Identifying a Heterogeneous Effect of Atrial Fibrillation Screening in Older Adults: A Secondary Analysis of the VITAL-AF Trial

Sachin J Shah, MD, MPH (1,2); Jay M. Iyer (3); Leila Agha, PhD (4); Yuchiao Chang, PhD (1); Jeffrey M. Ashburner, PhD, MPH (1); Steven J. Atlas, MD, MPH (1); David D. McManus, MD, ScM (5); Patrick T. Ellinor, MD, PhD (6), Steven A. Lubitz, MD, MPH (6); Daniel E. Singer, MD (1)

(1) Division of General Internal Medicine at Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
(2) Center for Aging and Serious Illness, Massachusetts General Hospital
(3) Departments of Molecular and Cellular Biology and Statistics, Harvard University, Cambridge, MA, USA
(4) Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
(5) Department of Medicine, UMass Chan Medical School, Worcester, MA, USA
(6) Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Contact
Sachin J Shah, MD, MPH
Massachusetts General Hospital
100 Cambridge St, Suite 1600
Boston, MA 02114
sshah@mgh.harvard.edu
617-977-4871

Words: 3410
Table and Figures: 6

ABSTRACT (298/300)

Background: One-time atrial fibrillation (AF) screening trials have produced mixed results; however, it is unclear if there is a subset for whom screening is effective. Identifying such a subgroup would support targeted screening.

Methods: We conducted a secondary analysis of VITAL-AF, a randomized trial of one-time, single-lead ECG screening during primary care visits. We tested two approaches to identify a subgroup where screening is effective. First, we developed an effect-based model using a T-learner. Specifically, we separately predicted the likelihood of AF diagnosis under screening and usual care conditions; the difference in probabilities was the predicted screening effect. Second, we used a validated AF risk model to test for a heterogeneous screening effect. We used interaction testing to determine if observed AF diagnosis rates in the screening and usual care groups differed when stratified by decile of the predicted screening effect and predicted AF risk.

Results: Baseline characteristics were similar between the screening (n=15187) and usual care (n=15078) groups (mean age 74 years, 59% female). In the effect-based analysis, in the highest decile of predicted screening effectiveness (n=3026), AF diagnosis rates were higher in the screening group (6.50 vs. 3.06 per 100 person-years, rate difference 3.45, 95%CI 1.62 to 5.28). In this group, the mean age was 84 years and 68% were female. The risk-based analysis did not identify a subgroup where screening was more effective. Predicted screening effectiveness and predicted baseline AF risk were poorly correlated (Spearman coefficient 0.13).

Conclusions: In a secondary analysis of the VITAL-AF trial, we identified a small subgroup where one-time screening was associated with increased AF diagnoses using an effect-based approach. In this study, predicted AF risk was a poor proxy for predicted screening

effectiveness. These data caution against the assumption that high AF risk is necessarily correlated with high screening effectiveness.

INTRODUCTION

The impetus to screen for atrial fibrillation (AF) is clear—AF is common and increases the risk of disabling strokes.[1,2] Among those 65 years old, the lifetime incidence of atrial fibrillation is 33%.[3] The goal of screening is to identify cases earlier than usual so stroke-preventive therapies, notably anticoagulation, can be used in appropriate patients to prevent ischemic stroke. However, randomized clinical trials (RCTs) of screening interventions have produced mixed results, and in 2022, the United States Preventive Services Task Force concluded that there was insufficient evidence to recommend routine screening for AF.[4–8]

One-time screening during routine clinical care is appealing because it is practical and is thought to identify individuals with high-burden AF.[9,10] Nevertheless, all but one RCT testing one-time screening in traditional care settings (e.g., primary care offices) have failed to show that screening identifies more cases of AF in 6 to 12 months.[5–8] However, it is unclear if there is a subset of people for whom screening is effective. While trials typically test for heterogeneity one subgroup at a time, newer methods allow for the examination of heterogeneity across multiple factors.[11–15] This approach has been successfully applied in clinical decision models such as the Dual Antiplatelet Therapy (DAPT) score, which guides the duration of dual antiplatelet therapy after coronary stenting, and the PFO-Associated Stroke Causal Likelihood (PASCAL) score, which guides patient selection for PFO closure.[16,17] Both models use multivariable prediction of treatment effect heterogeneity. Contemporary approaches also disentangle treatment effect and baseline disease risk.[12] For example, screening might be less efficacious for those at high baseline risk since they may also have a high probability of being identified through routine clinical care. Identifying a subset of patients where screening is effective could support future screening strategies, including targeted screening trials.

Our goal was to determine if one-time AF screening during routine clinical care is effective in a subset of older adults. To accomplish this, we conducted a secondary analysis of the VITAL-AF RCT, which demonstrated that single-lead ECG screening for adults 65 years and

older during their primary care office visits did not significantly change the rate of new AF diagnoses.[18] In this secondary analysis, we aimed to identify a subset of people in whom screening is effective using “effect-based” and “risk-based” approaches.

METHODS

Study Design and Participants

This is a secondary, post-hoc analysis of the VITAL-AF trial to identify if AF screening is effective in a subset of individuals. The design and primary results from VITAL-AF have been published.[18,19] In brief, VITAL-AF was a pragmatic, cluster-randomized trial that tested the effectiveness of single-lead ECG screening during primary care visits compared to usual care. The trial randomized 16 primary care practices (8 to screening and 8 to usual care) in the Massachusetts General Hospital Primary Care-Based Research Network between July 2018 and October 2019. We did not adjust for the cluster-randomized design in this study because the intracluster correlation was low (0.0013). For these analyses, individuals 65 years and older without a prior diagnosis of AF presenting for a primary care appointment were included and participants with missing predictors were excluded (**Supplemental Figure S1**). All analyses were conducted on an intention-to-screen basis (i.e., group assignment was based on the patient’s first visit to a study practice during the study period). The Mass General Brigham Institutional Review Board approved the research protocol. Participants provided informed consent to participate. The study was considered minimal risk, and a waiver of documentation of informed consent was granted.

Procedure

As previously described, screening was conducted when consenting patients placed their fingers on a single-lead AliveCor KardiaMobile ECG device (AliveCor Inc, Mountain View, CA). The screening resulted in one of five possible results, “Possible AF,” “Normal,”

“Unclassified,” “No analysis (Unreadable),” and “Patient Declined Screening.” All subsequent clinical management was determined by primary care clinicians, including follow-up 12-lead ECGs. Independent cardiologists reviewed all AliveCor tracings within 7 days and notified primary care clinicians if a prespecified actionable rhythm was identified.

Outcome

The primary outcome was an adjudicated incident AF diagnosis. Each clinical practice was enrolled for 12 months; participants were followed until the primary care practice to which they belonged completed participation. Thus, follow-up time was measured from each participant’s first visit date until the date their primary care practice completed participation or death, whichever came first. Potentially new AF diagnoses were identified from the electronic medical record using the same approach in both study arms by a centralized data repository.[20] Specifically, individuals with an International Classification of Diseases, 10th Revision code for atrial fibrillation or flutter or a 12-lead ECG with atrial fibrillation or flutter in the diagnostic statement were identified. These potential new AF diagnoses were then adjudicated by 2 research nurses with a cardiologist unaffiliated with the study resolving differences. The committee adjudicated events as “incident,” “prevalent,” or “not AF.”[21]

Predictors

We obtained patient characteristics from the electronic medical record, including demographics, medical diagnoses, medication use, physiological measures, and prior health use. We used measures obtained on the participant’s first visit date or the value most immediately prior. Candidate predictors were chosen because of their association with AF, cardiovascular disease, or health care utilization. Demographics included age, sex, and English language preference. Medical diagnoses included hypertension, myocardial infarction, coronary artery disease, diabetes, congestive heart failure, prior stroke, vascular disease, anemia,

bleeding history, chronic kidney disease, and tobacco smoking. Medications were grouped into one of the following categories: oral anticoagulants, rate control medications, antihypertensives, or antiarrhythmic medications. Physiological measures include systolic blood pressure, diastolic blood pressure, heart rate, height, and weight. Healthcare utilization measures included 12 lead ECGs in the prior year, indwelling cardiac rhythm device implanted in the prior 3 years (i.e., permanent pacemaker, implantable cardioverter defibrillator, or implantable loop recorder), and the number of primary care visits in the prior year.

Effect-based approach to heterogeneity

Our objective was to identify individuals for whom screening is effective by measuring the heterogeneity of AF screening using an effect-based approach.[12] To do so, we estimated the effect of screening for each individual given their observed characteristics, i.e., the conditional average screening effect.[22] Specifically, we used a T-learner wherein, for each individual, we estimated the probability of the outcome (i.e., new AF diagnosis) had they been randomized to screening and, separately, had they been randomized to usual care.[23]

Operationally, we first estimated the likelihood of the outcome conditional on being randomized to the screening. Using just the participants randomized to the screening practices, we fit a generalized linear model of the outcome as a function of all the predictors described above using a Poisson distribution offset by the log of the follow-up time to account for differential follow-up time.[24] This model was fit using the least absolute shrinkage and selection operator (LASSO), which we used to perform variable selection and regularization.[25] Then, we applied this model to the usual care participants to estimate the likelihood of the outcome had they been randomized to the screening. Next, we estimated the likelihood of the outcome conditional on being randomized to screening within the screening arm. To limit bias, we estimated the likelihood of the outcome within the intervention group using leave-one-out cross-validation.[26] That is, the likelihood of the outcome for an individual intervention participant was

estimated using a model that was fit with all intervention observations except that individual. These models were also fit using LASSO.

We repeated the same procedure to estimate the likelihood of the outcome under usual care. This process resulted in two predicted probabilities of AF diagnosis for each participant (n=30,265)—one had they been screened and one under usual care. The difference between these two is the conditional average screening effect. We then tabulated the observed outcome rate by randomization arm stratified by decile of conditional average screening effect.

In a sensitivity analysis, we used a survival model (i.e., time-to-atrial fibrillation diagnosis) to test the robustness of the modified Poisson regression model used in the primary analysis. The analysis demonstrated no difference in time-to-atrial fibrillation diagnosis by randomization arm supporting the primary modeling approach (**Supplemental Methods 1**).

Risk-based approach to heterogeneity

In a second approach, we measured the heterogeneity of AF screening using a risk-based approach. We used the CHARGE-AF risk score to estimate the baseline risk of developing AF.[27] CHARGE-AF is a risk model developed in the Atherosclerosis Risk in Communities study, Cardiovascular Health Study, and the Framingham Heart Study to predict incident AF. It uses age, race, height, weight, blood pressure, current smoking, use of antihypertensive medication, diabetes, myocardial infarction history, and heart failure as predictors. The CHARGE-AF model has been externally validated.[28] Because CHARGE-AF was developed externally, its performance may be disadvantaged compared to the internally developed effect score. In a sensitivity analysis, we determined if an internally optimized AF risk model performed better than the CHARGE-AF model. This internal AF risk model was developed using only the usual care patients.

Testing for heterogeneity

We assessed heterogeneity by testing the interaction between the randomization arm and the predicted screening effect. Specifically, we fit a generalized linear model where the outcome of new AF diagnosis was a function of the randomization arm, decile of predicted screening effect, and the interaction between the two. The model was fit using a Poisson distribution offset by the log of the follow-up time to account for differential follow-up time.[24] We tested the statistical significance of the interaction using the likelihood ratio test. For visual representation we plotted the observed outcome rate by randomization arm, stratified by decile of screening effect. We repeated this procedure to test for heterogeneity by predicted AF risk.

Patient characteristics by predicted effect and predicted risk

We used heatmaps to visualize participant characteristics by decile of predicted screening effect and predicted AF risk. We color-coded characteristics by their z-transformed value, where the darkest and lightest shade represent the highest and lowest value of a given patient characteristic, respectively.

Correlation of predicted effect and predicted risk

To determine the correlation between the conditional average screening effect and baseline risk, we plotted the percentile of predicted screening effectiveness against the percentile of predicted AF risk. We described the relationship using a locally estimated best-fit line with a span of 0.75.

RESULTS

Average effect of screening

We present the baseline characteristics of 30265 study participants in Table 1. Features were well balanced in the two arms of the trial. In 10333 person-years of follow-up, 238 people

in the usual care group were diagnosed with atrial fibrillation (2.30 per 100-person years). In 10284 person-years of follow-up, 262 people in the screening group were diagnosed with atrial fibrillation (2.55 per 100 person-years). These rates were not statistically significantly different (rate difference 0.24 per 100 person-years, 95%CI -0.18 to 0.67).

Model performance

The constituent models used to estimate the effect-based score discriminated well. The model to predict AF diagnosis under usual care and the model to predict AF diagnosis in screening both had a c-statistic of 0.73 (**Supplemental Figures S2**). The CHARGE-AF risk model also discriminated well; it had a c-statistic of 0.74 in both the control and intervention arms (**Supplemental Figures S3**).

Effect-based approach to screening heterogeneity

Participants whose predicted screening effect fell into the highest decile (estimated by the effect-based model) had a statistically significant increase in AF diagnoses due to their primary care practice being randomized to the screening arm (interaction p-value 0.038).
In **Figure 1**, we display the distribution of observed absolute rate differences in AF diagnosis in each decile of predicted screening effect. In the decile where screening was predicted to be the most efficacious, the observed rate of new atrial fibrillation diagnosis was higher in those randomized to intervention compared to those randomized to usual care (6.50 vs. 3.06 per 100 person-years, rate difference 3.45 per 100 person-years, 95% CI 1.62 to 5.28). In the remaining 9 deciles, the observed rates of AF diagnosis in the intervention and usual care groups were not significantly different. A sensitivity analysis that evaluated the screening effect model in a continuous fashion is consistent with a monotonic increase in observed treatment effect (**Supplemental Figure S4**).

Risk-based approach to screening heterogeneity

The risk model (i.e., CHARGE-AF) did not identify a subgroup in which one-time AF screening was effective (interaction p-value 0.46). In Figure 2, we display the distribution of observed absolute rate difference in AF diagnosis in each decile of predicted AF risk. The observed rates of AF diagnosis in the intervention and screening groups were not significantly different in any decile of predicted AF risk. In the highest risk decile, AF diagnosis rates were numerically higher in the screening (7.89 vs 5.77 per 100 person-years), but the difference was not statistically significant (rate difference 2.12, 95%CI -0.05 to 2.49). In a sensitivity analysis, we determined that an internally optimized risk model also did not identify a subgroup where screening was effective (**Supplemental Figure S5**).

Patient characteristics of effective screening groups

In **Figure 3**, we display patient characteristics by decile of screening effectiveness. Patients with an indwelling cardiac rhythm device placed in the prior 3 years, higher BMI, and greater number of PCP visits in the prior year were overrepresented in the lower deciles of screening effectiveness. Rate control medications, high systolic blood pressure, and smoking were more common in the highest deciles of screening effectiveness. Several characteristics displayed a U-shaped relationship, such as age, vascular disease, and congestive heart failure.

Patient characteristics of risk groups

In **Figure 4**, we display the patient characteristics by decile of risk estimated using the CHARGE-AF score. Black participants, Hispanic participants, and women were overrepresented in the lower deciles of risk. As expected, predictors used to calculate the CHARGE-AF score were more common in higher risk deciles, including older age; higher height, weight, or blood pressure; smoking history; White racial identity; antihypertensive medication use; or diagnosis of

diabetes, CHF, or prior MI. Among variables not directly used to calculate CHARGE-AF, we find that patients in the highest risk deciles were more likely to have had an indwelling cardiac rhythm device placed in the prior 3 years, male sex, chronic kidney disease, and anemia.

Relationship between predicted risk and predicted screening effect

The predicted screening effect and the predicted risk of new AF have a non-monotonic relationship. In **Figure 5**, we show a scatterplot of the relationship between the percentile of AF screening effectiveness against the percentile of baseline AF risk measured by the CHARGE-AF score. Predicted screening effectiveness and predicted baseline risk were poorly correlated (Spearman coefficient 0.13). A locally estimated best-fit line demonstrated a U-shaped relationship—the predicted risk of AF was high among both those with low and high predicted screening effectiveness.

DISCUSSION

In this secondary analysis of the VITAL-AF randomized trial, we identified a small subgroup in whom one-time screening was effective using an effect-based modeling approach. In the subset where screening appears to be effective, individuals are often in the 80s, have hypertension and vascular disease, are seen less often by their PCP and are women—a complete phenotype is available online.[29] Despite a trend toward greater effectiveness in the highest-risk group, a risk-based approach did not identify a subgroup for whom screening was statistically more effective. We also determined that predicted risk and predicted screening effectiveness were not well correlated. This suggests that when screening for AF with a brief one-time screen, predicted risk is an inadequate surrogate for predicted screening effectiveness.

An inspection of patient characteristics provides insights into the value of an effect-based approach. The 2% of study participants with an indwelling cardiac rhythm device (i.e.,

permanent pacemaker, implantable cardioverter defibrillator, or implantable loop recorder) are an important negative control. Because of the pragmatic study design, people with indwelling cardiac rhythm devices were included; however, because they monitor for AF continuously, mechanistically, one-time office-based AF should not be effective for people with such devices. In this study, the effect-based model appropriately concludes that screening would not be effective in people with indwelling cardiac rhythm devices. On the other hand, the risk model classified individuals with indwelling cardiac rhythm devices as among the highest risk of developing AF. Thus, using risk to guide screening would inappropriately prioritize individuals with indwelling cardiac rhythm devices. This example highlights how an effect-based approach can identify people for whom screening is effective, irrespective of their risk of AF.

The results of this study can inform future efforts at targeting screening in at least three ways. First, this study identifies a subgroup in which screening appears particularly effective. While VITAL-AF, D2AF, and Morgan & Mant found no average benefit of one-time screening, SAFE found an increase in AF diagnosis rate of 0.55% over 12 months.[6–8,18] In our current study, we identified a subgroup in whom screening was 3 to 10 times more effective than the average effect observed in SAFE. To demonstrate the value of one-time screening, future trials should consider targeting screening to the high-effectiveness phenotype. There are shortcomings to this approach, namely that the subset in whom screening is effective is small, and any targeting approach creates implementation challenges. To facilitate the identification of high-effectiveness patients, we have published a supplemental online code that can be incorporated into electronic medical records or screening trials.[29] Second, our findings that AF risk and screening effect are not well correlated indicate that screening trials targeting high-risk individuals may not be fruitful. Indeed, multiple trials are underway to test the value of screening by enrolling high-risk individuals.[30–33] Third, our findings highlight the potential for inequity using a risk-based approach, particularly a risk equation like CHARGE-AF that uses race as a predictor. In this study, while non-white and female participants were concentrated in the low-

risk deciles, both groups were more evenly distributed across the spectrum of screening effectiveness. Thus, well-meaning efforts to target screening to high-risk individuals may widen disparities.

There are a few possible reasons for the observed discordance between AF risk and screening effect. First, the effect model may identify people who do not fit physicians' heuristic for AF, so-called representation bias.[34,35] For example, non-White racial identity, women, and those with a language preference other than English were over-represented in the high-effectiveness group. These same characteristics were more prevalent in low-risk groups. Second, screening effectiveness may be a function of healthcare access and connectedness.[36] The average number of PCP visits in the prior year was lowest in the high-screening effectiveness group. This indicates that individuals who are often seen by their PCP may benefit less from one-time screening despite being at high risk for developing AF, presumably because physicians have an increased opportunity to detect heart rhythm abnormalities. It also highlights that "usual care" varies across patients. Since the degree of access to "usual care" varies, one-time screening may be more effective for people with limited healthcare access. Third, it is possible that those in the high-effectiveness group are more likely to develop asymptomatic AF, a hypothesis that could not be tested with the available data.

Finally, this study adds to the ongoing development of effect-based approaches to measuring the heterogeneous effect of cardiovascular interventions.[12–15] For the treatment and prevention of many cardiovascular conditions, like statins to prevent atherosclerotic heart disease and anticoagulants to prevent stroke in AF, clinicians are asked to identify a subset of people for whom treatment is particularly effective.[37,38] The conventional approach has been to model risk—that is, to develop an outcome model in untreated individuals and assume that this risk is tightly correlated with the effectiveness of the intervention, the true value of interest. This study demonstrates the vulnerability of this assumption. Prior studies have shown that risk-based models can effectively measure variation in treatment effect.[39,40] This study suggests

that, when possible, both risk-based and effect-based approaches should be tested when determining if an intervention should be targeted to a subset of individuals.

Our study design and data source have important limitations. First, the high-effectiveness subgroup identified in this study needs to be externally validated. While we took caution when identifying the subgroup (i.e., using regularized regression models and leave-one-out cross-validation), external validation in a separate RCT is necessary to test the robustness of the findings. Second, this analysis was not pre-specified in the statistical analysis plan of the main trial. Thus, the results should be regarded as exploratory and hypothesis-generating. Third, screening for atrial fibrillation is important in so far as it prevents ischemic strokes. While one-time screening tends to identify individuals with high-burden AF, the full study was not powered to detect the effect of screening on strokes and bleeding.

Prior trials of one-time screening have demonstrated mixed results. In a secondary analysis, we identified a small subgroup where one-time screening was associated with increased AF diagnoses using an effect-based modeling approach. Also, we determined that risk is a poor proxy for the effectiveness of AF screening. These data caution against the assumption that high risk is necessarily correlated with high effectiveness. Future screening efforts should focus on people for whom screening is projected to be most effective.

ACKNOWLEDGMENTS

Author contributions: Dr. Shah had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors listed have contributed sufficiently to the project to be included as authors, and all those who are qualified to be authors are listed in the author byline.

Conflict of Interest Disclosure: Dr. Shah reported funding from the National Institute on Aging/National Institutes of Health related to the conduct of this study (noted below). Dr. Atlas has received sponsored research support from Bristol Myers Squibb / Pfizer and American Heart Association (18SFRN34250007). Dr. Atlas has consulted for Boehringer Ingelheim, Bristol Myers Squibb, Pfizer, and Fitbit. Dr. Ashburner is supported by NIH grant K01HL148506, American Heart Association 18SFRN34250007, and has received sponsored research support from Bristol Myers Squibb / Pfizer. Dr. Ellinor is supported by grants from the National Institutes of Health (R01HL092577, R01HL157635), by a grant from the American Heart Association (18SFRN34110082, 961045), and by a grant from the European Union (MAESTRIA 965286). Dr. Ellinor has received sponsored research support from Bayer AG, Novo Nordisk, Pfizer, Bristol Myers Squibb and IBM; he has also served on advisory boards or consulted for Bayer AG. Dr. Lubitz is an employee of Novartis. Dr. Lubitz was previously supported by NIH grants R01HL139731 and R01HL157635, and American Heart Association 18SFRN34250007. Dr. Lubitz received sponsored research support from Bristol Myers Squibb, Pfizer, Boehringer Ingelheim, Fitbit, Medtronic, Premier, and IBM, and has consulted for Bristol Myers Squibb, Pfizer, Blackstone Life Sciences, and Invitae. Dr. Singer receives research support from Bristol Myers Squibb-Pfizer and has received consulting fees from Bristol Myers Squibb, Fitbit (Google), Medtronic, and Pfizer. Dr. McManus reports having received compensation from Fitbit for serving on the Fitbit Heart Study Advisory Board, from the Heart Rhythm Society for service as Editor, from Avania and NAMSA for serving on Data and Safety Monitoring Boards and has

received non-compensatory study support from Apple Computer, Fitbit, and Care Evolution. The remaining authors have nothing to disclose.

Funding: This analysis was funded by the NIA (K76AG074919). The VITAL-AF study was an investigator-initiated study funded by the Bristol Myers Squibb/Pfizer Alliance. Dr. Singer was supported, in part, from the Eliot B. and Edith C. Shoolman Fund of the Massachusetts General Hospital.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Supplemental Materials: Figures S1-S5, Supplemental Methods 1

REFERENCES

1. Parks AL, Jeon SY, Boscardin WJ, et al. Long-term individual and population functional outcomes in older adults with atrial fibrillation. *J Am Geriatr Soc*. 2021;69(6):1570-1578. doi:10.1111/jgs.17087

2. Fang MC, Go AS, Chang Y, et al. Long-term survival after ischemic stroke in patients with atrial fibrillation. *Neurology*. 2014;82(12):1033-1037.

3. Staerk L, Wang B, Preis SR, et al. Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. *BMJ*. Published online April 26, 2018:k1453. doi:10.1136/bmj.k1453

4. Kahwati LC, Asher GN, Kadro ZO, et al. Screening for Atrial Fibrillation: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2022;327(4):368. doi:10.1001/jama.2021.21811

5. Kaasenbrood F, Hollander M, Bruijn SH de, et al. Opportunistic screening versus usual care for diagnosing atrial fibrillation in general practice: a cluster randomised controlled trial. *Br J Gen Pract*. 2020;70(695):e427-e433. doi:10.3399/bjgp20X708161

6. Morgan S, Mant D. Randomised trial of two approaches to screening for atrial fibrillation in UK general practice. *Br J Gen Pract*. 2002;52(478):373-380.

7. Hobbs FDR, Fitzmaurice DA, Mant J, et al. A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study. *Health Technol Assess Winch Engl*. 2005;9(40):iii-iv, ix-x, 1-74. doi:10.3310/hta9400

8. Uittenbogaart SB, Gurp NV van, Lucassen WAM, et al. Opportunistic screening versus usual care for detection of atrial fibrillation in primary care: cluster randomised controlled trial. *BMJ*. 2020;370:m3208. doi:10.1136/bmj.m3208

9. Khurshid S, Healey JS, McIntyre WF, Lubitz SA. Population-Based Screening for Atrial Fibrillation. *Circ Res*. 2020;127(1):143-154. doi:10.1161/CIRCRESAHA.120.316341

10. Benjamin EJ, Go AS, Desvigne-Nickens P, et al. Research Priorities in Atrial Fibrillation Screening: A Report From a National Heart, Lung, and Blood Institute Virtual Workshop. *Circulation*. 2021;143(4):372-388. doi:10.1161/CIRCULATIONAHA.120.047633

11. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. Published online December 10, 2018:k4245. doi:10.1136/bmj.k4245

12. Dahabreh IJ, Kazi DS. Toward Personalizing Care: Assessing Heterogeneity of Treatment Effects in Randomized Trials. *JAMA*. 2023;329(13):1063. doi:10.1001/jama.2023.3576

13. Wang G, Heagerty PJ, Dahabreh IJ. Using Effect Scores to Characterize Heterogeneity of Treatment Effects. *JAMA*. Published online March 19, 2024. doi:10.1001/jama.2024.3376

14. Angus DC, Chang CCH. Heterogeneity of Treatment Effect: Estimating How the Effects of Interventions Vary Across Individuals. *JAMA*. 2021;326(22):2312. doi:10.1001/jama.2021.20552

15. Yarnell CJ, Fralick M. Heterogeneity of Treatment Effect — An Evolution in Subgroup Analysis. *NEJM Evid*. 2024;3(4):EVIDe2400054. doi:10.1056/EVIDe2400054

16. Yeh RW, Secemsky EA, Kereiakes DJ, et al. Development and Validation of a Prediction Rule for Benefit and Harm of Dual Antiplatelet Therapy Beyond 1 Year After Percutaneous Coronary Intervention. *JAMA*. 2016;315(16):1735. doi:10.1001/jama.2016.3775

17. Kent DM, Saver JL, Kasner SE, et al. Heterogeneity of Treatment Effects in an Analysis of Pooled Individual Patient Data From Randomized Trials of Device Closure of Patent Foramen Ovale After Stroke. *JAMA*. 2021;326(22):2277-2286. doi:10.1001/jama.2021.20956

18. 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*. Published online March 2, 2022:CIRCULATIONAHA.121.057014. doi:10.1161/CIRCULATIONAHA.121.057014

19. 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. doi:10.1016/j.ahj.2019.06.011

20. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a Research Patient Data Repository. *AMIA Annu Symp Proc AMIA Symp*. 2006;2006:1044.

21. Ashburner JM, Singer DE, Lubitz SA, Borowsky LH, Atlas SJ. Changes in Use of Anticoagulation in Patients With Atrial Fibrillation Within a Primary Care Network Associated With the Introduction of Direct Oral Anticoagulants. *Am J Cardiol*. 2017;120(5):786-791. doi:10.1016/j.amjcard.2017.05.055

22. Dahabreh IJ, Trikalinos TA, Kent DM, Schmid CH. Heterogeneity of Treatment Effects. In: *Methods in Comparative Effectiveness Research*. 1st Edition. Chapman and Hall/CRC; 2017:227-263.

23. Xu Y, Ignatiadis N, Sverdrup E, Fleming S, Wager S, Shah N. Treatment Heterogeneity with Survival Outcomes. In: *Handbook of Matching and Weighting Adjustments for Causal Inference*. 1st ed. Chapman and Hall/CRC; 2023:445-482. doi:10.1201/9781003102670-21

24. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol*. 2004;159(7):702-706. doi:10.1093/aje/kwh090

25. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

26. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: With Applications in R*. Springer New York; 2013.

27. 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 Cardiovasc Cerebrovasc Dis*. 2013;2(2):e000102. doi:10.1161/JAHA.112.000102

28. Himmelreich JCL, Lucassen WAM, Harskamp RE, Aussems C, Weert HCPM van, Nielen MMJ. CHARGE-AF in a national routine primary care electronic health records database in the Netherlands: validation for 5-year risk of atrial fibrillation and implications for patient selection in atrial fibrillation screening. *Open Heart*. 2021;8(1):e001459. doi:10.1136/openhrt-2020-001459

29. Iyer J. CATE Calculator. Accessed May 13, 2024. https://jaymiyer.shinyapps.io/Calculator_v2/

30. Sorlandet Hospital HF. *The NORwegian Atrial Fibrillation Self-SCREENing*. clinicaltrials.gov; 2023. Accessed December 31, 2023. https://clinicaltrials.gov/study/NCT05914883

31. Choi EK. *Atrial Fibrillation Screening With MobiCARE in Subjects With Health Care Examination*. clinicaltrials.gov; 2023. Accessed December 31, 2023. https://clinicaltrials.gov/study/NCT06036342

32. Etminani F, Sandgren E, Holm J, et al. Randomised, siteless study to compare systematic atrial fibrillation screening using enrichment by a risk prediction model with standard care in a Swedish population aged ≥ 65 years: CONSIDERING-AF study design. *BMJ Open*. 2024;14(1):e080639. doi:10.1136/bmjopen-2023-080639

33. Beerten SG, Proesmans T, Vaes B. The effect of a case-finding app on the detection rate of atrial fibrillation compared with opportunistic screening in primary care patients: protocol for a cluster randomized trial. *Trials*. 2021;22:525. doi:10.1186/s13063-021-05497-x

34. Mullainathan S, Obermeyer Z. Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*. *Q J Econ*. 2022;137(2):679-727. doi:10.1093/qje/qjab046

35. 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(1):e007698. doi:10.1161/CIRCEP.119.007698

36. Atlas SJ. Patient–Physician Connectedness and Quality of Primary Care. *Ann Intern Med*. 2009;150(5):325. doi:10.7326/0003-4819-150-5-200903030-00008

37. Joglar JA, Chung MK, Armbruster AL, et al. 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. Published online November 30, 2023:CIR.0000000000001193. doi:10.1161/CIR.0000000000001193

38. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140(11). doi:10.1161/CIR.0000000000000678

39. Buell KG, Spicer AB, Casey JD, et al. Individualized Treatment Effects of Oxygen Targets in Mechanically Ventilated Critically Ill Adults. *JAMA*. Published online March 19, 2024. doi:10.1001/jama.2024.2933

40. Goligher EC, Lawler PR, Jensen TP, et al. Heterogeneous Treatment Effects of Therapeutic-Dose Heparin in Patients Hospitalized for COVID-19. *JAMA*. 2023;329(13):1066. doi:10.1001/jama.2023.3651

TABLES and FIGURES

Table 1: Patient characteristics by randomization arm

Characteristic	Screening (n=15187)	Usual care (n=15078)
Age, mean (SD), yrs	74 (7)	74 (7)
Female sex, n (%)	9,068 (60%)	8,766 (58%)
Race, n (%)		
White	12,554 (83%)	12,490 (83%)
Black	791 (5.2%)	697 (4.6%)
Hispanic	331 (2.2%)	299 (2.0%)
Other	1,511 (9.9%)	1,592 (11%)
Systolic blood pressure, mean (SD), mmHg	131 (16)	131 (17)
Diastolic blood pressure, mean (SD), mmHg	75 (9)	74 (9)
Heart rate, mean (SD), bpm	75 (13)	74 (13)
Weight, mean (SD), kg	76 (18)	76 (18)
Height, mean (SD), cm	165 (10)	165 (10)
English language preference, n (%)	13,832 (91%)	13,411 (89%)
Current smoking, n (%)	717 (4.7%)	782 (5.2%)
Hypertension, n (%)	11,516 (76%)	11,443 (76%)
Coronary artery disease, n (%)	3,206 (21%)	3,128 (21%)
Diabetes, n (%)	3,620 (24%)	3,578 (24%)
Congestive heart failure, n (%)	1,586 (10%)	1,553 (10%)
Prior stroke, n (%)	1,269 (8.4%)	1,274 (8.4%)
Vascular disease, n (%)	2,983 (20%)	2,959 (20%)
Anemia, n (%)	3,736 (25%)	3,765 (25%)
Bleeding history, n (%)	5,587 (37%)	5,870 (39%)
Chronic kidney disease, n (%)	2,110 (14%)	2,374 (16%)
Antihypertensive medication, n (%)	10,632 (70%)	10,684 (71%)
Oral anticoagulant use, n (%)	340 (2.2%)	352 (2.3%)
Rate control medication use, n (%)	7,240 (48%)	7,275 (48%)
Antiarrhythmic medication use, n (%)	54 (0.4%)	55 (0.4%)
12 lead ECG in prior year, n (%)	4,884 (32%)	4,894 (32%)
Indwelling cardiac rhythm device prior 3 yrs, n (%)	305 (2.0%)	304 (2.0%)
PCP visits in the prior year, mean (SD)	4.1 (4.6)	4.6 (4.6)
CHARGE-AF score, mean (SD)	13.47 (0.95)	13.48 (0.96)

Legend
bmp – beats per minute, ECG – electrocardiogram, PCP – primary care physician. Indwelling cardiac rhythm devices include permanent pacemakers, implantable cardioverter defibrillators, and implantable loop recorders.

Figure 1: Heterogeneity of Screening Effect using an Effect-based Model

Legend
py–person-years. Model interaction p-value = 0.038

Decile of predicted efficacy	Usual care, observed events per 100py (cases/py)	Screening observed events per 100py (cases/py)	Observed rate difference (95% CI)
1 (lowest)	4.92 (58/1179)	4.48 (47/1050)	−0.44 (−2.24 to 1.36)
2	2.53 (28/1108)	2.01 (20/996)	−0.52 (−1.80 to 0.77)
3	1.35 (14/1038)	1.99 (20/1006)	0.64 (−0.48 to 1.76)
4	1.60 (16/1001)	1.81 (18/997)	0.21 (−0.94 to 1.35)
5	1.33 (13/976)	1.31 (13/991)	−0.02 (−1.04 to 1.00)
6	1.79 (17/952)	1.42 (14/985)	−0.36 (−1.49 to 0.77)
7	1.14 (11/967)	1.75 (18/1027)	0.62 (−0.44 to 1.67)
8	1.92 (19/989)	1.79 (19/1059)	−0.13 (−1.31 to 1.06)
9	2.78 (29/1043)	1.97 (21/1068)	−0.81 (−2.13 to 0.50)
10 (highest)	3.06 (33/1079)	6.50 (72/1107)	3.45 (1.62 to 5.28)

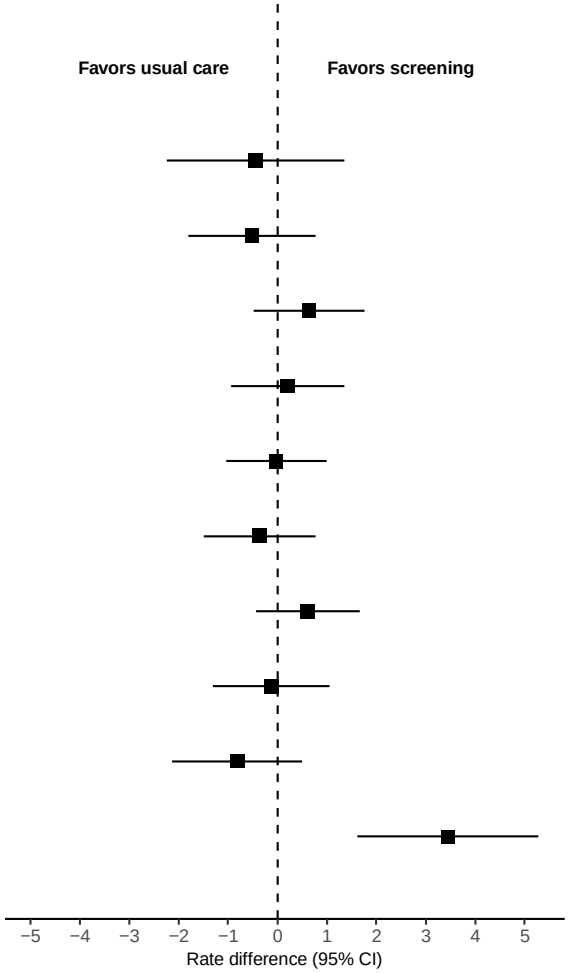


Figure 2: Heterogeneity of Screening Effect using a Risk-based Model

Legend
py–person-years. The baseline risk of incident AF was predicted using the CHARGE-AF score. Model interaction p-value = 0.46

Decile of predicted risk	Usual care, observed events per 100py (cases/py)	Screening observed events per 100py (cases/py)	Observed rate difference (95% CI)
1 (lowest)	0.11 (1/930)	0.44 (4/919)	0.33 (−0.15 to 0.80)
2	1.13 (11/977)	0.74 (7/940)	−0.38 (−1.25 to 0.48)
3	0.71 (7/985)	1.10 (11/997)	0.39 (−0.45 to 1.23)
4	1.20 (12/1000)	1.19 (12/1010)	−0.01 (−0.97 to 0.94)
5	1.26 (13/1029)	1.44 (15/1040)	0.18 (−0.82 to 1.18)
6	1.48 (15/1011)	1.96 (21/1070)	0.48 (−0.65 to 1.60)
7	2.55 (27/1057)	2.28 (24/1054)	−0.28 (−1.60 to 1.05)
8	4.15 (45/1083)	3.19 (34/1066)	−0.96 (−2.58 to 0.66)
9	3.61 (39/1081)	4.41 (49/1112)	0.80 (−0.87 to 2.47)
10 (highest)	5.77 (68/1178)	7.89 (85/1077)	2.12 (−0.05 to 4.29)

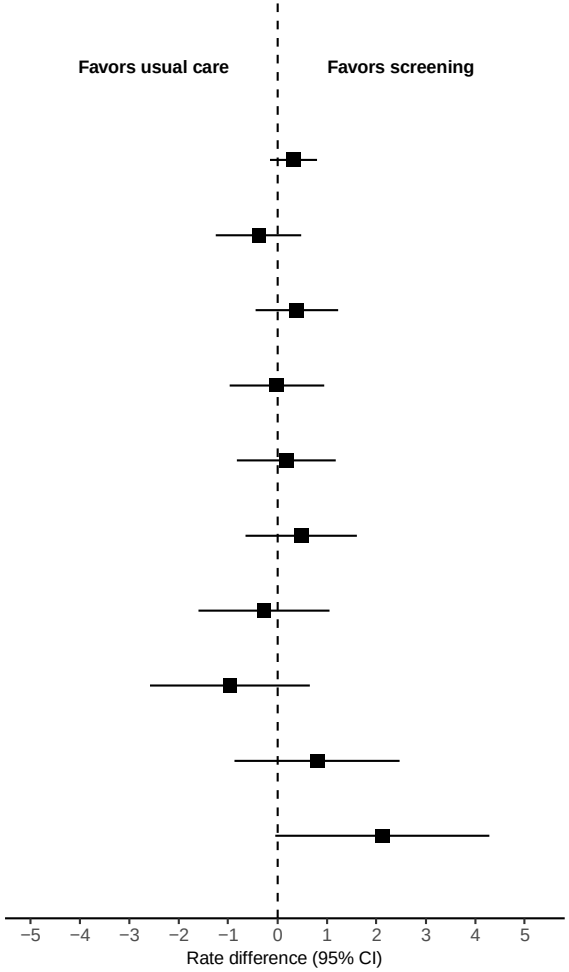
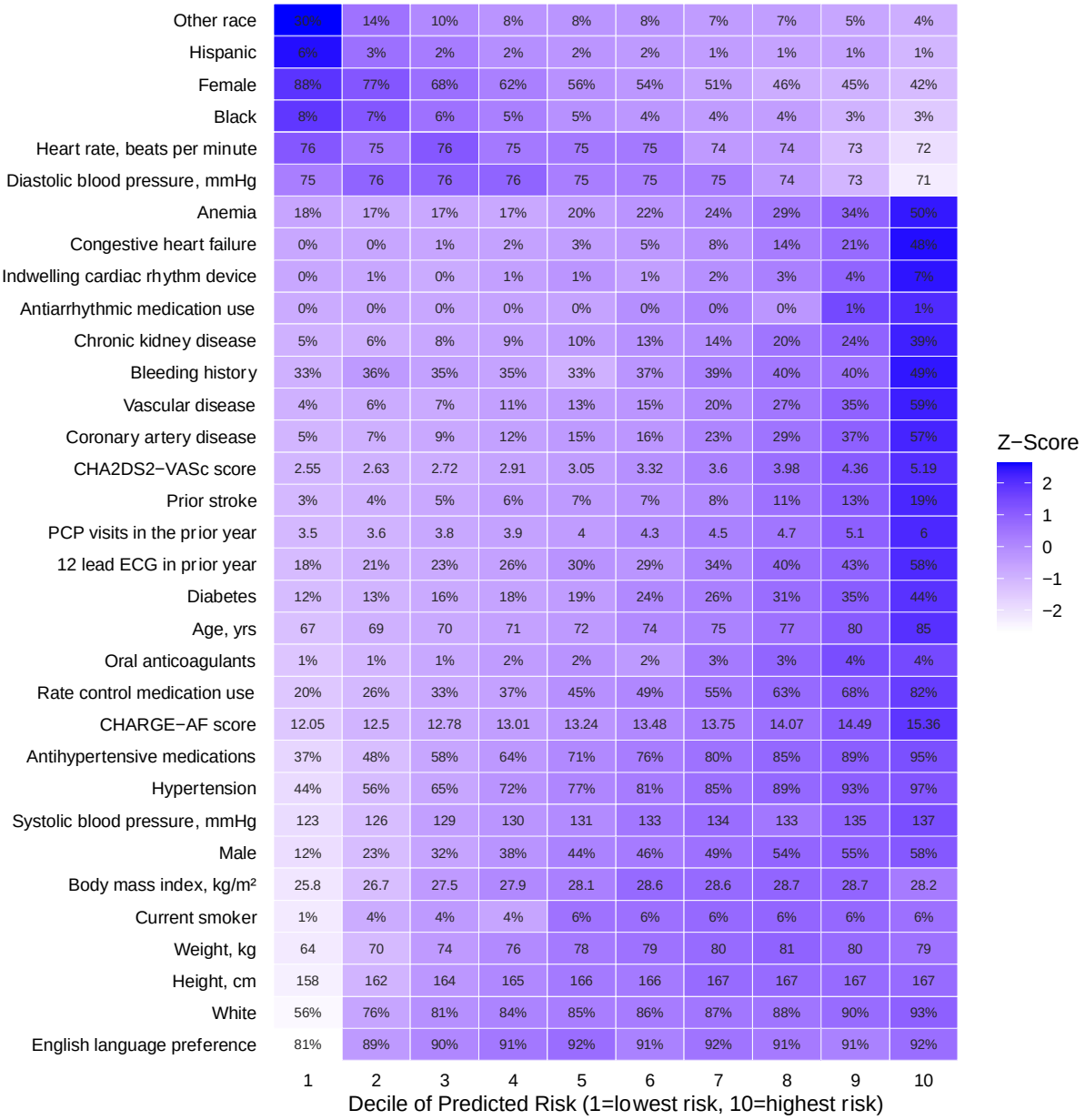


Figure 3: Heatmap of patient characteristics by decile of predicted screening effectiveness

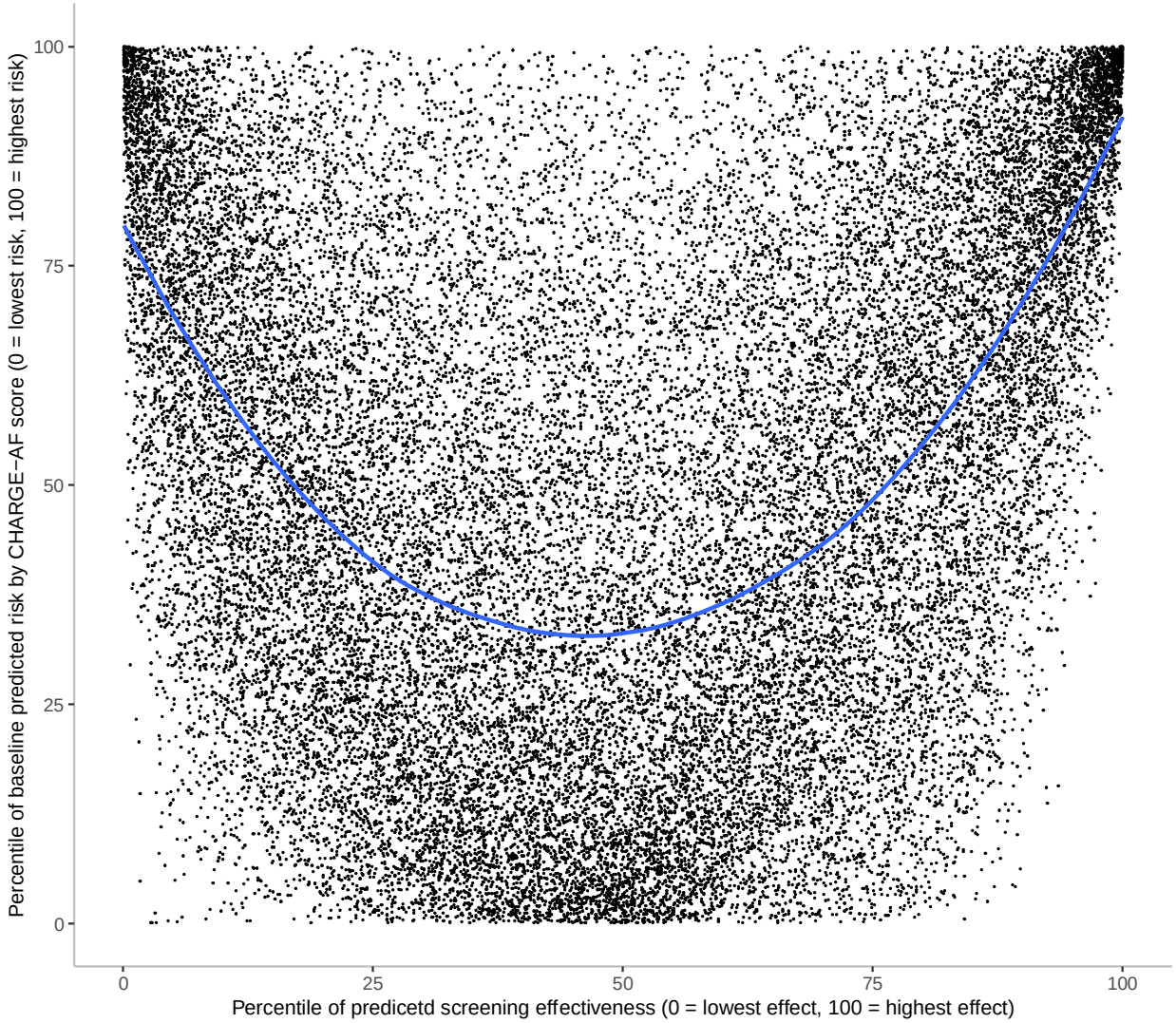
Legend: ECG – electrocardiogram, PCP – primary care physician, CHARGE-AF - Cohorts for Heart and Aging Research in Genomic Epidemiology atrial fibrillation. Indwelling cardiac rhythm devices include permanent pacemakers, implantable cardioverter defibrillators, and implantable loop recorders placed in the prior 3 years. We present the average for continuous measures, and for categorical measures, we present the percent. We color-coded characteristics by their z-transformed value where the darkest and lightest shade represents the highest and lowest value of a given patient characteristic, respectively.

Figure 4: Heatmap of patient characteristics by decile of CHARGE-AF score, predicted risk of AF



Legend: ECG – electrocardiogram, PCP – primary care physician, CHARGE-AF - Cohorts for Heart and Aging Research in Genomic Epidemiology atrial fibrillation. Indwelling cardiac rhythm devices include permanent pacemakers, implantable cardioverter defibrillators, and implantable loop recorders placed in the prior 3 years. We present the average for continuous measures, and for categorical measures, we present the percent. We color-coded characteristics by their z-transformed value where the darkest and lightest shade represents the highest and lowest value of a given patient characteristic, respectively.

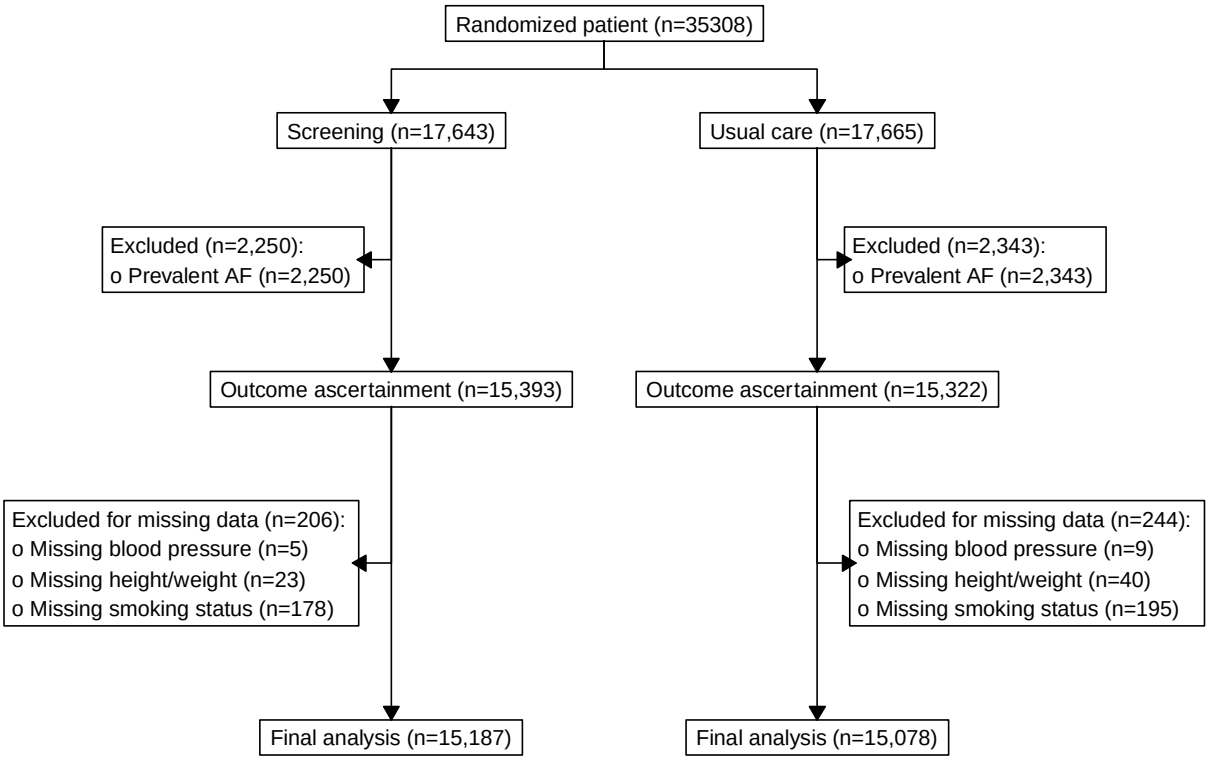
Figure 5: Scatter plot of individuals' predicted screening effect by predicted AF risk



Legend: Each dot represents a study participant and their percentile of screening effectiveness by their percentile of baseline risk. The blue line is a locally estimated best-fit line (i.e., LOESS) with a span of 0.75.

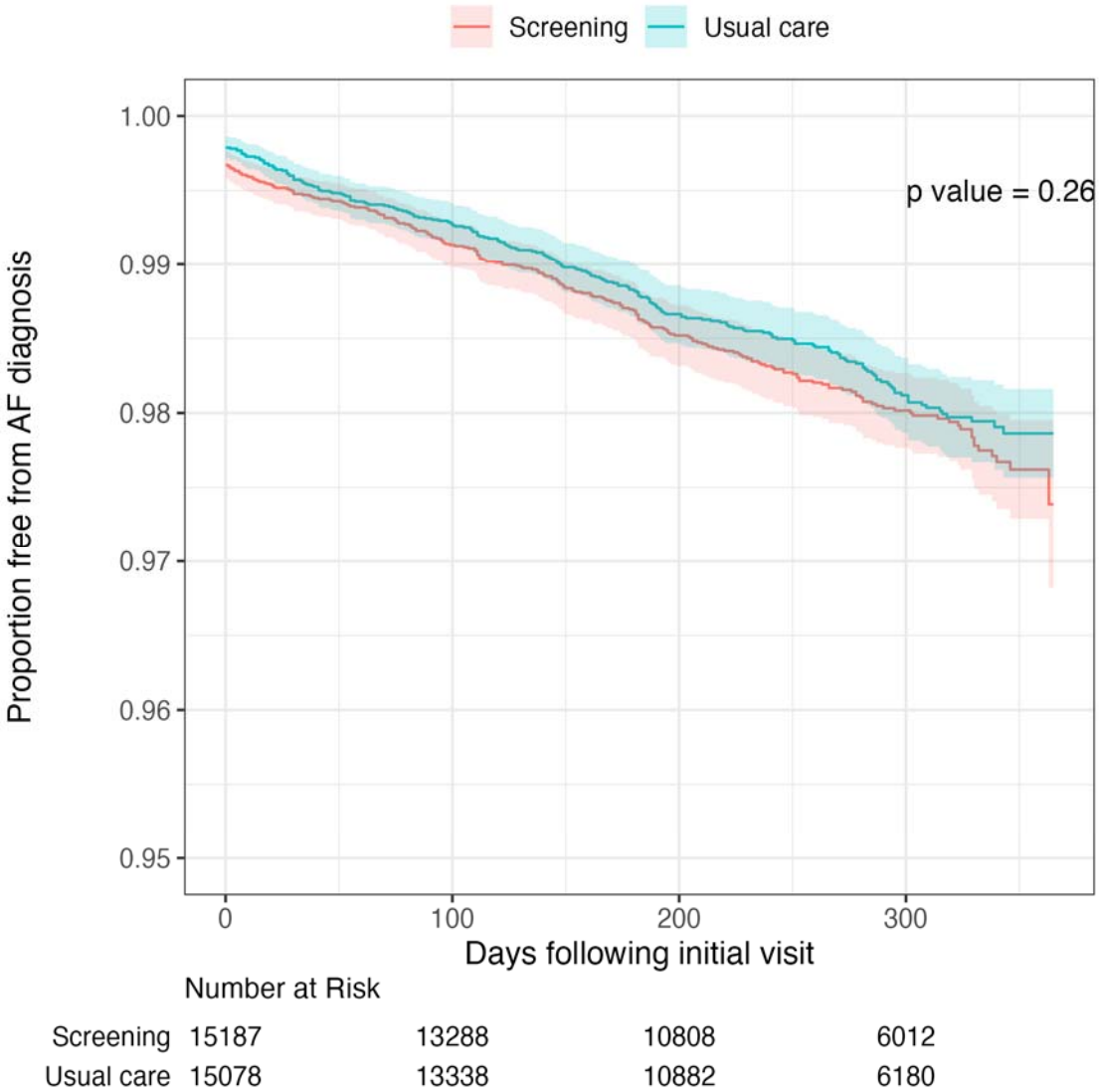
SUPPLEMENTAL MATERIAL

Figure S1: Consort diagram



Supplemental Methods 1

We used a survival model with time-to-atrial fibrillation as the outcome to test the robustness of the modified Poisson regression model used in the primary analysis. The analysis demonstrated no difference in time-to-atrial fibrillation diagnosis by randomization arm supporting the primary modeling approach. The Kaplan-Meier curve demonstrates no statistical or clinically meaningful difference in time-to-AF diagnosis.



Further, among those with a new diagnosis of AF during the study follow-up, there was no statistical or clinically meaningful difference in the distribution of time to AF diagnosis.

	Screening n = 262	**Usual care** n = 238	**p-value**
Time to AF diagnosis, median (Q1, Q3)	110 (14, 193)	118 (28, 192)	0.35

Figure S2: Receiver operator curves of constituent models used to calculate effect scores using leave one out cross validation

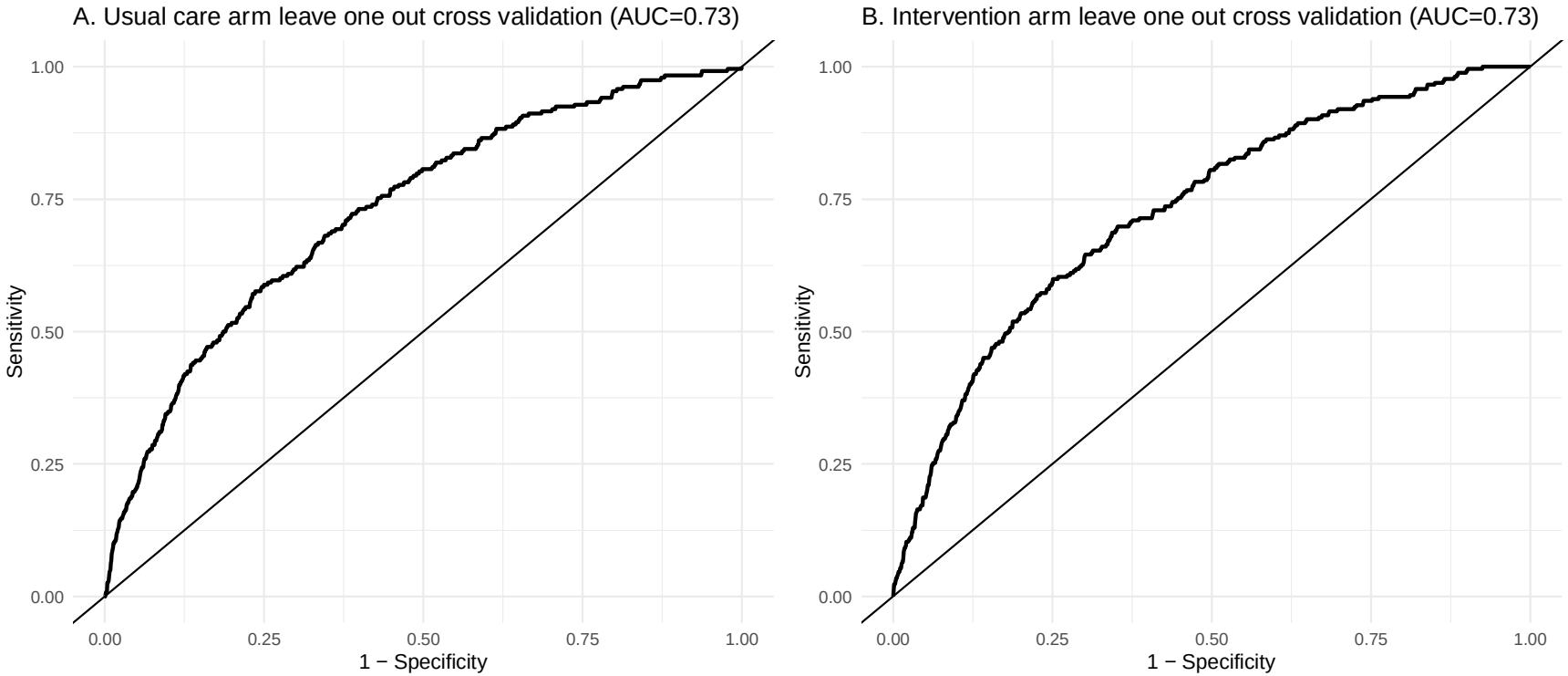


Figure S3: Receiver operator curves of CHARGE-AF risk model in usual care and intervention arm

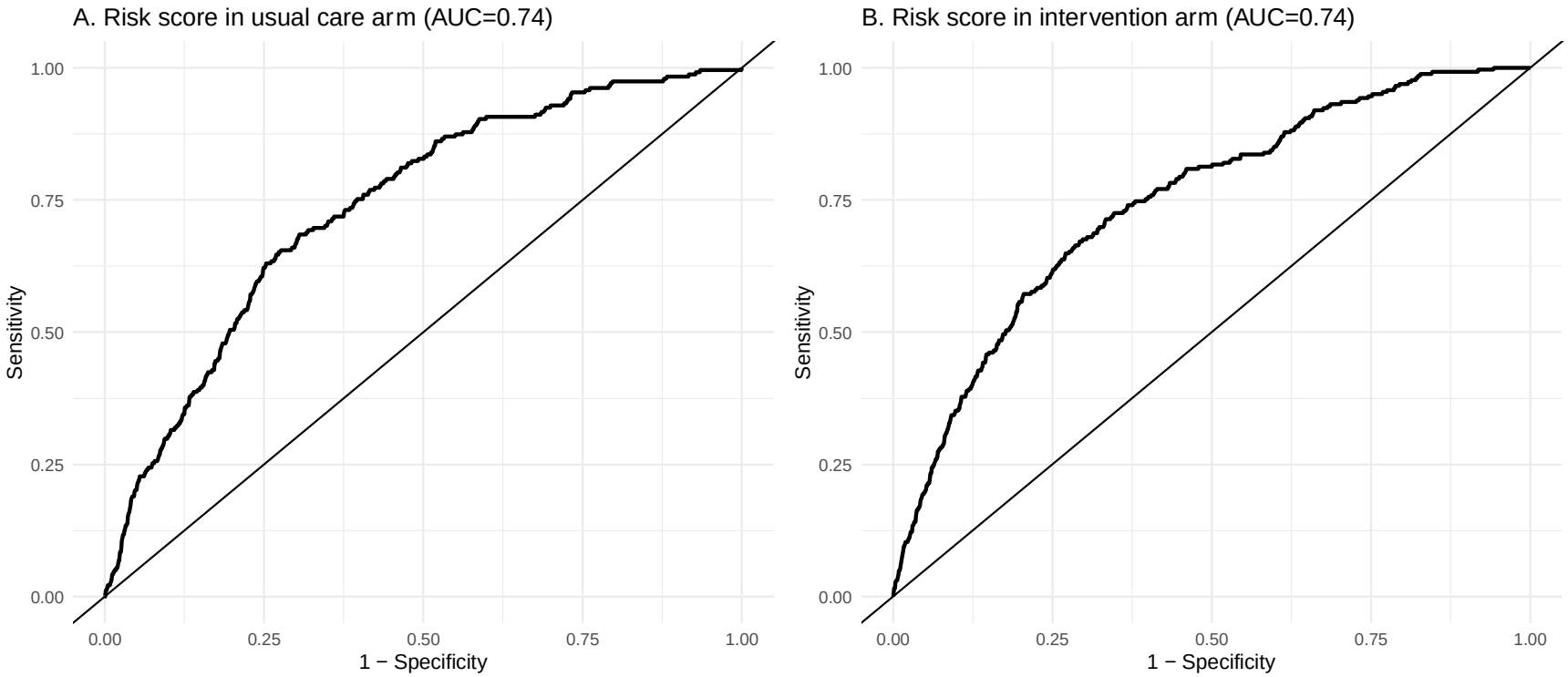
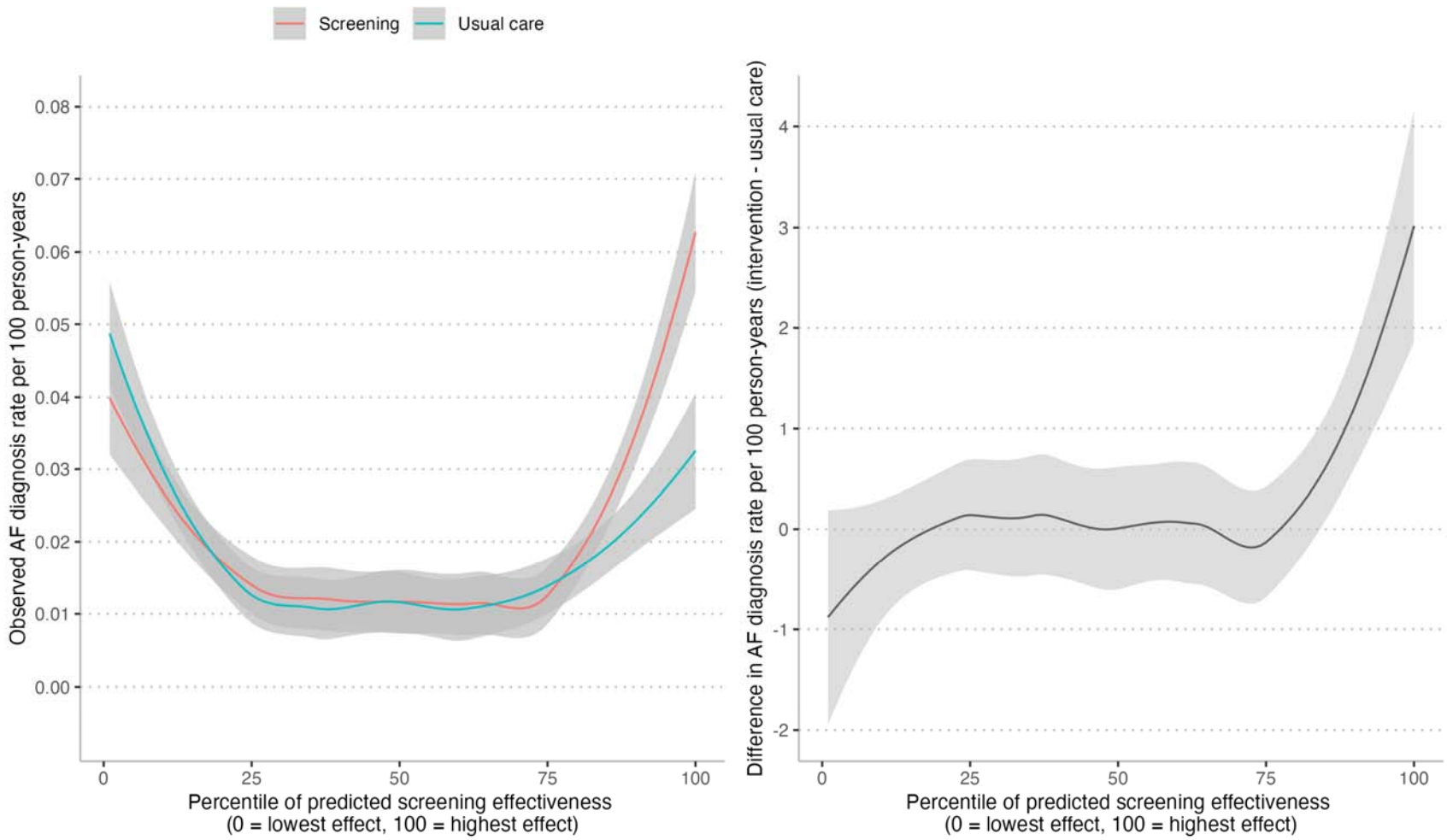


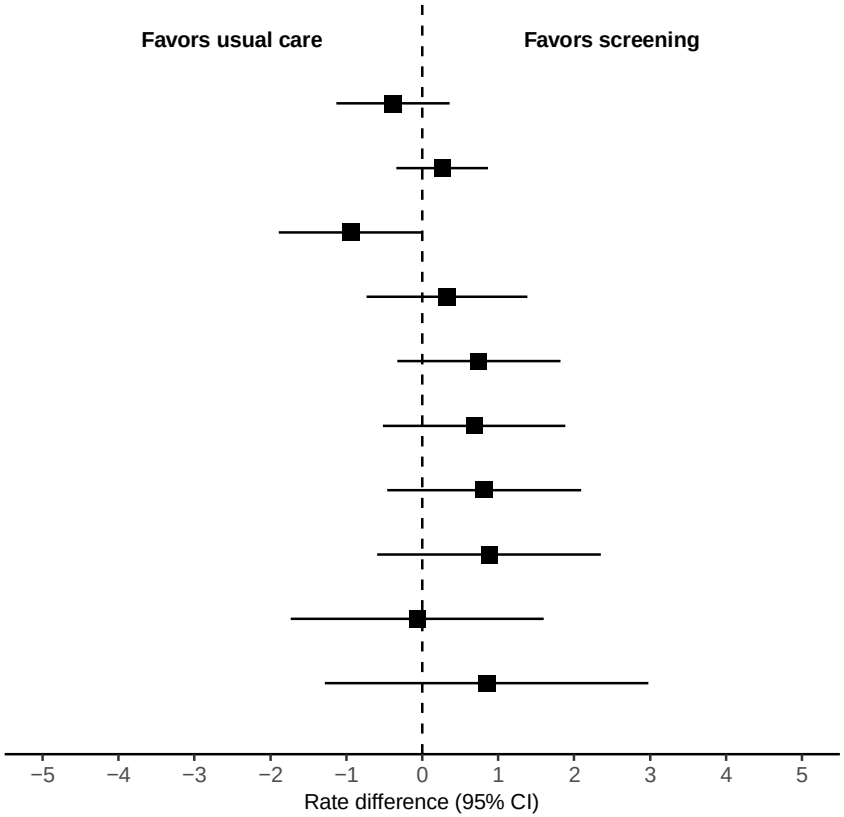
Figure S4: A Continuous Evaluation of the Effect-Based Model



Legend: This is a continuous representation of the data presented in Figure 1 (where it is presented as deciles of predicted screening effectiveness). The left panel displays the diagnosis rate by randomization arm by percentile of predicted screening effectiveness. The line is a locally estimated best fit line (i.e., LOESS) with a span of 0.75. The right panel displays the difference in AF diagnosis rates; values above 0 indicate percentiles where screening is more effective than usual care.

Supplement Figure S5: Heterogeneity of Screening Effect using an Internally Developed Risk-based Model

Decile of baseline risk	Usual care (observed events per 100py)	Screening (observed events per 100py)	Observed rate difference (95% CI)
1 (lowest)	0.83	0.44	−0.39 (−1.13 to 0.36)
2	0.33	0.59	0.26 (−0.34 to 0.87)
3	1.63	0.69	−0.94 (−1.89 to 0.01)
4	1.29	1.62	0.33 (−0.73 to 1.39)
5	1.18	1.92	0.74 (−0.33 to 1.82)
6	1.62	2.30	0.68 (−0.52 to 1.89)
7	1.85	2.66	0.82 (−0.46 to 2.09)
8	2.61	3.49	0.88 (−0.59 to 2.35)
9	4.02	3.95	−0.07 (−1.73 to 1.60)
10 (highest)	6.29	7.14	0.85 (−1.29 to 2.98)



Legend: The CHARGE-AF score has a potential disadvantage because it was developed externally, while the effect-based score was developed within the study. We tested the robustness of our primary finding by using an internally optimized risk model. Recall that we created a risk model to estimate the likelihood of AF under usual care when estimating the effect-base score (a “VITAL-AF risk score”). We find that the VITAL-AF risk score and the CHARGE-AF score are highly correlated (Spearman coefficient 0.85). Further, when we repeat the analysis, we find no heterogeneity of screening effect when using the internally developed risk score.