Dear Dr. Feeney,

To start we would like to thank you and the reviewers for the useful comments. We are very grateful for the opportunity to revise our manuscript. Our responses to the questions and comments of the editors and reviewers are detailed below.

One returning question throughout the responses regarded the test characteristics of the different screening methods. We consider these results as part of the diagnostic study nested within our trial. We prefer to publish these results separately from the results of the RCT, because mixing the trial results with the diagnostic characteristics of the screening instruments would make the manuscript less focused and too lengthy.

We hope our responses provide a clear answer to the gaps in the original manuscript. We strongly feel that the revision has improved our paper.

Sincerely,

On behalf of all authors,

Steven Uittenbogaart, MD
Corresponding author
Department of General Practice/Family Medicine
We performed both an intention-to-screen analysis (including all eligible patients, regardless of whether they were screened) and a per-protocol analysis (only including screened patients, i.e. roughly the percentage mentioned in the comment above). We will explain this choice below.

We randomised the practices (note: not the individual patients) to perform the screening programme or not. We choose this type of cluster randomisation to avoid contamination within practices. Inevitably, in the practices with a screening programme not all patients would be screened, because a part of those patients would not visit the practice in a year; some would refuse to participate; others would visit the practice at too busy a time; and some patients would move to nursing facilities or even die. In the latter case they were considered lost to follow-up, which partly explains the difference in mean age between screened and not screened patients.

As a consequence of our design, we had anticipated that a considerable proportion of eligible patients would not be screened, not only because of the aforementioned reasons, but also because we did use extensive exclusion criteria. We chose to compare the effect of screening between the two groups of potentially eligible patients as a whole in our intention-to-screen analysis (primary outcome). This choice was made to increase the generalizability of our findings to everyday practice situation.

In our intention-to-screen arm, we lost 2.6% (244/9,218) in follow-up. Of the eligible patients 55% (5,112/9,218) was not screened. So while the low screening rate might decrease the effect of the intervention [efficacy], it does not create selection bias in this analysis.

Our per-protocol analysis excludes the not-screened group. Table 1 shows that screened patients were younger and had less comorbidity than patients who were not screened. Therefore, we corrected for age, sex, and history of hypertension, diabetes mellitus, stroke (TIA and/or stroke), thromboembolism and heart failure in this analysis, as these are potential confounders.

In the Discussion section, we discuss the implications of the large proportion that was not screened.

We did not investigate why eligible patients were not screened. In the first paragraph of 'strengths and weaknesses' we discuss this issue.

From our experiences during the performance of the study, we conclude that the low inclusion rate was more due to organisational issues rather than patients not willing to participate. We presume that due to a high workload practices screened fewer patients than instructed. Additionally, a proportion of eligible patients did not visit the practice during the study year and could therefore not be screened. We do not know what proportion of eligible patients did not visit the practice during the study year. We are not able to analyse this proportion afterwards, because we cannot collect new data of these patients due to privacy issues.
3 Please describe in as much detail as possible what was the approach to detecting atrial fibrillation in the "usual care" arm. In the discussion you assume that your screening intervention was no better because "usual care" is particularly good in the Netherlands, or at least the practices in this trial.

In the Methods section we describe the 'usual care in the Netherlands'. We assume that usual care in the Netherlands is particularly good, due to the existence of structured disease management programmes. Patients with a cardiovascular disease or at increased risk for cardiovascular disease, patients with diabetes and patients with chronic respiratory disease participate in these programmes and visit the practice at least once a year. In general, over 65% of patients with a registered disease or risk factor, participate in these GP run programmes and visit the practice at least once a year (https://ineen.nl/assets/files/assets/uploads/20180528_Rapport_Transparante_Ketenzorg_2017.pdf, in Dutch language). A proportion of the patients with risk factors not included in these structured programmes receive follow-up by a specialist in the hospital. Patients in these programmes are mostly 65 years and over. The Dutch College of General Practitioners recommends assessing heart rhythm, with pulse palpation or sometimes with an ECG, in every patient when measuring the blood pressure. Moreover, the Dutch College of General Practitioners recommends assessing heart rhythm in every patient with shortness of breath, reduced ability to exercise, palpitations, dizziness, light-headedness, syncope, chest pain and TIA or stroke.

We adjusted the methods paragraph with these additional details. It now reads:

“The Dutch College of General Practitioners guideline for AF recommends assessing heart rhythm in every patient with shortness of breath, reduced ability to exercise, palpitations, dizziness, light-headedness, syncope, chest pain and TIA or stroke as part of the usual diagnostic work-up. Further recommendations are to assess the heart rhythm in each patient when measuring the blood pressure, a method of case-finding. Otherwise, screening is not recommended.”

4 Please give more information on those in the screening arm who were not screened. For example, what proportion had attended a GP appointment at some time in the year?

We describe differences between screened and not-screened patients in table 1. We do not know the proportion of patients not visiting the practice during the study year neither do we know the number of patients not willing to participate. See also our answer to remark 2.

5 One practice who terminated participation directly after randomisation does not appear to be shown in figure 1. Please double check this.

This was an oversight. Thank you for this remark. We corrected this error in the figure 1.

6 It seems surprising that the screened group had less comorbidity than the non-screened. Please comment on this and why it might be.
We reflect on this in the ‘Strengths and weaknesses’. We suggest that a proportion of eligible patients was not screened because they were not able to visit the practice due to frailty or having a pacemaker. This is a direct consequence of our choice of screening only patients who visited the practice.

Another explanation could be that younger and healthier patients were more likely to participate in screening – the so-called ‘worried-well’.

In the ‘Strengths and weaknesses’ we changed and added: “…but also meant that a proportion of our selected patients was not eligible for screening due to inability to visit the practice (e.g. due to frailty) or due to other reasons (e.g. having a pacemaker). The patients that participated in opportunistic screening were younger and had less comorbidity (‘worried well’) than patients who were not screened.”

7 It appears that secondary outcomes were not reported as registered, can you rectify this?

Indeed, we have registered another primary outcome and various secondary outcomes in the https://www.trialregister.nl/trial/4776. These will be topics for forthcoming publications. In the current paper we do report various additional analyses, which all support the primary analysis.

8 Please provide a PPI involvement statement and place it at the end of the methods and not in the end matter. Please provide a dissemination plan in the end matter. Please follow our guidance to authors for both areas. Please also consider acknowledging your participants as the research would not be possible without them.

Thank you for these corrections. We have adjusted our manuscript according to the guidance for authors. In addition we added a sentence to our PPI statement, because there was patient involvement in the grant application stage.

9 Dissemination

We added our dissemination plan in the end matter. "The project will lead to two PhD theses by GPs. Results have and will be presented at national and international conferences. We will publish written reports in international and national peer-reviewed medical journals. Evidence from this study will be used in national and international guidelines on atrial fibrillation and cardiovascular risk management. Our network includes a co-author of the Dutch NHG guideline on AF and a document reviewer for the 2010 ESC Guideline on AF. One co-authors (HS) is part of the advisory board of the ‘Staff Training Programme Cardiovascular Disease’ (Kaderopleiding Hart- en Vaatziekten), and member of the ‘GP Advisory Group Cardiovascular Diseases’ (Hart en Vaatziekten Huisartsen Adviesgroep, HartVaatHag). The D2AF study group is part of the AF-SCREEN international collaboration network. Thus, results will find their way to vocational training programmes and Continuing Medical Education (CME) courses in the Netherlands and abroad. Once we have published our results, we expect that
interest will be expressed by newspapers and lay magazines. Furthermore, the results of the study will be relevant for the Netherlands Heart Foundation (Nederlandse Hartstichting) and we expect that we will share information through their website, which is easily accessible for patients as well.”

Additional remark

We noticed in the comments of the reviewers that the terminology ‘single time point screening’ raised some questions. Therefore, we changed the term into ‘opportunistic screening’. The confusion about the term single time point screening was quite understandable, because including an additional Holter monitoring intervention made our intervention not a pure single time point intervention. We hope that this change clarifies some of the misunderstandings that were possibly raised.
The reviewer makes a valid point that the results of our trial are more likely to answer the question whether offering the intervention works. In addition, our per-protocol analysis inescapably suffers from some bias, despite our efforts to correct for this (see point 1 of Additional comments by the committee).

Although the proposed simulated analysis to estimate the actual power might be a good idea in general, we believe it has no added value to our study. We detected no clinically relevant effect in our trial, not even if all patients would have been screened. This is illustrated by the outcomes of the intention-to-screen (ITS) and per-protocol (PP) analyses. In the ITS analysis, the AF detection rates were 1.6% (144/8874) in the intervention group versus 1.5% (139/9102) in the control group, resulting in an odds ratio of 1.06 (95%CI 0.84, 1.35). This is much smaller than the odds ratio of 1.8 (2.3% versus 1.3%) that was indicated as clinically meaningful (see sample size calculation). If we look closer to the intervention group, then we see that the AF detection rate was 1.2% (48/4085) in the screened group, resulting in an adjusted OR of 0.86 (95%CI 0.61, 1.20), after correction for several potential confounders, such as age, sex, and history of hypertension, diabetes mellitus, stroke, thromboembolism and heart failure. In addition, the non-screened group had an AF detection rate of 2.0% (96/4789), which is higher than the rate in the screened group. This could (partially) be explained by the difference in characteristics of these two groups (see Table 1).

All these results, including the ones from the multiple imputation and the sensitivity analyses, show that there was no clinically relevant difference between intervention and care-as-usual group, which makes a power analysis less useful as power is only useful for detecting clinically relevant differences (non-relevant differences do not need to be detected with enough power and do not need to be statistically significant). If it is nevertheless desired, we can perform an IV analysis, but that will take us some additional time (without any expected change of results).

2 The near-future of AFib detection is prolonged monitoring, as by Zio patch, watches, etc. While the one-time screening technique will be valuable for a long time, I could use a little more context about what that might find. Maybe another line or 2 in the discussion, maybe an analysis if they think of one.
We changed our discussion to give more context to the findings of prolonged monitoring:

"Two week Holter monitoring detected an additional 1.5% of AF. Our findings demonstrate that single time point screening misses silent and paroxysmal AF. This is in line with STROKESTOP and ACE1950, studies screening respectively a 75- and 65-year old population using twice daily, intermittent ECG-recording for two weeks. In both studies, the detection rate at the index visit was 0.5%. In the two weeks of subsequent screening, an additional 2.5% and 0.9%, respectively were detected. In the mSTOPS-trial (mean age 73.5) a two-week monitoring using a Holter patch without an index visit detected newly detected AF in 4.7% (43/906)."

3 The comparison with literature was a bit long, but also missed something I’d find important, which is some attempt to disentangle if the results were due to "one-time AFib screening missed rare AFib" or "we screened for a problem that really isn’t at all common.” Could the discussion or another analysis help me work through that a touch?

We had a very sensitive method of screening, reflected by the fact that the 294 ECG’s in patients with three negative index tests did not detect any additional cases. If AF would have been present at the time of screening we would have detected it. Our detection rate was 0.63% in the screened patients. This is much lower than the detection rate of 1.44% (95% CI, 1.13%–1.82%) found in the review by Lowres et al (Plos Med 2019, https://doi.org/10.1371/journal.pmed.1002903). The SAFE study found a detection rate of 2.21% (1.68 – 2.88). This leads us to the conclusion that patients with undetected AF were hardly present in our population. However, prolonged screening detected an additional 1.5% of paroxysmal AF, all of whom had three negative index tests (pulse palpation, hand-ECG and blood pressure measurement) and a negative ECG, showing that single time point screening does miss paroxysmal and silent AF.

We have revised the discussion to be more on point and address the issue mentioned above: did we miss cases of undiagnosed AF or was it not there to miss? In short we can say it is a bit of both. We adjusted our ‘Comparison with literature’ section in several places.

4 If possible, I’d be curious to know how much of the AFib they found was due to each of the 3 screening techniques. (It’s possible they don’t have this. It’s also possible they’re saving it for a follow-up paper, which I’d find reasonable.)

We are currently in the process of finalizing an article describing test characteristics of the screening techniques we used in this trial. We are sure that it will satisfy the reviewer’s curiosity, but choose not to present these figures in this manuscript, because it would make this manuscript confusing, lengthy, and hard to read.
This paper presents the results of the D2AF randomised controlled trial of opportunistic screening for AF in general practices in the Netherlands. In many aspects it resembles the SAFE study performed in general practice in the UK a little over 10 years earlier. But the outcome was not the same, with no difference in rate of new AF between intervention and control practices. This is disappointing, and in retrospect, somewhat predictable. Appropriately, the authors have discussed a number of possibilities for the failure to demonstrate an effect of the screening strategy. While this might be interpreted as a failure of opportunistic screening, it might alternatively be interpreted as a success of the Netherlands health system in introducing opportunistic pulse/rhythm checks since 2013, with a contribution from the increased AF awareness brought about by availability of NOACs and their strong marketing by pharma.

One of the most important changes noted is that the prevalence of AF increased from 8% two 10% in all practices, and much higher than the earlier Lowres systematic review (ref #22) which could be indicative of an increased detection rate (amongst other possibilities). This would certainly impact on reducing efficacy of opportunistic screening. Also, the significantly lower age group of the actually screened patients would impact the detection rate which is extremely sensitive to age (Lowres et al PLOS Med 2019). That may in part explain why the per-protocol analysis showed an even lower rate of AF detection in the actually screened group, and should be included in discussion.

The reviewer raises a very justified question and is completely right that the lower age of the actually screened patients has an impact on the detection rate. In our discussion we added:

"Why did we not detect more new cases of AF with the our extensive and very sensitive screening protocol? This might be partly explained by the lower age of the actually screened patients, since detection rate is very dependent on the age of the population."

Only 45% of the intention-to-screen group actually received screening and this is an important limitation, as any effect is diluted by non-screened patients. Admittedly, the same limitation occurs with systematic screening in the SAFE study, as well as STROKESTOP studies with only 50% attending an invitation for screening). In the SAFE Study, it is important to note that 68% of manually flagged patients were actually screened by pulse palpation. The rate of EMR-prompted opportunistic screening in D2AF is quite a bit lower, but very similar to the upper limit of opportunistic screening we have observed in our own studies using electronic prompts (Orchard et al: BMC Family Practice 2019, Table 1, and Af-SMART, JAHA 2019), and was higher in rural practices. It was much lower in the recent Kaasenbrood study from the Netherlands (10.7%, ref #21).

We thank the reviewer for these constructive suggestions. We have added details to our study. In addition to these thoughts, although 68% of flagged patients were screened in the SAFE study, the follow-up of an irregular pulse with a 12-lead ECG was much higher in our study (92% vs 66%), so the complete screening figures are in reality quite similar. We altered our discussion to read:

"The intervention practices only screened 45% of eligible patients. This inclusion rate was higher than in the AF-SMART and the IDEAL-MD studies, but lower than the SAFE-study."
Also some information on proportion of rural vs metropolitan practices needs discussion. Reasons might include the somewhat more onerous tasks of having 3 types of screening test in rotating order, and a 12-lead ECG and 2 week Holter if negative (though only in 10%).

As to the question of rural vs metropolitan practices: this is of less importance in the Netherlands. With a population of 17 million in a country 1/185 the size of Australia, urbanization is high across the Netherlands. We did stratify practices by region (north vs south, with a higher level of urbanization in the north). In the regression analyses, region was not a significant factor.

In the Netherlands the central bureau for statistics (CBS) looks at urbanisation based on postal codes, https://www.cbs.nl/nl-nl/maatwerk/2018/30/kerncijfers-wijken-en-buurten-2018. Areas are labelled on a scale from 1 (not urban) to 5 (strongly urban). When looking at urbanisation levels of participating practices, we found no differences between the intention-to-treat (3.1) and usual-care group (3.0).

In the 10% random sample of negatives, only half (111 patients) received the 2 week Holter, and interesting to note that 4 were positive, however none of the 155 patients who had 14 day Holter following a negative 12-lead ECG, but a positive screening test (pulse, BP, or handheld ECG). This gives some indication of the additional information that a 2 week Holter period continuous ECG could provide: i.e. 4/266 = 1.5% on top of the rate with a single timepoint. This same 2-week continuous monitoring was used in mSTOPS, and is being used in the GUARD-AF study: they are more relevant than STROKESTOP or ACE studies, which used twice daily ECG rhythm snapshots for 2 weeks. Both studies mentioned using an external 2 week ECG patch should be discussed and referenced.

Again, we thank the reviewer for his constructive suggestion and added the mSTOPS-study in the discussion. However, in the mSTOPS study patients received a patch by post and AF was not ruled out by initial 12-lead ECG. The references we used, discuss the added value of prolonged screening after a negative initial 12-lead ECG. Because we do not know whether any selection took place when inviting patients for the patch monitoring, we have to be cautious when interpreting the findings.

We added the mSTOPS reference to our discussion. Because results for the GUARD-AF study are still pending, we did not use this reference.

“Two week Holter monitoring detected an additional 1.5% of AF. Our findings demonstrate that single time point screening misses silent and paroxysmal AF. This is in line with STROKESTOP and ACE1950, studies screening respectively a 75- and 65-year old population using twice daily, intermittent ECG-recording for two weeks. In both studies, the detection rate at the index visit was 0.5%. In the two weeks of subsequent screening, an additional 2.5% and 0.9%, respectively were detected. In the mSTOPS-trial (mean age 73.5) a two-week monitoring using a Holter patch without an index visit detected newly detected AF in 4.7% (43/906).”

Routine care in the Netherlands since 2013 has recommended pulse taking during BP checks which equates to opportunistic screening (ref 14). The statement on page 11 line 15 that no screening is recommended in Dutch guidelines is therefore not accurate. The authors do state that the advice to check the pulse, as well as marketing of NOACs,
will have increased AF awareness amongst control practices leading to a higher new AF rate than in the SAFE trial control arm, and this is quite likely.

We disagree somewhat in the use of terminology. We think that rhythm assessment during BP-measurement equates case-finding, because it does not target every patient presenting in the GP’s office, as systematic screening would. We adjusted our methods section to read:

"Further recommendations are to assess the heart rhythm in each patient when measuring the blood pressure, a method of case-finding. Otherwise, screening is not recommended."

6 Blinding of the control practices to the 200 patients selected is unlikely to have had any impact on the Hawthorne effect. Just by being in the study, it Is quite likely that the Hawthorne effect was operating in the Control practices. This could be checked by looking in the non-selected patients in the Control practices to determine if new AF incidence rate was the same. But it would be interesting to look at data from other practices not in the study (if possible) to see what the AF prevalence and annual new AF diagnosis rate was during the study period. It might be something that has happened across the board in the Netherlands.

The reviewer is right that by just being in the study the Hawthorne effect may have been operating in Control practices. We made efforts to minimise it. In the control practices neither GPs nor patients were aware of who was selected. Therefore, we cannot check for a possible Hawthorne effect in the control practices by checking the files of non-flagged patients, as both would be influenced by the same effect.

We agree that it would be interesting to look at data of other practices.

7 It is possible that new AF was missed in the active screening arm. How many of the patients with 1 positive screening test had a 12-lead ECG performed immediately after that test? If delayed, it could mean that paroxysmal AF found at the initial screening was missed. Additionally, how many of those with at least 1 positive screening test had a positive MyDiagnostick handheld ECG? Given the known specificity, I cannot believe that only 26/448 handheld ECG positives showed AF on the 12-lead ECG if done immediately afterwards (i.e. false positive rate of 94%). This would be worth isolating. There also needs to be a review of the lead 1 ECG rhythm strips compared to lead 1 on the subsequent 12-lead in the 422 patients who might otherwise be considered “false positive”.

Instructions to our sub-investigators were to immediately follow-up with an ECG during the same appointment and we provided the instruments and facilities to do that. We trained the ancillary staff of participating practices on how to perform an ECG. So practice assistants were able to perform an ECG while the GP was able to continue with his/her activities.

We added the following sentences to the Methods section:

"We instructed to perform the 12-lead ECG immediately after the index tests. The ECG result was transferred digitally and assessed by an experienced assessor, supervised by a cardiologist. A second cardiologist re-assessed all AF-diagnosed ECG’s and a random sample of negative ECG’s. In case of disagreement, a third cardiologist decided. The GP’s office received the ECG and a report of the assessment by the cardiologist.”
We collected the time of performance of the three index tests and of the performance of the 12-lead ECG in the electronic case report form (eCRF). We added the following text to the Results section:

"The median time between registration of the performance of the first index test and the performance of the 12-lead ECG was 26 minutes."

Figure 2 shows that n=113 patients had a positive MyDiagnostick result (red light). For your information (not written down in this study): of those 113 patients indeed only n=26 had AF on the 12-lead ECG (false positive rate 87/113). In patients with a negative MyDiagnostick (green light), no AF was seen on the 12-lead ECG (false negative rate 0). For further details we refer to our diagnostic study showing results of our screening methods in more detail. This study will be offered for publication soon.

Thank you for your suggesting for a review of the one lead ECG compared to the lead one on the subsequent 12-lead. We are currently working on it.

---

8 On page 11 stated that the diagnosis of AF is accepted only if confirmed in a hospital letter. Is it possible that some AF diagnoses were made in general practice only and not referred to hospital? This is certainly the case in our study in the UK (Martinez et al, Heart, 2015), where 39% of incident AF diagnoses were made in general practice, and only a proportion were sent to hospital.

---

Diagnoses of AF were made both in general practices with an ECG available or in hospitals. We specified our methods to read: "We accepted a diagnosis of AF if it was confirmed with an ECG – either made in primary or secondary care. We also accepted or description of the diagnosis of AF in a hospital letter.”
1 This is a very interesting and timely pragmatic trial of AFib screening. The topic is impactful. The findings indicate that point-of-care screening in primary care practices in the Netherlands may not identify many new cases of AFib. The reasons for this are not addressed in the manuscript, but major limitations of the study may have limited the effectiveness of the screening intervention.

We thank the reviewer for appreciating our study. We agree that on first sight there may be limitations. However, we extensively discuss our finding that opportunistic screening did not yield more AF than usual care. In the discussion-paragraph ‘comparison with literature’, we question: “Why did we not detect more new cases of AF by opportunistic screening compared with usual care?” Then we elaborately try to answer this important and justified question.

2 First, the proportion of individuals actually screened in the screening arm was low (45%). Second, the fact that 3 index screening tests were used inherently made the test a bit less pragmatic, perhaps onerous in busy primary care practices, and may have biased the screened sample in favor of healthier participants either more motivated to participate, less critical medical issues to address at their encounters, or with more time to spare at their encounters – indeed the profile of screened patients favored a healthier and younger cohort than the non-screened individuals in the screening arm. I have several queries:

   The risk of contamination is not well-addressed. Can the authors clarify the geographic proximity of the clinics? Can the authors address whether clinicians sometimes rotated in both the screening and usual care clinics? What type of education was provided to clinicians in the trial over the course of the study?

In a small country as the Netherlands, geographical proximity of participating practices is inevitable. However, we recruited across the whole country in order to minimise geographical influence. In the recruitment of practices, we did not allow multiple practices under the same roof to participate separately.
Most clinicians however work in one practice and do not rotate at all. By exception, one or two locums might have been rotating in both screening and usual care practices but the numbers will definitely have been very low and not effecting the results of our study. Education was not provided during the study year. During the study year, we stimulated intervention practices by telephone and email to include patients. Control practices were not contacted during the study year.

3 Cross-overs: Can the authors address whether patients sometimes "cross-over" between screening and usual care clinics and how these were handled?

Cross-over of patients was nearly impossible. We preselected patients at the beginning of the study year of a practice. In the Netherlands, patients register with one general practitioner who coordinates the care of the patient. In order for a patient to cross-over, he or she would have to unregister at his or her current general practice, subsequently register at another practice which would have to participate in our study and start with the study after the previous practice did. On top of that, the patient would again have to be in the selection of 200 patients. We cannot check if this has happened because we have no identifying information of patients due to privacy regulations. However, we believe chances of cross over happening, are very small.

4 Missing data: I did not follow the description of missingness and what exactly was imputed.

We imputed the outcome AF in case of loss of follow-up, for instance if a patient moved or died before the end of the study year. As a sensitivity analysis, we performed multiple imputation (MI) of missing outcome data, where the missing outcome (AF, yes/no) was imputed using the fully conditional specification (FCS) method, where missing AF was imputed using the association between AF and group, age, sex, and the stratification
variables that was observed in the group of patients with complete data. After the creation of five complete datasets using MI with 20 iterations, the analysis was performed on each complete dataset, after which the results were pooled and compared with those from the original analysis. Accordingly, we have added this information in the manuscript.

5 Temporal effects: Since awareness of AFib is likely to have changed substantially over time, were there any treatment interactions with time?

Our trial ran between September 2015 and August 2018 for a one-year period in each individual practice; intervention practices and control practices started at the same time. If there had been temporal effects in this relatively short period, these effects would have been equal for both arms in this cluster-randomised trial.

6 Was there evidence of any heterogeneity in screening uptake in the intervention arm of the study?

There was considerable variation in the inclusion rate of patients between practices. The results section reads:

"In the intention-to-screen group, 4,106 out of 9,218 patients (44.5%) actually participated in the screening protocol. The percentage of screened patients varied per practice, ranging from 6.7% to 65.8%. Two practices discontinued during the study year, both because of organisational issues; both were included in the analysis."

7 Electronic prompt: was it guaranteed that the electronic prompts were viewed at each visit? Is it possible the electronic prompts were dismissed or ignored? When was the electronic prompt surfaced, and how was attention to this ensured?

Although electronic health record systems differed between practices, in each system the electronic prompts were visible each time the patient file was opened. Electronic prompts can be ignored and previous studies showed that when ignoring was not possible, participation of practices in a study (and during real practice) would diminish dramatically ((Arts et al, PLOSE ONE 2017, https://doi.org/10.1371/journal.pone.0170974)). We had no means to observe to what extent prompts were ignored, but we do believe it occurred. We monitored the inclusion rates of newly started practices and if these were low, we contacted the sub-investigator to motivate them to act on the electronic prompts.

8 Performance of screening modalities: Can the authors provide a Venn diagram of AFib cases, or a PPV, for each of the index screening modalities?

We are currently preparing a separate paper describing the diagnostic test characteristics of the used methods. A Venn diagram is a good suggestion, for which we are grateful.
Again, we thank the reviewer for these questions. Either a practice nurse, practice assistant or general practitioner executed the screening protocol. In most practices, the GP delegated this task to either practice nurse or assistant. The interpretations of pulse palpation was done by the person who performed the screening. The blood pressure monitor and handheld-ECG had automated software that made interpretation straightforward.

We have added additional information on interpretation of the 12-lead ECGs made as part of the screening protocol. In general, the GP received the report within a few working days. We added:

"The ECG result was transferred digitally and assessed by an experienced assessor, supervised by a cardiologist. A second cardiologist re-assessed all AF-diagnosed ECG’s and a random sample of negative ECG’s. In case of disagreement, a third cardiologist decided. The GP’s office received the ECG and a report of the assessment by the cardiologist."

The rate of anticoagulation in the Afib cases in the screening and usual care arms is an important issue and will be analysed and published in a different paper in the near future.