



Bruk av Apple Watch til påvisning av arytmi

Desember 2020

Innhold

| | |
|---|-----------|
| Oversikt | 3 |
| Innledning | 3 |
| PPG-basert påvisning av arytmi | 3 |
| Teknisk informasjon og funksjonsbeskrivelse | 3 |
| Preklinisk utvikling..... | 5 |
| Klinisk validering | 5 |
| Apple Heart Study | 5 |
| Utformingen til AHS-delstudie..... | 5 |
| Resultater..... | 6 |
| EKG-basert påvisning | 6 |
| Teknisk informasjon og funksjonsbeskrivelse | 6 |
| Preklinisk utvikling..... | 7 |
| Klinisk validering | 7 |
| Eksperimentets utførelse | 7 |
| Studie for klinisk validering av EKG 1.0..... | 8 |
| Resultater..... | 8 |
| Studie for klinisk validering av EKG 2.0..... | 12 |
| Resultater | 14 |
| Konklusjoner | 17 |

Oversikt

Apple Watch-kunder har tilgang til to medisinske programvarefunksjoner for påvisning av hjerterytmie som for eksempel atrieflimmer (AFib): funksjonen for varsling om uregelmessig hjerterytmie og EKG-appen.

På Apple Watch Series 1 og nyere analyserer funksjonen for varsling om uregelmessig hjerterytmie pulldata som registreres av sensoren for fotopletysmografi (PPG), for å identifisere forekomster av uregelmessig hjerterytmie som tyder på AFib. Når Apple Watch registrerer en episode, får brukeren en varsling. På Apple Watch Series 4, Series 5 og Series 6 kan EKG-appen generere et elektrokardiogram (EKG) tilsvarende et elektrokardiogram med én avledning, og deretter klassifisere bølgeformen som sinusrytme (SR), atrieflimmer (AFib), rask puls, langsom puls eller ikke-avgjørende. Med EKG 2.0 er andre klassifiseringer som for eksempel AFib med rask puls og dårlig registrering også tilgjengelige.

Dette dokumentet gir en grundig innføring i disse funksjonene – deriblant testing og validering.

Innledning

AFib, en type uregelmessig hjerterytmie der forkamrene slår ujevnt og noen ganger også raskt, er blant de mest utbredte årsakene til slag. AFib er imidlertid ofte asymptomatisk. Mange kan derfor være rammet av atrieflimmer uten å være klar over det. Kombinasjonen av slagrisiko, manglende symptomer, effektive farmakologiske behandlinger som reduserer slagrisikoen til et minimum, og veksten i markedet for forbrukerelektronikk som potensielt kan registrere AFib, har ført til økt interesse for tidlig registrering av AFib utenfor en klinisk ramme.

Med watchOS 5.1.2 og nyere kan Apple Watch Series 1 og nyere registrere uregelmessig puls som tyder på AFib, ved hjelp av PPG-signaler kombinert med en algoritme. I tillegg til denne PPG-baserte identifiseringsalgoritmen har Apple Watch Series 4, Series 5 og Series 6 også en elektrisk hjertermåler som kombinert med EKG-appen gir muligheten til å generere og analysere et elektrokardiogram tilsvarende et EKG med én avledning.

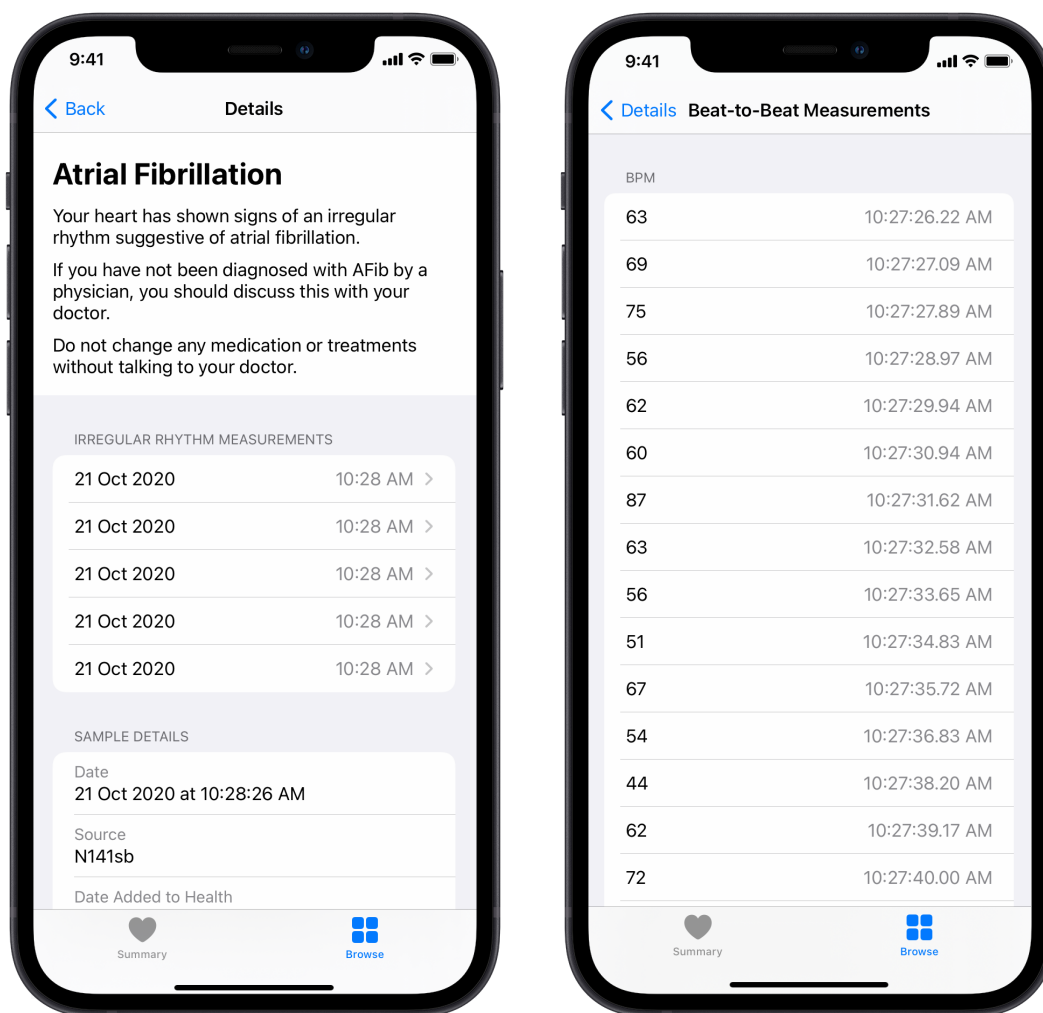
PPG-basert påvisning av arytmi

Teknisk informasjon og funksjonsbeskrivelse

Apple Watch har en optisk pulsmåler med grønne LED-lys som sammen med lyssensitive fotodioder registrerer puls i blodvolumet gjennom håndleddet ved hjelp av fotopletysmografi. Disse sensorene og de underliggende algoritmene danner grunnlaget for registrering av hjerterytmie og hjerterytmievariabilitet på Apple Watch Series 1 og nyere. For å registrere hjerterytmievariabilitet registrerer Apple Watch et takogram – en graf over tidsintervallet mellom hjerteslag – med to til fire timers mellomrom. Med watchOS 5.1.2 og nyere kan brukere også slå på en funksjon for registrering av arytmi basert på disse takogrammene. Før de kan ta i bruk funksjonen for varsling om uregelmessig hjerterytmie på Apple Watch, må brukerne fullføre registreringen i Helse-appen på en sammenkoblet iPhone. Her lærer de hvordan de bruker funksjonen, og får informasjon om AFib. Du finner mer informasjon om funksjonen på <https://support.apple.com/no-no/HT208931>.

Hvis PPG-basert påvisning av arytmi er aktivert, klassifiserer Apple Watch hvert takogram ved hjelp av en merkevarebeskyttet algoritme for å avdekke om brukeren kan ha uregelmessig hjerterytme. Hvis det påvises uregelmessig hjerterytme, utløses en serie av hyppigere registreringer av takogrammer – så ofte som mulig innenfor et tidsrom på minimum 15 minutter – og analyser. Apple Watch kan bare registrere og analysere takogrammer når brukeren er tilstrekkelig i ro til at det kan gjennomføres en måling. Brukeren overvåkes altså ikke konstant ved hjelp av algoritmen, men målingene gjennomføres i stedet opportunistisk når signalet er godt nok til at det kan gjennomføres registrering og analyser. Hvis fem av seks etterfølgende takogrammer – medregnet det første – klassifiseres som uregelmessige i løpet av 48 timer, varsles brukeren om mulig arytmi. I tillegg til å motta varslingen har brukeren tilgang til mer informasjon knyttet til de uregelmessige takogrammene i Helse-appen (illustrasjon 1). Hvis to takogrammer klassifiseres som regelmessige før grensen blir nådd, nullstilles syklusen. Deretter går registreringen av takogrammer tilbake til basisfrekvensen for målinger med to timers mellomrom.

Illustrasjon 1: Visning av uregelmessige hjerterytmemålinger i Helse-appen



I Helse-appen kan brukerne se tidspunktene når algoritmen har registrert et takogram med uregelmessig hjerterytme som har bidratt til at det ble utløst en varsling (venstre side). Ved å trykke på en dato og et klokkeslett kan brukeren visualisere målingene av hvert enkelt hjerteslag som er beregnet i takogrammet.

Preklinisk utvikling

Før kliniske tester ble det gjennomført studier for å utvikle algoritmen for PPG-basert påvisning og vurdere resultatene for algoritmen under ulike forhold og ved ulike former for atferd hos brukerne. Eksempler på slike forhold og atferdsvarianter er dyp pusting, bilkjøring, skjelvende hender, hender i bevegelse, redusert perfusjon i hånd eller håndledd, bruk om natten, hurtig ventrikulær respons hos personer med AFib og andre arytmier. Disse studiene ble gjennomført blant over 500 personer med AFib og 2300 kontrollpersoner.

Ettersom PPG er avhengig av lysabsorpsjonsevne, ble algoritmen for påvisning av arytmi testet på ulike hudtyper og hudtoner for å sikre at justeringen for hudtone i sensorplattformen var tilstrekkelig for algoritmene som benyttes til å påvise arytmi. Melanin har høy absorpsjonsevne ved bølgelengden som benyttes av det grønne LED-lyset på Apple Watch, slik at det potensielt er vanskeligere å gjøre pulsmålinger på mørkere hud. For å ta hensyn til dette justerer sensorplattformen i Apple Watch LED-strømmen (og dermed også lyseffekten), fotodiodeforsterkningen (lyssensitiviteten) og målefrekvensen for å sikre tilstrekkelig signalamplitude for ulike hudtoner.

Av valideringshensyn ble 1,3 millioner takogrammer fra 1124 deltakere (51 prosent kvinner) med ulike hudtyper og hudtoner analysert (basert på Fitzpatrick-hudtyper og måling av hudfarge ved håndleddet med spektrofotometer). Ettersom en av de største tekniske utfordringene var signalamplituder hos personer med mørk hud, hadde nesten fem prosent av deltakerne i studien hud i kategori VI på Fitzpatrick-skalaen. Gjennom valideringen av algoritmen ble det ikke avdekket vesentlige forskjeller i sensitivitet eller spesifisitet knyttet til ulike hudtyper og hudtoner.

Klinisk validering

Apple Heart Study

Apple Heart Study (AHS) er en prospektiv pragmatisk enkeltarmstudie som ble gjennomført virtuelt for å vurdere algoritmen for varsling om uregelmessig puls i Apple Watch med hensyn til å identifisere arytmi som tyder på AFib. Deltakere i studien som oppfylte grensen på fem av seks etterfølgende uregelmessige takogrammer, mottok en varsling på iPhone og Apple Watch. De fikk tilbud om en nettkonsultasjon med en lege knyttet til studien og motta en håndholdt hjertemonitor (ePatch fra BioTelemetry, Inc. i Conshohocken i Pennsylvania i USA). Deltakerne ble bedt om å gå med ePatch-monitoren i opptil sju dager. Data samlet inn fra deltakerne ble likevel regnet som tilstrekkelig så lenge data for minst én time kunne analyseres.

Detaljerte resultater fra AHS ble publisert i november 2019 i *New England Journal of Medicine* (Perez, Marco V., et al. «Large-scale assessment of a smartwatch to identify atrial fibrillation.» *New England Journal of Medicine* 381.20 (2019): 1909–1917).

Utformingen til AHS-delstudie

Det ble gjennomført en delstudie av dataene fra et utvalg av deltakerne i AHS for å fastsette om klassifiseringsalgoritmen for takogrammer (individuelle målinger eller stikkprøver) og algoritmen for bekreftelsesyklus (varslingsgrense på fem av seks takogrammer) hadde tilfredsstillende positiv prediktiv verdi (PPV) sammenlignet med ePatch-målinger med hensyn til påvisning av uregelmessig hjerterytm som tyder på AFib. Forskerne som deltok i AHS, var klar over delstudien, de påfølgende analysene og at dataene ble sendt inn til den amerikanske etaten Food and Drug Administration (FDA), men fikk ikke opplysninger om resultatene fra delstudien mens AHS pågikk. Kontrollkomiteen som godkjente AHS, kom frem til at det ikke var behov for at de skulle følge opp denne delstudien. Alle deltakere i AHS samtykket også i at dataene deres fra studien kunne brukes i delstudien.

Data for delstudien ble samlet inn fra personer som deltok i AHS mellom 30. november 2017 og 22. juni 2018. Deltakerne i denne delstudien mottok en varsling om uregelmessig hjerterytme fra appen som ble undersøkt i AHS. Deretter mottok og gikk de med en ePatch-hjertemonitor, som ble brukt til å gjøre målinger til analyse. De første uregelmessige takogrammene som ledet frem til den første varslingen og potensielt den første kontakten med nettkonsultasjonen med en lege, ble ikke analysert i forbindelse med denne delstudien. Analysene omfattet bare takogrammer med uregelmessig hjerterytme og varslinger som ble utløst mens brukeren gikk med ePatch-monitoren knyttet til studien.

To uavhengige EKG-eksperter med amerikansk sertifisering innenfor kardiologi eller elektrofysiologi evaluerte EKG-utskrifter og klassifiserte disse som sinusrytme, AFib, annen uregelmessig hjerterytme eller uleselig resultat. I tilfeller der de to ekspertene ikke var enige, ble kurven vurdert av en tredje ekspert med tilsvarende kvalifikasjoner. Ekspertene kjente ikke til klassifiseringen i takogrammene. EKG-klassifiseringen fra ekspertene og klassifiseringen utført av algoritmer i takogrammene ble oversendt på en sikker måte til en statistiker knyttet til studien for analyse.

Resultater

Av de 226 deltakerne i delstudien som mottok en varsling om arytmi og gikk med en ePatch-monitor i rundt én uke, fikk 41,6 prosent (94 deltakere) påvist AFib av ePatch-monitoren. Av deltakerne som både gikk med Apple Watch og en ePatch-monitor, mottok 57 av de 226 deltakerne en varsling om AFib. For disse deltakerne ble fem av seks etterfølgende takogrammer altså klassifisert som uregelmessige. Av disse 57 deltakerne fikk 78,9 prosent (45 deltakere) også påvist AFib i dataene for ePatch-monitoren, og 98,2 prosent (56 deltakere) fikk påvist AFib eller andre klinisk relevante arytmier. Disse resultatene viser at varslingene gir riktig påvisning av AFib i de fleste tilfeller, men at varslinger i enkelte tilfeller kan identifisere forekomster av andre typer arytmier enn AFib. Det ble ikke rapportert alvorlige bivirkninger knyttet til utstyret.

EKG-basert påvisning

Teknisk informasjon og funksjonsbeskrivelse

Apple Watch Series 4, Series 5 og Series 6 har en integrert titaniumelektrode i Digital Crown og et løvtynt lag med krom, silikon og karbonnitrid på safirkristallen på baksiden av Apple Watch. EKG-appen leser av og registrerer de elektriske signalene som kontrollerer hjertet, fra brukerens fingerspiss (Digital Crown) og håndleddet (baksiden av Apple Watch), noe som gir en lukket krets. Før brukerne kan ta i bruk EKG-appen på Apple Watch, må de fullføre registreringen i Helse-appen på en sammenkoblet iPhone. Her lærer de hvordan de bruker funksjonen, og får informasjon om AFib. Brukerne utfører en EKG-måling ved å åpne EKG-appen installert på Apple Watch og legge en finger på den motsatte hånden av den de har Apple Watch på, på Digital Crown i 30 sekunder. Polariteten bestemmes av hvilket håndledd brukeren har plassert Apple Watch på, slik dette er angitt i innstillingene.

Etter at EKG-målingen er utført, klassifiserer den merkevarebeskyttede algoritmen målingen som SR, AFib eller ikke-avgjørende i EKG 1.0. EKG 2.0, der dette er tilgjengelig, har ytterligere klassifiseringer som for eksempel AFib med rask puls (100–150 slag/min) og skiller også mellom dårlige og ikke-avgjørende målinger. Disse rytmeklassifiseringene – snittpuls, brukerrapporterte symptomer og bølgeform – legges inn i Helse-appen. All denne informasjonen lagres og kan deles av brukeren i PDF-format fra appen på brukerens sammenkoblede iPhone. Du finner mer informasjon om hvordan du bruker dette på <https://support.apple.com/no-no/HT208955>.

Preklinisk utvikling

Algoritmen for påvisning og klassifisering av EKG-signal ble også testet i flere studier før den kliniske valideringen ble innledet. Algoritmen for sensorene og klassifiseringen ble testet for ulike etnisiteter, omkretser på håndledd, kroppsmasseindeks (BMI), alder, arytmier som ikke var AFib, hvor stramt båndet var festet, stillinger og svetting/aktivitetsnivå. Rundt 2500 personer deltok i disse testene, hvorav 25 prosent tidligere hadde fått diagnosen AFib eller en annen uregelmessig hjerterytme.

Økt frekvens av uleselige EKG-målinger representerte den største variasjonen i algoritmene. Bakenforliggende faktorer var lav signalamplitude (som et resultat av avvik i høyre-aksen, spesielt hos personer med lav BMI, eller når det ble registrert svetting under testing som følge av trening) og bevegelsesartefakter som resultat av brukernes atferdsmønster. Apple Watch bruker tørre elektroder som er mekanisk solide og korrosjonsbestandige og dermed egner seg for bærbare enheter. Tørre elektroder – og spesielt tørre elektroder som plasseres på armer og ben – har imidlertid en tendens til å produsere støy lik den som er beskrevet ovenfor, sammenlignet med midlertidige gel-elektroder brukt i klinisk utstyr.

I tillegg til faktorene ovenfor førte enkelte arytmier som ikke var AFib, til svært ulik funksjonalitet i algoritmen sammenlignet med deltakere med sinusrytme. Disse forholdene og resultatene er beskrevet i tabell 1 nedenfor.

Tabell 1. Arytmier som ikke er AFib, men som påvirker algoritmen (EKG 2.0)

| Arytmi | Variasjon |
|---|--|
| V/H grenblokk | 7,9 % av forsøkene klassifisert som AFib |
| Førstegrads atrioventrikulært blokk | 10,2 % av forsøkene ble ikke klassifisert, av de klassifiserte forsøkene ble 2,4 % klassifisert som AFib |
| Bigemini | 92,5 % ble ikke klassifisert |
| Hyppe ventrikulære ekstrasystoler | 48,8 % ble ikke klassifisert, av de klassifiserte forsøkene ble 24,1 % klassifisert som AFib |
| Hyppe supraventrikulære ekstrasystoler | 19,8 % ble ikke klassifisert, av de klassifiserte forsøkene ble 23,5 % klassifisert som AFib |
| Supraventrikulær takykardi | 7,7 % ble klassifisert som AFib |
| Blandet rytme | 46,5 % ble ikke klassifisert, av de klassifiserte forsøkene ble 29,7 % klassifisert som AFib |
| Rask/langsom puls (utenfor området 50–150 slag/min) | 94,8 % ble ikke klassifisert |

Klinisk validering

Ekspérimentets utførelse

Apple sponset to studier utført på ulike forskningssentre med mål om å validere hvorvidt EKG-appen kan 1) generere en EKG-bølgeform lik et EKG fra én enkelt avledning i en EKG-standardmåling med tolv avledninger, og 2) bruke en klassifiseringsalgoritme for hjerterytme til å klassifisere enten SR eller AFib ved hjelp av EKG-et med én avledning.

Studie for klinisk validering av EKG 1.0

Det ble gjennomført en studie for klinisk validering av resultatene fra EKG 1.0-appen. De primære endepunktene i studien var sensitiviteten til klassifiseringsalgoritmen for rytme i påvisningen av AFib og spesifisiteten i registreringen av SR. En ekstern kontrollkomité godkjente protokollen, samtykkeskjemaet og alt annet relevant materiale i forkant av deltakerregistreringen, og alle deltakerne ga skriftlig samtykke før de deltok i studien.

Deltakerne i studien omfattet personer med diagnostisert AFib og personer uten kjente uregelmessigheter i hjerterytmen. Deltakerne ble bedt om å registrere tre EKG-målinger med én avledning ved hjelp av EKG-appen. Medarbeidere tilknyttet studien registrerte samtidig tre EKG-målinger med tolv avledninger ved hjelp av en klinisk enhet godkjent av FDA (GE Healthcare CardioSoft ECG). De første forsøkene ble vurdert for evaluering og analyse. Deltakerne fikk hjelp med å sette på Apple Watch riktig og ble bedt om å holde armene i ro, eventuelt å hvile armene på et bord eller på benene. De fikk også muligheten til å øve seg på å gjøre målinger i forkant av forsøkene.

For å teste punkt 1 la tre uavhengige og kvalifiserte fagpersoner rytmeutskrift generert fra 140 tilfeldig utvalgte deltakere (70 med AFib og 70 med SR) over den tilsvarende utskriften fra én enkelt avledning fra det kliniske utstyret. Dette ble gjort for å kunne foreta en visuell sammenligning av morfologien i de seks etterfølgende PQRST-kompleksene. Fagpersonene klassifiserte hver utskrift som godkjent eller ikke godkjent basert på en visuell vurdering av morfologisk likhet. Fagpersonene ble også bedt om å måle R-amplituden fra isoelektrisk grunnlinje til nærmeste millimeter for de første to QRS-kompleksene i utskriftene, både fra referanseenheden og EKG-appen, og å vurdere likheten mellom disse to.

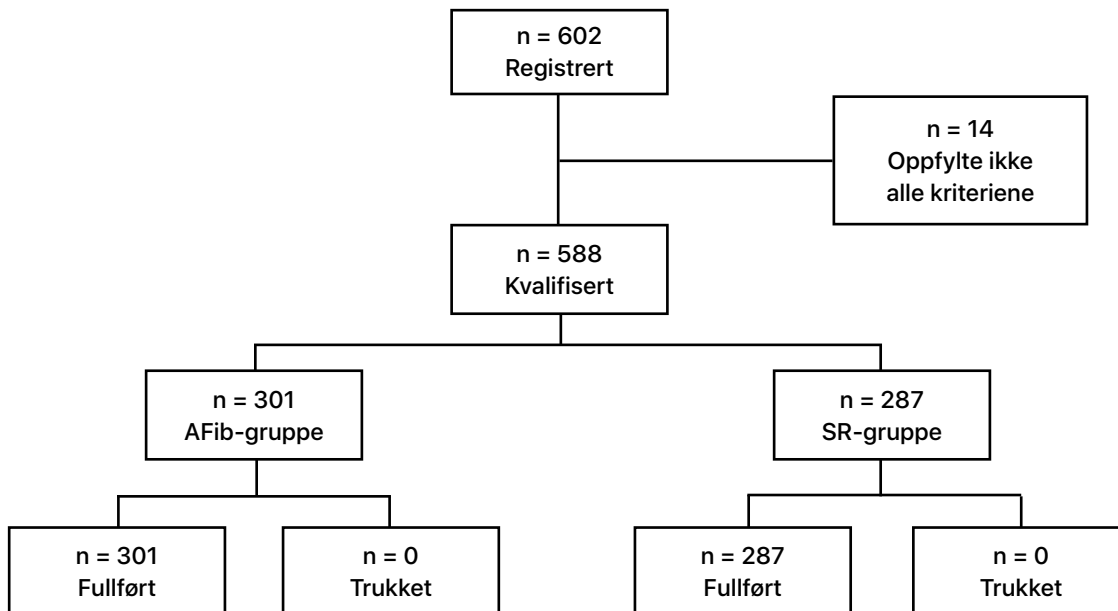
For punkt 2 ble hver EKG-utskrift fra målinger med tolv avledninger vurdert av tre uavhengige og kvalifiserte kardiologer med amerikansk sertifisering, som ikke hadde tilgang til informasjon om studien. De klassifiserte rytmene som enten SR, AFib, annet (alt som ikke var SR eller AFib innen parameterne for hjerterytmen) eller ikke lesbar (det kunne ikke stilles en diagnose ettersom utskriften ikke inneholdt tilstrekkelig informasjon for avlesing). Algoritmen for EKG-appen klassifiserte de genererte EKG-utskriftene som SR, AFib, ikke klassifiserbar eller ikke lesbar. Fagpersonene beregnet EKG-appens sensitivitet og spesifisitet i klassifiseringen av SR og AFib (for EKG-utskrift som kunne klassifiseres) sammenlignet med en kardiologs tolkning av EKG-målinger med tolv avledninger. En uavhengig og kvalifisert kardiolog (med sertifisering fra USA) som ikke hadde tilgang til informasjon om studien, ble deretter bedt om å klassifisere EKG-utskrift produsert av EKG-appen, i de samme kategoriene.

En ensidig nedre konfidensgrense på nøyaktig 97,5 prosent ble utregnet separat for sensitivitet og spesifisitet for analysene knyttet til de primære endepunktene. Hvis den nedre grensen for sensitivitet overskred 90 prosent, ble nullhypotesen avvist til fordel for sensitivitet som overskred 90 prosent. Hvis den nedre grensen for spesifisitet overskred 92 prosent, ble nullhypotesen avvist til fordel for spesifisitet som overskred 92 prosent.

Resultater

Studien hadde 602 deltakere, hvorav 588 var kvalifisert i henhold til kriteriene. Av de 588 kvalifiserte deltakerne ble 301 personer med egenrapportert AFib plassert i AFib-gruppen og 287 personer uten egenrapportert AFib plassert i SR-gruppen. Hensikten med denne inndelingen var utelukkende å oppnå riktig registrering. De som evaluerte resultatene, ble ikke informert om gruppene, og forekomsten eller fraværet av AFib var utelukkende basert på EKG-målingene utført i løpet av testingen. 14 deltakere som fullførte studien, men ikke ble registrert i en gruppe, var ikke kvalifisert for studien ettersom de tidligere hadde fått påvist paroksysmal AFib, men AFib ble ikke påvist i EKG-målingene under screeningen. Alle kvalifiserte deltakere fullførte studien (illustrasjon 2). Ingen bivirkninger ble rapportert i løpet av studien.

Illustrasjon 2: Flytskjema over deltakernes status



Tre uavhengige, sertifiserte fagpersoner fant visuelle morfologiske likheter mellom bølgeformen fra EKG-appen og EKG-referansemålingene med én avledning i 98,4 prosent av de analyserte utskriftene i AFib-gruppen og 100 prosent i SR-gruppen (tabell 2). Den samlede andelen deltakere med godkjente målinger var 99,2 prosent (97,5 prosent nedre konfidensgrense = 95,7 prosent). Utskriftene ble utelukket hvis det ikke kunne identifiseres seks etterfølgende slag (PQRST-komplekser) uten artefakter i begge utskriftene (fra EKG-appen og fra referanseenheten).

Tabell 2. Sammenligning av bølgeform

| Karakteristikk | AFib-deltakere (n = 61) | SR-deltakere (n = 65) | Totalt (n = 126) | Nedre konfidensgrense* | P-verdi** |
|--|-------------------------|-----------------------|------------------|------------------------|-----------|
| Antall godkjente sammenlignede utskrifter (EKG-app og referanseenhets) | 60 | 65 | 125 | | |
| Antall lesbare sammenlignede utskrifter (EKG-app og referanseenhets) | 61 | 65 | 126 | | |
| Andel av utskrifter godkjent | 60/61 (98,4 %) | 65/65 (100 %) | 125/126 (99,2 %) | 95,7 % | < 0,0001 |
| Antall sammenlignede utskrifter utelatt | 8 | 5 | 13 | | |

*Nedre eksakt grense for ensidig binomial konfidensintervall på 97,5 prosent for samlet antall

**Test av hypotese for vellykket resultat > 0,8

Forkortelser: AFib = atrieflimmer, SR = sinusrytme

For ytterligere bekreftelse på at bølgeformene fra EKG-appen og referanseenhets var like, ble EKG-utskriftene fra appen som var klassifisert av kardiologer uten informasjon om studien, sammenlignet med referanseutskriftene klassifisert av kardiologene (tabell 3). Prosentandelen samsvar mellom klassifiseringen av utskriftene fra testenheten med AFib- og SR-resultatene fra referanseenhets var henholdsvis 100 prosent og 99,1 prosent. Utskrifter som ikke var lesbare, ble ikke inkludert i denne analysen.

Tabell 3. Klassifisering av utskrifter fra EKG-appen og referanseenheten

| Karakteristikk | Totalt (n = 522) |
|--|--------------------------|
| Endelig resultat for EKG-referanse = AFib | 263 |
| Klassifisering av EKG-utskrifter fra app = AFib | 239/263 (90,9 %) |
| Klassifisering av EKG-utskrifter fra app = SR | 0/263 (0,0 %) |
| Klassifisering av EKG-utskrifter fra app = annet | 0/263 (0,0 %) |
| Klassifisering av EKG-utskrifter fra app = ikke lesbar | 24/263 (9,1 %) |
| % samsvar med resultat fra AFib-referanse* | 239/239 (100,0 %) |
| Endelig resultat for EKG-referanse = SR | 244 |
| Klassifisering av EKG-utskrifter fra app = AFib | 0/244 (0,0 %) |
| Klassifisering av EKG-utskrifter fra app = SR | 232/244 (95,1 %) |
| Klassifisering av EKG-utskrifter fra app = annet | 2/244 (0,8 %) |
| Klassifisering av EKG-utskrifter fra app = ikke lesbar | 10/244 (4,1 %) |
| % samsvar med resultat fra SR-referanse* | 232/234 (99,1 %) |
| Endelig resultat for EKG-referanse = annet | 15 |
| Klassifisering av EKG-utskrifter fra app = AFib | 0/15 (0,0 %) |
| Klassifisering av EKG-utskrifter fra app = SR | 3/15 (20,0 %) |
| Klassifisering av EKG-utskrifter fra app = annet | 12/15 (80,0 %) |
| Klassifisering av EKG-utskrifter fra app = ikke lesbar | 0/15 (0,0 %) |
| % samsvar med resultat fra annet-referanse* | 12/15 (80,0 %) |

*Utskrifter som ikke var lesbare, ble utelatt
 Forkortelser: AFib = atrieflimmer, SR = sinusrytme

Til sammen 485 av 602 parede utskrifter fra EKG-appen og referanseenheten ble ansett som klassifiserbare. De resterende parene hadde utskrifter fra EKG-appen eller referanseenheten som ikke kunne klassifiseres eller ikke var lesbare. Tabell 4 viser fordelingen i AFib- og SR-gruppene.

Tabell 4. Endelig klassifiseringsresultat for algoritmen i EKG-appen og referanseutskriften

| Klassifisering via algoritmen i EKG-appen | Klassifisering via referanseutskriften | | | | |
|---|--|------|-------|--------------|--------|
| | SR | AFib | Annet | Ikke lesbare | Totalt |
| SR | 238 | 4 | 4 | 1 | 247 |
| AFib | 1 | 236 | 2 | 2 | 241 |

Tabell 4. Endelig klassifiseringsresultat for algoritmen i EKG-appen og referanseutskriften

| Klassifisering via algoritmen i EKG-appen | Klassifisering via referanseutskriften | | | | |
|---|--|------|-------|--------------|--------|
| | SR | AFib | Annet | Ikke lesbare | Totalt |
| Ikke klassifiserbare | 6 | 7 | 6 | 0 | 19 |
| Ikke lesbare | 18 | 30 | 1 | 0 | 49 |
| Resultat fra enhet ikke rapportert* | 32 | 13 | 1 | 0 | 46 |
| Totalt | 295 | 290 | 14 | 3 | 602 |

*Resultater ble ikke rapportert basert på kriterier fastsatt på forhånd (for eksempel manglende synkronisering) for alle deltakere bortsett fra én
Forkortelser: AFib = atrieflimmer, SR = sinusrytme

Klassiferingen via algoritmen i EKG-appen oppnådde 98,3 prosent sensitivitet og 99,6 prosent spesifisitet (tabell 5). Hvis analysen utvides til å inkludere de 2,4 prosentene (7 av 290) og 2,0 prosentene (6 av 295) med utskrifter som ikke kunne klassifiseres av enheten i AFib- og SR-gruppene, var sensitiviteten 95,5 prosent (95 prosent KI: 92,2 prosent, 97,8 prosent) og spesifisiteten 97,1 prosent (95 prosent KI: 94,2 prosent, 98,8 prosent). Disse resultatene samsvarer med endepunktene som ble fastsatt i forkant av denne studien. I tillegg var 12,2 prosent (68 av 556) av målingene ikke-avgjørende (ikke lesbare eller kunne ikke klassifiseres) og kunne ikke klassifiseres som SR eller AFib. Hvis analysen utvides til å inkludere de ikke-avgjørende registreringene, klassifiserte EKG-appen SR riktig for 90,5 prosent (238 av 263) av deltakerne med SR, og AFib for 85,2 prosent (236 av 277) av deltakerne med AFib. De kliniske valideringsresultatene gjenspeiler forsøk i et kontrollert miljø. Bruk av EKG-appen utenfor et kontrollert miljø kan resultere i et høyere antall ikke-avgjørende utskrifter og utskrifter som ikke kan klassifiseres.

Tabell 5. Sensitivitets- og spesifisitetsanalyse – klassifiserbare utskrifter

| Parameter | Verdi | Nedre konfidensgrense* | P-verdi** |
|---|------------------|------------------------|-----------|
| Endelig resultat for EKG-referanse = AFib (n) | 240 | | |
| Resultat for EKG-appenhet = AFib | 236/240 (98,3 %) | | |
| Resultat for EKG-appenhet = SR | 4/240 (1,7 %) | | |
| Sensitivitet | 236/240 (98,3 %) | 95,8 % | < 0,0001 |
| Endelig resultat for EKG-referanse = SR (n) | 239 | | |
| Resultat for EKG-appenhet = AFib | 1/239 (0,4 %) | | |
| Resultat for EKG-appenhet = SR | 238/239 (99,6 %) | | |
| Spesifisitet | 238/239 (99,6 %) | 97,7 % | < 0,0001 |

*Nedre eksakt binomial ensidig konfidensgrense

**Test av hypotese for sensitivitet > 0,9 og spesifisitet > 0,92

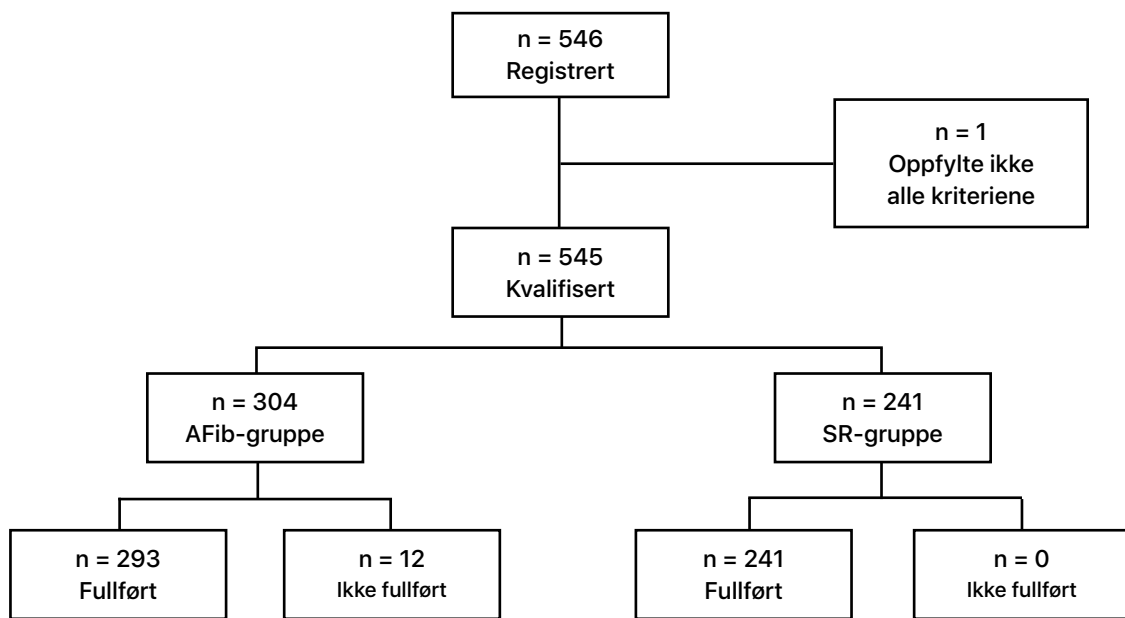
Forkortelser: AFib = atrieflimmer, SR = sinusrytme

Studie for klinisk validering av EKG 2.0

Det ble gjennomført enda en studie for å støtte og validere algoritmer i EKG 2.0-appen (testenheten), med utvidelse av det klassifiserbare pulsområdet (50–150 slag/min) og innføring av nye klassifiseringsresultater (SR, SR med rask puls, AFib, AFib med rask puls, ikke-avgjørende måling og dårlig måling). Målet med denne studien var å vurdere resultatene for testenheten. Spesifisitet og sensitivitet var de primære endepunktene som ble vurdert. Sekundære endepunkter var riktig klassifisering av følgende kategorier hos deltakerne i spesifikke lesbare og klassifiserbare EKG-utskrifter: NSR (puls på 50–150 slag/min), SR ved SR på samtidig EKG med tolv avledninger, AFib (puls på 50–99 slag/min), AFib ved AFib på samtidig EKG med tolv avledninger, sinustakykardi (puls på 100–150 slag/min), rask puls ved SR på samtidig EKG med tolv avledninger og AFib med rask puls (puls på 100–150 slag/min) ved AFib på samtidig EKG med tolv avledninger. I tillegg vurderte fagpersonene samsvaret mellom bølgeformen i EKG-appen og én avledning fra et EKG med tolv avledninger – målt med tilfredsstillende samsvar mellom morfologien i PQRST-kompleksene og amplitude i R-bølgeformen – tilsvarende vurderingen som ble gjort i studien for klinisk validering av EKG 1.0.

Gjennomføringen av denne studien ble godkjent av den relevante kontrollkomiteen ved de respektive forskningssentrene for den prospektive studien ved ulike forskningssentre i USA. På hvert enkelt forskningssenter innhentet forskerne godkjenning fra kontrollkomiteen, samtykkeskjema og deltakermateriell før studien ble innledet.

Illustrasjon 3: Flytskjema over deltakernes status



Deltakerne i studien omfattet personer med normal SR under screening – uten tidligere påvist AFib – og personer med kjent vedvarende, permanent eller kronisk AFib som fikk påvist AFib under screeningen.

Alle deltakere ble bedt om å ta et EKG med tolv avledninger og samtidig et separat EKG med én avledning med Apple Watch. For hver enkelt deltaker ble det gjennomført tre forsøk. Prosessen ble gjentatt under treningsøkter. Deltakerne man anså for å være i god form, syklet på en ergometersykkel i fem minutter for å nå målområdet for puls. Både forsøkene der deltakerne var i ro og fysisk aktive, ble vurdert for evaluering og analyse.

To uavhengige, kvalifiserte kardiologer med amerikansk sertifisering, som ikke hadde tilgang til informasjon om studien, foretok puls- og rytmediagnoser av EKG-målingen med tolv avledninger. I tilfeller der de to ekspertene ikke var enige, gjorde en tredje ekspert en vurdering. Puls ble beregnet for hver EKG med tolv avledninger. Ekspertene registrerte puls og valgte pulsdiaagnosekoden som samsvarte med pulsen som ble målt i EKG-referansemålingen. Følgende diagnoser for hjerterytme ble vurdert i dataene for EKG-målingene med tolv avledninger: SR, AFib, supraventrikulær takykardi (SVT), annen abnormal rytme (hyppige premature atriekontraksjoner, hyppige premature ventrikkellkontraksjoner, atrieflutter, ventrikkeltakykardi, ventrikkelflimmer, annengrads atrioventrikulært blokk type I, annengrads atrioventrikulært blokk type II, tredjegrads atrioventrikulært blokk og andre varianter) samt målinger som ikke kunne tolkes. Tre eksperter uten tilgang til informasjon om studien vurderte bølgeformen under evalueringen av EKG-data fra EKG-utskriftsparene. De første seks etterfølgende tydelig lesbare PQRST-kompleksene uten artefakter med samsvar mellom utskriften for evaluering fra deltakerens testenheter og referanseenheter ble identifisert av en av ekspertene og brukt av de to andre ekspertene. Utskrifter uten seks etterfølgende slag ble utelatt.

For analyse av det primære endepunktet implementerte man bootstrapping for å oppnå tosidige konfidensintervaller på 95 prosent for sensitivitet og spesifisitet, ettersom data fra de samme studiedeltakerne ble samlet inn i ro og etter fysisk aktivitet. Deltakerne som fikk påvist AFib (for sensitivitet) eller SR (for spesifisitet) i minst ett evaluert resultat og hadde et klassifiserbart resultat for algoritmen (SR eller AFib), ble tilfeldig utvalgt med erstatninger. 2,5. og 97,5. persentil av distribusjonen av beregningene for bootstrapping representerer grensene i det tosidige konfidensintervallet på 95 prosent. Hvis den nedre grensen for både sensitivitet og spesifisitet overskred det på forhånd fastsatte resultatmålet knyttet til disse beregningene (90 prosent for sensitivitet og 92 prosent for spesifisitet), ble nullhypotesen avvist til fordel for den alternative hypotesen. For sammenligning av bølgeformer ble data fra det første forsøket med tilfeldig utvalgte deltakere – i ro og etter fysisk aktivitet – brukt til å prøve hypotesene for vurdering av bølgeformer ved hjelp av bootstrapping med tilfeldige uttrekk. Hvis den 2,5. persentilen av distribusjonen for bootstrapping for andelen med samsvarende morfologi overskred 80 prosent, ble nullhypotesen (andelen med samsvarende morfologi var lavere enn 80 prosent) avvist. Det samme kriteriet ble brukt under evalueringen av samsvarsandelen for amplitude i R-bølgeformen.

Resultater

Til sammen 546 deltakere ble registrert i studien. 305 personer ble registrert i AFib-gruppen, og 241 personer ble registrert i gruppen med normal sinusrytme (NSR). Én deltaker i AFib-gruppen ble utelatt i henhold til kvalifiseringskravene. Av de 546 deltakerne fullførte 534 (293 i AFib-gruppen og 241 i NSR-gruppen) studien. 12 av deltakerne i AFib-gruppen fullførte ikke studien.

Tre fagpersoner foretok individuell kontroll av analyseutvalget for vurdering av bølgeformer. Evalueringen omfattet par med utskrifter fra testenheten og referanseenheden. Til sammen 91 deltakere (100 prosent) hadde et lesbart par med målinger da de var i ro, 87 deltakere (100 prosent) hadde et lesbart par med målinger etter fysisk aktivitet, og 93 deltakere (100 prosent) hadde minst ett utskriftspar fra de var i ro eller etter fysisk aktivitet.

Det klassifiserbare analyseutvalget besto av alle deltakere som hadde par med lesbare, evaluerte resultater fra EKG 2.0-appen og EKG-måling med tolv avledninger. Data fra til sammen 512 deltakere (n = 279 i AFib-gruppen og n = 233 i NSR-gruppen) ble regnet som klassifiserbare. Analyseutvalget for vurdering av bølgeformer – definert som tilfeldig utvalgte deltakere med lesbare utskriftspar fra EKG-appen og én avledning i EKG-måling med tolv avledninger – besto av totalt 93 deltakere (n = 48 i AFib-gruppen og n = 45 i NSR-gruppen).

Tabell 6 viser dataene for ulike klassifiseringer fra EKG 2.0-appen og de evaluerte resultatene fra EKG-måling med tolv avledninger for sammenslåtte data for målingen i ro og etter fysisk aktivitet. SR- og AFib-kategoriene viser en høy grad av samsvar mellom de evaluerte resultatene for EKG-måling med tolv avledninger og enheten med EKG-appen. For utskrifter fra EKG-måling med tolv avledninger evaluert som SR (n = 470) ble 436 resultater også klassifisert som SR i målinger fra enheten med EKG-appen. For utskrifter fra EKG-måling med tolv avledninger evaluert som AFib (n = 521) ble 474 resultater også klassifisert som AFib i målinger fra enheten med EKG-appen. Enheten med EKG-appen klassifiserte 9 resultater som SR, 37 resultater som AFib og 13 resultater som ikke-avgjørende. 9 resultater ble klassifisert som dårlige målinger av enheten med EKG-appen og som annen uregelmessig hjerterytme (SVT eller annet) i EKG-målingen med tolv avledninger.

Tabell 6. Tabell over ulike klassifiseringer fra EKG-appen og evaluerte resultater fra EKG-måling med tolv avledninger (i ro og etter fysisk aktivitet) – klassifiserbart analyseutvalg

| Empiriske observasjoner med evaluerte resultater for EKG-måling med tolv avledninger | | | | | | |
|--|-------------------|---------------------|---|---------------|-------------------|-------------|
| Måling fra enhet med EKG-app | SR (50≤HR≤150) | AFib (50≤HR≤150) | Annet (SVT eller annet med HR 50–150, HR<50, HR>50) | Ikke tolkbart | Ikke tilgjengelig | Totalt |
| SR [SR (50–99), NSR sinustak. (100–150)] | 436 | 7 | 9 | 0 | 2 | 454 |
| AFib (50≤HR≤150) | 3 | 474 | 37 | 0 | 6 | 520 |
| Ikke-avgjørende [inkl. HR<50, HR>150] | 10 | 13 | 13 | 0 | 1 | 37 |
| Dårlig måling | 14 | 23 | 9 | 0 | 3 | 49 |
| Ikke tilgjengelig | 7 | 4 | 1 | 1 | 0 | 13 |
| Totalt | 470 | 521 | 69 | 1 | 12 | 1073 |

Forkortelser: AFib = atrieflimmer, SR = sinusrytme, HR = hjerterytme, SVT = supraventrikulær takykardi

For EKG 2.0-appen var beregnet spesifisitet 99,3 prosent og beregnet sensitivitet 98,5 prosent (tabell 7). De nedre konfidensgrensene for beregnet spesifisitet (98,4 prosent) og beregnet sensitivitet (97,3 prosent) overgikk de på forhånd fastsatte resultatmålene knyttet til disse beregningene. Derfor ble nullhypotesen for spesifisitet og sensitivitet avvist til fordel for den alternative hypotesen, og de primære endepunktene for spesifisitet og sensitivitet var oppfylt. Beregnet spesifisitet i ro var 99,1 prosent, og beregnet sensitivitet i ro var 98,5 prosent. Beregnet spesifisitet etter fysisk aktivitet var 99,5 prosent, og beregnet sensitivitet etter fysisk aktivitet var 98,6 prosent. For de sammenslåtte dataene for deltakere i ro og etter fysisk aktivitet ble det gjennomført en tilleggsanalyse der «ikke-avgjørende»-klassifiseringen inngikk i beregningen. Den resulterende beregnede spesifisiteten var 97,1 prosent, mens den beregnede sensitiviteten var 96,0 prosent.

Det ble avdekket en høy grad av samsvar mellom de evaluerte pulsresultatene for EKG-måling med tolv avledninger og målingene fra enheten med EKG-appen. Forekomsten av feilklassifisering var lav ved sammenligning med resultatene i referanseutskriftene fra EKG-måling med tolv avledninger. Tabell 8 viser sammenfattede resultater som bekrefter at de sekundære endepunktene for riktig klassifisering av pasienter med NSR (puls på 50–99 slag/min) som SR ved SR på samtidig EKG-måling med tolv avledninger, AFib (puls på 50–99 slag/min) ved AFib på samtidig EKG-måling med tolv avledninger, rask puls ved sinustakykardi (puls på 100–150 slag/min) på samtidig EKG-måling med tolv avledninger og AFib med rask puls (puls på 100–150 slag/min) ved AFib på samtidig EKG-måling med tolv avledninger på en lesbar og klassifiserbar utskrift fra EKG-appen. Prosentandelen av SR (puls på 50–99 slag/min) som ble riktig klassifisert, var 98,3 prosent. For dette endepunktet oversteg den nedre konfidensgrensen på 96,6 prosent det på forhånd fastsatte resultatmålet på 90 prosent. Prosentandelen av AFib (puls på 50–99 slag/min) som ble riktig klassifisert, var 98,9 prosent. For dette endepunktet oversteg den nedre konfidensgrensen på 97,5 prosent det på forhånd fastsatte resultatmålet på 81 prosent. Prosentandelen av SR (puls på 100–150 slag/min) som ble riktig klassifisert, var 90,7 prosent. For dette endepunktet oversteg den nedre konfidensgrensen på 86,7 prosent det på forhånd fastsatte resultatmålet på 81 prosent. Prosentandelen av AFib (puls på 100–150 slag/min) som ble riktig klassifisert, var 83,0 prosent. For dette endepunktet oversteg den nedre konfidensgrensen på 77,8 prosent det på forhånd fastsatte resultatmålet på 75,5 prosent.

Tabell 7. Resultater for sensitivitet og spesifisitet (resultater for primært endepunkt) – klassifiserbart analyseutvalg

| Parameter | Verdi | Tosidig 95 % konfidensintervall ved bootstrapping* |
|---|-----------------------------|--|
| Sammenslått (i ro og etter fysisk aktivitet) | | |
| Referanseresultat = AFib (puls på 50–150 slag/min) | 481 | |
| EKG-app = AFib (puls på 50–150 slag/min) | 474/481 (98,5 %) | |
| EKG-app = SR (puls på 50–150 slag/min) | 7/481 (1,5 %) | |
| Sensitivitet | 474/481 (98,5 %) | (97,3 %, 99,6 %) |
| Referanseresultat = SR (puls på 50–150 slag/min) | 439 | |
| EKG-app = AFib (puls på 50–150 slag/min) | 3/439 (0,7 %) | |
| EKG-app = SR (puls på 50–150 slag/min) | 436/439 (99,3 %) | |
| Spesifisitet | 436/439 (99,3 %) | (98,4 %, 100,0 %) |
| *Resultater innhentet med nye uttrekk ved bootstrapping Forkortelser: AFib = atrieflimmer, SR = sinusrytme, HR = hjerterytme | | |

Tabell 8. Prosentandeler for rytmer riktig klassifisert som AFib og SR

| Parameter | Verdi | Konfidensintervall ved bruk av bootstrapping** |
|--|------------------|---|
| Sammenslått (i ro og etter fysisk aktivitet)* | | |
| Referanse for riktig klassifisert SR (puls på 50–99 slag/min) | 230/234 (98,3 %) | (96,6 %, 99,6 %) |
| Referanse for riktig klassifisert AFib (puls på 50–99 slag/min) | 272/275 (98,9 %) | (97,5 %, 100 %) |
| Referanse for riktig klassifisert SR (puls på 100–150 slag/min) | 186/205 (90,7 %) | (86,7 %, 94,6 %) |
| Referanse for riktig klassifisert AFib (puls på 100–150 slag/min) | 171/206 (83,0 %) | (77,8 %, 88,0 %) |
| *Beregningen av riktig klassifisering omfatter klassifiserbar AFib og SR i EKG 2.0-appen **Konfidensintervall innhentet med nye uttrekk ved bootstrapping Forkortelser: AFib = atrieflimmer, HR = hjerterytme, SR = sinusrytme | | |

I tillegg viste en sammenligning mellom algoritmen i EKG 2.0-appen og én avledning fra EKG-måling med tolv avledninger at den samlede prosentandelen godkjente morfologiretultater for sammenslåtte målinger i ro og etter fysisk aktivitet var 100 prosent, med til sammen fem utelatte resultater. Den nedre konfidensgrensen var 97,9 prosent, noe som oversteg den på forhånd fastsatte andelen godkjente morfologiske resultater på 80 prosent. Den samlede prosentandelen for godkjente sammenslåtte evalueringsresultater for R-bølgen i ro og etter fysisk aktivitet var 97,2 prosent. Den nedre konfidensgrensen var 93,6 prosent, noe som oversteg den på forhånd fastsatte samsvarsandelen for amplitude i R-bølgen på 80 prosent. Gruppene med målinger i ro og etter fysisk aktivitet hadde tilsvarende R-bølger.

Konklusjoner

Apple Watch-kunder har nå tilgang til to valgfrie funksjoner for påvisning av uregelmessig hjerterytme: funksjonen for varsling om uregelmessig hjerterytme (tilgjengelig på Apple Watch Series 1 og nyere) og EKG-appen (tilgjengelig på Apple Watch Series 4, Series 5 og Series 6).

Funksjonen for varsling om uregelmessig hjerterytme er en programvare som klassifiserer takogrammer som registreres opportunistisk i bakgrunnen, og varsler brukere som har aktivert funksjonen, om uregelmessig hjerterytme. Apple Heart Study har vist at av deltakerne som mottok en varsling mens de gikk med Apple Watch og en hjertemonitor, fikk 78,9 prosent også påvist AFib på hjertemonitoren, og 98,2 prosent fikk påvist AFib eller andre klinisk relevante arytmier. Disse resultatene viser at varslingene gir riktig påvisning av AFib i de fleste tilfeller, men at varslinger i enkelte tilfeller kan identifisere forekomster av andre typer arytmier enn AFib.

Apple Watch Series 4, Series 5 og Series 6 er utstyrt med en elektrisk hjertemåler som – sammen med EKG-appen og algoritmen – genererer en EKG-bølgeform lik et EKG med én enkelt avledning og kan klassifisere bølgeformen for å identifisere AFib. Den merkevarebeskyttede algoritmen som klassifiserer EKG-ene, hadde en sensitivitet på over 98 prosent og en spesifisitet på mer enn 99 prosent i påvisningen av AFib sammenlignet med EKG-målinger utført ved hjelp av en referanseenheter og tolket av uavhengige kliniske eksperter.

Personer som ønsker å bruke disse to funksjonene, må gjennomgå opplæring i brukergrensenettet for hvordan de tolker resultatene. De får også informasjon om at resultatene ikke alene kan brukes til diagnostisering, og om algoritmenes begrensninger. Begge funksjonene har fått klassifiseringen «De Novo» eller «510K» av FDA i USA.