The year in cardiovascular medicine 2021: digital health and innovation

Panos E. Vardas[1], Folkert W. Asselbergs[2], Maarten van Smeden[3], Paul Friedman[4]

Affiliations:

1. Heart Sector, Hygeia Hospitals Group, HHG Athens, Greece; European Heart Agency, ESC, Brussels, Belgium

2. Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom

3. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands.

4. Norman Blane Billie Jean Harty Chair, Mayo Clinic Department of Cardiovascular Medicine Honoring Robert L. Frye, M.D., Mayo Clinic, Rochester, MN, US

Address for correspondence: Panos Vardas, MD, PhD

Sector, Hygeia Hospitals Group

5, Erithrou Stavrou, Marousi, Athens, Greece

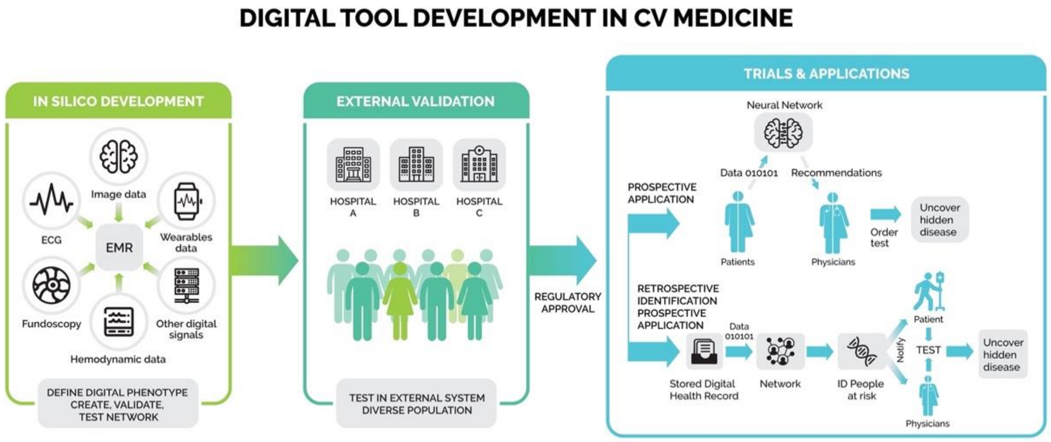
Postal Code 15123

Tel: +30 210 6867691 / +30 210 686 7937

Fax:+30 210 6869926

E-mail: pvardas@hygeia.gr

Graphical Abstract



Keywords: AI-ECG, AI-Wearables, Digital Health, Cardiovascular medicine, Big data, Machine learning

Introduction

Digital health, a broad-spectrum concept that has received a significant boost as a result of the Covid-19 pandemic, is growing exponentially, flexing its muscles with scientific breakthroughs and associated publications, while also driving trends and developments in industry.

For cardiovascular medicine in particular, during the last year an impressive number of authoritative new publications have confirmed previous research findings and proposed new innovative ideas and practices related to the diagnostic and therapeutic management of cardiovascular diseases, with the promise of ground-breaking developments during the coming years, for both cardiovascular sciences and care.

In the year 2021, as in the years immediately preceding, the field of digital health has been flooded with publications referring to the diverse applications of artificial intelligence (AI), from supervised to unsupervised learning, focusing mainly on the diagnostic capabilities of this impressive new technology.

Furthermore, the role of machine learning algorithms in the development of clinical prognostic models for risk assessment and early warning systems represents a rapidly evolving field that may be expected to have a catalytic effect by improving the prediction of medium- and long-term clinical outcomes.

Indeed, the prospects seem to be excellent.

Nonetheless, some questions still remain. Apart from the *in silico* design and development, the explainability of the machine learning algorithms and their validation methodology need to be more solidly confirmed in well-designed longitudinal studies, as well as in clinical practice before these algorithms find their way into the guidelines.

Beyond the field of AI—though often closely connected with it—developments in wearable devices have commandeered a significant part of the recent scientific literature, highlighting emerging new possibilities for the fuller monitoring and treatment of cardiovascular diseases and their related risk factors.

The technological developments in wearables—especially as they expand to cover not only the needs of fitness, but also those of diagnosis and monitoring of cardiovascular diseases—will obviously require more substantial regulation to ensure device reliability, backed by well-organised studies that will highlight their cost-effectiveness so that insurance companies may be persuaded they should be reimbursable.

AI-enabled cardiovascular diagnostic tools, techniques & methodologies

A new era in ECG analysis

The application of AI to the ECG has seen significant advances recently, and has developed in two broad categories: 1) tools to automate ECG interpretation, extending human capabilities via massive scalability, important as mobile form factors permit signal acquisition; and 2) algorithms to identify conditions not visible to human readers by training networks to identify multiple, complex, nonlinear patterns in the ECG signal to find occult disease (confirmed using other tests such as imaging), or impending disease. In contrast to automation tools in which a human overread provides a gold standard, algorithms identifying occult or future conditions require additional patient information.

Several groups have used large, labelled data sets to train neural networks to accurately apply diagnostic codes to single-lead and multiple-lead ECGs. Hannun[1] et al used 91,232 single-lead ECGs from a wearable patch to train a network to provide

12 rhythm classes, and found that the network outperformed the average cardiologist's read. Subsequently, two mega trials using smart watches based on PPG technology enrolled 419,297 and 246,541 patients to screen for AF in under 9 months.[2,3]

These trials confirmed the ability to massively enrol subjects and acquire data, at the cost of high rates of early dropout, low yield of disease (<0.5% in both studies), and with limited clinical characterisation of the study subjects. Ongoing trials will assess these tools in the context of patients selected for arrhythmia risk. Finally, there have been recent reports of interesting research that aimed to develop and validate an AI-enabled ECG algorithm capable of comprehensive 12-lead ECG analysis comparable to that of practicing cardiologists.[4]

Furthermore, the AI-ECG has identified occult and manifest cardiac conditions, including ventricular dysfunction,[5] peripartum cardiomyopathy,[6] amyloid heart disease[7] and pulmonary hypertension,[8] as well as non-cardiac conditions such as hyperkalaemia and cirrhosis.[9,10] In addition, special algorithms have been used for the early diagnosis of valvular diseases such as asymptomatic or oligosymptomatic severe aortic stenosis and mitral regurgitation,[11,12,13] left ventricular hypertrophy,[14,15] myocardial infarction[16,17] and a number of other conditions. Common findings in these studies include a strong clinical performance (AUC often above 0.90) and detection of disease months to years ahead of the clinical diagnosis.

The significance of these findings remains to be evaluated, taking into account the scalability of electrocardiography and hence the contribution of AI to its further and more substantial utilisation.

The ECG is an ever-present diagnostic tool that has served medical practitioners for more than a century. With the support of deep-learning AI techniques it is clearly entering a new era, in which it may prove to be a powerful detector of subclinical and clinical cardiac diseases, going beyond the boundaries of human observation. There can be no doubt that, when the previous capabilities of the ECG are combined with the evolving features of wearable devices such as smartphones, the chances of a much broader and pluralistic diagnostic process will increase rapidly.

The AI-ECG and clinical trials

Clinical trials are essential to demonstrate the ability of novel digital tools like the AI-ECG to improve human health. Factors to consider in evaluating the quality of AI-ECG studies are listed in Table 1. A framework for the assessment of how well AI-ECG clinical trials can predict meaningful outcomes, based on whether the trials are single-centre or multicentre, prospective or retrospective, is shown in Table 2. It is likely that level 3 or higher would be required for regulatory approval, allowing for variation in specific tests and regional differences. There is a pressing need for additional clinical trials to assess AI-ECG tools. A search of clinicaltrials.gov on Oct 8, 2021, for trials utilising the terms “artificial intelligence” and “ECG” returned 27 studies, with only 5 completed.

The first AI-ECG prospective trial published, the Eagle study,[18] demonstrated how digital, pragmatic trials can effectively and rapidly enrol subjects, and how the AI-ECG can positively impact clinical practice. It randomised 120 primary care teams from 45 clinics or hospitals in Minnesota and Wisconsin to an intervention arm (clinicians have access to AI-ECG results screening for left ventricular dysfunction when routinely ordering a clinical ECG) or a control arm (no AI results). Despite the development of the pandemic, over 22,000 patients were enrolled in 8 months, and the AI-ECG increased the diagnosis in the overall cohort (OR 1.32, p=0.007). The test performance

(AUC=0.92) matched that of the initial retrospective cohort (0.93).[19] Interestingly, the overall utilisation of echocardiography was similar in both groups but in the intervention group more echocardiograms were ordered for patients with a positive AI-ECG (38.1% control vs. 49.6% intervention, $p<0.001$), suggesting that the AI-ECG did not lead to more echocardiograms, but to better selection of patients to undergo imaging.

Cardiovascular imaging

Imaging has been the frontrunner in the application of AI in healthcare, because of the repetitive nature of imaging processing and evaluation. AI may improve imaging quality—and thereby scan and dose time—and assist in segmentation, processing and analysis.[20] Furthermore, most data are retrieved from a single standardised data source, making it more accessible for large scale analyses. During the pandemic, critics were pointing out that, despite massive efforts, AI had no impact on the care of COVID-19 patients, while simple straightforward randomised controlled trials did save lives.[21] However, this clearly shows only one side of the coin. The pandemic led to a greater burden on radiology resources, as CT scans were carried out routinely in all patients. AI is key in all parts of the imaging pipeline, including acquisition, processing, and analyses.[22,23] Furthermore, a plethora of papers have been published during the pandemic, showing the prognostic value of calcium score measurements in COVID-19 chest CT-scans.

Those measurements can be automated using deep learning,[24] providing clinicians with information, not only about the pulmonary status of COVID-19 patients, but also their cardiovascular risk.[25] AI will enable automated analyses of routine chest CT exams for opportunistic cardiovascular screening, allowing early preventive treatment. All these developments, together with the notable FDA clearance of a new technology to identify strokes on brain CT scans enabled by AI, hold out the prospect of a bright future in medical diagnostics.[26,27]

Retinal photography to detect cardiovascular disease

Another imaging application that can determine risk across a wide range of diseases is retinal photography. Retinal photography is a non-invasive imaging modality that aides in the diagnosis and treatment of major eye diseases, but can also provide information on the human vasculature and therefore cardiovascular disease. Prior manual coded studies have shown that retinal vascular abnormalities are predictive for cardiovascular disease.[28] Deep learning can extend this knowledge through automation and detection of more subtle signs that are not clearly visible to the human eye. Several large-scale studies have been published recently, focusing on the predictive value of features extracted from retinal photographs. Studies have shown that deep learning algorithms can predict levels of biomarkers such as haemoglobin to detect anaemia,[29] as well as age, sex, body composition and creatinine levels,[30] although external validation is warranted before this can be widely adopted in population screening. Another interesting study investigated the predictive capability of deep-learning–enabled coronary artery calcium (CAC) scores derived from retinal scan data.[31] CT scans and retinal measurements were performed on the same day and the score derived from retinal images showed an AUC of 0.74 for predicting CAC>0. Although higher than other single risk factors, like age, sex and cholesterol, the added predictive value in the multivariable clinical model was limited (AUC from 0.782 to 0.784). However, the CAC score derived from retinal scans showed a similar performance in predicting cardiovascular outcomes to CAC measured by CT scan (both AUC 0.71). Furthermore, the authors showed in the UK Biobank that this retinal-based

CAC score could improve risk stratification in those with borderline or intermediate risk.

Disadvantages exist. Home-based tests are not yet available, and images with poor quality were excluded in the reported analyses, which is likely to limit the external validity. Real-world data are necessary to estimate the added value in population screening, and the development of mobile applications for self-tests is needed[32] before implementation on a large scale. These deep learning applications are, however, already useful in those who already undergo regular retinal scans, such as diabetic patients, to screen for retinopathy.[33]

To close this section, at least a brief mention should be made of the diagnostic capability and cost-effectiveness of the combined imaging approach, where the use of AI and MRI yields the atheroma index of the coronary arteries or peripheral vessels as a byproduct of the primary diagnostic evaluation of other organs.[34]

Automation of imaging processing

While the application of AI in cardiovascular imaging for clinical decision making is still in its infancy, the use of AI to automate imaging processing in other fields, such as ophthalmology as discussed above, oncology and dermatology, has already matured. However, several promising studies using different imaging modalities have recently been published and show that cardiology is able to catch up with the other disease domains. A large international collaborative study showed that the coefficient of variation in measuring left ventricular wall thickness by cardiovascular magnetic resonance was significantly lower for machine learning in comparison to human experts.[35] This study involved a cohort of patients with hypertrophic cardiomyopathy, where variations in wall thickness measurements directly impact clinical decision making by affecting the calculation of sudden death risk and thereby the indication for preventive ICD implementation.

Another recent example of automation is the International Society for Heart and Lung Transplantation's grading of endomyocardial biopsies in heart transplant patients.[36] The authors compared histological grading performed by expert pathologists with a computer-assisted automated pipeline and showed similar performance of the Computer-Assisted Cardiac Histologic Evaluation (CACHE) grader in comparison to the pathologist (Figure 1). Moreover, they showed only limited attenuation of the performance when it was applied to an external validation dataset, indicating good generalisability across different scanning and tissue preparation protocols. International collaborative efforts in the field of transplant research have been hampered by variations in grading by individual centres, which increase the noise-to-signal ratio in the detection of biologically meaningful results when datasets from individual centres are merged. CACHE-enabled automated grading can play an instrumental role in advancing the field of transplant research.

Finally, AI will increasingly be applied in the field of echocardiography. Prior studies have shown that AI can identify different echo views, can segment cardiac structures, estimate ejection fraction[37,38] and diagnose diseases such as cardiac amyloidosis.[39] Recently, a study from Stanford also showed that deep learning algorithms are able to detect pacemaker or ICD leads and, interestingly, are able to predict age, sex, height and weight based on echo images.[40] Furthermore, they used gradient-based sensitivity mapping methods to highlight the regions of interest for human interpretation. Visualisation methods to unlock the so-called "black box" algorithms are essential if healthcare professionals are to fully adopt the results generated by AI models. These algorithms will support untrained professionals with the interpretation of echocardiograms when cardiological expertise is of limited availability. A recent study showed that deep learning can even help untrained nurses to perform limited echocardiograms for standard evaluation of left and right ventricular size and pericardial effusion, enabling the use of echocardiograms in non-cardiological settings, such as primary care, COVID wards or remote areas.[41] However, before its widespread implementation, additional studies regarding safety and generalisability are warranted.

Big data and prognostic models for cardiovascular risk prediction

Machine learning for risk prediction

Clinical risk prediction modelling based on machine learning has been an active field of research. During the first months of the pandemic, hundreds of such models were developed.[42] Clinical prediction models are commonly developed to inform physicians about the probability of a certain disease being present (diagnosis), or to predict a certain health state in the future (prognosis), for individual patients, and to use that knowledge in the care of those patients.[43] By applying machine learning techniques that can use complex data relationships between predictors and outcome without the need for the modeller to pre-specify them, the expectation is that the accuracy of predictions will improve compared to traditional risk prediction modelling approaches, and that its application will be less labour intensive at the bedside.

Improvements in predictive accuracy are, however, not guaranteed.[44] For instance, a study that developed machine learning models to predict the risk of death after acute myocardial infarction (AMI) found that machine learning models were not uniformly superior to a traditional logistic regression approach in a cohort of 755,402 AMI patients.[45] In fact, of the three models used, two were superior to the logistic regression model for risk stratification. In addition, those two models were much better calibrated across patient groups based on age, sex, race and mortality risk, and thus better suited for risk prediction. In contrast, the third model, based on a neural network, was found to be inferior to the logistic regression model used in the study. There may be pragmatic reasons for this inferiority, but they are probably related to the methodology used and in particular the sample sizes of each of the study's populations.

Nonetheless, in other settings machine learning approaches have yielded promising results. One such study developed models to predict the risk of death, myocardial infarction, and major bleeding after an acute coronary syndrome (ACS). The machine learning-based models were developed from a cohort with 19,826 adult ACS patients and were shown to predict the risk with high AUCs on external validation, at 1 year (AUCs: 0.81 to 0.92) and 2 years (AUCs: 0.84 to 0.93).[46]

Early warning systems

Early warning systems are prognostic predictive models that aim to inform physicians about important future health outcomes. Often, these early warning systems are used to monitor patients and to update these predictions over time. For instance, to predict circulatory failure in patients admitted to intensive care, a machine learning model was developed that made a new prediction for every patient every 5 minutes.[47] The early warning systems developed were shown to yield high AUCs, between 0.88 and 0.94. However, these models also produced 2 to 3 alarms per patient per day. This may result in so-called "alarm fatigue", which can lead to inadequate responses and may even impact patient safety.[48] Hence, for these early warning systems and other risk prediction models used to guide clinical decisions, it is essential to ensure safety and effectiveness in improving patient outcomes, for instance through an RCT comparing the early warning system to standard of care. One such an RCT evaluated a machine learning-based early warning system for pending intraoperative hypotension.[49] This early warning system updates every 20 seconds the probability of a hypotensive event in the next 15 minutes (warning when estimated probability >85%) based on the arterial pressure waveform.[50] In an RCT with 60 adult elective non-cardiac surgery patients, the early warning system, in combination with a hemodynamic diagnostic guidance and treatment protocol, reduced the median total time of hypotension per patient from 32.7 minutes under standard of care to 8 minutes.

Big data: representativeness and algorithmic fairness

Access to large and diverse databases with electronic health records creates important new research opportunities. Such large databases include the Clinical Practice Research Datalink (CPRD), with highly detailed data from over 5 million individuals representative of the UK population. Using the CPRD data, one interesting study developed and validated several machine learning-based risk prediction models for predicting the risk of familial hypercholesterolaemia in primary care patients.[51] These prediction models were shown to have high AUCs of around 0.89. The large scale and representativeness of large databases also allows for studying specific groups that may otherwise be difficult to study. For instance, one study compared cardiovascular disease incidences and outcomes in homeless individuals using a linkage between CPRD, hospital episode statistics and the Office of National Statistics for mortality data.[52] This study showed that homeless individuals have a 1.8 times higher risk of developing cardiovascular disease and are 1.6 times more likely to die within 1 year after cardiovascular disease diagnosis, compared to similar individuals who are not homeless. Finally, large and diverse databases, where minority groups are also well represented, are essential to ensure that the algorithms developed are fair,[53] i.e. do not systematically disadvantage certain groups of individuals. This requires evaluation of the performance of the algorithms in important subgroups. For instance, a recent study on atherosclerotic cardiovascular disease risk prediction showed a comparable performance of existing pooled cohort equations and newly developed machine-learning based models in Asian and Hispanic subgroups, for which the performance was so far uncertain.[54]

Wearable devices in cardiovascular risk assessment, cardiovascular disease prevention, diagnosis and management

Wearables in AF risk assessment and management

The role of physical activity as a modifiable risk factor for the development of atrial fibrillation (AF) was studied recently in a well-organised prospective study,[55] which included 93,669 participants from the UK Biobank prospective cohort, without a prevalent history of AF, who wore a wrist-based triaxial accelerometer for one week. The sensor captured acceleration at 100 Hz with a dynamic range of ±8 g. The primary outcome of the study was incident AF.

According to the findings of the study, greater accelerometer-derived physical activity is associated with a lower risk of incident AF and stroke, after adjustment for clinical risk factors (Figure 2). Wearable sensors may enable both objective assessment of physical activity and modification of AF risk through targeted feedback. The authors consider that future preventive efforts to reduce AF risk may be most effective if they target adherence to objective activity thresholds.

Another study[56] that aimed to investigate the association between changes in physical activity and the onset of AF reported similar findings. A total of 1410 participants from the general population were studied (46.2% women, mean age

74.7±4.1 years), with risk factors but with no prior AF diagnosis, who underwent continuous monitoring for AF episodes along with daily accelerometric assessment of physical activity, using an implantable loop recorder, over an average period of 3.5 years.

According to the findings of the study, intra-individual changes in physical activity were associated with the onset of AF episodes, as detected by continuous monitoring, in a high-risk population. For each person, a 1-h decrease in daily physical activity during the previous week increased the odds of AF onset the next day by ~25%, while the strongest association was seen in the group with the lowest activity overall.

Apart from these two recent and revealing studies of the relationship between a person's physical activity and the occurrence of AF, a significant number of ongoing or recently published studies have evaluated the capabilities of wearables, focusing on the relationship between the individual clinical outcome and the burden of recorded episodes of clinical or subclinical AF.[57]

Wearables in HF assessment and management

Heart failure (HF), a fast growing disease internationally, also has a long-standing affinity with wearable technology, since the pathophysiology of the disease and its clinical consequences require close and continuous long-term monitoring. Indeed, wearables offer a unique opportunity to assess patients' status and a number of indicators closely, outside the classical settings. In patients with HF, data from consumer wearables, such as physical activity step count or heart rate, but also more intense monitoring of such factors as pulmonary artery pressure or fluid retention, have long been the target of these evolving devices.

When we look at the findings and messages of the most recent relevant studies, those of the Link-HF multicentre study by Stehlik et al,[58] which evaluated the accuracy of non-invasive remote monitoring in predicting rehospitalisation for HF, were quite revealing. This was a study of 100 patients with heart failure, aged 68.4±10.2 years (only 2% female). The investigators showed that multivariate physiological telemetry from a wearable sensor, in combination with machine learning analytics, can accomplish accurate early detection of impending rehospitalisation with a predictive accuracy comparable to that of implantable devices. The authors emphasise, however, that the clinical efficacy and generalisability of this low-cost non-invasive approach to rehospitalisation mitigation still needs further testing.

Looking at the issues more broadly, apart from the use of modern electronic technology for continuous haemodynamic monitoring in HF patients, it has become clear that such technology can and should be used for education and support in these patients' therapeutic management.[59]

The EPIC-HF study (Electronically Delivered Patient-Activation Tool for Intensification of Medications for Chronic Heart Failure with Reduced Ejection Fraction) evaluated patients from a diverse health system who had HF and reduced ejection fraction, randomising them to usual care versus patient activation tools. The tools—a 3-minute video and a 1-page checklist—encouraged patients to work collaboratively with their clinicians to "make one positive change" in their HF medication.

The findings were clear. A patient activation tool delivered electronically before the cardiology clinic visit enhanced clinicians' intensification of guideline-directed medical therapies.

ST-segment elevation myocardial infarction.

The vast majority of wearable devices currently offer single-lead electrocardiographic (ECG) recording, which allows the detection of AF and, more rarely, other arrhythmias to a satisfactory extent. However, such ECG recordings cannot reliably detect ST/T changes due to regional myocardial ischaemia. Nevertheless, a good many expectations have been invested in this possibility, as ECG recording by wearables, backed by telemonitoring to detect the early signs of myocardial ischaemia, could limit its often destructive effects.

Muhlestein J. et al.,[60] in their relatively recent publication, reviewed the feasibility of combining serial smartphone single-lead recordings to create a virtual 12-lead ECG capable of reliably diagnosing ST-elevation myocardial infarction. The study included 200 subjects (mean age 60 years, 43% female).

For all interpretable pairs of smartphone ECGs, compared with standard 12-lead ECGs (n=190), the sensitivity, specificity, and positive and negative predictive values for STEMI or STEMI equivalent (LBBB) achieved by the smartphone were 0.89, 0.84, 0.70 and 0.95, respectively. The authors concluded that a 12-lead equivalent ECG constructed from multiple serial single-lead recordings from a smartphone can identify STEMI with a good correlation to a standard 12-lead ECG.

Similarly to the previous study, a prospective study[61] also investigated the feasibility and accuracy of a smartwatch in recording multiple electrocardiographic leads and detecting ST-segment changes associated with acute coronary syndromes, compared with a standard 12-lead ECG. A commercially available smartwatch was used in 100 participants. The watch was placed in different body positions to obtain 9 bipolar ECG tracings (corresponding to Einthoven leads, II and III and precordial leads V1-V6), which were compared with a simultaneous standard 12-lead ECG.

To a significant extent there was agreement between the findings of the smartwatch tracings and the standard ECGs for the identification of a normal ECG, ST-segment changes, and no ST-segment elevation.

The findings of the two previous studies give cause for optimism that, in the near future, the technical difficulties will be overcome, so that the recording of wearable devices will gain sufficient reliability for the recording of ischaemic changes on the ECG.

Conclusions

Digital health stands poised to transform cardiovascular medicine, much as echocardiographic imaging has upended stethoscope-based auscultation for diagnosis. Work published in 2021 has advanced this hope, and engaged an ever-widening group of stakeholders, critical to ensure proper evaluation of this important technology that may touch so many lives. Digital health's great promise in no small measure stems from its ability to endow extant medical tests (ECG, fundoscopy, imaging) and EHR data that are known to practitioners and integrated into workflows with new superpowers, and to draw massively scalable data from wearables into the fold. This integration will accelerate adoption, and impact care.

Before the promise of digital health can bear fruit to improve human health, a major gap must be addressed – the paucity of clinical trials to address outcomes. The "black box" issue and lack of explainability are widely discussed concerns that may not be solved in the short term, but may be mitigated or overcome with robust evidence from prospective clinical trials. Data management processes to prevent overwhelming an already taxed health care system are mandatory. Further development of novel hybrid regulatory strategies recognising software as a medical device coupled to consumer hardware are prerequisites to exponentially driving data availability. With broad input from clinicians, industry, regulators, and patients; attention to privacy and human rights; diligent testing, validation and oversight; and prospective trial data, digital health promises an exciting and healthy future, as opposed to a brave new world.

Disclosure

Conflict of Interest: P.V. has teaching fees from Servier International and consultancy fees from Hygeia Hospitals Group, HHG and ESC/ European Heart Agency, Brussels. P.F. is co-inventor of AI-ECG algorithms, and he and Mayo Clinic may benefit from their commercialisation. F.A is supported by UCL Hospitals NIHR Biomedical Research Center. M.v.S. has nothing to declare.

Acknowledgments

The authors would like to thank the following for their contributions to the preparation of the manuscript: Dr. Anastasia Xintarakou, Mr. Philip Lees, Mrs. Alexandra Kourlampa. Thanks also to Professors E. Peyster and S. Khurshid for permission to use the figures.

References

1. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65–69. doi: 10.1038/s41591-018-0268-3

2. Perez M V., Mahaffey KW, Hedlin H, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019;381:1909–1917. doi: 10.1056/nejmoa1901183

3. Guo Y, Wang H, Zhang H, et al. Mobile Photoplethysmographic Technology to Detect Atrial Fibrillation. J Am Coll Cardiol 2019;74:2365–2375. doi: 10.1016/j.jacc.2019.08.019

4. Kashou AH, Ko W-Y, Attia ZI, et al. A comprehensive artificial intelligence–enabled electrocardiogram interpretation program. Cardiovasc Digit Heal J 2020;1:62–70. doi: 10.1016/j.cvdhj.2020.08.005

5. Adedinsewo D, Carter RE, Attia Z,et al. Artificial Intelligence-Enabled ECG Algorithm to Identify Patients with Left Ventricular Systolic Dysfunction Presenting to the Emergency Department with Dyspnea. Circ Arrhythmia Electrophysiol 2020;13:e008437. doi: 10.1161/CIRCEP.120.008437.

6. Adedinsewo DA, Johnson PW, Douglass EJ,et al. Detecting cardiomyopathies in pregnancy and the postpartum period with an electrocardiogram-based deep learning model. Eur Hear J - Digit Heal 2021;, ztab078, doi: 10.1093/ehjdh/ztab078

7. Grogan M, Lopez-Jimenez F, Cohen-Shelly M, et al. Artificial Intelligence–Enhanced Electrocardiogram for the Early Detection of Cardiac Amyloidosis. *Mayo Clin Proc*, doi: 10.1016/j.mayocp.2021.04.023. Published online ahead of

print 02/07/21

8. Kwon J myoung, Kim KH, Medina-Inojosa J, et al. Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography. J Hear Lung Transplant 2020;39:805–814. doi: 10.1016/j.healun.2020.04.009

9. Kwon J myoung, Jung MS, Kim KH, et al. Artificial intelligence for detecting electrolyte imbalance using electrocardiography. Ann Noninvasive Electrocardiol 2021;26:e12839. doi: 10.1111/anec.12839

10. Galloway CD, Valys A V., Shreibati JB, et al. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia from the Electrocardiogram. JAMA Cardiol 2019;4:428–436. doi: 10.1001/jamacardio.2019.0640

11. Kwon JM, Lee SY, Jeon KH, et al. Deep learning–based algorithm for detecting aortic stenosis using electrocardiography. J Am Heart Assoc 2020;9:e014717. doi: 10.1161/JAHA.119.014717

12. Cohen-Shelly M, Attia ZI, Friedman PA, et al. Electrocardiogram screening for aortic valve stenosis using artificial intelligence. Eur Heart J 2021;42:2885–2896. doi: 10.1093/eurheartj/ehab153

13. Kwon J myoung, Kim KH, Akkus Z, et al. Artificial intelligence for detecting mitral regurgitation using electrocardiography. J Electrocardiol 2020;59:151–157. doi: 10.1016/j.jelectrocard.2020.02.008

14. Attia ZI, Kapa S, Noseworthy PA, Lopez-Jimenez F, Friedman PA. Artificial Intelligence ECG to Detect Left Ventricular Dysfunction in COVID-19: A Case Series. Mayo Clin Proc 2020;95:2464–2466. doi: 10.1016/j.mayocp.2020.09.020

15. Kwon JM, Jeon KH, Kim HM, et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. Europace 2020;22:412–419. doi: 10.1093/europace/euz324

16. Cho Y, Kwon J myoung, Kim KH, et al. Artificial intelligence algorithm for detecting myocardial infarction using six-lead electrocardiography. Sci Rep 2020;10:1–10. doi: 10.1038/s41598-020-77599-6

17. Makimoto H, Höckmann M, Lin T, et al. Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction. Sci Rep 2020;10:1–9. doi: 10.1038/s41598-020-65105-x

18. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med 2021;27:815–819. doi: 10.1038/s41591-021-01335-4

19. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med 2019;25:70–74. doi: 10.1038/s41591-018-0240-2

20. Antoniades C, Oikonomou EK. Artificial intelligence in cardiovascular imaging—principles, expectations, and limitations. *Eur Heart J*, doi: 10.1093/eurheartj/ehab678. Published online ahead of print 24/09/21

21. RECOVERY Collaborative Group; Horby P, Lim WS, Emberson JR, et al. Dexamethasone in Hospitalized Patients with Covid-19. N Engl J Med 2021;384:693–704. doi: 10.1056/nejmoa2021436

22. Shamout FE, Shen Y, Wu N, et al. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. npj Digit Med 2021;4:1–11. doi: 10.1038/s41746-021-00453-0

23. Shi F, Wang J, Shi J, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. IEEE Rev Biomed Eng 2021;14:4–15. doi: 10.1109/RBME.2020.2987975

24. Eng D, Chute C, Khandwala N, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. npj Digit Med 2021;4:1–13. doi: 10.1038/s41746-021-00460-1

25. Zeleznik R, Foldyna B, Eslami P, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. Nat Commun 2021;12:1–9. doi: 10.1038/s41467-021-20966-2

26. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Heal 2020;2:e486–e488. doi: 10.1016/S2589-7500(20)30160-6

27. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392:2388–2396. doi: 10.1016/S0140-6736(18)31645-3

28. Seidelmann SB, Claggett B, Bravo PE, et al. Retinal Vessel Calibers in Predicting Long-Term Cardiovascular Outcomes: The Atherosclerosis Risk in Communities Study. Circulation 2016;134:1328–1338. doi: 10.1161/CIRCULATIONAHA.116.023425

29. Mitani A, Huang A, Venugopalan S, et al. Detection of anaemia from retinal

fundus images via deep learning. Nat Biomed Eng 2020;4:18–27. doi: 10.1038/s41551-019-0487-z

30. Rim TH, Lee G, Kim Y, et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. Lancet Digit Heal 2020;2:e526–e536. doi: 10.1016/S2589-7500(20)30216-8

31. Rim TH, Lee CJ, Tham YC, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. Lancet Digit Heal 2021;3:e306–e316. doi: 10.1016/S2589-7500(21)00043-1

32. Tran T, Huang NT, Montezuma SR. Smartphone Funduscopy - How to Use Smartphone to Take Fundus Photographs. https://eyewiki.aao.org/Smartphone_Funduscopy_-_How_to_Use_Smartphone_to_Take_Fundus_Photographs (02 March 2018)

33. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. Lancet Digit Heal 2021;3:e10–e19. doi: 10.1016/S2589-7500(20)30250-8

34. Canton G, Hippe DS, Chen L, et al. Atherosclerotic burden and remodeling patterns of the popliteal artery as detected in the magnetic resonance imaging osteoarthritis initiative data set. J Am Heart Assoc 2021;10:18408. doi: 10.1161/JAHA.120.018408

35. Augusto JB, Davies RH, Bhuva AN, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. Lancet Digit Heal 2021;3:e20–e28. doi: 10.1016/S2589-7500(20)30267-3

36. Peyster EG, Arabyarmohammadi S, Janowczyk A, et al. An automated computational image analysis pipeline for histological grading of cardiac allograft rejection. Eur Heart J 2021;42:2356–2369. doi: 10.1093/eurheartj/ehab241

37. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice: Feasibility and diagnostic accuracy. Circulation 2018;138:1623–1635. doi: 10.1161/CIRCULATIONAHA.118.034338

38. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. npj Digit Med 2018;1:1–11. doi: 10.1038/s41746-018-0065-x

39. Goto S, Mahara K, Beussink-Nelson L, et al. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. Nat Commun 2021;12:1–12. doi: 10.1038/s41467-021-22877-8

40. Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. NPJ Digit Med 2020;3:10. doi: 10.1038/s41746-019-0216-8

41. Narang A, Bae R, Hong H, et al. Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. JAMA Cardiol 2021;6:624–632. doi: 10.1001/jamacardio.2021.0185

42. Wynants L, Calster B Van, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ 2020;369:26. doi: 10.1136/bmj.m1328

43. Smeden M van, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. J Clin Epidemiol 2021;132:142–145. doi: 10.1016/j.jclinepi.2021.01.009

44. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22. doi: 10.1016/j.jclinepi.2019.02.004

45. Khera R, Haimovich J, Hurley NC, et al. Use of Machine Learning Models to Predict Death after Acute Myocardial Infarction. JAMA Cardiol 2021;6:633–641. doi: 10.1001/jamacardio.2021.0122

46. D'Ascenzo F, Filippo O De, Gallone G, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. Lancet 2021;397:199–207. doi: 10.1016/S0140-6736(20)32519-8

47. Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med 2020;26:364–373. doi: 10.1038/s41591-020-0789-4

48. Wilken M, Hüske-Kraus D, Klausen A, et al. Alarm fatigue: Causes and effects. Stud. Health Technol. Inform. 2017;243:107-111. doi: 10.3233/978-1-61499-808-2-107

49. Wijnberge M, Geerts BF, Hol L, et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension during Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA 2020;323:1052–1060.

doi: 10.1001/jama.2020.0592

50. Hatib F, Jian Z, Buddi S, et al. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology 2018;129:663–674. doi: 10.1097/ALN.0000000000002300

51. Akyea RK, Qureshi N, Kai J, Weng SF. Performance and clinical utility of supervised machine-learning approaches in detecting familial hypercholesterolaemia in primary care. npj Digit Med 2020;3:1–9. doi: 10.1038/s41746-020-00349-5

52. Nanjo A, Evans H, Direk K, et al. Prevalence, incidence, and outcomes across cardiovascular diseases in homeless individuals using national linked electronic health records. Eur Heart J 2020;41:4011–4020. doi: 10.1093/eurheartj/ehaa795

53. Mhasawade V, Zhao Y, Chunara R. Machine learning and algorithmic fairness in public and population health. Nat Mach Intell 2021;3:659–666. doi: 10.1038/s42256-021-00373-4

54. Ward A, Sarraju A, Chung S, et al. Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. npj Digit Med 2020;3:1–7. doi: 10.1038/s41746-020-00331-1

55. Khurshid S, Weng LC, Al-Alusi MA, et al. Accelerometer-derived physical activity and risk of atrial fibrillation. Eur Heart J 2021;42:2472–2483. doi: 10.1093/eurheartj/ehab250

56. Bonnesen MP, Frodi DM, Haugan KJ, et al. Day-to-day measurement of physical activity and risk of atrial fibrillation. Eur Heart J 2021;42:3979–3988. doi: 10.1093/eurheartj/ehab597

57. Bonnesen MP, Diederichsen SZ, Isaksen JL, et al. Atrial fibrillation burden and cognitive decline in elderly patients undergoing continuous monitoring. Am Heart J 2021;242:15–23. doi: 10.1016/j.ahj.2021.08.006

58. Stehlik J, Schmalfuss C, Bozkurt B, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: The link-hf multicenter study. Circ Hear Fail 2020;13:e006513. doi: 10.1161/CIRCHEARTFAILURE.119.006513.

59. Allen LA, Venechuk G, McIlvennan CK, et al. An electronically delivered patient-activation tool for intensification of medications for chronic heart failure with reduced ejection fraction the EPIC-HF trial. Circulation 2021;143:427–437. doi: 10.1161/CIRCULATIONAHA.120.051863

60. Muhlestein JB, Anderson JL, Bethea CF, et al. Feasibility of combining serial smartphone single-lead electrocardiograms for the diagnosis of ST-elevation myocardial infarction: Smartphone ECG for STEMI Diagnosis. Am Heart J 2020;221:125–135. doi: 10.1016/j.ahj.2019.12.016

61. Indolfi C, Spaccarotella CAM, Polimeni A, et al. Multichannel Electrocardiograms Obtained by a Smartwatch for the Diagnosis of ST-Segment Changes. JAMA Cardiol 2020;5:1176–1180. doi: 10.1001/jamacardio.2020.3994

Figure legends

Figure 1: An overview of the 'Computer-Assisted Cardiac Histologic Evaluation-Grader' multi-centre validation experiment. The Computer-Assisted Cardiac Histologic Evaluation-Grader performance was compared to both the grade of record and to independent pathologists performing re-grading, demonstrating non-inferiority to expert pathologists, generalizability to external datasets, and excellent sensitivity and negative predictive value. Reproduced by permission, from Peyster EG et al.[36]

Figure 2: Cumulative risks of atrial fibrillation (upper panel) and stroke (lower panel) stratified by adherence to physical activity recommendations, as it is validated by accelerometer-derived physical activity. Reproduced by permission, from Khurshid S. et al.[55]

Tables

Table 1: Factors to consider in evaluating AI-ECG studies

1. Data label accuracy: robustness of data labels used for training and testing
    a. Proxy labels (EMR report of "chest pain") vs. gold standard labels (physician described angina, troponin levels, serial ECGs)
    b. Number of subjects for whom labels available
    c. Absence in labels of false distractors (e.g. all ECGs from patients with condition taken at one hospital, using an acquisition system different than that used in controls, so that network may identify differences in ECG machines rather than disease)
2. Risk of bias: cohort creation and controls
    a. Controls not identical to cases in all conditions except the desired AI differentiator, most commonly in demographics (example: using adult controls for pediatric ECGs with WPW to train a network)
    b. Controls and cases taken from public data sets (difficult to know details regarding absence/presence of conditions, poor phenotyping)
    c. Use of only subsets of larger data sets, introducing potential bias – need for racial, ethnic, and geographic diversity in data sets (example: initial face recognition AI trained using only Caucasians, mislabeled African Americans as primates)
    d. Inappropriate exclusion of data at the patient or signal feature level will bias results (examples: exclusion of signals on the basis of artifact of those same exclusions won't be used in real world implementation; or exclusion of patients with hypertension when creating an AI ECG screen for hypertension)

e. Temporal shifts – training using data acquired in the remote past and application to recent data sets
f. Commercial interest and backgrounds of engineers creating AI tools (potential bias)

3. Overfitting/ lack of generalizability
    a. Overly complex AI ECG network with a small number of samples (the results are not generalizable to other populations).
    b. Most datasets for AI-ECG training number in the tens of thousands or more, although exceptions exist

Table 2: Proposed Categories of Clinical Trials to Assess the AI-ECG

Study Category	Description of Population Used to TEST an AI-ECG network	Study Design	Strengths	Limitations
1	Public data set	Retrospective	Inexpensive, rapid	Unreliable phenotyping, high risk of bias, limited clinical utility
2	Single Center: same hospital/clinic used to acquire data, but different patients	Retrospective	Rich data sets to phenotype patients, rapid, relatively inexpensive, robust proof of concept approach	Risk of bias, underrepresenting important populations
3	Multicenter: different hospital system used to test AI, than one used to create	Retrospective	Lower risk of bias, potential for greater diversity among subjects, test types, potential to rapidly and meaningfully assess tests	Need to confirm labels assessed in systematic, similar manner across sites (example: assessment of EF by echo)
4	Single center: same hospital used to test AI, different patients	Prospective	Assesses AI, impact on workflow, adoption by clinicians, clinical impact	Greater technical infrastructure required, more expensive, greater time requirement
5	Multicenter	Prospective (may use retrospective ECGs to prospectively enroll patients)	Prospective trial but with accelerated enrollment, by screening large dataset of stored ECGs; potential for portal/email study invitations and pragmatic design, statistical robust, potential to minimize bias	Greater technical requirements, time, expense

Figures

Figure 1

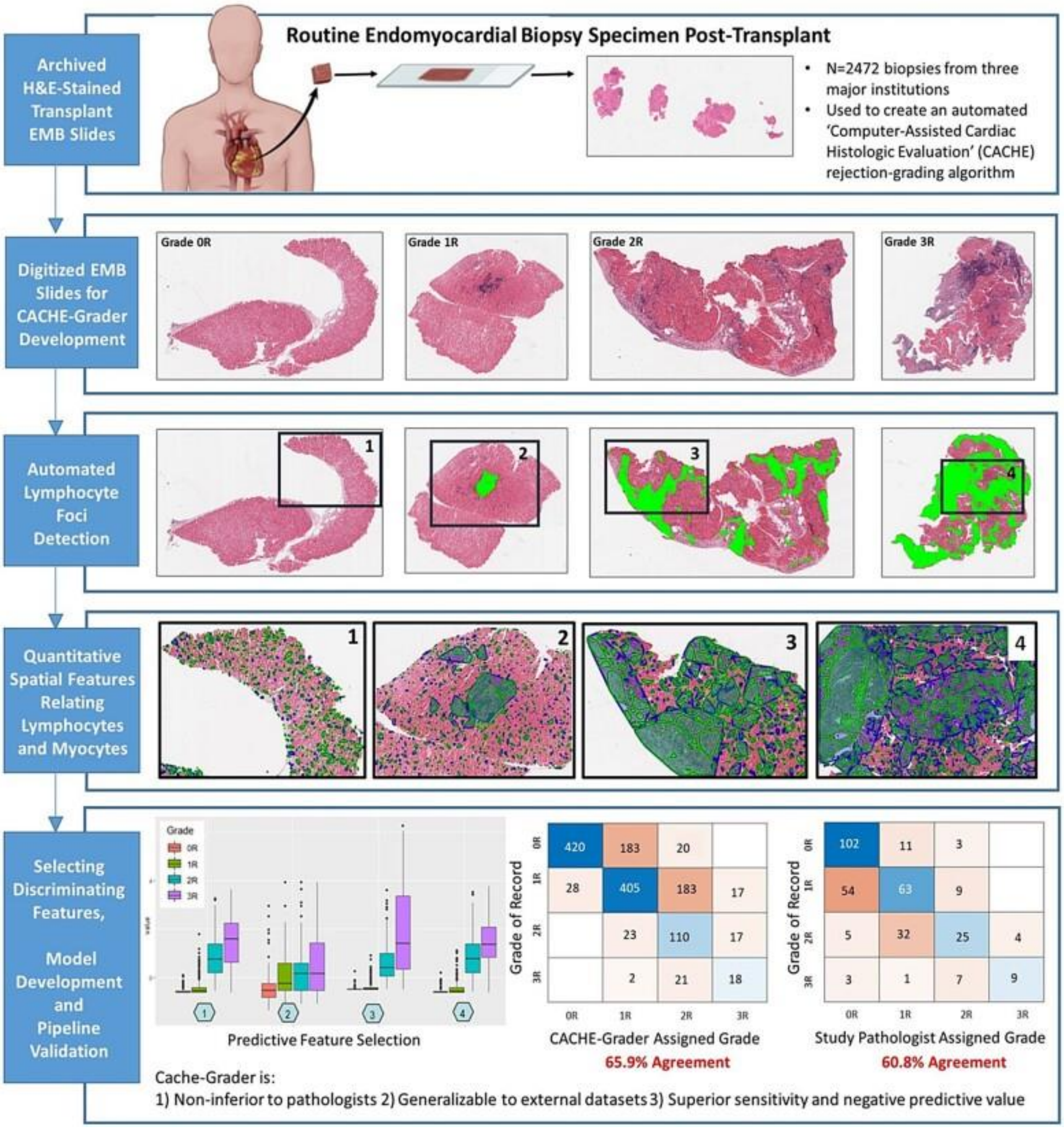


Figure 2

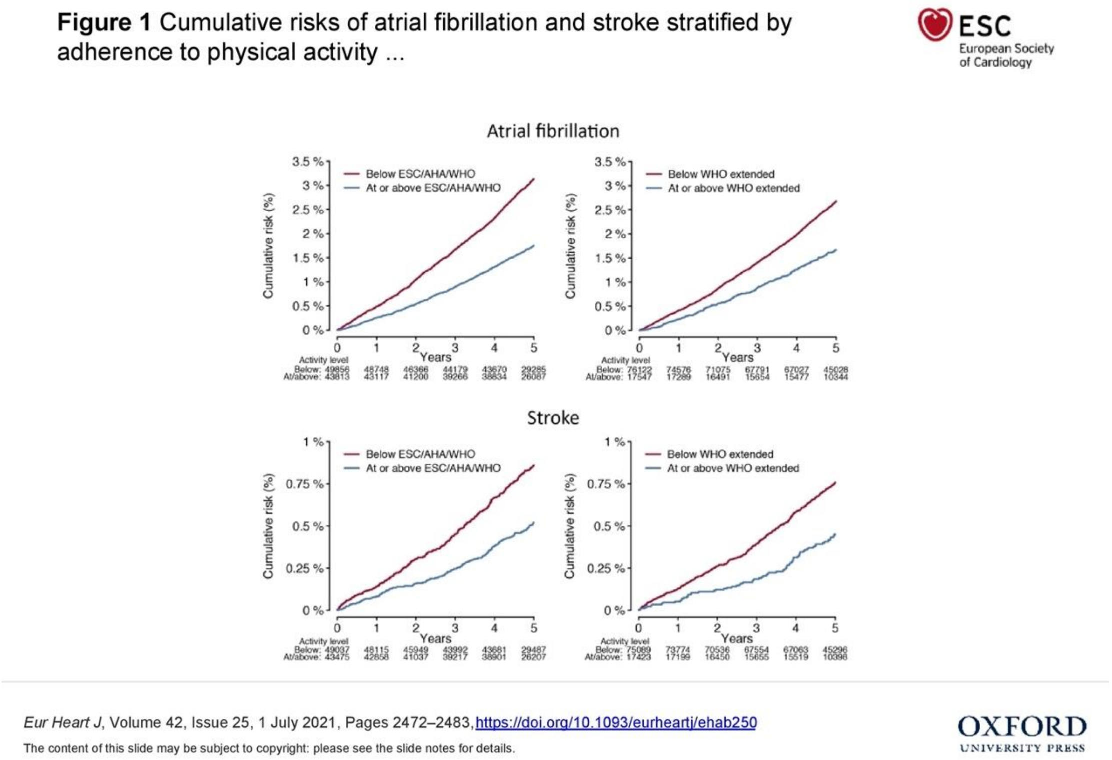


Figure 1 Cumulative risks of atrial fibrillation and stroke stratified by adherence to physical activity ...
ESC
European Society of Cardiology
Atrial fibrillation
Below ESC/AHA/WHO
At or above ESC/AHA/WHO
Below WHO extended
At or above WHO extended
Cumulative risk (%)
Years
Activity level
Stroke
Eur Heart J, Volume 42, Issue 25, 1 July 2021, Pages 2472–2483, https://doi.org/10.1093/eurheartj/ehab250
The content of this slide may be subject to copyright: please see the slide notes for details.
OXFORD
UNIVERSITY PRESS