two versus one encounter coded as AF in ambulatory data sources within a single year.[4,11–13] This consistently increased specificity but decreased sensitivity. A single study compared 2 versus 1 year for case ascertainment in Veterans Affairs outpatient records, finding greater sensitivity with only slightly reduced specificity.[13] Overall in that study, 2 diagnoses over 2 years were optimal for detecting AF.[13] Only one study from Canada examined more complex algorithms including cardioversion codes and pharmacy dispensations for antiarrhythmic drugs.[11]

Characteristics of AF

Prevalent and incident AF were assessed in two-thirds and one-thirds of studies, respectively (Table 1). Incident cases were typically defined by exclusion of prior AF diagnoses since the records began, or methods were not specified. The incidence and prevalence of AF varied markedly depending on the population studied, from 0.3% to 55%.[14,15] The incidence and prevalence was highest in studies following cardiac surgery (32–36%) and stroke (10–28%), lower in general hospitalizations (7–9%) and lowest in unselected outpatients (0.3–1%). No study distinguished between persistent and paroxysmal AF. Three studies reported from 7.0% to 20.4% of the coded AF to be transient.[4,16,17] Two defined transient as a single episode without recurrence,[4,16] while two added precipitants including cardiac surgery and/or hyperthyroidism.[16,17]

Gold standard for diagnosis of AF

With the exception of two studies, medical chart review was considered the gold standard by which history of AF was classified (Table 2). Of these, 13 studies specifically included ECG review, of which 2 employed ECG alone for confirmation of AF.[18,19] No prospective protocols or frequencies for ECG were reported. A median of 11 ECGs per patient with AF was noted in a Swedish outpatient setting.[15] Only 4 studies mentioned use of longer term rhythm monitoring such as Holter, although these results may also have been available in medical record review.

Sensitivity and specificity

Half the studies (n=12) sampled an unrestricted population including those without AF to assess sensitivity of case definitions and these ranged from 57% to 93%, median 81% (Table 3). The pooled random effects estimate for sensitivity was 80% (95% CI 72–86%) with significant heterogeneity (Q 439, I^2 97.7%, Tau^2 0.08). One-third of the studies (n=8) reported specificity. Estimates were consistently high in ambulatory, hospitalized and mixed populations, ranging from 91% to 99%, median 99% (Table 3). The pooled random effects estimate for specificity was 98% (95% CI 96–99%).

Positive and negative predictive value

Positive predictive value was reported in nearly all studies (n=22), and was the only parameter reported in half the studies (n=12). The PPV ranged from 71% to 99%, median 93% (Table 3). The pooled random effects estimate was 88% (95% CI 82–94%) with significant heterogeneity (Q 4997, I^2 99.6%, Tau^2 0.02) (Figure 1). The pooled estimate was similar in ambulatory, hospitalized and mixed populations (respectively: 87% (78–96%); 87% (79–95%); 90% (85–94%)). Negative predictive value was reported in 8 studies. Estimates were consistently high, ranging from 86% to 99%, median 98% (Table 3). The pooled random effects estimate was 97% (95% CI 94–99%).

Discussion

This analysis reports several key findings. The overall specificity and NPV of an AF diagnosis using the ICD case definitions was high, 98% and 97%, respectively. The sensitivity and PPV were lower though reasonable, 80% and 88%, respectively. Only half the studies sampled patients with and without an assigned diagnosis of AF to determine the sensitivity of the case definitions and thus the proportion potentially missed by using administrative data. Half the studies confirmed AF using electrocardiography as the gold standard, while the remainder employed medical record review, alternative databases (like primary care EMRs) and/or patient questionnaires. Only 3 studies reported all accuracy parameters and included rhythm monitoring in the gold standard definition.[15,17,20]

Sensitivity

High sensitivity improves case finding as it more completely captures a population, increases the estimated incidence and prevalence and enhances generalizability. This is particularly relevant when estimating the burden of disease and to reduce bias when studying health inequalities. Sensitivity is also important when defining AF as an

Table 2 Characteristic and validation of atrial fibrillation in studies

Validated	ICD code(s) *	Number of encounters required to diagnose (and over what time period)	AF confirmation gold standard
Mixed data			
Frost 07[28]	427.93–94, I48	1	Chart, ECG, telemetry, event recorder
Norberg 13[29]	I48	1	Chart, ECG, Holter, CIED
Baturova 14[15]	427D, I48	1	Chart, ECG
Tu 16[11]	427.3x, I48 427.x physician bill	1 Hosp/ED or 1MD + cardioversion or 1 rhythm control or 1 MD + OAC (1 year)	Primary care chart (physician notes and/or ECG)
Navar-Boggan 15[4]	427.31	1 inpatient or 2 outpatient/ED visits (1 year)	Chart, ECG
Sung 16[12]	427.31	1 inpatient or 2 outpatient visits (2 years)	Database registry
Sundboll 16[30]	427.93–94, I48	1	Chart
Rix 12[31]	427.93–94, I48	1	Chart, ECG, electrophysiologist review
Frost 05[32]	427.93–94, I48	1	Chart, ECG, telemetry or Holter
Smith 10[33]	427.92, 427D, I48	1	Chart, ECG, electrophysiologist review
Ambulatory data			
Brophy 04[34]	427.3, 427.31	1	Chart, ECG
Borzecki 04[13]	427.3	≥1 or ≥2 (1 or 2 years)	Chart
Go 01/00[16,18]	427.31	1	ECG
Ruigomez 05/02[14,35]	OXMIS codes	1	Questionnaire to general practitioner
Hospital data			
Quon 04[36]	427.3x	1 (inpatient)	Chart
Kokotailo 05[20]	Not stated ICD9	1 (inpatient)	Chart, ECG, Holter
Thigpen 15[19]	427.31	1 (inpatient)	ECG
Shen 08[37]	427.3x 1st 3 codes	1 (inpatient)	Chart, ECG
Shireman 04[38]	427.31	1 (inpatient)	Chart
Chotchaisuwatana[22]	I48	1 (inpatient)	Chart
Munkholm 15[39]	Not stated	1 (inpatient)	Chart
Walkey 11[40]	427.3x	1 (inpatient)	Chart
Alonso 09[17]	427.31	1 (inpatient)	Chart, ECG
Hravnak 01[41]	427.31	1 (inpatient)	Text search, prescriptions

Abbreviations: CIED, cardiac implantable electronic device; ECG, electrocardiogram; ED, emergency department; ICD, International Classification of Diseases; MD, doctor of medicine; OAC, oral anticoagulation.

Table 3 Codes and accuracy for identifying atrial fibrillation

Validated	Coding position	n Sample	AF	True positive	False positive	PPV	Not AF	True negative	False negative	NPV	SN	SP
Mixed data												
Frost 07[28]	1°/2°	174	174	172	2	99	–	–	–	–	–	–
Norberg 13[29]	–	2274	2196	2119	77	96	–	–	155	–	93	–
Baturova 14[15]	1°/2°	666	188	152	36	81	482	446	32	92	83	93
Tu 16[11]	–	7500	218	155	63	71	7308	7245	37	99	81	99
Navar-Boggan 15[4]	–	300	300	287	13	96	–	–	–	–	–	–
Sung 16[12]	–	6469	666	474	192	71	5818	5626	177	97	73	97
Sundboll 16[30]	1°/2°	97	97	92	5	95	–	–	–	–	–	–
Rix 12[31]	–	284	284	262	22	93 All / 65 ED / 94 IP/OP	–	–	–	–	–	–
Frost 05[32]	–	116	116	112	4	97	–	–	–	–	–	–
Smith 10[33]	–	100	100	97	3	97	–	–	–	–	–	–
Ambulatory data												
Brophy 04[34]	–	3366	–	2619	–	–	–	–	747	–	78	78
Borzecki 04[13]	–	1176	69	59	10	86≥1/1	1103	1093	14	99≥1/1	80≥1/1, 67≥2/1, 86≥1/2, 74≥2/2	99≥1/1, 99≥2/1, 97≥1/2, 99≥2/2
Go 01/00[16,18]	–	50	50	39	11	78	–	–	–	–	–	–
Ruigomez 05/02[14,35]	–	1888	1888	1763	125	93	–	–	–	–	–	–
Hospital data												
Quon 04[36] Prev	–	1200	102	96	6	94	1098	1081	17	98	85	99
Kokotailo 05[20]	–	137	18	17	1	94	116	115	4	97	81	99
Thigpen 15[19]	Any	1706	1706	1489	217	87	–	–	–	–	–	–
Shen 08[37]	Any	100	100	96	4	96	–	–	–	–	–	–
Shireman 04[38]	1°/2°	38,924	38,924	27,674	11,250	71	–	–	–	–	–	–
Chotchaisuwatana[22]	–	193	193	169	24	88	–	–	–	–	–	–
Munkholm 15[39]	–	1381	458	378	80	83	878	798	125	86	75	91
Quon 04[36] Incident	–	1200	7	6	1	86	1193	1186	7	99	46	99
Walkey 11[40]	–	163	163	147	16	90	–	–	–	–	–	–
Alonso 09[17]	1°	1546	169	135	34	80	1377	1351	26	98	84	98
Hravnak 01[41]	–	260	–	148	–	–	–	–	112	–	57	–

Abbreviations: AF, atrial fibrillation; NPV, negative predictive value; PPV, positive predictive value; SN, sensitivity; SP, specificity; ED, Emergency department population; IP, inpatient population; OP, outpatient population.
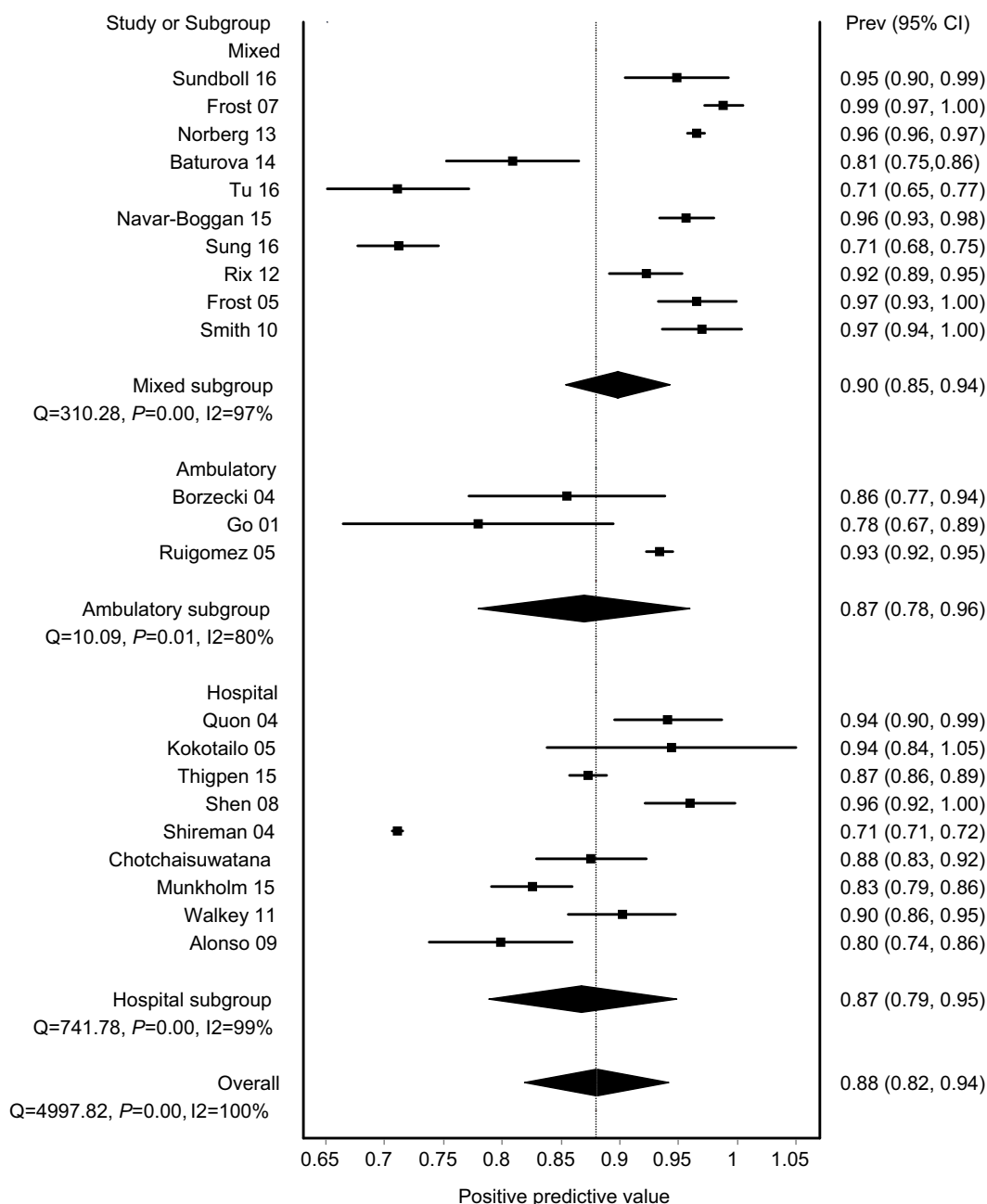
Study or Subgroup		Prev (95% CI)
Mixed		
Sundboll 16		0.95 (0.90, 0.99)
Frost 07		0.99 (0.97, 1.00)
Norberg 13		0.96 (0.96, 0.97)
Baturova 14		0.81 (0.75, 0.86)
Tu 16		0.71 (0.65, 0.77)
Navar-Boggan 15		0.96 (0.93, 0.98)
Sung 16		0.71 (0.68, 0.75)
Rix 12		0.92 (0.89, 0.95)
Frost 05		0.97 (0.93, 1.00)
Smith 10		0.97 (0.94, 1.00)
Mixed subgroup		0.90 (0.85, 0.94)
Q=310.28, P=0.00, I2=97%		
Ambulatory		
Borzecki 04		0.86 (0.77, 0.94)
Go 01		0.78 (0.67, 0.89)
Ruigomez 05		0.93 (0.92, 0.95)
Ambulatory subgroup		0.87 (0.78, 0.96)
Q=10.09, P=0.01, I2=80%		
Hospital		
Quon 04		0.94 (0.90, 0.99)
Kokotailo 05		0.94 (0.84, 1.05)
Thigpen 15		0.87 (0.86, 0.89)
Shen 08		0.96 (0.92, 1.00)
Shireman 04		0.71 (0.71, 0.72)
Chotchaisuwatana		0.88 (0.83, 0.92)
Munkholm 15		0.83 (0.79, 0.86)
Walkey 11		0.90 (0.86, 0.95)
Alonso 09		0.80 (0.74, 0.86)
Hospital subgroup		0.87 (0.79, 0.95)
Q=741.78, P=0.00, I2=99%		
Overall		0.88 (0.82, 0.94)
Q=4997.82, P=0.00, I2=100%		

Positive predictive value

Figure 1 Positive predictive values of atrial fibrillation (AF) algorithms stratified by population type.

exposure. Misclassification of exposure (eg, AF) as non-exposure (eg, no AF) attenuates the association with outcomes such as stroke.[21] By contrast, sensitivity is less of a concern when defining AF as an outcome, for example in pharmacovigilance studies. In these circumstances, estimates of relative risk are not biased providing misclassification occurs to the same degree in exposed and non-exposed patients.

Sensitivity is reduced when cases are missed and AF is misclassified as normal (ie, false negatives). This occurs in two circumstances. First, when recording or coding is incorrect. Second, when correctly recorded and coded diagnoses are missed in time or space. Examining shorter time frames may miss infrequent encounters, as evidenced in the Veterans Affairs study where sensitivity increased using a 2 versus 1 year period for case ascertainment.[13] Information also flows by varying amounts and rates through health systems. Although the median time from AF on ECG to diagnosis in the Swedish Patient Register was 16 days, this time lapse exceeded 6 months in one-third of patients.[15]

Sensitivity may be viewed from different perspectives: local, horizontal level of care (eg, primary care), vertical (eg, health maintenance organization) or global (entire health care system). Examining a single health care setting may miss encounters meeting the AF case definition in another. For example, hospitalization data alone misses patients managed entirely in the community, causing under-estimates of prevalence rates and over-estimates of adverse outcome rates. There were insufficient studies to accurately compare sensitivity between ambulatory, hospital and mixed populations. However, one of the mixed population studies did compare the accuracy of coding between primary care, secondary care or both together. In that study from Ontario, the sensitivity was 45%, 39% and 75% for hospitalization, emergency department or outpatient data sources alone, respectively, and 83% combining the three sources.[11]

The true population incidence and prevalence may also be influenced by access to rhythm monitoring (ECG, Holter, event or loop recorder), reporting standards (eg, training, quality assurance) and information transfer (eg, interface to electronic medical record). These factors are potentially more challenging in community than in hospital settings, particularly relevant to measuring inequalities, and difficult to quantify. None of the included studies described these aspects of access.

Positive predictive value

Since sensitivity and specificity are typically inversely related, higher sensitivity reduces specificity, which increases false positives and lowers PPV. The impact on PPV is magnified for diseases with a relatively low prevalence such as AF. A high PPV ensures persons identified truly have AF (fewer false positives). This is central to adjudicating treatment uptake, which will appear inappropriately withheld if patients with sinus rhythm are misclassified as having AF, unless OAC is prescribed for an alternate reason.[4]

A PPV value exceeding 85–90% suggested adequate for research purposes.[19,20] The reasons for false positives were rarely explored.[22] Potential scenarios include: 1) miscoding eg, allergic rhinitis was written as "AR" and coded as AF;[22] 2) rhythm misinterpretation such as atrial tachycardia; 3) misreporting if based on medical history alone; and 4) AF defined by an intervention shared with other conditions eg, cardioversion. PPV is also highly dependent on disease prevalence: as many studies focused on older or high-risk individuals they may overestimate

the true PPV for that case definition if applied in a younger population.

Oral anticoagulation is the only treatment to improve survival in patients with AF, and thus a key quality indicator. Although the overall PPV was high (88%), the specificity and PPV to identify AF requiring anticoagulation (as opposed to any AF) could be lower for several reasons. First, up to 10% of incident AF is isolated with a defined precipitant, low recurrence, and may not require anticoagulation.[4,16,23] Only three studies reported or excluded such patients.[4,16,17] Second, anticoagulation adjudication requires accurate coding of embolic and bleeding risk factors, which like AF exhibit high specificity but are under-reported.[12,22] More subjective bleeding risks such as frailty and falls are particularly difficult to quantify, although a recently described frailty score based on administrative data (the Hospital Frailty Risk Score) has been described.[24] Finally, patient preferences are major drivers of anticoagulation decisions but are never captured in administrative databases.

Atrial fibrillation phenotype and coding considerations

The disease spectrum (permanent, persistent, paroxysmal, isolated unprovoked or provoked episodes) was rarely reported yet also impacts accuracy of AF detection. Permanent or persistent AF is associated with greater comorbidity and hence health care encounters during which arrhythmia is continuously present. By contrast, isolated or paroxysmal AF may be under-represented by health care encounters. Treatment including rate versus rhythm control and anticoagulation also varies based on symptoms, AF duration, risk of recurrence and thromboembolism.[25] The accuracy of administrative data to identify AF requiring anticoagulation is thus further lessened by limited phenotypic characterization.[4] The AF phenotype may also impact the "gold standard" for diagnosing AF, whereby paroxysmal AF is missed by ECG alone but detected by chart review. In the only study examining this issue, ECG review did not improve sensitivity of AF detection over diagnosis codes alone.[4]

Most developed health systems collect reasons for hospital and ambulatory encounters for administration, service planning, quality improvement and reimbursement. A single primary or most responsible diagnosis is typically assigned, while conditions complicating or prolonging stay are coded in multiple secondary positions, sometimes

further categorized as pre-existing or de novo disease. Differences in coding accuracy, treatment and prognosis are reported between primary and secondary positions for conditions such as heart failure.[26,27] To our knowledge, such differences have not been explored in patients with AF, and no study identified by our search compared coding positions.

Strengths and limitations

Several strengths and limitations merit consideration. Our analysis is contemporary, included varied health systems, ICD-8 to ICD-10 codes, and both ambulatory and hospital populations. However, most studies originated from North America or Scandinavia, and examined ICD codes in administrative data sources. This potentially limits generalizability to other health care systems. There was significant heterogeneity in terms of population, prevalence of AF and reported accuracy parameters. Most studies assessed accuracy in restricted cohorts as opposed to the broader population.

Directions for future research

Health service researchers and administrators may interpret administrative data using either our pooled estimates or locally relevant studies from among those identified. Jurisdictions would ideally conduct nationally representative validation studies to provide estimates specific to their populations and data sources. These should examine existing codes and test new case definition algorithms in all data sources with differences in coding practices and diagnostic accuracy (eg, hospitalization, emergency department, ambulatory primary and secondary care), and in scenarios with varying disease prevalence. Though challenging and costly, random sampling of representative populations is essential to define sensitivity, enhance generalizability and reduce bias when studying inequality. To understand the true disease burden, algorithms should combine primary and secondary care data sources.

More complex algorithms utilizing advanced analytics such as natural language processing and machine learning to mine free-text medical records merit investigation. Potential avenues include integrating corroboratory data such as medications and procedures, and temporal and spatial coding patterns. Future work should investigate the optimal gold standard including rhythm monitoring, electronic data sources and chart review. The reasons for false positives and negatives need to be explored in detail, as does the impact of AF phenotype and coding position.

Finally, the accuracy of embolic and bleeding risk factor case definitions requires further validation in order to adjudicate appropriateness of anticoagulation management choices.

Conclusion

The overall accuracy of AF identification was reasonable for system planning and surveillance of prevalence, quality and outcomes. However, there is a marked disconnect between the volume of publications in these domains, and those examining the underpinning data. Sensitivity and PPV were the least accurate parameters with greatest uncertainty in terms of evidence and interpretation. This potentially underestimates the burden of disease and may bias estimates of outcomes and treatment quality. The optimal AF case definition should consider the purpose of the study and the data sources available. Health service administrators, researchers and clinicians should be mindful of these factors, and work together to refine our use of electronic data.

Abbreviations

AF, atrial fibrillation; AFL, atrial flutter; ICD, International Classification of Diseases; NPV, negative predictive value; PPV, positive predictive value; Sn, sensitivity; Sp, specificity.

Author contributions

All authors contributed toward data analysis, drafting and revising the paper, gave final approval of the version to be published and agree to be accountable for all aspects of the work.

Disclosure

References

1. Andrade JG, Verma A, Mitchell LB, et al. 2018 focused update of the Canadian Cardiovascular Society guidelines for the management of atrial fibrillation. *Can J Cardiol*. 2018;34:1371–1392. doi:10.1016/j.cjca.2018.08.026

2. Andrade J, Khairy P, Dobrev D, Nattel S. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ Res*. 2014;114:1453–1468. doi:10.1161/CIRCRESAHA.114.303211

3. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65:343–349 e342. doi:10.1016/j.jclinepi.2011.09.002

4. Navar-Boggan AM, Rymer JA, Piccini JP, et al. Accuracy and validation of an automated electronic algorithm to identify patients with atrial fibrillation at risk for stroke. *Am Heart J*. 2015;169:39–44 e32. doi:10.1016/j.ahj.2014.09.014

5. Jensen PN, Johnson K, Floyd J, et al. A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:141–147. doi:10.1002/pds.2317

6. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.

7. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629–634. doi:10.1136/bmj.315.7109.629

8. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50:1088–1101.

9. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560. doi:10.1136/bmj.327.7414.557

10. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36:1–48. doi:10.18637/jss.v036.i03

11. Tu K, Nieuwlaat R, Cheng SY, et al. Identifying patients with atrial fibrillation in administrative data. *Can J Cardiol*. 2016;32:1561–1565. doi:10.1016/j.cjca.2016.06.006

12. Sung SF, Hsieh CY, Lin HJ, et al. Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. *Int J Cardiol*. 2016;215:277–282. doi:10.1016/j.ijcard.2016.04.069

13. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual*. 2004;19:201–206. doi:10.1177/106286060401900504

14. Ruigomez A, Johansson S, Wallander MA, Rodriguez LA. Incidence of chronic atrial fibrillation in general practice and its treatment pattern. *J Clin Epidemiol*. 2002;55:358–363.

15. Baturova MA, Lindgren A, Carlson J, Shubik YV, Bertil Olsson S, Platonov PG. Atrial fibrillation in patients with ischaemic stroke in the Swedish national patient registers: how much do we miss? *Europace*. 2014;16:1714–1719. doi:10.1093/europace/euu165

16. 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;285:2370–2375. doi:10.1001/jama.285.18.2370

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;158:111–117. doi:10.1016/j.ahj.2009.05.010

18. Go AS, Hylek EM, Phillips KA, et al. Implications of stroke risk criteria on the anticoagulation decision in nonvalvular atrial fibrillation: the Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Circulation*. 2000;102:11–13. doi:10.1161/01.cir.102.1.11

19. Thigpen JL, Dillon C, Forster KB, et al. 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. doi:10.1161/CIRCOUTCOMES.113.000371

20. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005;36:1776–1781. doi:10.1161/01.STR.0000174293.17959.a1

21. White E. The effect of misclassification of disease status in follow-up studies: implications for selecting disease classification criteria. *Am J Epidemiol*. 1986;124:816–825. doi:10.1093/oxfordjournals.aje.a114458

22. Chotchaisuwatana S, Jedsadayanmata A, Chaiyakunapruk N, Jampachaisri K. Validation of electronic medical database in patients with atrial fibrillation in community hospitals. *J Med Assoc Thai*. 2011;94:686–692.

23. Glazer NL, Dublin S, Smith NL, et al. Newly detected atrial fibrillation and compliance with antithrombotic guidelines. *Arch Intern Med*. 2007;167:246–252. doi:10.1001/archinte.167.3.246

24. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet*. 2018;391:1775–1782. doi:10.1016/S0140-6736(18)30668-8

25. Macle L, Cairns J, Leblanc K, et al. 2016 focused update of the Canadian Cardiovascular Society guidelines for the management of atrial fibrillation. *Can J Cardiol*. 2016;32:1170–1185. doi:10.1016/j.cjca.2016.07.591

26. Kucharska-Newton AM, Heiss G, Ni H, et al. Identification of heart failure events in medicare claims: the Atherosclerosis Risk in Communities (ARIC) study. *J Card Fail*. 2016;22:48–55. doi:10.1016/j.cardfail.2015.07.013

27. Shoaib A, Farag M, Nasir M, et al. Is the diagnostic coding position of acute heart failure related to mortality? A report from the Euro Heart Failure Survey-1. *Eur J Heart Fail*. 2016;18:556–563. doi:10.1002/ejhf.505

28. Frost L, Andersen LV, Vestergaard P, Husted S, Mortensen LS. Trend in mortality after stroke with atrial fibrillation. *Am J Med*. 2007;120:47–53. doi:10.1016/j.amjmed.2005.12.027

29. Norberg J, Backstrom S, Jansson JH, Johansson L. Estimating the prevalence of atrial fibrillation in a general population using validated electronic health data. *Clin Epidemiol*. 2013;5:475–481. doi:10.2147/CLEP.S53420

30. 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:e012832. doi:10.1136/bmjopen-2016-012832

31. 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;46:149–153. doi:10.3109/14017431.2012.673728

32. Frost L, Vestergaard P. Caffeine and risk of atrial fibrillation or flutter: the Danish diet, cancer, and health study. *Am J Clin Nutr*. 2005;81:578–582. doi:10.1093/ajcn/81.3.578

33. Smith JG, Platonov PG, Hedblad B, Engstrom G, Melander O. Atrial fibrillation in the Malmo Diet and Cancer study: a study of occurrence, risk factors and diagnostic validity. *Eur J Epidemiol*. 2010;25:95–102. doi:10.1007/s10654-009-9404-1

34. Brophy MT, Snyder KE, Gaehde S, et al. Anticoagulant use for atrial fibrillation in the elderly. *J Am Geriatr Soc*. 2004;52:1151–1156. doi:10.1111/j.1532-5415.2004.52314.x

35. Ruigomez A, Johansson S, Wallander MA, Garcia Rodriguez LA. Predictors and prognosis of paroxysmal atrial fibrillation in general practice in the UK. *BMC Cardiovasc Disord*. 2005;5:20. doi:10.1186/1471-2261-5-17

36. Quan H, Parsons GA, Ghali WA. Assessing accuracy of diagnosis-type indicators for flagging complications in administrative data. *J Clin Epidemiol*. 2004;57:366–372. doi:10.1016/j.jclinepi.2003.01.002

37. Shen AY, 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;39:2736–2743. doi:10.1161/STROKEAHA.107.508580

38. Shireman TI, Howard PA, Kresowik TF, Ellerbeck EF. Combined anticoagulant-antiplatelet use and major bleeding events in elderly atrial fibrillation patients. *Stroke*. 2004;35:2362–2367. doi:10.1161/01.STR.0000141933.75462.c2

39. Munkholm SB, Jakobsen CJ, Mortensen PE, Lundbye-Christensen S, Andreasen JJ. Validation of post-operative atrial fibrillation in the Western Denmark Heart Registry. *Dan Med J*. 2015;62:A5162.

40. Walkey AJ, Wiener RS, Ghobrial JM, Curtis LH, Benjamin EJ. Incident stroke and mortality associated with new-onset atrial fibrillation in patients hospitalized with severe sepsis. *JAMA*. 2011;306:2248–2254. doi:10.1001/jama.2011.1615

41. Hravnak M, Hoffman LA, Saul MI, et al. Atrial fibrillation: prevalence after minimally invasive direct and standard coronary artery bypass. *Ann Thorac Surg*. 2001;71:1491–1495. doi:10.1016/s0003-4975(01)02477-8

Supplementary materials

Table S1 PRISMA checklist

Section/topic	#	Checklist item	Report page #
Title			
Title	1	Identify the report as a systematic review, meta-analysis or both.	1
Abstract			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes and study design (PICOS).	4
Methods			
Protocol	5	Indicate if a review protocol exists, if and where it can be accessed (eg, Web address), and, if available, provide registration information including registration number.	5
Eligibility	6	Specify study characteristics (eg, PICOS, length of follow-up) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Selection	9	State the process for selecting studies (ie, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Collection	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (eg, PICOS, funding sources) and any assumptions and simplifications made.	5
Bias in studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (eg, risk ratio, difference in means).	5
Synthesis	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (eg, I^2) for each meta-analysis.	6
Bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A

(Continued)

Table S1 (Continued).

Section/topic	#	Checklist item	Report page #
Results			
Selection	17	Give numbers of studies screened, assessed for eligibility and included in the review, with reasons for exlusions at each stage, ideally with a flow diagram.	Figure S1
Characteristics	18	For each study, present characteristics for which data were extracted (eg, study size, PICOS, follow-up period) and provide the citations.	Table 1
Bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 1
Synthesis	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7,8
Bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional	23	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, meta-regression).	N/A
Discussion			
Summary	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (eg, health care providers, users and policy makers).	8–10
Limitations	25	Discuss limitations at study and outcome level (eg, risk of bias), and at review level (eg, incomplete retrieval of identified research, reporting bias).	11
Conclusion	26	Provide a general interpretation of the results in the context of other evidence, implications future research.	11
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (eg, supply of data); role of funders for the systematic review.	13

Table S2 Search strategy

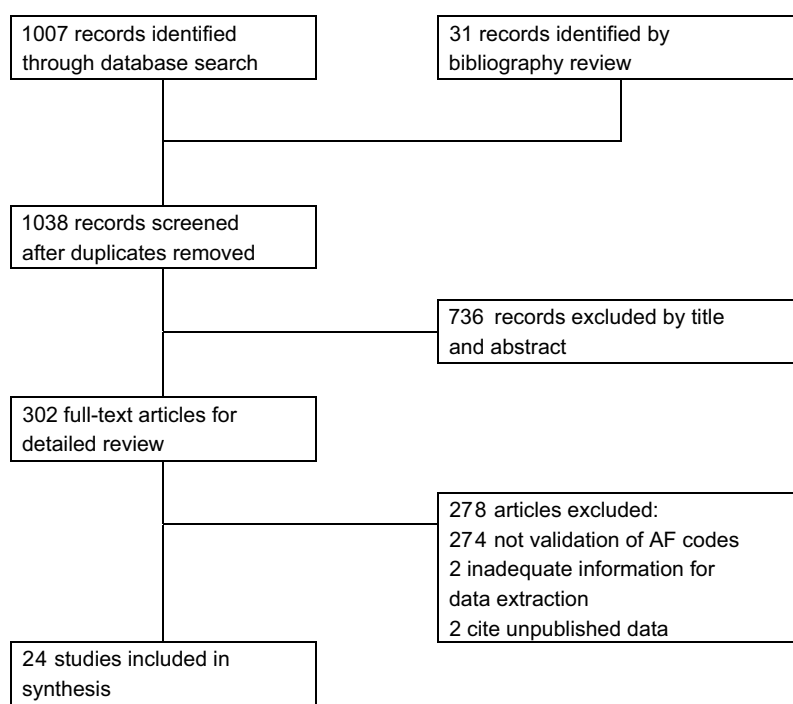| |
|---|
| 1) *atrial fibrillation |
| 2) Atrial fibrillation.ti,ab. |
| 3) or/1–2 |
| 4) *registries/ or *records as topic/ or *databases, factual/ or *database management systems/ or *epidemiologic studies/ |
| 5) (administrative or registr* or database* or claims or health maintenance organization or population-based).ti,ab. |
| 6) or/4–5 |
| 7) *validation studies/ or *data accuracy/ or *predictive value of tests/ |
| 8) (sensitivity or specificity or predictive value or accuracy or abstract* or identif*).ti,ab. |
| 9) or/7–8 |
| 10) 3 and 6 and 9 |
| 11) Limit 10 to humans |
| 12) Limit 11 to english language |
| 13) Limit 12 to yr ="2000-Current" |
| 14) 13 not exp newborn/ not exp infant/ not exp child/ not exp adolescent/ |
| 15) 14 not (comment or editorial or note or letter or interview or lectures or personal narratives or biography or autobiography or addresses or patient education handout or interactive tutorial or news or newspaper article or historical article or webcasts or video-audio media or portraits or twin study or retraction of publication or retracted publication or published erratum or duplicate publication or case reports or legal cases or guideline or conference abstract or English abstract or clinical conference or congresses or meta-analysis or randomized controlled trial or clinical trial or clinical trial, phase I or clinical trial, phase II or clinical trial, phase III or clinical trial, phase IV or controlled clinical trial).pt. |
| 16) 15 not catheter ablation/ not transcatheter aortic valve replacement/ not septal occluder device/ not antibodies, monoclonal/ |
| 17) 16 not (ablation or pulmonary vein or Amplat* or watchman or pacemaker or defibrillator or single-chamber or dual-chamber or resynchronization or gene or genes or genet* or ibrutinib).ti,ab. |

```
┌─────────────────────────┐        ┌─────────────────────────┐
│ 1007 records identified │        │ 31 records identified by│
│ through database search │        │ bibliography review     │
└─────────────────────────┘        └─────────────────────────┘

          ┌─────────────────────────┐
          │ 1038 records screened   │
          │ after duplicates removed│
          └─────────────────────────┘
                                    ┌─────────────────────────┐
                                    │ 736  records excluded by │
                                    │ title and abstract      │
                                    └─────────────────────────┘
          ┌─────────────────────────┐
          │ 302 full-text articles for│
          │ detailed review         │
          └─────────────────────────┘
                                    ┌─────────────────────────┐
                                    │ 278 articles excluded:  │
                                    │ 274 not validation of AF│
                                    │ codes                   │
                                    │ 2 inadequate information│
                                    │ for data extraction     │
                                    │ 2 cite unpublished data │
                                    └─────────────────────────┘
          ┌─────────────────────────┐
          │ 24 studies included in  │
          │ synthesis               │
          └─────────────────────────┘
```

Figure S1 Flow diagram of study selection.