JACC: ADVANCES VOL. 2, NO. 8, 2023

ORIGINAL RESEARCH

RHYTHM DISORDERS AND ELECTROPHYSIOLOGY

Performance of Single-Lead Handheld Electrocardiograms for Atrial Fibrillation Screening in Primary Care

The VITAL-AF Trial

Shaan Khurshid, MD, MPH,[a,b] Yuchiao Chang, PhD,[c,d] Leila H. Borowsky, MPH,[c,d] David D. McManus, MD,[e] Jeffrey M. Ashburner, PhD, MPH,[c,d] Steven J. Atlas, MD, MPH,[c,d] Patrick T. Ellinor, MD, PhD,[a,b] Daniel E. Singer, MD,[c,d] Steven A. Lubitz, MD, MPH[a,b]

ABSTRACT

BACKGROUND Handheld single-lead electrocardiographic (1L ECG) devices are increasingly used for atrial fibrillation (AF) screening, but their real-world performance is not well understood.

OBJECTIVES The purpose of this study was to quantify the diagnostic test characteristics of 1L ECG automated interpretations for prospective AF screening.

METHODS We calculated the diagnostic test characteristics of the AliveCor KardiaMobile 1L ECG (AliveCor, US) algorithm using unblinded cardiologist overread as the gold standard using single 30s tracings administered by medical assistants among individuals aged ≥65 years participating in the VITAL-AF trial (NCT03515057) of population-based AF screening embedded within routine primary care.

RESULTS A total of 14,230 individuals (mean age 74 ± 7 years, 60% women, 82% White) had 31,376 tracings reviewed by 13 cardiologists. A total of 24,906 (79.6%) tracings had an AliveCor interpretation of *normal*, 5,046 (16.1%) were *unclassified*, 797 (2.5%) were *possible AF*, and 573 (1.8%) were *no analysis*. Cardiologists read 808 (2.6%) tracings as AF. AliveCor *possible AF* had a PPV of 51.7% (95% CI: 47.8%-55.6%). AliveCor *normal* had an NPV of 99.8% (95% CI: 99.7%-99.8%). The AliveCor algorithm had an overall sensitivity of 51.0% (95% CI: 47.1%-54.9%) and a specificity of 98.7% (95% CI: 98.6%-98.9%). AliveCor tracings interpreted as *unclassified* (PPV 5.9%, 95% CI: 5.1%-6.7%) and *no analysis* (PPV 6.5%, 95% CI: 4.6%-8.9%) had low predictive values for AF and were increasingly prevalent at older ages (13.7% for age 65-69 years to 28.1% for age ≥85 years, $P < 0.01$).

CONCLUSIONS In an older primary care population undergoing AF screening with handheld 1L ECGs, automated algorithm interpretations were sufficiently accurate to exclude the presence of AF but not to establish an AF diagnosis. (JACC Adv 2023;2:100616) © 2023 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; [b]Cardiology Division, Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, Massachusetts, USA; [c]Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; [d]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; and the [e]Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2023.100616

ABBREVIATIONS AND ACRONYMS

1L ECG = single-lead electrocardiogram

AF = atrial fibrillation

CHARGE-AF = Cohorts for Aging and Genomic Epidemiology Atrial Fibrillation

ECG = electrocardiogram

MGH = Massachusetts General Hospital

NPV = negative predictive value

PPV = positive predictive value

Undiagnosed atrial fibrillation (AF) is an important cause of stroke. AF screening can detect undiagnosed AF,[1-4] and use of oral anticoagulants can prevent strokes. However, oral anticoagulants also increase bleeding risk, and AF screening may lead to preferential detection of very low-burden AF, whose association with stroke risk is not well understood.[5] Therefore, it remains unclear whether population-based screening for AF leads to net clinical benefit.[2] As a result, consensus guidelines offer conflicting endorsements, with cardiology societies from Europe[6] and Australia/New Zealand[7] providing a Class I recommendation for AF screening among individuals aged >65 and ≥65 years, respectively, but the United States Preventive Services Task Force concluding that there is insufficient evidence for or against AF screening.[8] Newer technologies, such as single-lead electrocardiograms (1L ECGs) obtained using mobile devices[2] that are marketed directly to consumers and may be used in real-world care represent a contemporary approach to screening.[9] However, the ability of such devices to facilitate accurate AF detection is not well understood.

Most 1L ECG sensors employ automated AF detection algorithms, which may enable more resource-efficient screening by reducing the need for manual interpretation.[2] Although current AF detection algorithms demonstrate reasonable accuracy in small testing samples enriched for individuals with known AF,[10] their performance when deployed prospectively for population-based screening in clinical practice is not well understood. Additionally, automated algorithms may return equivocal interpretations (eg, unclassified), whose frequency and potential clinical implications have not been adequately described. A better understanding of the performance of 1L ECG-based automated AF detection is critical to inform future AF screening efforts and define the extent to which automated 1L ECG interpretations require manual clinician overreading or confirmatory follow-up testing.

VITAL-AF was a cluster-randomized trial (NCT03515057) that assessed the efficacy of AF screening with 30-second 1L ECGs among primary care patients aged ≥65 years. In the current analysis, we quantified the diagnostic test characteristics of the AliveCor automated 1L ECG interpretation on tracings taken during AF screening among individuals without prevalent AF, using manual cardiologist review as the gold standard.

METHODS

TRIAL DESIGN AND PARTICIPANTS. The design, conduct, and primary outcome results of VITAL-AF have been published previously.[3,11] Briefly, VITAL-AF recruited patients from 16 primary care practices within the Massachusetts General Hospital (MGH) Practice-Based Research Network. VITAL-AF was a pragmatic cluster randomized trial, in which practices were randomized in a 1:1 ratio to AF screening versus usual care. Patients were eligible for inclusion if they were aged ≥65 years and attended an outpatient clinic appointment at a participating primary care practice with a primary care physician, nurse practitioner, or physician's assistant. The trial enrolled patients between July 31, 2018 and October 8, 2019. In the current analysis, we specifically focus on individuals who had ≥1 1L ECG screening performed after randomization to a screening practice (ie, the *per-protocol* screening population) who did not have a prior diagnosis of AF at the time of screening. The research protocol was approved by the Mass General Brigham Institutional Review Board.

AF SCREENING INTERVENTION. Eligible individuals visiting intervention practices were offered AF screening with the AliveCor Kardia (AliveCor, US) 1L ECG at each encounter during the assessment of vital signs. The 1L ECG was administered by practice medical assistants who received dedicated training in the use of the Kardia device prior to study start as well as monthly refreshers. Each tracing underwent automated interpretation using the AliveCor AF detection algorithm (KardiaAI version 1), resulting in one of 5 possible interpretations (Supplemental Table 1). In addition to the automated interpretation generated by the AliveCor device, all 1L ECG tracings were later reviewed by one of the 13 study cardiologists. Cardiologists reviewing 1L ECG tracings were not blinded to the AliveCor automated interpretation, mirroring the usual process of clinical ECG reading in which an automated interpretation is typically provided.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received March 10, 2023; revised manuscript received June 22, 2023, accepted July 16, 2023.

Although not required by the study protocol, 12-lead ECGs could be performed at clinician's discretion. When obtained, 12-lead ECGs were adjudicated by clinical cardiologists at MGH as part of routine clinical workflows. In addition to review by a study cardiologist, potentially actionable 1L ECG findings (eg, new AF, sinus arrest ≥3 seconds) were also reviewed by study electrophysiologists and triggered notification of the primary care provider. Potentially actionable findings are listed in Supplemental Table 2.

ASSESSMENT OF 1L ECG AUTOMATED INTERPRETATION PERFORMANCE. For the primary analysis, we assessed the AliveCor 1L ECG automated algorithm's performance using cardiologist review as the gold standard. For the purposes of this analysis, we collapsed the 5 observed AliveCor interpretations into 4 classes, as follows: AF = *possible AF*; normal = *normal*; unclassified = *unclassified*; and too short or noise = *no asnalysis* (Supplemental Table 1). Since cardiologist review of 1L ECG tracings may represent an imperfect reference standard,[12] we performed secondary analyses in which we compared the 1L ECG automated algorithm interpretation to: 1) a subset of tracings reviewed by 2 board-certified cardiac electrophysiologists (separate from the cardiologist reviewers); and 2) the 12-lead (12L) ECG interpretation for the n = 2,230 individuals who had a same-day 12L ECG performed. We also assessed the effect by reversing progressively greater fractions of cardiologist interpretations. In accordance with VITAL-AF study protocols, all same-day 12L ECGs were performed at the discretion of treating clinicians (ie, were not part of the trial protocol) and were interpreted by clinical cardiologists at MGH separate from the VITAL-AF cardiologist reviewers. An overview of the study design is provided in the (**Central Illustration**).

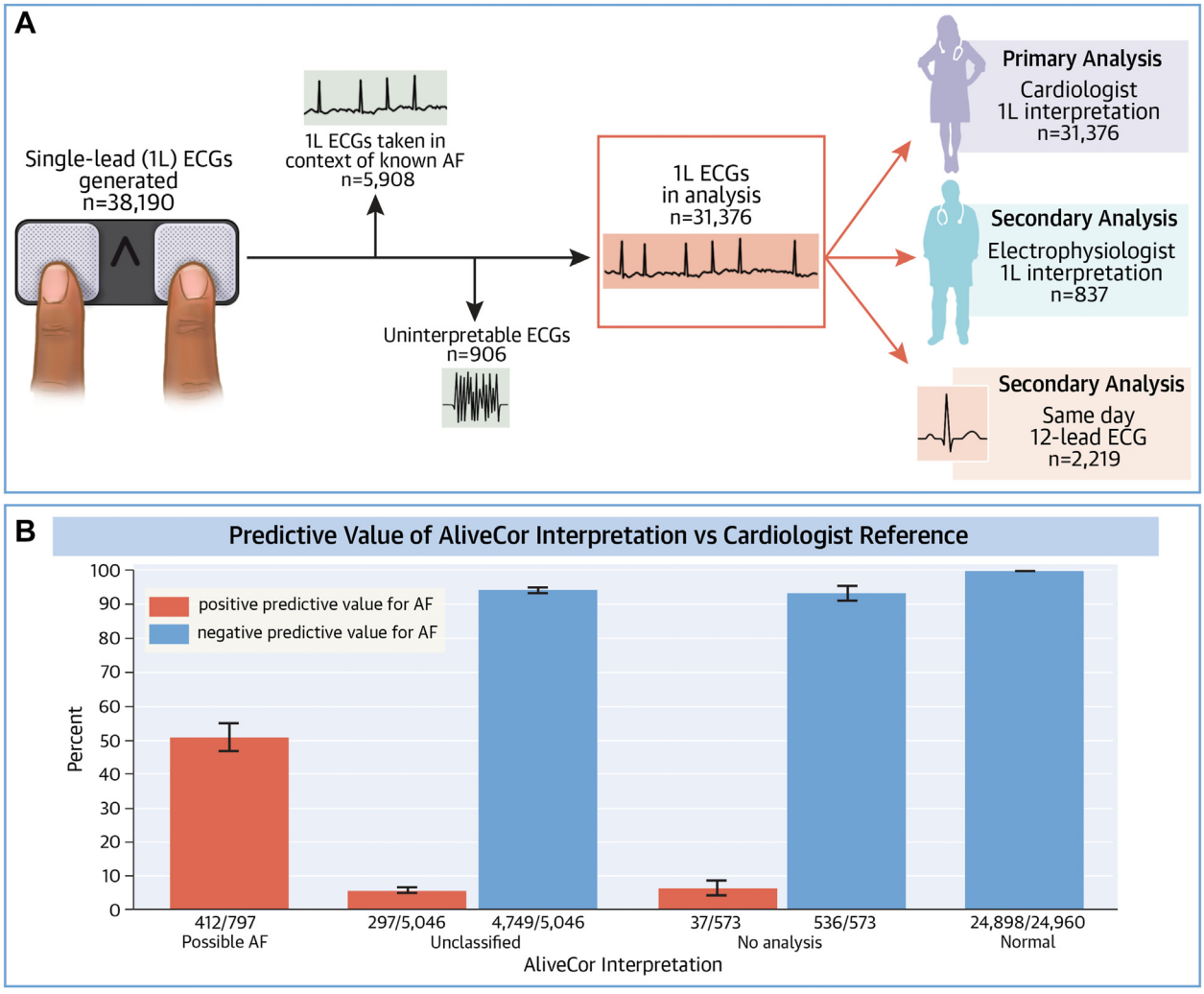
STATISTICAL ANALYSIS. To quantify the performance of the 1L ECG automated algorithm interpretation for detecting AF, we calculated the positive predictive value (PPV) and the negative predictive value (NPV) for each AliveCor reading. For this analysis, a cardiologist's review result of AF was considered AF, while reviews not indicating AF (ie, normal or other) were classified as non-AF. We also tabulated the distribution of 1L automated interpretations among tracings read as AF and separately tracings read as not AF by the cardiologist reviewer. For calculation of 1L ECG automated algorithm sensitivity, specificity, and predictive values, we decided *a priori* to analyze equivocal interpretations (ie, *unclassified* and *no analysis*) in 2 ways. First, we considered equivocal interpretations as an AliveCor non-AF interpretation (ie, a negative test result), therefore defining only AliveCor *possible AF* tracings as AF. Second, we considered equivocal interpretations as an AliveCor AF interpretation (ie, a positive test result), therefore defining AliveCor *possible AF*, *unclassified*, or *no analysis* tracings as AF by the automated algorithm.[12] Since the true frequency of AF on equivocal tracings is unknown, we performed secondary analyses in which 25%, 50%, and 75% of equivocal tracings were randomly assigned as positive test results, as well as an analysis in which equivocal tracings were excluded.

While administering the 1L ECG, medical assistants were instructed to repeat the screening procedure once if the results of the initial screen were either *unclassified* or *no analysis*. Therefore, since individuals could have multiple tracings performed in the context of a single encounter, in the primary analysis we calculated test characteristics for individual tracing results, but in a secondary analysis we quantified test characteristics at an encounter level (see below). The generalized estimating equations with independent working correlation structure approach was used to account for the repeated measures data structure (multiple tracings per individual) when estimating the 95% CIs for test characteristics. Since the within-cardiologist intraclass correlation was low, we did not make any adjustment for clustering by cardiologist.

To assess whether the AliveCor algorithm's performance may differ on the basis of key sociodemographic factors or clinically estimated AF risk, we repeated the primary analysis within subgroups divided by age (ie, 65-70 years, 70-75 years, 75-80 years, 80-85 years, and ≥85 years), sex, race (White or other race), and categories of predicted 5-year AF risk estimated using the Cohorts for Aging and Genomic Epidemiology Atrial Fibrillation (CHARGE-AF) score (ie, <5%, 5%-10%, 10%-25%, 25%-50%, and ≥50%). To calculate predicted 5-year AF probabilities, CHARGE-AF scores were converted to probabilities using the equation: $1-0.9718412736^{\exp(\Sigma\beta X-12.5815600)}$, where $\Sigma\beta X$ is the individual's CHARGE-AF score.[13]

We performed several sensitivity analyses to assess the robustness of our findings. First, we repeated the main analyses described above using: 1) electrophysiologist review; and 2) cardiologist interpretation of a 12-lead ECG performed the same day as alternative reference standards (**Central Illustration**). Second, we assessed 1L ECG performance on a per-encounter (as opposed to per-tracing

CENTRAL ILLUSTRATION **Study Overview and Predictive Values of AliveCor Algorithm Interpretations**



Khurshid S, et al. JACC Adv. 2023;2(8):100616.

(A) An overview of the current study. A total of 38,190 handheld 1L ECG tracings were generated in the context of screening within the VITAL-AF trial. Of these, we excluded tracings performed among individuals with known AF at the time of screening, as well as those overread as uninterpretable by cardiologist readers, resulting in 31,376 tracings in the primary analysis (see text). We quantified the test characteristics of the AliveCor 1L ECG algorithm against cardiologist overread (primary analysis), against electrophysiologist overread (secondary analysis), and against clinical interpretation of a 12-lead ECG performed on the same day as screening (secondary analysis). (B) The positive predictive values (PPVs) (red) and negative predictive values (NPVs) (green) for the 4 AliveCor automated 1L ECG interpretations (x-axis), using cardiologist overread as the gold standard. Only PPV is depicted for *possible AF*, which always denotes a positive test, while only NPV is depicted for *normal* which always denotes a negative test. Since *unclassified* and *no analysis* represent equivocal findings, both PPV and NPV are shown. AF = atrial fibrillation; ECG = electrocardiogram.

level), utilizing the last tracing taken during a study visit in which screening was performed. Third, we calculated test characteristics when reversing 0.5%, 1%, 1.5%, 2%, and 5% of cardiologist 1L ECG interpretations to account for expected misclassification rates.[14]

We considered 2-sided P values <0.05 to indicate statistical significance. All analyses were performed

TABLE 1 Baseline Characteristics of the Primary Analysis Sample (N = 14,230)[a]

Age	73.9 ± 6.7
Female	8,474 (59.6%)
Race	
White	11,701 (82.2%)
Black	769 (5.4%)
Hispanic	321 (2.3%)
Other	1,168 (8.2%)
Unknown	271 (1.9%)
Active smoker	644 (4.6%)
Systolic blood pressure, mm Hg	131 ± 16
Diastolic blood pressure, mm Hg	75 ± 9
Body mass index, kg/m^2	27.8 ± 5.4
Antihypertensive medication use	7,566 (53.2%)
Hypertension	10,756 (75.6%)
Diabetes	3,450 (24.2%)
Myocardial infarction	1,007 (7.1%)
Coronary heart disease	2,989 (21.0%)
Heart failure	1,467 (10.3%)
Stroke	1,177 (8.3%)
Vascular disease[b]	2,768 (19.5%)
CHA_2DS_2-VASc score	3.4 ± 1.4
CHARGE-AF score	13.5 ± 0.9

Values are mean ± SD or n (%). [a]All characteristics obtained at study enrollment. [b]Defined as myocardial infarction, peripheral arterial disease, or aortic plaque.

CHARGE-AF = Cohorts for Aging and Research in Genomic Epidemiology-Atrial Fibrillation.[13]

TABLE 2 Distribution of AliveCor Algorithm Interpretations Using Cardiologist Overread as Reference Overall and Stratified by Age

		Cardiologist Interpretation	
	Alivecor Interpretation	**AF (n = 808)**	**No AF (n = 30,568)**
Overall	Possible AF	412 (51.0%)	385 (1.3%)
	Normal	62 (7.7%)	24,898 (81.5%)
	Unclassified	297 (36.8%)	4,749 (15.5%)
	No analysis	37 (4.6%)	536 (1.8%)
		Cardiologist Interpretation	
		AF (n = 123)	**No AF (n = 9,801)**
Age 65-69 y	Possible AF	54 (43.9%)	65 (0.7%)
	Normal	10 (8.1%)	8,435 (86.1%)
	Unclassified	50 (40.7%)	1,161 (11.8%)
	No analysis	9 (7.3%)	140 (1.4%)
		Cardiologist Interpretation	
		AF (n = 174)	**No AF (n = 8,537)**
Age 70-74 y	Possible AF	93 (53.4%)	82 (1%)
	Normal	17 (9.8%)	7,059 (82.7%)
	Unclassified	60 (34.5%)	1,263 (14.8%)
	No analysis	4 (2.3%)	133 (1.6%)
		Cardiologist Interpretation	
		AF (n = 187)	**No AF (n = 6,099)**
Age 75-79 y	Possible AF	88 (47.1%)	85 (1.4%)
	Normal	14 (7.5%)	4,911 (80.5%)
	Unclassified	75 (40.1%)	990 (16.2%)
	No analysis	10 (5.3%)	113 (1.9%)
		Cardiologist Interpretation	
		AF (n = 154)	**No AF (n = 3,638)**
Age 80-84 y	Possible AF	79 (51.3%)	72 (2%)
	Normal	8 (5.2%)	2,771 (76.2%)
	Unclassified	60 (39%)	709 (19.5%)
	No analysis	7 (4.5%)	86 (2.4%)
		Cardiologist Interpretation	
		AF (n = 170)	**No AF (n = 2,493)**
Age ≥85 y	Possible AF	98 (57.6%)	81 (3.2%)
	Normal	13 (7.6%)	1,722 (69.1%)
	Unclassified	52 (30.6%)	626 (25.1%)
	No analysis	7 (4.1%)	64 (2.6%)

Value are n (%).

AF = atrial fibrillation.

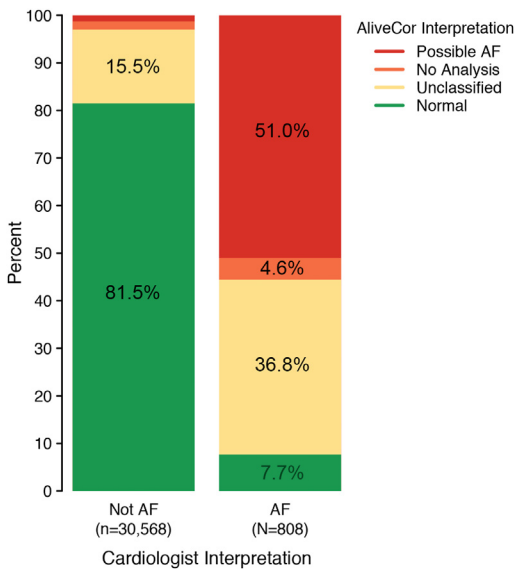
using R v4.0[15] and SAS v9.4 (SAS Institute, Cary NC).

RESULTS

Among 18,199 unique patients with visits to intervention practices during the study period, 16,496 individuals had at least one 1L ECG tracing (median 2 per individual, quartile 1: 1, quartile 3: 3), with a total of 38,190 1L ECG tracings generated. We excluded 5,908 tracings (15.5%) generated among individuals with prior AF known at the time of screening (Supplemental Table 3) as well as an additional 906 tracings (2.4%) manually reviewed as uninterpretable, resulting in 31,376 tracings representing 14,230 individuals and 29,497 encounters in the primary analysis (**Central Illustration**). Of the 906 tracings excluded for a manual review of uninterpretable, 54% had an AliveCor reading of *no analysis* (Supplemental Table 4). The mean age was 73.9 years, and 59.6% were female. Other baseline characteristics are listed in **Table 1**. The median time from screening to overread was 10 hours (quartile 1: 5, quartile 3: 18), and each reader interpreted a median of 2,414 (quartile 1: 1,974, quartile 3: 2,678) tracings.

Of the 31,376 tracings in the primary analysis, 808 (2.6%) were read by a cardiologist as AF. A total of 24,960 (79.6%) tracings had an AliveCor interpretation of *normal*, 5,046 (16.1%) *unclassified*, 797 (2.5%) *possible AF*, and 573 (1.8%) *no analysis*. The distribution of AliveCor automated interpretations stratified by cardiologist overread is shown in **Table 2** and **Figure 1**.

FIGURE 1 Distribution of AliveCor Algorithm Interpretations Stratified by Cardiologist Overread



Depicted is the distribution of AliveCor interpretations (see legend), stratified by cardiologist overread as AF (right bar) vs non-AF (left bar). The proportion of tracings represented by each AliveCor interpretation within each stratum of cardiologist overread is depicted within the corresponding bar, with the exception of not AF–*possible AF* (1.3%) and not AF–*no analysis* (1.8%), which are unlabeled for graphical purposes. AF = atrial fibrillation.

When equivocal interpretations (ie, *unclassified* or *no analysis*) were considered negative results (ie, no evidence of AF), the sensitivity of the automated algorithm for detecting AF was 51.0% (95% CI: 47.1%-54.9%) and the specificity was 98.7% (95% CI: 98.6%-98.9%). The PPV of *possible AF* was 51.7% (95% CI: 47.8%-55.6%), and the NPV was 99.8% (95% CI: 99.7%-99.8%). Assessed individually, PPV for *unclassified* (5.9%, 95% CI: 5.1%-6.7%) and *no analysis* (6.5%, 95% CI: 4.6%-8.9%) was low but not negligible (**Central Illustration**). When equivocal interpretations were considered positive test results, test characteristics changed substantially. The sensitivity of a positive result increased to 92.3% (95% CI: 90.3%-94.0%), and the specificity decreased to 81.5% (95% CI: 80.8%-82.1%). The PPV of a positive test decreased to 11.6% (95% CI: 10.7%-12.7%), and NPV was stable at 99.8% (95% CI: 99.7%-99.8%) (**Table 3**). Test characteristics were intermediate when 25%, 50%, and 75% of equivocal tracings were considered positive tests (Supplemental Table 5). Sensitivity was 86.9% (95% CI: 83.5%-89.7%), specificity was 98.5% (95% CI: 98.3%-98.6%), PPV was 51.7% (95% CI: 47.8%-55.6%), and NPV was 99.8% (95% CI: 99.7%-99.8%) when equivocal tracings were excluded.

In general, the frequency of equivocal results increased with older age among screened patients (n = 1,360 [13.7%] for age 65-69 years to n = 749 [28.1%] for age ≥85 years, $P < 0.01$) (**Figure 2**). Despite an increasing AF prevalence, the proportion of equivocal tracings with a cardiologist review result of AF remained generally similar across strata of age (roughly 4%-8%), and therefore the higher frequency of equivocal results observed among older individuals appeared driven largely by a greater number of equivocal tracings read as non-AF (**Figure 2**, Supplemental Figure 1). As a result, the specificity of the AliveCor algorithm generally decreased with age, particularly when equivocal results were considered positive tests (**Figure 3**). In contrast, the sensitivity of the algorithm was generally stable with greater age. Consistent with a rising prevalence of AF with older age, the PPV of the algorithm increased and the NPV decreased with age (**Figure 3**).

In secondary analyses, patterns of variation in AliveCor diagnostic performance were similar across categories of predicted AF risk (Supplemental Tables 6 and 7), and performance was largely consistent across strata of sex and race (Supplemental Table 8). When assessing algorithm results against electrophysiologist review and same-day 12L ECG as alternative reference standards in subsets of tracings, the sensitivity of *possible AF* was moderately higher (72.8%, 95% CI: 66.3%-78.4%; 12L ECG 87.7%, 95% CI: 75.8%-94.2%), while the PPV of *possible* AF was lower (44.3%, 95% CI: 39.5%-49.3%; 12L ECG 14.5%, 95% CI: 11.2%-18.6%). Similar to cardiologist review, the NPV of *normal* was very good using both electrophysiologist review (92.7%, 95% CI: 79.6%-97.6%) and same-day 12L ECG (99.8%, 95% CI: 99.4%-100%) (Supplemental Tables 9 and 10). Results were similar when algorithm performance was assessed at the encounter level as opposed to the tracing level (Supplemental Tables 11 and 12). When progressively greater fractions of cardiologist interpretations were reversed, specificity, PPV, and NPV estimates remained largely stable even at reversal rates up to 5%. Although sensitivity estimates were more variable, the sensitivity of *possible AF* remained in the range of 20% to 50% (Supplemental Table 13).

DISCUSSION

In a primary care screening trial including approximately 14,000 individuals without known AF who underwent over 30,000 handheld 1L ECG tracings administered by trained medical assistants, we found

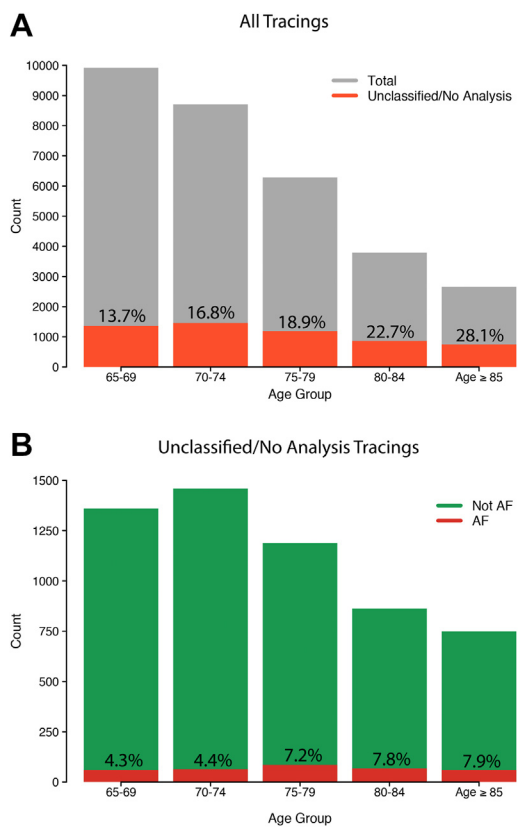
TABLE 3 Test Characteristics of AliveCor Algorithm Interpretations of 1L ECG Tracings Using Cardiologist Overread as Reference Overall and Stratified by Age

	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
Overall (N = 31,376)				
Possible AF	51.0% (47.1%-54.9%)		51.7% (47.8%-55.6%)	
Possible AF or unclassified or no analysis	92.3% (90.3%-94.0%)		11.6% (10.7%-12.7%)	
Normal		81.5% (80.8%-82.1%)		99.8% (99.7%-99.8%)
Normal or unclassified or no analysis		98.7% (98.6%-98.9%)		98.7% (98.5%-98.9%)
Unclassified			5.9% (5.1%-6.7%)	94.1% (93.3%-94.9%)
No analysis			6.5% (4.6%-8.9%)	93.5% (91.1%-95.4%)
Age 65-69 y (n = 9,924)				
Possible AF	43.9% (34.4%-53.8%)		45.4% (36.1%-55.0%)	
Possible AF or unclassified or no analysis	91.9% (85.4%-95.6%)		7.6% (6.1%-9.6%)	
Normal		86.1% (85.0%-87.1%)		99.9% (99.8%-99.9%)
Normal or unclassified or no analysis		99.3% (99.1%-99.5%)		99.3% (99.1%-99.5%)
Unclassified			4.1% (3.0%-5.7%)	95.9% (94.3%-97.0%)
No analysis			6.0% (3.0%-11.7%)	94.0% (88.3%-97.0%)
Age 70-74 y (n = 8,711)				
Possible AF	53.4% (45.1%-61.6%)		53.1% (44.6%-61.5%)	
Possible AF or unclassified or no analysis	90.2% (85.0%-93.8%)		9.6% (7.9%-11.6%)	
Normal		82.7% (81.4%-83.9%)		99.8% (99.6%-99.9%)
Normal or unclassified or no analysis		99.0% (98.8%-99.2%)		99.1% (98.8%-99.3%)
Unclassified			4.5% (3.3%-6.2%)	95.5% (93.8%-96.7%)
No analysis			2.9% (1.1%-7.4%)	97.1% (92.6%-98.9%)
Age 75-79 y (n = 6,286)				
Possible AF	47.1% (39.2%-55.1%)		50.9% (42.3%-59.4%)	
Possible AF or unclassified or no analysis	92.5% (87.5%-95.6%)		12.7% (10.5%-15.3%)	
Normal		80.5% (79.0%-82.0%)		99.7% (99.5%-99.8%)
Normal or unclassified or no analysis		98.6% (98.2%-98.9%)		98.4% (97.9%-98.7%)
Unclassified			7.0% (5.3%-9.2%)	93.0% (90.8%-94.7%)
No analysis			8.1% (4.3%-15.0%)	91.9% (85.0%-95.7%)
Age 80-84 y (n = 3,792)				
Possible AF	51.3% (42.4%-60.1%)		52.3% (43.6%-60.9%)	
Possible AF or unclassified or no analysis	94.8% (90.0%-97.4%)		14.4% (11.9%-17.3%)	
Normal		76.2% (73.9%-78.3%)		99.7% (99.4%-99.9%)
Normal or unclassified or no analysis		98.0% (97.5%-98.5%)		97.9% (97.3%-98.4%)
Unclassified			7.8% (5.9%-10.3%)	92.2% (89.7%-94.1%)
No analysis			7.5% (3.7%-14.8%)	92.5% (85.2%-96.3%)
Age ≥85 y (N = 2,663)				
Possible AF	57.6% (48.7%-66.1%)		54.7% (46.5%-62.7%)	
Possible AF or unclassified or no analysis	92.4% (87.3%-95.5%)		16.9% (14.1%-20.2%)	
Normal		69.1% (66.1%-71.9%)		99.3% (98.7%-99.6%)
Normal or unclassified or no analysis		96.8% (95.9%-97.4%)		97.1% (96.2%-97.8%)
Unclassified			7.7% (5.5%-10.6%)	92.3% (89.4%-94.5%)
No analysis			9.9% (4.6%-19.9%)	90.1% (80.1%-95.4%)

AF = atrial fibrillation.

that the 1L ECG automated interpretation had moderate PPV for new AF, with 52% of tracings identified as AF by the algorithm confirmed by cardiologist readers. An AliveCor interpretation of *normal* had excellent NPV, with only 0.2% of tracings ultimately overread as AF. Equivocal AliveCor interpretations were common, representing nearly 20% of tracings, and roughly 4% to 8% of such tracings were reviewed as AF. Age of screening was an important determinant of diagnostic accuracy, with older age associated with decreasing specificity of AliveCor algorithm, due in large part to a higher frequency of equivocal interpretations ultimately overread by cardiologists as non-AF. Although current US clinical practice guidelines do not specifically endorse AF screening,[8] use of consumer devices for arrhythmia detection is nevertheless growing increasingly prevalent in primary care.[16] In this context, our findings provide important evidence that 1L ECG tracings with an automated interpretation of

FIGURE 2 Frequency of *Unclassified* and *No Analysis* Tracings Stratified by Age of Screening



(A) The frequency of equivocal tracings (ie, *unclassified* or *no analysis*, orange) as a fraction of the total tracings taken (gray) within each subgroup of age (x-axis). (B) The frequency of tracings overread as AF (red) as a fraction of the total number of equivocal tracings (green) observed within each subgroup of age (x-axis). Percentages depict the minority fraction (ie, proportion of equivocal tracings in A and proportion of tracings read as AF in B). AF = atrial fibrillation.

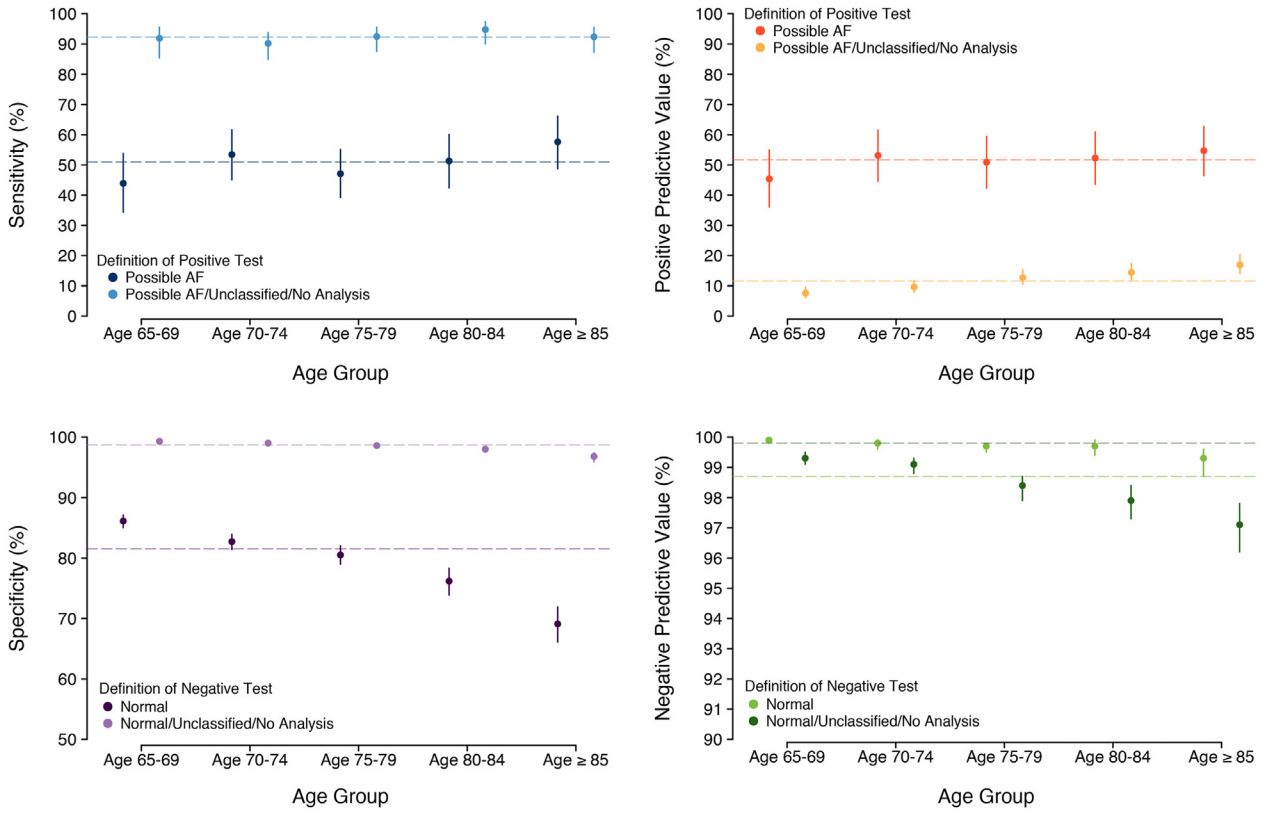
normal appear useful for ruling out the presence of AF, while all other interpretations merit confirmatory testing.

Our results extend previous findings by quantifying the diagnostic performance of a commonly used 1L ECG automated AF-detection algorithm when used in a large primary care screening setting.[6,7] Of note, multiple early studies of 1L ECG algorithms have assessed performance in small samples enriched for patients with known AF and with inconsistent handling of equivocal results, each of which may lead to inflated estimates of diagnostic performance. In a hospital-based sample of 200 patients (38 with known AF), Rajakariar et al[12] found that a wrist-worn wearable-based algorithm analyzing 30s 1L ECG tracings had a PPV of 54.8% for AF when unclassified tracings were treated as negative results, although they reported nearly 95% sensitivity against an immediately subsequent 12-lead ECG reference. Similarly, in a validation set of 204 patients (48 with known AF), Lau et al[14] observed 98% sensitivity and 97% specificity for AF using a handheld 1L ECG. In contrast to prior evaluations, our analysis included a large sample of primary care patients without known AF and therefore likely provides a more accurate estimation of how 1L ECG algorithms may perform when applied prospectively in clinical practice.

Our findings suggest that an automated 1L ECG interpretation of AF possesses a reasonable PPV for AF, but is not sufficiently accurate to replace manual overread or confirmatory testing. Overall, only 52% of tracings interpreted as AF by the automated algorithm were confirmed to show AF upon cardiologist review. As expected, we observed that the PPV of *possible AF* was influenced by pretest probability. For example, the PPV of *possible AF* increased from 45.4% among individuals aged 65 to 69 years to 54.8% among individuals aged ≥85 years. Given the effects of pretest probability on the predictive value of algorithmic screening, future work is warranted to assess whether focusing AF screening efforts on individuals estimated to be at high short-term AF risk (eg, based on clinical risk factors or biomarkers[17]), or less likely to receive regular AF-related medical care (eg, persons of color, underinsured)[18,19] may increase screening efficiency. Of note, current consensus guidelines from the United States[20] and Europe[6] support the clinical diagnosis of AF on the basis of a rhythm strip but do not offer specific guidance for 1L ECG findings based on automated algorithms. Our findings generally support recommendations from a recent European expert consensus statement[21] suggesting that abnormal findings on 1L ECG automated algorithms should be confirmed prior to taking clinical action and likely merit consultation with a cardiologist or rhythm specialist. Optimal management of an isolated 1L ECG finding of AF that cannot be confirmed with follow-up monitoring requires future study, particularly given emerging evidence linking AF burden and stroke risk.[5]

In contrast to the limited accuracy of the algorithm for positive results, we found that an automated 1L ECG interpretation of *normal* was very accurate in identifying tracings that did not represent underlying AF. Specifically, the NPV of *normal* was above 99% both overall and across all subgroups of age. Likewise, we observed very good accuracy for excluding AF across subgroups of AF risk, and also when

FIGURE 3 **Effect of Unclassified and No Analysis Tracings on 1L ECG Diagnostic Performance Stratified by Age of Screening**



Depicted are the test characteristics (sensitivity, top left; positive predictive value [PPV], top right; specificity, bottom left, negative predictive value [NPV], bottom right) of the AliveCor 1L ECG algorithm using cardiologist overread as the gold standard. In each plot, the relevant metric is plotted across age group (X-axis), and stratified by whether equivocal tracings (ie, *unclassified* or *no analysis*) are considered positive vs negative results. In each plot, the hashed horizontal line depicts the corresponding estimate in the overall sample. ECG = electrocardiogram.

utilizing same-day 12L ECG and electrophysiologist overread as alternative reference standards. Although randomized trials have suggested varying estimates of the yield of incident AF with population-based screening, the proportion of individuals diagnosed with new AF is consistently a minority despite varying methods of selecting screening candidates (eg, age thresholds).[2] As a result, the ability to accurately exclude the presence of AF on 1L ECG in an automated fashion may increase the efficiency of population-based screening. For example, a future staged screening protocol might infer the absence of AF solely based on automated 1L ECG interpretations, whereas any positive or equivocal results would trigger clinician review or confirmatory testing.

Our results suggest that efforts to reduce equivocal results may greatly increase the efficiency and accuracy of automated algorithms for AF screening. Consistent with some prior observations,[12] a substantial proportion of tracings in our study were labeled with equivocal interpretations (ie, *unclassified* or *no interpretation*) by the 1L ECG automated algorithm. As a result, the test characteristics of the algorithm changed substantially based on whether equivocal findings were considered positive results (ie, indicative of potential AF), negative results (ie, not indicative of potential AF), or excluded entirely. For example, when considering equivocal findings as indicative of AF, the sensitivity of the 1L ECG algorithm increased from 51% to 93%, while the specificity decreased from 99% to 82%. Analogous findings were observed with PPV and NPV. We observed a greater proportion of equivocal reads among older individuals, which may be driven by a higher frequency of non-AF-related abnormalities (eg, tremor, premature atrial contractions, atrial tachycardias), which are more common with older age[22] and are known to adversely affect the accuracy of 1L ECG algorithms.[23]

Since AF screening may be most effective in elderly individuals,[4,24] the benefit of automated algorithms is likely to be maximized if such algorithms can offer greater accuracy and lower frequency of equivocal findings when applied among older individuals at risk for AF. To this end, future work is needed to elucidate specific mechanisms underlying equivocal tracings. In the meantime, given that roughly 5% to 10% of tracings with AliveCor equivocal interpretations were ultimately reviewed as AF, equivocal findings appear sufficiently high-risk to merit confirmatory testing for possible AF.

Our study should be interpreted in the context of its design. First, we utilized a unique resource of over 30,000 1L ECG tracings reviewed by cardiologists to define the reference standard. Although a simultaneously performed 12-lead ECG may have provided a more accurate gold standard, routine acquisition of such tracings was not performed and would have been incompatible with the design of the VITAL-AF trial,[11] a pragmatic intervention embedding 1L ECG screening as part of routine primary care visits. Nevertheless, a previous analysis has shown that cardiologist interpretations of 1L ECGs are ~95% accurate against a simultaneous 12L ECG in a sample with 25% AF prevalence. Although we submit a 5% misclassification rate in our sample would be unlikely because of our 10-fold lower AF prevalence, even when up to 5% of tracing classifications were reversed, PPV, specificity, and NPV remained largely stable, and sensitivity remained in the range of 20% to 50%.[14] Furthermore, we also performed analyses utilizing electrophysiologist overread and same-day 12L ECG (when performed for clinical reasons) as alternative reference standards and observed largely similar results. Since not all tracings underwent electrophysiologist overread or had a same-day 12-lead ECG available, however, we acknowledge that these secondary analyses are subject to bias (eg, greater prevalence of AF). Use of single cardiologist's review as the reference standard also prevented us from assessing the performance of the AliveCor algorithm on the 2% of tracings that were manually uninterpretable or quantifying any potential effects of inter-reader variability. Third, we analyzed automated interpretations obtained using the AliveCor KardiaAI version 1. Although this was the current version available for consumer use throughout the VITAL-AF study period, our results may not generalize to other 1L ECG algorithms or devices or subsequent versions of AliveCor KardiaAI. Nevertheless, our findings suggest that marked improvements in diagnostic accuracy would be required before an automated algorithm result would be sufficient to establish an AF diagnosis. Fourth, we are unable to report specific mechanisms of equivocal AliveCor interpretations, although we submit that 1L ECG tracings obtained by trained medical assistants likely represent higher-quality tracing acquisition than regular consumer use. Fifth, cardiologist readers received individual tracings for review, which were not explicitly batched according to encounter or patient. Therefore, we submit that it is unlikely that the presence of multiple tracings from the same patient on the same day would have influenced interpretations, but we cannot rule out this possibility. Sixth, our results reflect those in a population comprising primarily White individuals who receive regular primary care. As a result, our findings may not generalize to other populations.

In summary, in an analysis including over 30,000 tracings obtained from roughly 14,000 primary care patients without prior known AF who underwent handheld 1L ECG screening within the context of a large randomized trial, we found that the AliveCor 1L ECG automated interpretation had moderate PPV for new AF and very good NPV for the absence of AF when using cardiologist review as the gold standard. Equivocal interpretations were common, became even more frequent with older age, and substantially influenced diagnostic performance. Abnormal findings on automated 1L ECG algorithms, including equivocal results, merit confirmation prior to taking clinical action, while normal interpretations appear sufficiently accurate to exclude the presence of AF.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Ellinor is supported by grants from the National Institutes of Health (1RO1HL092577, 1R01HL157635, 1R01HL157635); by a grant from the American Heart Association Strategically Focused Research Networks (18SFRN34110082); and by a grant from the European Union (MAESTRIA 965286). Dr Singer is supported, in part, by the Eliot B. and Edith C. Shoolman Fund of Massachusetts General Hospital. This study was investigator-initiated and funded by the Bristol-Myers Squibb-Pfizer Alliance. Dr Ellinor has received sponsored research support from Bayer AG and IBM Research; and has served on advisory boards or consulted for Bayer AG, MyoKardia, and Novartis. Dr Atlas has received sponsored research support from Bristol-Myers Squibb-Pfizer; has received sponsored research funding from Bristol Myers Squibb-Pfizer; and has consulted for Boehringer Ingelheim, Bristol Myers Squibb, Pfizer, and Fitbit. Dr Lubitz is an employee of Novartis Institutes for BioMedical Research. Dr Singer has received research support from Bristol Myers Squibb-Pfizer; and has received consulting fees from Bristol Myers Squibb-Pfizer, Fitbit, and Google. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Steven A. Lubitz, Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, 55 Fruit Street, GRB 109, Boston, Massachusetts 02114, USA. E-mail: slubitz@mgh.harvard.edu.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Automated 1L ECG algorithm interpretations appear sufficiently accurate to exclude AF but not to establish an AF diagnosis.

COMPETENCY IN PATIENT CARE: Abnormal 1L ECG findings, including equivocal results, merit confirmatory testing prior to taking clinical action.

TRANSLATIONAL OUTLOOK: Future staged AF screening protocols may infer the absence of AF solely based on automated 1L ECG interpretations, whereas any positive or equivocal results would trigger clinician review or confirmatory testing.

REFERENCES

1. Mant J, Fitzmaurice DA, Hobbs FDR, et al. Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. *BMJ*. 2007;335:380.

2. Khurshid S, Healey JS, McIntyre WF, Lubitz SA. Population-based screening for atrial fibrillation. *Circ Res*. 2020;127:143-154.

3. Lubitz SA, Atlas SJ, Ashburner JM, et al. Screening for atrial fibrillation in older adults at primary care visits: the VITAL-AF randomized controlled trial. *Circulation*. 2022;145(13):946-954.

4. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. *Circulation*. 2015;131:2176-2184.

5. Kaplan RM, Koehler J, Ziegler PD, Sarkar S, Zweibel S, Passman RS. Stroke risk as a function of atrial fibrillation duration and CHA2DS2-VASc score. *Circulation*. 2019;140:1639-1646.

6. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). *Eur Heart J*. 2021;42:373-498.

7. Brieger D, Amerena J, Attia JR, et al. National Heart Foundation of Australia and the Cardiac Society of Australia and New Zealand: Australian Clinical Guidelines for the Diagnosis and Management of Atrial Fibrillation 2018. *Heart Lung Circ*. 2018;27:1209-1266.

8. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Screening for atrial fibrillation with electrocardiography: US Preventive Services Task Force recommendation statement. *JAMA*. 2018;320:478-484.

9. Khurshid S, Chen W, Singer DE, et al. Comparative clinical effectiveness of population-based atrial fibrillation screening using contemporary modalities: a decision-analytic model. *J Am Heart Assoc*. 2021;10:e020330.

10. Guo Y, Wang H, Zhang H, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. *J Am Coll Cardiol*. 2019;74:2365-2375.

11. Ashburner JM, Atlas SJ, McManus DD, et al. Design and rationale of a pragmatic trial integrating routine screening for atrial fibrillation at primary care visits: the VITAL-AF trial. *Am Heart J*. 2019;215:147-156.

12. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. *Heart*. 2020;106:665-670.

13. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.

14. Lau JK, Lowres N, Neubeck L, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. *Int J Cardiol*. 2013;165:193-194.

15. Anon. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; 2015.

16. Al-Alusi MA, Khurshid S, Wang X, et al. Trends in consumer wearable devices with cardiac sensors in a primary care cohort. *Circ Cardiovasc Qual Outcomes*. 2022;15:e008833.

17. Kemp Gudmundsdottir K, Fredriksson T, Svennberg E, et al. Stepwise mass screening for atrial fibrillation using N-terminal B-type natriuretic peptide: the STROKESTOP II study. *Europace*. 2020;22:24-32.

18. Heckbert SR, Austin TR, Jensen PN, et al. Differences by race/ethnicity in the prevalence of clinically detected and monitor-detected atrial fibrillation: MESA. *Circ Arrhythm Electrophysiol*. 2020;13:e007698.

19. Tamirisa KP, Al-Khatib SM, Mohanty S, et al. Racial and ethnic differences in the management of atrial fibrillation. *CJC Open*. 2021;3:S137-S148.

20. January CT, Wann LS, Calkins H, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines and the Heart Rhythm Society. *Heart Rhythm*. 2019;16(8):e66-e93.

21. Svennberg E, Tjong F, Goette A, et al. How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. *Europace*. 2022;24:979-1005.

22. Khurshid S, Choi SH, Weng L-C, et al. Frequency of cardiac rhythm abnormalities in a half million adults. *Circ Arrhythm Electrophysiol*. 2018;11:e006273.

23. Kwon S, Hong J, Choi E-K, et al. Deep Learning approaches to detect atrial fibrillation using photoplethysmographic signals: algorithms development study. *JMIR Mhealth Uhealth*. 2019;7:e12770.

24. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, et al. Screening for atrial fibrillation in the older population: a randomized clinical trial. *JAMA Cardiol*. 2021;6(5):558-567.

KEY WORDS atrial fibrillation, mobile technologies, randomized controlled trials, screening

APPENDIX For supplemental tables and a figure, please see the online version of this paper.