Imperial College London

Towards a Tricorder: Clinical, Health Economic, and Ethical Investigation of Point-of-Care Artificial Intelligence Electrocardiogram for Heart Failure

Patrik Bachtiger

MBBS BSc MRCP (UK) MSc MBA MPH

Supervisor: Professor Nicholas Peters

Department: National Heart and Lung Institute

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy (PhD)

First submitted: September 24th, 2022

Revisions submitted: January 16th, 2023

Final submission: January 24th, 2023

Statement of Originality and Contribution

In accordance with Imperial College London regulations, I declare that all the work within this thesis is my own, unless specifically stated otherwise in the text. For all chapters, I benefitted from feedback from my supervisor.

For Chapter 2, I led on defining the research questions; wrote the grant that secured NIHR funding; designed and authored all aspects of the research protocol; determined the statistical analysis plan; authored all the study materials; designed the data collection instruments (including REDcap database); secured Health Research Authority ethical approval; secured research and data governance approval for international data transfers; formulated the statistical analysis plan; managed all device logistics; managed all clinical, academic (international), funder, and patient/public involvement stakeholder relations; trained all operators contributing to recruitment across all seven sites; project managed the entire study; and led on writing the first draft of the manuscript accepted by The Lancet Digital Health. I directly supervised recruitment of over 200 patients; team members listed under Acknowledgements conducted the remainder independently (with me always available to troubleshoot). Statistical support with R and Python was generously provided by Dr Camille Petri.

For Chapter 3, I led on the formulation of the research question; mapped the logic model for the study; defined the clinical and health economic parameters (codes) for inclusion; designed the propensity-score-matching approach; formulated the analysis plan; and presented the study protocol to the Discover Research Access Group for approval. Moulesh Shah at Imperial College Health Partners performed data extraction using SQL. Dr Camille Petrie provided support with statistical analyses in R. Gareth Hooper and Samir Khan validated the final approach for health economic

analysis. I led on writing the manuscript in press with BMJ Health and Care Informatics with Dr Mihir Kelshiker as co-first author.

For Chapter 4, I independently reviewed the literature, defined the subject areas for analysis, and compiled the ethical principles to focus on. I wrote the first draft, receiving guidance on structure for a normative ethical analysis from Dr Daniel B. Kramer. Additional feedback was provided by listed co-authors and delegates at a writing workshop during the 2022 Harvard Law School Petrie Flom Center for Health Law Policy, Biotechnology, and Bioethics annual conference.

Copyright Declaration

Dedication

Dedicated to my partner, my mother, my family, my patients; and my younger self (we did it!).

Acknowledgements

The defining privilege of my PhD has been the opportunity to collaborate with a talented and diverse group of friends and colleagues. Most importantly, my huge thanks to my supervisor, Professor Nicholas Peters, who has afforded me a wealth of opportunity to develop as a clinician scientist. I am extremely proud of what we have achieved together and can only hope, in my post-PhD career, to emulate his capacity for support, vision, and incisive scientific inquiry.

My thanks also to the clinical research fellows who have made coming to work a joy. Dr Mihir Kelshiker's intellectual and collegial support and friendship have been invaluable. Dr Camille Petrie effortlessly untangled my statistical questions and helped me pursue the best analytical approaches – making sure that the ensuing pages do justice to a large patient recruitment effort. I am also grateful for support from Melanie Almonte, Dr Manisha Gandhi, Dr Alex Noar, Abdullah Alrumayh, Alexei Peters, Dr Alex Adamson, and Dr Kiran Patel; and all the medical students I had the pleasure of supervising during my PhD (Seri Park, Fran Scott, Jake Griffin, Rayyan Islam, and Akila Wickramasinghe).

Thanks also to senior colleagues for PhD support: Prof Darrel Francis and Dr Ramzi Khamis as my advisors, and Dr Daniel Keene, Dr Carla Plymen, Prof Jamil Mayet, Dr Phil Molyneaux, and Prof Jennifer Quint. Special thanks to Dr Daniel B. Kramer from Harvard Medical School, whose unique command of the intersection of clinical science and ethics has broadened my own horizons and enabled my passion for transatlantic collaboration. Here I am also grateful for the enduring support and inspiration from friends and mentors from my MPH programme. My thanks also to further collaborators in the US, at the Mayo Clinic and the team at Eko Health.

AstraZeneca funded the team at Imperial College Health Partners and enabled my access to the Discover-NOW research platform. My thanks here to Dr Tahereh Kamalati and Moulesh Shah.

To the often-unsung heroes in the echocardiography departments, who supported a heroic patient recruitment drive despite pandemic pressures: Nina Bual, Wing-See Cheung, Harpreet Sahemey, Regine Alquero, Alfa Ali, and Bianca Dumea; all and many others contributed to the pandemic resilience of my study, which became one of the highest and fastest recruiting regional research activities – during an otherwise challenging time for science.

This research could not have been conducted without generous funding support from Imperial Health Charity, the National Institute for Health and Care Research, and the British Heart Foundation. My thanks also to the Harvard Law School Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, and the AIMed Global Summit organisers for generous travel grants.

Lastly, I am indebted to over one thousand NHS patients who willingly gave up their time to be participants in my studies. Their contribution has already underpinned substantial expansion of the research presented in these pages.

Abstract

Heart failure (HF) is an international public health priority and a focus of the NHS Long Term Plan. There is a particular need in primary care for screening and early detection of heart failure with reduced ejection fraction (HFrEF) – the most common and serious HF subtype, and the only one with an abundant evidence base for effective therapies. Digital health technologies (DHTs) integrating artificial intelligence (AI) could improve diagnosis of HFrEF. Specifically, through a convergence of DHTs and AI, a single-lead electrocardiogram (ECG) can be recorded by a smart stethoscope and interrogated by AI (AI-ECG) to potentially serve as a point-of-care HFrEF test. However, there are concerning evidence gaps for such DHTs applying AI; across intersecting clinical, health economic, and ethical considerations. My thesis therefore investigates hypotheses that AI-ECG is 1.) Reliable, accurate, unbiased, and can be patient self-administered, 2.) Of justifiable health economic impact for primary care deployment, and 3.) Appropriate across ethical domains for deployment as a tool for patient self-administered screening.

The theoretical basis for this work is presented in the Introduction (Chapter 1). Chapter 2 describes the first large-scale, multi-centre independent external validation study of AI-ECG, prospectively recruiting 1,050 patients and highlighting impressive performance: area under the curve, sensitivity, and specificity up to 0·91 (95% confidence interval: 0·88–0·95), 91·9% (78·1–98·3), and 80·2% (75·5–84·3) respectively; and absence of bias by age, sex, and ethnicity. Performance was independent of operator, and usability of the tool extended to patients being able to self-examine. Chapter 3 presents a clinical and health economic outcomes analysis using a contemporary digital repository of 2.5 million NHS patient records. A propensity-matched cohort was derived using all patients diagnosed with HF from 2015-2020 (n = 34,208). Novel findings included the unacceptable reality that 70% of index HF diagnoses are made through hospitalisation; where index diagnosis

through primary care conferred a medium-term survival advantage and long-term cost saving (£2,500 per patient). This underpins a health economic model for the deployment of AI-ECG across primary care. Chapter 4 approaches a normative ethical analysis focusing on equity, agency, data rights, and responsibility for safe, effective, and trustworthy implementation of an unprecedented at-home patient self-administered AI-ECG screening programme. I propose approaches to mitigating any potential harms, towards preserving and promoting trust, patient engagement, and public health.

Collectively, this thesis marks novel work highlighting AI-ECG as tool with the potential to address major cardiovascular public health priorities. Scrutiny through complimentary clinical, health economic, and ethical considerations can directly serve patients and health systems by blueprinting best-practice for the evaluation and implementation of DHTs integrating AI – building the conviction needed to realise the full potential of such technologies.

Publications related to this thesis

Bachtiger P, Petri CF, Scott FE, Ri Park S, Kelshiker MA, Sahemey HK, Dumea B, Alquero R, Padam PS, Hatrick IR, Ali A, Ribeiro M, Cheung WS, Bual N, Rana B, Shun-Shin M, Kramer DB, Fragoyannis A, Keene D, Plymen CM, Peters NS. *Artificial Intelligence for Point-of-Care Heart Failure Screening During ECG-Enabled Stethoscope Examination: Independent Real-World Prospective Multicentre External Validation Study.* **The Lancet Digital Health [Impact Factor 36.6]**. 2022 Jan 6. PMID: 34998740

Bachtiger P, Plymen CM, Pabari PA, Howard JP, Whinnett ZI, Opoku F, Janering S, Faisal AA, Francis DP, Peters NS. *Artificial intelligence, data sensors and interconnectivity: future opportunities for heart failure.* **Cardiac Failure Review** [Impact Factor 2.5]. 2020 Mar;6. PMID: 32514380

Bachtiger P, Kelshiker MA, Petrie CP, Gandhi M, Shah M, Kamalati T, Stephens J, Hooper G, Khan S, Barton C, Plymen C, Kramer DB, Peters NS. *Comparative Survival and Health Economic Outcomes in Heart Failure Diagnosed at Hospital Admission versus Community Settings: A Propensity-Matched Analysis.* BMJ Health and Care Informatics [Impact Factor 2.28], *in press,* 2023.

Bachtiger P, Kelshiker MA, Moe M, Kramer DB, Peters NS. *Point-of-Care Diagnosis of Cardiovascular Disease in the Home Using Artificial Intelligence.* **Book chapter** in edited volume titled "Diagnosing in the Home: The Ethical, Legal, and Regulatory Challenges and Opportunities of Digital Diagnostics and Therapeutics Outside of Traditional Clinical Settings." Compiled by The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics Annual Conference, **Harvard Law School.** Cambridge University Press, *in press*.

International abstracts

Bachtiger P, Petri CF, … Peters NS. *Artificial Intelligence For Point-of-care Heart Failure Screening During ECG-enabled Stethoscope Examination: Independent Real-world Prospective Multicenter External Validation Study.* AHA21. Circulation 2021. **Late-breaking science, oral presentation.**

Bachtiger P, Scott FE, ... Peters NS. *Multicentre validation of point-of-care screening tool for heart failure: single-lead ECG recorded by smart stethoscope predicts low ejection fraction using artificial intelligence.* European Heart Journal. 2021. **Poster presentation.**

Bachtiger P, Park S, … Peters NS. *Point-of-Care Detection of Heart Failure with an AI-ECG Stethoscope: Independent Real-World Prospective Multicentre External Validation Study.* AI Med Global Summit 2022, San Francisco. **Oral presentation.**

Invited presentations

Bachtiger P, Kelshiker MA, … Peters NS. *Point-of-Care Diagnosis of Cardiovascular Disease in the Home Using Artificial Intelligence.* The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics Annual Conference, **Harvard Law School,** on "Diagnosing in the Home: The Ethical, Legal, and Regulatory Challenges and Opportunities of Digital Diagnostics and Therapeutics Outside of Traditional Clinical Settings".

Health policy contributions

Commissioned expert technical content author. *The Evaluation Playbook: Artificial Intelligence Award for Health & Care Technologies*. National Institute for Health and Care Research, NHSX, NHS Accelerated Access Collaborative. FutureGov on behalf of Department of Health and Social Care. 2020

Expert contributor. National Institute for Health and Care Excellence, *NICE Evidence Standards Framework for Digital Health Technology update*, August 2022

Invited contributions

Invited speaker and panelist, *Digital Health REWIRED*, 2022

National Media Coverage

Bawden, T. *New AI stethoscope can diagnose heart failure in 15 seconds*. **i News.** 5 Feb 2022. Available: https://inews.co.uk/news/science/new-ai-stethoscope-can-diagnose-heart-failure-15-seconds-1443766

Prizes/Awards

Winner, Faculty of Clinical Informatics Rising Star Award, London, 2021

Winner, Best Research Presentation, AIMed Global Summit, San Francisco, 2022

Best BSc Project Supervisor, Cardiovascular Sciences BSc, 2022

Finalist, Rising Star Award, AIMed Global Summit, San Francisco, 2022

Finalist (team member), Health Service Journal Awards, Primary and Community Care Innovation of the Year, 2022

Finalist (team member), Health Service Journal Awards, Digitising Patient Care, 2022

Additional publications during my PhD

Bachtiger P, Adamson A, Quint JK, Peters NS. *Belief of having had unconfirmed Covid-19 infection reduces willingness to participate in app-based contact tracing.* **Nature Partner Journals Digital Medicine**. 2020 Nov 6;3(1):1-7. PMID: 33299071

Bachtiger P, Adamson A, Chow JJ, Sisodia R, Quint JK, Peters NS. *The impact of the COVID-19 pandemic on the uptake of influenza vaccine: UK-wide observational study.* **JMIR** Public Health and Surveillance. 2021 Apr 14;7(4):e26734. PMID: 33651708

Bachtiger P, Adamson A, Maclean WA, Kelshiker MA, Quint JK, Peters NS. *Determinants of Shielding Behaviour During the COVID-19 Pandemic and Associations with Wellbeing in > 7,000 NHS Patients: 17-week Longitudinal Observational Study.* **JMIR** Public Health and Surveillance. 2021 Jul 15. PMID: 34298499

Bachtiger P, Peters NS, Walsh SL. *Machine learning for COVID-19—asking the right questions.* **The Lancet Digital Health.** 2020 Aug 1;2(8):e391-2. PMID: 32835197

Davies HJ, **Bachtiger P**, Williams I, Molyneaux PL, Peters N, Mandic D. *Wearable In-Ear PPG: Detailed Respiratory Variations Enable Classification of COPD.* IEEE Transactions on Biomedical Engineering. 2022 Jan 25. PMID: 35077352

Research grants awarded

NIHR Artificial Intelligence in Health & Social Care Award, *DUO-EF: Point-of-Care Screening for Heart Failure Using an AI Stethoscope*, 2020, **£446,315** (co-applicant and lead author)

Innovate at Imperial Award, Imperial Health Charity, 2021, **£75,000** (co-applicant and lead author)

NIHR Invention for Innovation FAST Award, 2022, **£45,000** (co-applicant and co-author)

NIHR Invention for Innovation Challenge Award, *TRICORDER: Point-of-Care Cardiovascular Disease Screening With Artificial Intelligence Enabled Stethoscope*, 2022, **£1,181,666** (co-applicant and co-lead author)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AF	Atrial fibrillation
AHSN	Academic Health Sciences Network
AI-ECG	Artificial intelligence electrocardiogram (for detection of reduced LVEF)
BNP	Brain natriuretic peptide
CI	Confidence interval
CKD	Chronic kidney disease
CNN	Convolutional neural network
Covid-19	SARS-CoV-2 virus
CT	Computed tomography
DHT	Digital health technology
ECG	Electrocardiogram
FDA	Food and drug administration
GP	General practitioner
HF	Heart failure
HFmrEF	Heart failure with moderately reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
IQR	Interquartile range
LVEF	Left ventricular ejection fraction
MHRA	Medicines and Healthcare products Regulatory Agency
ML	Machine learning
NHS	National Health Service
NIHR	National Institute for Health and Care Research
NT-proBNP	N-terminal pro-brain natriuretic peptide
ONS	Office for National Statistics
SaMD	Software as a medical device
SD	Standard deviation

1 Introduction

This chapter outlines the public health, scientific, ethical, legal, and regulatory background relevant to addressing screening for heart failure with point-of-care artificial intelligence – specifically through use of a single-lead electrocardiogram acquired with a smart stethoscope. Throughout, I emphasise the evidence gaps that, if addressed, could contribute to underpinning system-wide deployment of such technology. I conclude with an outline of my PhD's aims and hypotheses for empirical chapters, designed to contribute towards addressing the outlined evidence gaps.

1.1 Overview of Heart Failure

1.1.1 Definition

Heart failure (HF) is not a single pathological diagnosis, but a clinical syndrome typified by cardinal symptoms – classically breathlessness, ankle swelling, and fatigue – that may be accompanied by signs on examination (e.g. pulmonary crackles, elevated jugular venous pressure, and peripheral oedema). It is due to a functional and/or structural abnormality of the cardiac tissues that results in elevated intracardiac pressures and/or inadequate cardiac output.[1] Identification of the causal mechanism of cardiac dysfunction is an essential component in the diagnosis of HF, since the specific pathology informs subsequent treatment approaches. Myocardial dysfunction (disorder of the heart muscle itself) is the most common cause – this can be either diastolic (during cardiac filling), systolic (during cardiac emptying), or both. However, commonly pathology of the endocardium, heart rhythm, conduction, and valves, can also precipitate or contribute to HF.[2]

1.1.2 Heart Failure Subtypes

HF has traditionally been segmented into distinct phenotypes based on the measurement of left ventricular ejection fraction (LVEF), most-commonly calculated through ultrasound-based echocardiography (echo). This rationale is underpinned by foundational trials for HF treatment, which demonstrated significantly improved outcomes in patients with LVEF≤40%.[3]

Summarised in figure 1.1, the latest European Society of Cardiology (ESC) guidelines (2021) therefore classifies HF into three categories by LVEF: heart failure with preserved ejection fraction (HFpEF, LVEF ≥50%), heart failure with moderately reduced ejection fraction (HFmrEF, LVEF 41-49%), and heart failure with reduced ejection fraction (HFrEF, LVEF ≤40%). Across these subtypes,

there is substantial variation across patient characteristics, outcomes, and available guideline-directed therapies. HFrEF is the most common subtype.[4] HFrEF also has the worst symptom burden and clinical outcomes, but notably is the only subtype with an abundant evidence base for prognostically-beneficial drug therapies, implanted cardiac devices, and rehabilitation programmes.[5-7] For patients with HFrEF, there is strong evidence to show that pharmacological therapy with beta-blockers, angiotensin converting enzyme inhibitors (ACEi), angiotensin receptor blockers (ARB) or angiotensin receptor neprilysin inhibitors (ARNi), mineralocorticoid receptor antagonists (MRAs), and more-recently, sodium-glucose co-transporter 2 inhibitors (SGLT2i) can greatly reduce mortality, hospitalisation, and improve quality of life.[8-12] Early initiation of these therapies through prompt diagnosis improves prognosis.[13] To date, no trials have convincingly shown that these medications reduce mortality in patients with HFpEF, though some may reduce hospitalisations. Drug studies specifically targeted at patients with HFmrEF are lacking. Instead, most guideline recommendations are drawn from sub-group analyses of larger trials which suggest some possible reduction in mortality.[1]

Unlike the more straightforward diagnosis of HFrEF, where LVEF≤40% observed on echocardiography is sufficient, HFmrEF generally requires four features to be present: (i) symptoms with or without signs of HF, (ii) LVEF of 41-49%. (iii) elevated natriuretic peptides, and (iiii) relevant structural heart disease (left ventricle hypertrophy or left atrial enlargement).[14] HFpEF poses similar diagnostic challenges, requiring elevated natriuretic peptides and specific echo criteria in the presence of normal (preserved) LVEF>50%.[15]

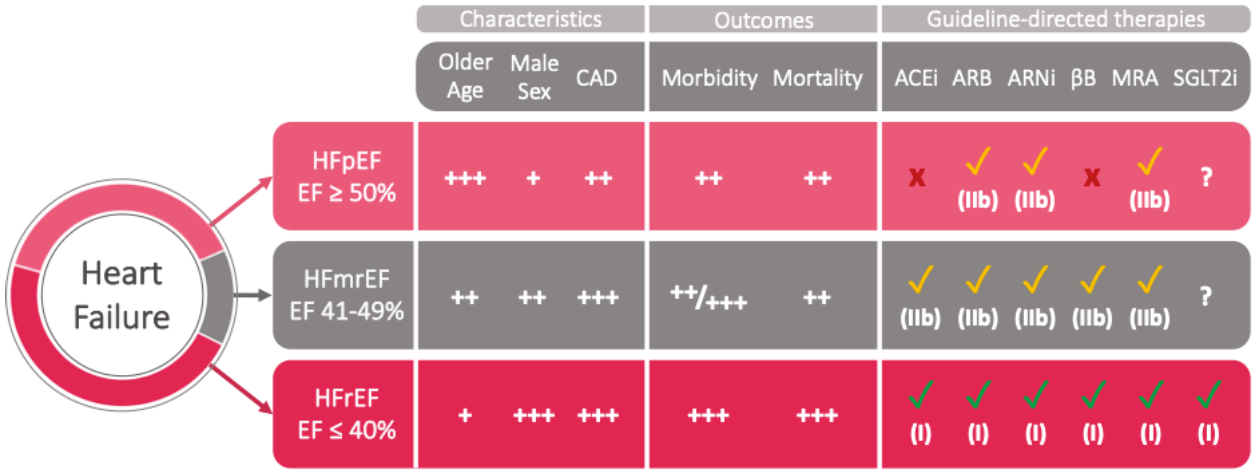


Figure 1.1 HF subtypes, characteristics, outcomes, and availability of guideline directed therapies.

Green tick: class I evidence (highest level), recommended/indicated treatment; Orange tick: class IIb evidence, may be considered (usefulness/efficacy less well established)

1.1.3 Epidemiology and Natural History of Heart Failure

HF remains the only cardiovascular disease where incidence and prevalence continue to rise, principally due to an ageing population.[16] Between 1-2% of the global population are estimated to be living with known HF, with around 1-9 in 1000 people newly diagnosed each year. Up to 6% of the general population is estimated to have asymptomatic HF.[17] In the UK, HF is the most common reason for hospital admission in patients over 65, accounting for one million inpatient bed days per annum.[18] For patients, the symptoms of HF often result in a markedly reduced quality of life.[19] Additionally, the mortality associated with HF is estimated to be between 53-67% at 5-years after diagnosis, higher than the mortality for most cancers.[20,21]

1.1.4 Heart Failure Health Economic Burden

The increasing burden of HF is estimated to cost the UK National Health Service (NHS) approximately £2 billion per year, consuming over 2% of its annual budget. Projections indicate that these figures are likely to rise by 50% over the next 25 years.[22] A systematic review conducted from 2004 to 2016 identified cost-of-illness studies and estimated a lifetime cost for HF at $126,819 per patient. Studies included in this review showed large variations in methodological approaches and therefore ultimately wide variations in cost estimates, highlighting a need for more robust research measuring the health economic consequences of HF. [23] This is particularly necessary for framing evaluations of interventions intended to mitigate the far-reaching challenges that HF poses to health systems. Though it is the main driver of cost, interpreting the health economic burden of HF principally through units of hospitalisation has limitations. Studies have shown that post-HF diagnosis, there is an average of one hospital admission per year, of which two-thirds are attributable to non-cardiovascular comorbidities.[24-26] The complex and common picture of multimorbidity within HF therefore forms part of the need for more holistic approaches to measuring cost, where consideration of patient experience and the life-course cost implications of a HF diagnosis would better serve policy makers.

1.1.5 Heart Failure as Priority for the NHS Long Term Plan

The NHS Long Term Plan, published in 2019, was drawn up by frontline staff, patient groups, and national experts to define key challenges and set an ambitious but realistic vision for the next ten years of health and social care in the UK.[27] The Plan highlights HF as an explicit priority area, prompted by a UK-wide study of place of index diagnosis of HF between 2010-2013 by Bottle et al. These authors found that "80% of new HF diagnoses were made in a hospital setting, despite 41% of

these patients having been seen in primary care exhibiting symptoms of heart failure within the previous five years."[28] This highlights a substantial missed opportunity for early detection, and a clear failing of the established approach to a diagnostic pathway for HF.

1.1.6 Heart Failure Diagnostic Pathway in NHS primary care

The National Institute for Health and Care Excellence (NICE) clinical guidelines for the management of HF recommend that all patients with symptoms are screened using a brain natriuretic peptide (BNP) or N-terminal pro-brain natriuretic peptide (NT-proBNP) blood test[22], with the latter increasingly preferred due to more frequent use during monitoring studies.[29] Patients may undergo additional investigations including chest X-rays and 12-lead electrocardiogram (ECG). Based on the level of natriuretic peptide, patients are triaged to referral for specialist assessment and echocardiogram, either within 6 weeks, or 2 weeks if highly elevated (figure 1.2).

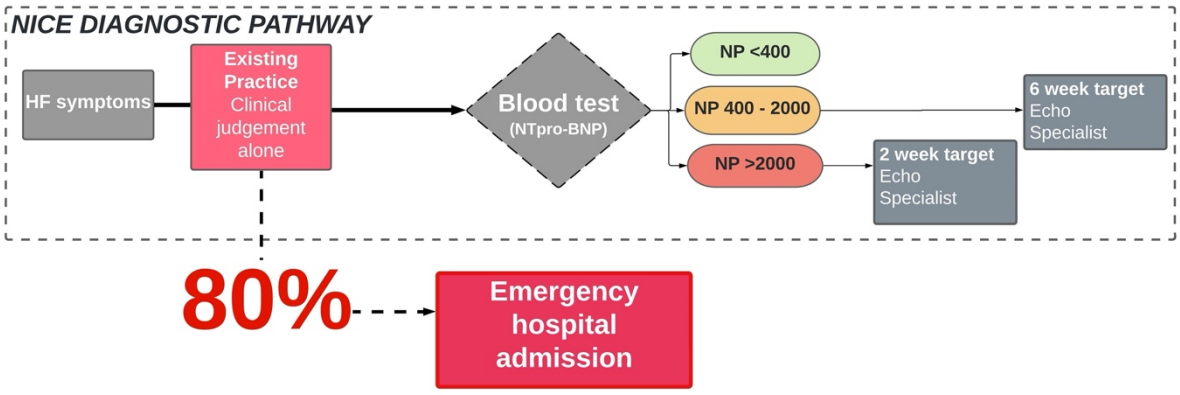


Figure 1.2 Current NICE pathway for HF diagnosis.

1.1.7 Limitations of natriuretic peptides

Use of natriuretic peptide as a screening tool for suspected HF has limitations, including requiring an invasive blood draw. Results are not instantly available, and each test costs around £15-30, which may deter use unless clinical suspicion is strong.[30] Depending on the chosen threshold/cut-off, natriuretic peptide testing for acute HF has generally shown sensitivity above 90%, but with consistently inferior specificity closer to 78%.[30] The lower specificity can in part be explained by natriuretic peptide levels also being raised in atrial fibrillation (AF), chronic kidney disease (CKD), and increasing age, leading to challenges in interpreting the result.[1] Conversely, concentrations may be disproportionately low in obese patients.[31]

Synthesis of criteria considered by UK National Screening Committee
The screening programme should respond to a recognised need
The objectives of screening should be defined at the outset.
There should be a defined target population
There should be scientific evidence of screening programme effectiveness.
The programme should integrate education, testing, clinical services, and programme management.
There should be quality assurance, with mechanisms to minimize potential risks of screening.
The programme should ensure informed choice, confidentiality, and respect for autonomy.
The programme should promote equity and access to screening for the entire target population.
Programme evaluation should be planned from the outset.
The overall benefits of screening should outweigh the harm.

Table 1.1 Criteria considered by the National Screening Committee. From Andermann et al. (2008).

Based on these observations, natriuretic peptides have fallen short in assessment of their potential use as part of national general screening programmes for HF.[32] Despite the overwhelming

public health burden, there has yet to be a screening test developed that can meet the stringent criteria of the UK National Screening Committee[33] in a way that would justify exploration of a programmatic approach for increasing the detection of HF. The Screening Committee's criteria draw on a synthesis of screening criteria proposed over the last 50 years (table 1.1), building on the "classic" screening criteria first proposed by Wilson and Jungner.[34]

1.1.8 Challenges in diagnosing HF in primary care

Diagnosis and management of HF in primary care has been criticised based on evidence of underdiagnosis[35,36] and no improvement in survival over nearly 15 years.[37] Healthcare professionals may fail to follow guidelines and lack confidence in initiating downstream investigations and treatments.[38] This is reported to have remained unchanged despite new guidelines and incentivisation. The reality remains that identification, diagnosis, and management of HF in primary care poses substantial challenges.[39] The condition manifests in myriad ways, with patients varying in their health-seeking behaviours and thus receiving variable primary care intervention. From the general practitioner (GP) perspective, the diagnosis is made more difficult by the lack of specificity of HF symptoms; there is confusion with respiratory conditions (up to a third have chronic obstructive pulmonary disease), limited time per appointment, and limited access to investigations and low confidence in interpreting results of these. While clinical guidelines exist, perceived information overload, as well as a belief that they do not apply to all patients, means that GPs do not always use them and not all will be aware of them.[38] The ESC updated their HF guidelines in 2021 with a new recommendation to refer patients directly for echocardiography if clinical suspicion for HF is sufficiently high i.e., to not delay on account of waiting for blood test results and other investigations. Though this recommendation is intended to increase the number of patients receiving the gold

standard diagnostic test for HF in good time (and maximise benefit from prompt commencement of therapies), this clashes against the reality of a nationwide shortage of echocardiographers and stands to compound subsequently long waiting lists.[40] Notably, echocardiography is also an imperfect test, requiring substantial technical skill and relying on often subjective, visual estimation of parameters such as LVEF. Such measurements are therefore well-established to exhibit inter- and intra-operator variability.[41]

Considering all these challenges, it is perhaps unsurprising but nonetheless striking that 80% of HF is diagnosed through hospital admission despite frequently preceding symptomatic primary care encounters. Only 24% of patients who do present to their GP with HF symptoms follow the NICE recommended pathway – with only 4% completing their diagnostic odyssey to time and target.[28] This cements HF's status as one of the major public health priorities of this decade, with policymakers, health systems, and patient groups all advocating for the need to apply innovative solutions where other approaches have thus far failed. Among the potential solutions, "digital transformation" is emphasised as one of the core pillars of the NHS Long Term Plan, set to underpin innovative approaches to care and anchored in a broad set of deliverables, including "providing digital services and tools to give people more control over their own health and the care they receive from the NHS." Specifically for HF, these solutions collectively fall within the realm of digital health technologies (DHTs).

1.2 Digital Health Technologies for Heart Failure

1.2.1 Policy Priorities for DHTs and Evidence Gaps

Broadly defined, digital health technologies (DHTs) encompass apps, programs, and software, which may be standalone or combined with other products such as medical devices (hardware).[42] DHTs commonly generate high volumes of data, which unlocks the opportunity to apply artificial intelligence (AI). The NHS Long Term Plan singles out DHTs with AI as a key driver for digital transformation. Specifically, "use of decision support and AI to help clinicians in applying best practice, eliminate unwarranted variation across the whole pathway of care, and support patients in managing their health and conditions." However, for the most part, widespread implementation of such DHTs is currently limited by unaddressed evidence gaps. Particularly for DHTs with an AI component, there is an expressed need for large-scale external clinical validation studies[43,44]; demonstration of a compelling health economic upside[45,46]; and scrutiny of the wider, unique ethical implications posed by use of these types of technologies.[47,48] Importantly, proven efficacy across these domains alone will not necessarily translate to high uptake and usage by health systems, highlighting the need to also understand the usability and acceptability of such DHTs across healthcare professionals and patients – and specifically how to operationalise these tools within clinical pathways.[49] NICE has generated an Evidence Standards Framework intended to provide a standardised approach to guiding developers and commissioners on the levels of evidence needed for the clinical and health economic evaluation of DHTs by health and care systems.[50] Validation of any such framework is reliant on practical, real-world application. Given the accumulating number of DHTs using AI for application to HF and cardiovascular disease more broadly, this disease area represents opportunities for focused evidence generation that could blueprint best practices for use of DHTs in other settings and disease areas.

1.2.2 Overview of Artificial Intelligence in Medicine

DHTs have enabled the opportunity to collect low-cost data, often passively and at massive scale across populations. The subsequent datasets qualify as 'big data', characterised as high volume, high velocity and/or high variety information assets that require new forms of processing to enable enhanced discovery, insight, decision-making, and process optimisation.[51] Following several decades of dormancy, the combination of massive amounts of data alongside advances in computing power triggered a resurgence and evolution of AI methods that can now analyse big data to confer previously inaccessible insights.

AI itself is a broad term that encompasses machine-based data processing to achieve objectives that typically require human-level cognitive function, such as recognising images. Broadly speaking, building an AI model takes an approach that first presents a "training" dataset from which specific patterns can be learned. Models require subsequent validation and testing of whether appropriate learning has occurred by processing further independent input data.

This last decade has seen significant AI breakthroughs using machine learning (ML) and, more specifically, deep learning. Deep learning is a subfield of ML that uses neural networks – a computing architecture inspired by biological brains – with many layers (hence 'deep') to learn a function between a set of inputs and outputs.[52] The training of neural networks can be achieved by presenting a set of input data with corresponding output labels – so-called "supervised" machine learning. The model then learns certain rules by applying and adjusting the network weights to minimise an error function until the model outputs are as close as possible to the actual data values. The strength of deep neural networks lies in using their ability to identify novel relationships in the data independent of features selected by a human. Ultimately, ML aims to learn from data to correctly answer a question, which is

different from conventional computer programming; that is, hardcoding the answer into the system. Complex datasets can now be mined with the potential to identify patterns and novel representations of data beyond direct human capability.

Through AI and ML, paradigms across all sectors of society are being disrupted. In medicine, some of the most high-profile early AI research has been across ophthalmology[53], dermatology[54], radiology[55], intensive care[56], and mental health.[57] Many of these studies focus on algorithm-enhanced risk prediction, diagnosis, and treatment selection, but there is also substantial enthusiasm for AI potentially liberating clinical staff from tedious administrative tasks to spend more time with patients. Thought leaders in AI as applied to medicine argue that it will "make healthcare human again" through increasing automation of any process that prevents clinical staff from working at the top of their license.[58]

1.2.3 Evidence gaps and concerns for AI in medicine

Though enthusiasm for AI in medicine abounds, among the major reservations around widescale deployment of AI technology in health and social care is the current lack of a standardised and proven approach to mitigating unique pitfalls posed by this technology. For example, algorithmic bias remains a principal concern for the safe and ethical deployment of AI. Indeed, there has been a recent shift in emphasis from reporting impressive performance results to active investigation of algorithmic errors and failure modes.[59-61] The trade-off for using powerful ML approaches to solving complex challenges is the so-called "black box" problem. Particularly for deep learning, these AI models establish complex and opaque mathematical relationships between the input data and the output predictions, with little to no human control over how predictions are generated. Although this enables powerful learning patterns from the data, there is also a risk of learning spurious correlations:

relationships that appear useful in training but prove unreliable when applied to real-world datasets. For example, a deep learning system might learn to detect surgical skin markings to diagnose skin cancer, rather than looking for features related to the lesion itself.[61]

Perhaps more seriously, there is societal concerns that use of AI risks creating both new and deepening of existing inequities.[62] For medical practice, this is on a background of established, persistent bias against racial and ethnic minority groups, leading to unequal access to care and unequal health. Much of this discriminatory behaviour is already encoded in the medical datasets that form the substrate for building AI algorithms.[63] An AI system could easily learn to perpetuate such racial biases, which could lead to the underdiagnosis of disease or less frequent treatment recommendations in racial and ethnic minority groups.[64] This is a particular worry for applications of AI within cardiovascular disease, given the pronounced sociodemographic gradient that already persists: patients with minority ethnic status, lower income, and lower educational attainment all suffer disproportionately – including with HF.[65,66]

Beyond the immediate risks of algorithmic bias, there are wider concerns around the implementation of AI tools in how they are operationalised. As a DHT, AI algorithms are usually built as part of a wider software programme, which may itself be built into a platform have has an element of hardware e.g. integrating sensors that collectively form the "product" as a whole. Such technology is inherently highly connected; to the internet, to cloud-based infrastructure, via Bluetooth, and therefore potentially across a wide variety of end-user devices. This poses a potential vulnerability and a new frontier of challenges for ensuring delivery of safe healthcare. The NHS continues to be heavily targeted by cyber attacks[67-69], but robust standards and recommendations for mitigating DHTs being associated with security and data breaches, for example, through cyber threat modelling[70,71], are yet to be defined. Further still, AI can itself be harnessed to serve as nefarious malware. For example, an AI

model has been shown to be able to either remove cancerous nodules from a CT scan of a patient's lungs or insert cancerous nodules into images of healthy lungs – fooling radiologists to make incorrect diagnoses in nearly 100% of cases.[72]

In the presence of readily available, ostensibly accurate and safe AI technology for healthcare applications, concerns have also been raised about the preparedness of the NHS' workforce to use these. To address this, the Department of Health and Social Care Commissioned the Topol Review, broadly framed to address "preparing the healthcare workforce to deliver the digital future" of the NHS.[73] Among its many recommendations, the Review advocates for medical staff to receive education on the foundations of AI technologies as part of their training. Lastly and perhaps most importantly, given some of the challenges described, concern for AI technology also extends to whether patients and the wider public are willing to accept and trust in its use as part of decision-making for their health. Asymmetric uptake of AI across patient populations risks adding a new dimension of inequity to the already existing "digital divide"[74,75], where certain groups with protected characteristics risk being left behind – originally by not being able to, but here by not being *willing* to participate in clinical care where optimal outcomes are heavily reliant on AI technology.

Overall, consideration of these pitfalls for DHTs using AI highlights the need for robust, holistic evaluation of such technologies before pursuing large-scale deployment. These deployments pose a specific form of "ethical debt"; the reality that some ethical problems with AI DHTs will only manifest and be detected after they are deployed.[76] Scrutiny therefore needs to be sustained after deployment. Ethical debt relates to the problem of "drift" with AI algorithms. ML algorithms use data describing historical episodes to make ahead-of-time predictions of clinical outcomes. However, clinical settings are dynamic environments and the underlying data distributions characterising

episodes can change with time (data drift), and so can the relationship between episode characteristics and associated clinical outcomes (concept drift).[77]

By definition, the AI drift problem is one that occurs over time, typically in the range of months or years. Drift is detected after the AI solution has made faulty determinations – and individuals are wrongly flagged or not flagged. Likewise, problems with fairness are associated with trend analysis – which requires a substantial amount of deployment data. The impact on unsuspecting individuals can range from relatively innocuous, such as waiting longer for a routine appointment, to potentially harmful, such as receiving the wrong diagnosis or being denied essential treatment. Practically, this means deployed AI technologies need robust frameworks for sustained monitoring to ensure their safety.[78]

1.2.4 Regulatory requirements for AI in medicine

From a regulatory perspective in the UK, the Medicines and Healthcare products Regulatory Agency (MHRA) is responsible for ensuring that medicines and medical devices work and are acceptably safe within the scope of their labelled indications. Medical "products" that draw on the output of an AI algorithm to deliver their intended purpose are commonly referred to as "software as a medical device" (SaMD). The MHRA grants SaMDs authorisation for UK market entry in a similar fashion to the European Medicines Agency for the European Union, and the Food and Drug Administration (FDA) in the United States. Post-Brexit, the UK has moved away from *Conformitè Europëenne* marking, instead requiring UK Conformity Assessed (UKCA) for new products, though the underlying risk-based classification system remains similar to international counterparts.[79] In practice, most technology with an AI component would be considered relatively low-risk if the focus is diagnostic, with treatment applications, especially invasive/implantable technology, or those that

are explicitly life-sustaining, considered highest risk. One implication of this risk tiering is that unlike a new implanted cardiac device such as a novel pacemaker or coronary stent, diagnostic technology generally would not be expected to demonstrate safety and effectiveness prior to legal marketing through a large trial with hard clinical endpoints.

The speed at which SaMD technologies with an AI component are being developed has left regulators struggling to keep pace.[80] Accordingly, responsible regulatory bodies in the UK, US, Europe, and beyond are all embarking on defining new frameworks for evaluating the safety of AI-driven medical products. For example, the current benchmark for approval requires developers to demonstrate good model performance on a varied dataset, and in a real-world setting, but with no explicit definition of what constitutes the parameters of optimum validation data.[81] Under most current regulation, once an AI algorithm is approved, the model will remain fixed – defeating the distinguishing advantage of many algorithms in their ability to 'learn' (improve) throughout their lifecycle. This current inflexibility not only restricts clinical utility but can also infringe on patient safety. For instance, an algorithm trained in 2019 to recognize pneumonia on a chest radiograph will not be able to differentiate it from SARS-CoV-2 virus (Covid-19) infection (an example related to AI drift).

1.2.5 AI for Cardiovascular Disease

For the most part, AI technologies have focused on tasks that humans can already do, with the principal aim of performing at the level of experts or, where possible, surpassing them. There are still relatively few AI technologies that go to the next level and yield previously unimaginable, novel, and clinically actionable insight beyond human capability using forms of routinely available clinical

data. However, cardiovascular disease is starting to accumulate examples of AI overcoming confines of conventional scientific inquiry.[82-84]

Multiple aetiological factors, complex disease mechanisms, and heterogeneity in clinical presentations make cardiovascular disease detection and prognostication a daunting task. However, this disease area has benefitted from being at the forefront of generating and utilising complex high-dimensional data that results from integration of multi-modal data, including electronic medical records, mobile health devices, waveform signals, and imaging data – collectively offering extensive opportunities for data-driven discovery and research. Although traditional statistical approaches for risk stratification have been developed and are well established for cardiovascular disease, many of these models show limitations when practically applied for individualised risk prediction. Newer approaches using ML have emerged as a potential solution to addressing this.[85] In terms of cardiovascular data inputs to achieve this purpose, the electrocardiogram (ECG) continues to distinguish itself as a rich "digital biomarker" – defined here as a "physiological or behavioural measure collected through connected digital tools."[86]

1.3 AI applied to the Electrocardiogram

1.3.1 Electrocardiogram as a Universal Tool

The ECG is a ubiquitous tool in clinical practice, having been used by clinicians for decades. The established standard in clinical practice is the 12-lead ECG, requiring attachment of ten electrodes to a patient's chest and limbs to create a total of 12 vectors, which can subsequently be displayed (usually on a paper printout) as 12 ECG traces that collectively represent a comprehensive picture of the heart's electrical activity across key anatomical structures. This informs on the structural and

physiological condition of the heart, as well as lending actionable diagnostic clues for systemic conditions (e.g. toxic drug effects, electrolyte imbalances). The ECG offers the advantage of being a relatively simple, low-cost, and quick test that is available in most health systems. More recently, DHTs that offer the ability to record single- and increasingly multiple-lead ECGs have entered the consumer market in the form of watches, on-body patches, or hand-held devices.[87]

The protocol for recording ECGs is well-established, but human interpretation is highly subjective and dependent on experience and expertise. To mitigate this, interpretation by computer algorithms has existed in clinical practice for several years. Such interpretations draw on prespecified rules and algorithms that require manual pattern or feature recognition. Such automation is not a new concept in cardiology; first attempts at automated ECG interpretation date back to the 1970s.[88] However, these algorithms are incapable of capturing the rich complexity and nuances within ECGs. More recently, the ECG has been one of the stand-out inputs for ML research that has rapidly taken novel technology from bench to bedside – and further towards now being increasingly common within consumer wearable health technology.[89]

Application of AI to the ECG most commonly uses convolutional neural networks (CNNs), a form of deep learning already widely applied to image processing, speech recognition, and computer vision. CNNs can readily analyse routine ECG waveforms with far greater accuracy than previous, more basic, traditional rule-based computer interpretation approaches. Fully automated CNNs often surpass human-level performance at ECG interpretation.[84] The increasing number of useful applications of AI to the ECG stem from this being an ideal substrate for AI models with deep-learning architecture. As well as being widely used, and therefore offering large datasets, ECG waveforms yield raw data that can be reproduced, and stored and transferred in digitised format. This has resulted in an increasing number of institutions and health systems worldwide amassing vast

databanks of ECGs. In combination with wider clinical datasets and coupled with computational power, this unlocks the utility of AI-enhanced interpretation of ECG (henceforth referred to as AI–ECG) as a tool for superhuman detection of ECG signatures and patterns that elucidate specific pathologies. Among these, to give two prominent examples, AI-ECG algorithms have been developed for AF and hypertrophic cardiomyopathy (HCM).[90,91]

More recently, application of AI-ECG has been extended to detect a reduction in LVEF.[92] Considering the gold-standard incumbents of natriuretic peptide testing and echocardiography, the traditional 12-lead ECG should also form an early part of investigation for HF. Here, AI-ECG poses a potentially quick, cheap, and scalable screening test for reduced LVEF. The foundational research for developing this technology was conducted by teams at the Mayo Clinic (Rochester, Minnesota, US). This research institution and healthcare provider has been digitising ECGs for several decades, having since amassed one of the world's largest data vaults of ECG waveforms. Such datasets are further enriched by extensive additional variables (labels) across clinical and other categories. AI-ECG for detection of reduced LVEF highlights an application of AI that yields "superhuman" insights; some of the best-known applications of AI to the ECG, such as detection of heart rhythm abnormality, can be matched by the well-trained human eye.[93] Conversely, accurately detecting reduced LVEF from just a 12-lead ECG is beyond human capability.

1.3.2 Development of AI-ECG for Reduced LVEF

Previous studies have applied deep learning methods to highlight the 12-lead ECG as an accurate digital biomarker for changes in LVEF. Most of these studies relate to the Mayo Clinic's AI-ECG machine learning algorithm for detecting low LVEF from 12-lead ECGs. Since its invention and publication in 2019, this algorithm has been through further clinical validation and initial real-

world deployment, as summarised in figure 1.3. This AI-ECG algorithm uses a CNN, trained on 35,970 independent pairings of 12-lead ECGs and corresponding echocardiograms from the proprietary Mayo Clinic digital data vault.[92] Following initial development and validation, performance of the algorithm has been further validated with 12-lead ECGs (with paired echocardiography results for LVEF ground truth) from further US and Russian cohorts.[94,95]

Generalisability remains an important consideration prior to clinical application of AI-ECG; is the model's diagnostic performance consistent across various populations? There is the potential for AI-ECG to reflect, perpetuate, and even exacerbate racial and ethnic disparities already well established in current cardiovascular clinical care.[96] Overall, the impact of race and ethnicity on ECG analysis via machine learning remains largely unknown. AI-ECG therefore needs to be validated in diverse populations to ensure their diagnostic performance is maintained across racial and ethnic subgroups. Most recently, AI-ECG was deployed and evaluated through a pragmatic cluster randomised trial within Mayo Clinic primary care practices. This highlighted an increased rate of LVEF ≤50% detection in the intervention group vs control group (2.1% vs. 1.6%; odds ratio [OR] 1.31 [1.01–1.61]).[97] Though AI-ECG for detection of reduced LVEF from 12-lead ECG has had substantially more external clinical validation than comparable technologies, these studies have been in largely ethnically White populations. Indeed, the original training dataset of over 35,970 patients was drawn from a >90% White population living in the Midwest of the US.[98]

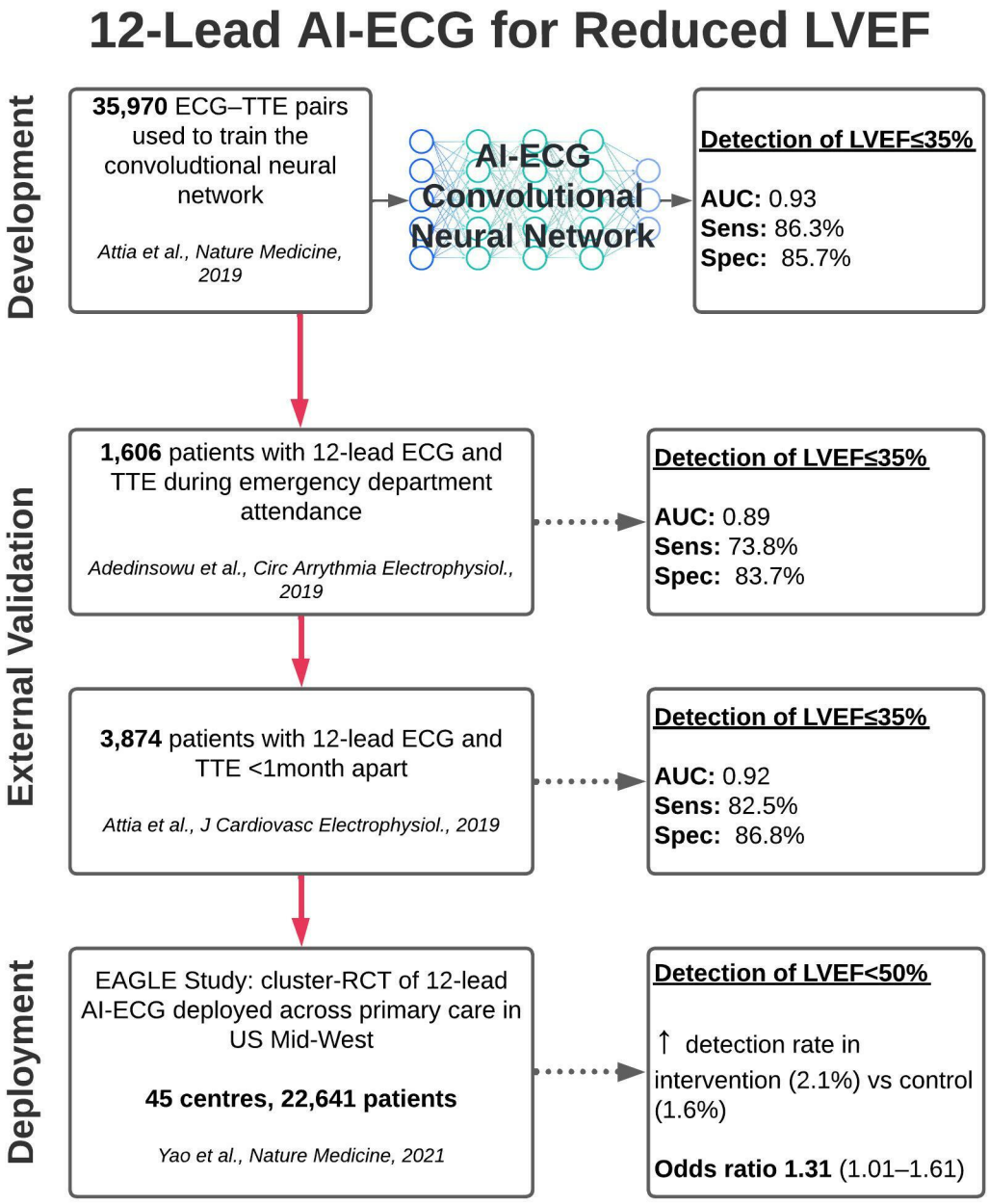


Figure 1.3 Development of AI-ECG for detection of reduced LVEF from 12-lead ECGs.

AUC: Area Under Curve; ECG: Electrocardiogram; TTE: Transthoracic echo; Sens: Sensitivity; Spec: Specificity.

1.3.3 Limitations of 12-lead ECG

Though a universally familiar tool in clinical practice, access to 12-lead ECG is far from universal. Some studies suggest that in lower income countries only 10% of clinics offer this service to patients.[99] The cost of a 12-lead ECG machine can run to several thousands of pounds. Even in the NHS GP practice setting, where this tool is commonly available, its use relies on machine maintenance, competence in applying the electrodes and performing the test, and expertise in interpreting the 12-lead trace output. Given the reality of primary care appointments in the NHS being mostly curtailed to less than ten-minute timeslots[100], patients often need to return to have their ECG recorded in a separate encounter, possibly at another facility. This introduces clinical risk through delays in clinical decision making, as well as inconveniencing patients and clinicians alike.

1.3.4 Single-Lead ECG Devices

Given some of the outlined potential shortcomings of 12-lead ECG, the development of simple, quick, and accurate single-lead ECG tools has resulted in these now being increasingly commonplace in clinical practice. Among the most popular devices has been the Kardia Mobile (AliveCor, US), the latest version of which is credit-card sized, with two electrodes for placement of fingers from right and left hand.[101] This creates a vector mirroring lead I of a standard 12-lead ECG. Though useful for inferring rhythm disturbances and elucidating potentially worrisome ECG morphologies, the obvious limitation is that these devices provide a less comprehensive picture of the heart's electrical activity than 12-lead ECG. Though notably, a newer model from Alivecor (Kardia 6L), features three electrodes, two on the front and one on the back, with the latter rested on the patient's knee – creating sufficient vectors to output a six-lead ECG.[102]

1.3.5 Adoption Challenges of Single-lead ECG Technology

The increasing evidence generation and traction for single-lead ECG use in the detection of heart rhythm disturbances resulted in NICE recommending Kardia as a personal ECG device to support at-home monitoring and detection of AF.[103] However, evidence for impact of such readily available technologies at the primary care interface has been mixed. One programme led by England's Academic Health Science Networks (AHSNs) was born out of a drive for system-wide procurement of mobile ECG solutions. NHS England, the central government organisation responsible for health and social care in England, identified £500,000 to purchase a large volume of Kardia devices (>6,000) for large-scale deployment. A subsequent evaluation of this programme highlighted the need to mitigate against early abandonment of the technology. Around two thirds of participating health professionals were low users (<25 recordings in 14 months) and had negative perceptions of the programme overall. Doctors were most likely to abandon use early compared with other staff groups.[104]

The variable success of this AHSN project illustrates some of the key challenges around non-adoption, abandonment, and scaling up, spreading, and sustaining use of DHTs.[105] The track record of DHT implementation programmes, especially those that require major changes in organisations or the wider care system, is poor. This highlights the importance of prioritising the understanding the clinical workflow and how a DHT fits into "business as usual."[106,107] Particularly in primary care, consultations are highly time pressured such that any additional new process that creates extra work and deviates from familiar workflows is likely to suffer from low uptake and poor adherence. This may extend to performing a thirty second ECG with a Kardia-like device, realistically taking several minutes to perform end-to-end. In considering this workflow challenge, some manufacturers have sought to integrate and overlay single-lead ECG capability within tools that are already universal to

routine clinical practice. For example, every clinician, particularly in primary care, owns and uses a stethoscope. Through the convergence of sophisticated hardware, software, and connectivity, a new category of DHTs in the form of "smart" stethoscopes has emerged, now capable of recording digital biomarkers including single-lead ECG.

1.4 Evolution of the Stethoscope

1.4.1 A brief history

René Laennec (1781–1826) was a French physician credited with inventing the stethoscope in 1816.[108] In the intervening two hundred years, the fundamental design has changed little, still relying on noise conduction from a small diaphragm within a bell-shaped end piece along tubing and into the user's ears via earpieces. Recent rapid innovation both in terms of the quality and miniturisation of sensors technologies has led to the development of smart stethoscopes that retain the same form factor as the original but are enhanced with technology that provides substantial additional insights during clinical examination.

On example of such a technology, the Eko DUO (Eko Health, Berkeley, CA, US), features both a digital microphone for recording of phonocardiograms (PGC, or heart sounds) as well as two electrodes for single-lead ECG capture. Both waveforms can be captured upon application of the device to a patient's chest e.g. such as during the universal, clinically familiar auscultation over the aortic, pulmonary, tricuspid, and mitral values as part of a standard cardio-respiratory examination. Connectivity via Bluetooth enables ECG and PPG waveforms to be displayed for real-time clinical insight on the user's smartphone or tablet. Further connectivity to the internet allows operation of a cloud-based platform for live-streaming of waveform examination. Furthermore, AI algorithms can

be operationalised within such cloud-based architecture and interrogate waveforms for diseases of interest (figure 1.4).

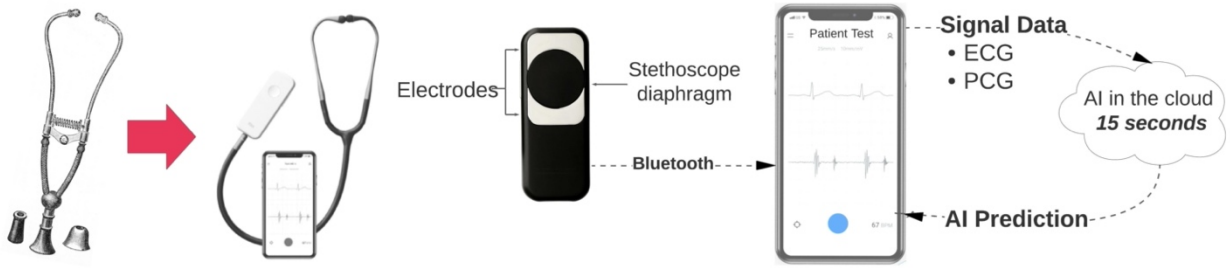


Figure 1.4 Evolution of the stethoscope.

Left, drawing of 19th century stethoscope, with fundamental engineering unchanged in most stethoscopes in current use. Right of this, the Eko DUO smart stethoscope with digital microphone and two electrodes for capture of single-lead ECG; waveforms are transmitted to users' smartphones and onward to the Cloud for interrogation by AI algorithms.

1.4.2 Patient Self-Administered Stethoscope Examination

The connectivity of such technology highlights the opportunity for decentralised clinical examination, where a patient could avoid the need to attend for an in-person exam at all but instead hold a device to their chest under remote supervision via telemedicine. A playbook for this has already been established for single-lead ECG by market-leading technologies such as the Kardia. A device positioned as a stethoscope with in-built single-lead ECG capability poses a tool for more comprehensive remote cardio-respiratory assessment, joining other well-established investigations initiated by patients in their own home, such as measurement of blood pressure.[109] However, deployment of a home-based screening tool combining hardware, AI, and a cloud-based digital

platform for administration – and all anchored in patient self-administration – raises distinct clinical, health economic, and ethical challenges for safe, effective, and trustworthy implementation.[110,111]

1.5 Point-of-Care AI-ECG for Reduced LVEF Using Single-Lead ECG: Opportunity and Evidence Gaps

The systematic recording, streaming, and storage of cardiac waveforms within cloud-based infrastructure by devices such as the Eko DUO also brings the opportunity for interrogating these recordings with AI algorithms. Neural networks such as AI-ECG for reduced LVEF can readily operate within cloud-based architecture and return interpretations to the end-user – subject to adequate connectivity – instantaneously.

The key limitation is whether AI-ECG, originally designed for and thus far only validated and applied to 12-lead ECG, can sustain adequate, clinically useful performance through interpretation of only a single-lead ECG input. If this were proven to be the case, collectively the Eko DUO with AI-ECG would represent a compelling point of care test for reduced LVEF that would be quick to perform, non-invasive, and relatively cheap. By executing the test as part of a stethoscope examination, this test draws on a cornerstone of the routine clinician-patient encounter without creating additional work.

Addressing the evidence gap for this context carries substantial complexity due to the performance of AI-ECG being reliant on several important factors. These include whether the Eko DUO's single-lead ECG recording is of sufficient signal quality to even attempt AI-ECG interpretation, as well as the question of which position a recording should be taken from, given the device's ability to record ECG waveforms from placement anywhere on the patient's chest (or indeed

handheld by emulating use of a Kardia device). Considering potential barriers to uptake, use of such a device would ideally be investigated for ease-of-use across different healthcare professionals. This ought to extend further to evaluating use by lay persons, given the outlined opportunity for such tools to facilitate decentralised, self-administered examination by patients themselves. Scrutiny of the outlined concerns around performance biases in AI potentially driving up health inequities would require a sufficiently large number of patient participants from diverse socioeconomic and ethnic backgrounds to reassure potential users.

From the perspective of NHS funders, a novel tool, regardless of its putative effectiveness, will generally struggle to achieve large-scale commissioning in the absence of a strong clinical case being matched by a persuasive health economic argument. Given the outlined need to increase detection of HF in primary care, any tool's ability to deliver this would ideally be framed against a contemporary picture of the health economic benefits for achieving index diagnosis through primary care vs. the currently dominating mode of hospital admission.

Lastly, real-world use of point-of-care AI-ECG for reduced LVEF using single-lead ECG raises untested assumptions around ethical challenges. Particularly in the absence of precedent, any NHS stewardship of a novel clinical pathway enabling patient self-administered screening with AI-ECG would require guardrails that preserve and promote trust, patient engagement, and public health.

1.6 Chapter Summary

This introductory chapter detailed the clinical background, health economic challenges, and health policy priorities around HF, focusing specifically on the substantial opportunity to increase detection at the primary care interface. I detailed the emergence of DHTs using AI to target cardiovascular disease as potential solutions and described the development of AI applied to the ECG (AI-ECG). This carries the opportunity for an impactful point-of-care test through integration with a routinely used clinical tool such as a smart, ECG-enabled stethoscope. The chapter ended by expanding on the evidence gaps around clinical validity, health economic impact, and ethical acceptability, concluding that these need to be addressed to maximise the uptake and impact of such a tool.

1.7 Scope of Thesis, Aims, and Hypotheses

Based on the outlined Introduction, the principal aim of this thesis is to investigate the potential application of AI-ECG for detection of HFrEF using single-lead ECG – towards highlighting a potential tool for mitigating the public health burden posed by HF. Establishing meaningful uptake and large-scale impact of such technologies requires synergistic investigation across the outlined clinical, health economic, and ethical evidence gaps. Subsequently, this thesis tests the following specific, intersecting aims and hypotheses, all anchored in AI-ECG applied to single-lead ECG recorded during a smart stethoscope examination.

1.7.1 Aims

1. To conduct a large-scale independent, prospective, external validation study of AI-ECG for reduced LVEF (≤40%) applied to single-lead ECG, recorded during a smart stethoscope examination with NHS patients attending for echocardiography (ground truth for comparing performance of AI-ECG).
2. To measure the contemporary clinical and health economic implications of HF by the route to index diagnosis (through primary care or hospital admission) and estimate the potential cost savings of increasing rates of diagnosis through point-of-care AI-ECG deployment in primary care.
3. To evaluate the ethical implication of extending this technology to patient self-administered screening and specific policy recommendations to blueprint best practice.

1.7.2 Hypotheses

1. Detection of reduced LVEF using AI-ECG applied to single-lead ECG recorded by a smart stethoscope is reliable, accurate, operator-independent, unbiased, and suitable for patient self-administration.
2. The mechanism for index diagnosis of HF remains dominated by hospital admission, with worse long-term clinical and health economic outcomes compared to patients first diagnosed in primary care – to sufficiently justify the cost of deploying point-of-care AI-ECG
3. Patient self-administration of AI-ECG raises ethical challenges including considerations of equity, agency, data rights, and ultimately responsibility for safe, effective, and trustworthy implementation.

Each of these aims and hypotheses is addressed in the ensuing empirical chapters. Each has its own discrete and detailed sub-section on materials and methods, rather than a standalone chapter for this, to avoid repetition. This thesis concludes with a synthesis of the results from the ensuing chapters and sets out a roadmap for future work.

2 Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during AI-ECG stethoscope examination: prospective, observational, multicentre, external validation study

This chapter describes the results of my 1,050-patient prospective external validation study for the detection of HFrEF using AI-ECG applied to single-lead ECG recorded by a smart stethoscope. This is the first study of its kind, highlighting that AI-ECG is accurate, reliable, operator-independent, and unbiased. Results presented in this chapter have also been published in The Lancet Digital Health (Bachtiger et al., 2022) and are reproduced here under a Creative Commons BY 4.0 license.

2.1 Abstract

2.1.1 Background

Most patients who have heart failure with reduced ejection fraction (HFrEF), when left ventricular ejection fraction (LVEF) is 40% or lower, are diagnosed in hospital. This is despite prior presentations to primary care with symptoms. In this study, I aimed to test an artificial intelligence (AI) algorithm applied to a single-lead ECG (AI-ECG), recorded during an ECG-enabled stethoscope examination to validate a potential point-of-care screening tool for LVEF≤40%.

2.1.2 Methods

I conducted a prospective, observational, multicentre study of a convolutional neural network (AI-ECG) that was previously validated for the detection of reduced LVEF using 12-lead-ECG as input. I used AI-ECG re-trained to interpret single-lead ECG input alone. Patients aged ≥18 years attending for transthoracic echocardiogram (TTE) in London (UK) were recruited. All participants had 15 s of supine, single-lead ECG recorded at the four standard anatomical positions for cardiac auscultation, plus one handheld position, using a single-lead ECG-enabled stethoscope. TTE-derived percentage LVEF was used as ground truth. The primary outcome was performance of AI-ECG at classifying reduced LVEF (LVEF ≤40%), measured using metrics including the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, with two-sided 95% CIs. Secondary outcome measures included algorithmic bias, operator independence, and feasibility of patient self-administered examination. The primary outcome was reported for each position individually and with an optimal combination of AI-ECG outputs (interval range 0–1) from two

positions using a rule-based approach and the best of several classification models. This study is registered with ClinicalTrials.gov, NCT04601415.

2.1.3 Results

Between Feb 6, 2021, and May 27, 2021, my study prospectively recruited 1,050 patients (mean age 62 years [SD 17·4], 535 [51%] male, 432 [41%] non-White). 945 (90%) had an ejection fraction of at least 40%, and 105 (10%) had LVEF≤40%. Across all positions, single-lead ECGs were most frequently of adequate quality for AI-ECG interpretation at the pulmonary position (979 [93·3%] of 1,050). Quality was lowest for the aortic position (846 [80·6%]). AI-ECG performed best at the pulmonary valve position (p=0·02), with an AUC of 0·85 (95% CI 0·81–0·89), sensitivity of 84·8% (76·2–91·3), and specificity of 69·5% (66·4–72·5). Diagnostic odds ratios did not differ by age, sex, or non-White ethnicity. Taking the optimal combination of two positions (pulmonary and handheld positions, the rule-based approach resulted in an AUC of 0·85 (0·81–0·89), sensitivity of 82·7% (72·7–90·2), and specificity of 79·9% (77·0–82·6). Using AI-ECG outputs from these two positions, a weighted logistic regression model with l2 regularisation resulted in an AUC of 0·91 (0·88–0·95), sensitivity of 91·9% (78·1–98·3), and specificity of 80·2% (75·5–84·3). There was no observed algorithmic bias or operator dependence. Results from clinician and patient self-administered recordings were strongly correlated (intraclass correlation coefficient >0.8).

2.1.4 Conclusion

A deep learning system applied to single-lead ECGs acquired during a routine examination with an ECG-enabled stethoscope can detect LVEF≤40%. These findings highlight the potential for

inexpensive, non-invasive, workflow-adapted, point-of-care screening. This may translate to earlier diagnosis and benefits through early commencement of prognostically beneficial treatment. The primary care interface poses a compelling area for deployment, with the possibility of further decentralising use of the technology through patient self-administered examination.

2.2 Introduction

The escalating worldwide burden of heart failure (HF) is compounded by late diagnosis, which both worsens patients' prognoses and increases costs for health systems, primarily through avoidable hospital admissions.[28,112,113] In the UK, the National Health Service (NHS) Long Term Plan emphasises this shortcoming in care, highlighting that "80% of heart failure is currently diagnosed in hospital, despite 40% of patients having symptoms that should have triggered an earlier assessment".[27] Among these patients, over 50% have heart failure with reduced ejection fraction (HFrEF), designated by an echocardiography derived left ventricular ejection fraction (LVEF) of 40% or lower.[114] The prognosis for patients with an LVEF≤40% continues to improve with advancements in cost-effective drug and device therapies, where timely commencement maximises benefits.[6-8] There is, therefore, an important unmet need for inexpensive and practical point-of-care screening for an LVEF≤40%. Nearly half of patients diagnosed with HF through hospital admission have previously seen their primary care physician with HF symptoms, therefore the outlined tool would be particularly compelling for deployment in the primary care setting.

Through the application of AI, the 12-lead ECG has been described as an accurate digital biomarker for the stages of LVEF compromise. Previous research by the Mayo Clinic showed that a convolutional neural network (CNN), trained on 12-lead ECGs labelled with corresponding ECG-derived LVEF, could detect LVEF of 35% or lower with 86·3% sensitivity and 85·7% specificity.[92] This AI-ECG model has since been externally validated with 12-lead ECGs in further Midwestern (US) cohorts[94], and in a Russian population (sensitivity 80·8% and specificity 67·3%).[95] Most recently, a cluster randomised controlled trial made AI-ECG accessible for 12-lead ECG interpretation in a cohort of Mayo Clinic primary care practices, highlighting an increase in the diagnosis of LVEF of 50% or lower (odds ratio [OR] 1·32 (1·01–1·61).[97]

The emergence of ECG-enabled stethoscopes, capable of recording single-lead ECGs during contact for routine auscultation, highlights an opportunity to apply AI-ECG for rapid, easy point-of-care screening. Beyond accuracy of the algorithm when using single-lead ECG alone, this is contingent on these inputs being easy to record and being consistently of adequate quality for attempting AI-ECG interpretation. Secondarily, translation into clinical practice requires scrutiny for algorithmic bias and operator independence. As an ostensibly easy-to-perform test, there is also the opportunity to evaluate whether patients can perform the test on themselves (self-examination), thereby highlighting an opportunity for decentralised screening.

I therefore aimed to investigate whether AI-ECG, retrained to use single-lead ECG as input, could interpret recordings from an AI-ECG stethoscope at anatomical sites established within routine clinical examination, and whether LVEF≤40% could be detected in a previously untested, diverse population. My study tested the principal hypothesis that LVEF≤40% could be detected at or above the clinically meaningful accuracy of previous 12-lead ECG studies (sensitivity >81% and specificity >67%), demonstrating that a universal cornerstone of patient encounters – the stethoscope examination – could provide a point-of-care screening opportunity.

2.3 Methods

2.3.1 Study design, participants, and ethical approval

I designed a prospective, multicentre external validation study, where patients were recruited across seven NHS sites, including hospitals and community health centres (table 2.1), that performed transthoracic echocardiography (TTE) in London, UK. Patients were recruited by 15 operators (six clinicians, six sonographers, and three senior medical students), all of whom received the same training. All adult (aged ≥18 years) patients attending for TTE were eligible to participate (inpatients and outpatients).

Patients were attending for TTE as part of their routine clinical care, having been referred by clinicians for various standard TTE indications, such as investigation of symptoms (e.g., breathlessness, peripheral oedema, fatigue, and chest pain) and screening (e.g., due to hypertension, arrhythmia, stroke, or suspected valve disease). Patients were not excluded based on any reason for TTE referral or patient clinical characteristics. Written informed consent was obtained from all participants before enrolment and a patient information sheet was offered (consent form and patient information sheet included in supplementary appendix). This study was approved by the UK Health Research Authority (reference 21/LO/0051). This study is registered with ClinicalTrials.org (NCT04601415).

Study Site (London, UK)	Setting	Patient group
Hammersmith Hospital	Hospital	Inpatient and outpatient
St. Mary's Hospital	Hospital	Inpatient and outpatient
Charing Cross Hospital	Hospital	Inpatient and outpatient
Maida Vale Heart Health Centre	Community	Outpatient
Hanwell Community Clinic	Community	Outpatient
Parkview Medical Centre	Community	Outpatient
St Charles Community Hospital	Community	Outpatient

Table 2.1 Hospital and community recruitment sites across London

2.3.2 AI-ECG algorithm architecture

The AI model design for 12-lead AI-ECG has been previously described.[9] To summarise, the model uses a CNN, trained on 35,970 independent pairings of 12-lead-ECG and echocardiograms from the proprietary Mayo Clinic digital data vault (figure in 1.3 in Introduction). CNN's, which are commonly applied to images, operate such that the convolutions can be used to extract very subtle patterns in a data set. Each 12-lead ECG was considered a 12×5,000 (that is, 12 leads by 10-s duration sampled at 500Hz) matrix, where the first dimension represents a spatial dimension and the second represents a temporal one.[115] ECG analysis is mostly a visual task. While the signal is a time series, it is pseudocyclical, and its main features are morphologic.[116-118] To enable detection of patterns in these features, AI-ECG uses architectures that were based on convolutional layers for feature extraction. For my study, I tested the Mayo Clinic's single-lead ECG version of AI-ECG, which uses the same CNN model architecture as the original 12-lead model but this time retrained with each single-lead ECG, extracted from the original, 35,970 12-lead patient dataset. This study is enabled by the AI-ECG model having been licensed for research to Eko Health (Berkeley, CA, US), who manufactures the

Eko DUO single-lead-ECG-enabled smart stethoscope (detailed further below in this Methods section) used in this study.

2.3.3 AI-ECG Output

For each single-lead ECG input, the output for the CNN is a continuous value between 0 and 1, indicating the probability of the condition of interest being present i.e., LVEF of the specified cut-off. The closer the AI-ECG input is to 1, the more likely the model "thinks" that the condition of interest is present. For this study I decided to focus on evaluating AI-ECG performance for detection of LVEF≤40% i.e. HFrEF, given that this is both the most common subtype and because prognostically beneficial therapies are disproportionately abundant. The AI-ECG threshold can be adjusted along the 0–1 scale (figure 2.1, panel C) according to the LVEF cutoff of interest and trade-off in performance (e.g., sensitivity *vs* specificity). On the basis of the chosen threshold, recordings are subsequently classified as positive (LVEF ≤40%) or negative (LVEF>40%).

2.3.4 Single-lead ECG recording with Eko DUO

This study used a smart stethoscope capable of recording a single-lead ECG (Eko DUO, Eko Health, Berkeley, CA, US). The device records 15 seconds of single-lead ECG upon the press of a button on the device or within the accompanying mobile phone app (connected by Bluetooth). Only the first 10 seconds are analysed by the AI-ECG model. A separate deep learning classifier previously trained on signal quality annotations, validated based on ground truth determined by a plurality vote of three cardiologists, categorised single-lead ECG recordings as either adequate or inadequate quality to attempt AI-ECG interpretation. AI-ECG will interpret any adequate quality ECG waveform and

produce a prediction, regardless of the position or orientation from which the single-lead ECG is being recorded by the Eko DUO (figure 2.1, panel B).

2.3.5 Regulatory status of Eko DUO and AI-ECG

The Eko DUO smart stethoscope is a Class IIa CE marked medical device. AI-ECG falls under the regulatory category of software as a medical device (SaMD) and currently does not have regulatory approval. The data generated through this study would be admissible as part of a regulatory submission to the UK MHRA.

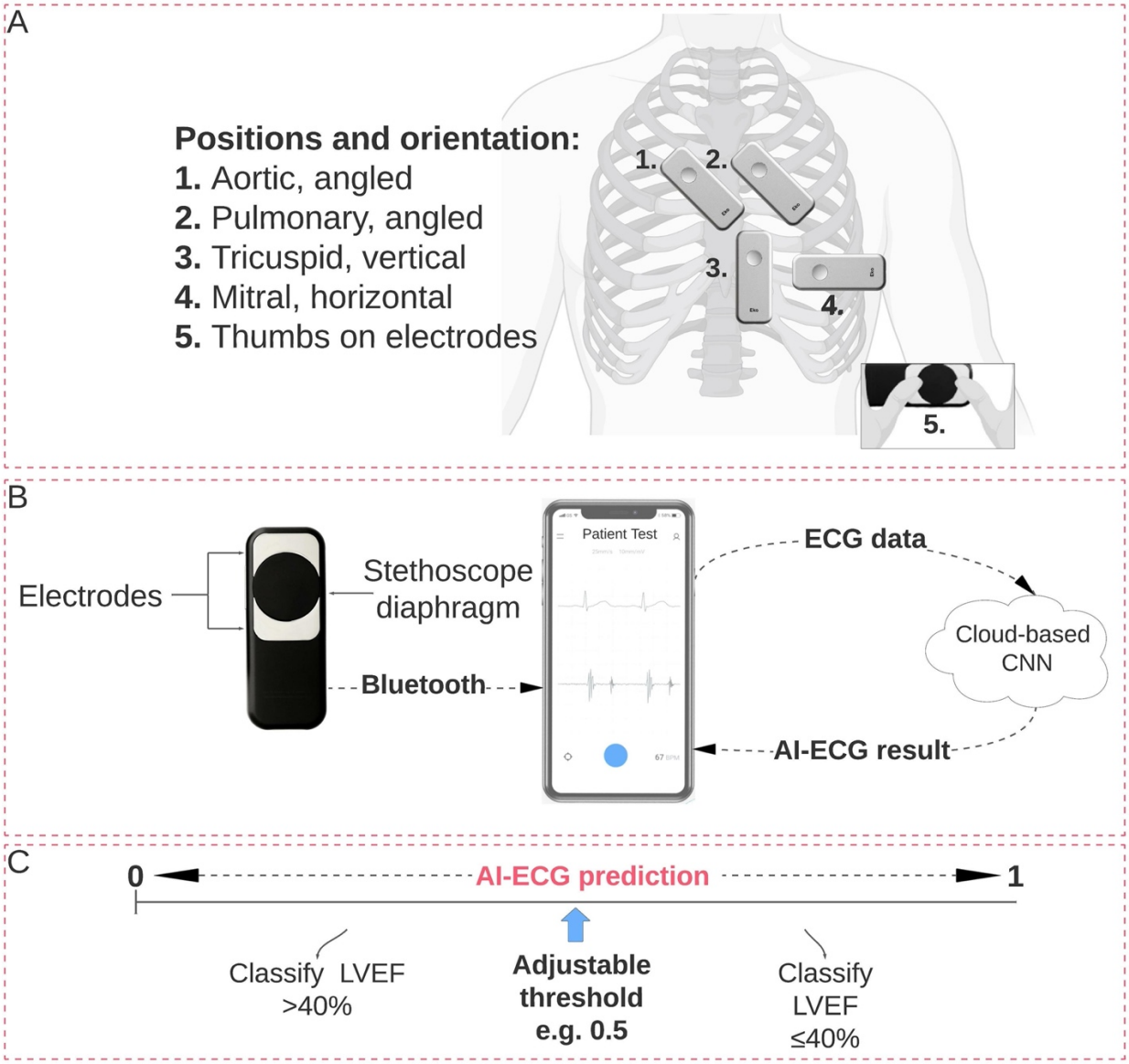


Figure 2.1 Eko DUO recordings of single-lead ECG interrogated by AI-ECG.

Panel A: Recording positions and orientation of the Eko DUO across aortic, pulmonary, tricuspid, and mitral positions for cardiac auscultation, with additional hand-held position. Panel B: Eko DUO patient-facing side with sensors, where ECG waveforms are transmitted to app on smartphone via Bluetooth, and onwards via internet connectivity for cloud-based interpretation by AI-ECG. Modified from Bachtiger et al. 2022.

2.3.6 Main Study Protocol

Patient recruitment started on Feb 6, 2021. All participants had 15 seconds of supine, single-lead ECG recorded via the two electrodes on the patient-facing side of the Eko DUO (figure 2.1, panel A). All recordings were made within 24 hours of TTE; almost all were recorded during the same clinical encounter. Members of the research team were unaware of participants' LVEF at the time of recording and remained blinded to these results for the duration of the study.

The AI-ECG stethoscope has two electrodes on the patient-facing side of the device. Placement on a patient's chest (or handheld) creates a vector for recording ECGs. For simplicity and following a familiar clinical workflow, positions were recorded in sequence at standard anatomical landmarks for auscultation of the aortic, pulmonary, tricuspid, and mitral valves, and at one handheld position. The single, fixed angulation specified for each position was reached via clinical consensus of what was most intuitive and captured various vectors across the five positions. Aortic and pulmonary positions were recorded holding the device angled to the left, with the tricuspid position in a vertical and mitral position in a horizontal orientation (figure 1, panel A). Although precordial placement is not identical to electrode positioning for 12-lead ECG, the vectors explored were similar. For example, the pulmonary valve position most closely resembles lead II of a 12-lead ECG. Heart sounds (phonocardiograms [PCG]) were automatically recorded at the same time but did not serve as inputs for AI-ECG. For the handheld position, patients were asked to place a thumb on each of the two electrodes, with the left thumb on the exploring electrode, such that this represented Lead 1 of a standard 12-lead ECG.

2.3.7 Reproducibility of clinician recording vs. patient self-recording

To investigate whether patients could perform a self-recording over the pulmonary position (selected due to this representing commonly-used lead II rhythm strip), an unselected, opportunistic subset of participants (target n = 50) was also recruited. Participants were asked to follow simple instructions (included in supplementary material 10.4) to place the device on their chest in the pulmonary position and record three fifteen second single-lead ECG traces, lifting the device off their chest between each recording and placing it anew. Following this, a clinician (same individual each for all) would similarly perform three such recordings on the same patient. Patients would always go first to avoid bias of them copying the positioning of the physician.

Figure 2.2 Member of the clinical research team examining a patient recruited for the study using an Eko DUO.

Written informed consent to show image received from both patient and member of research team.

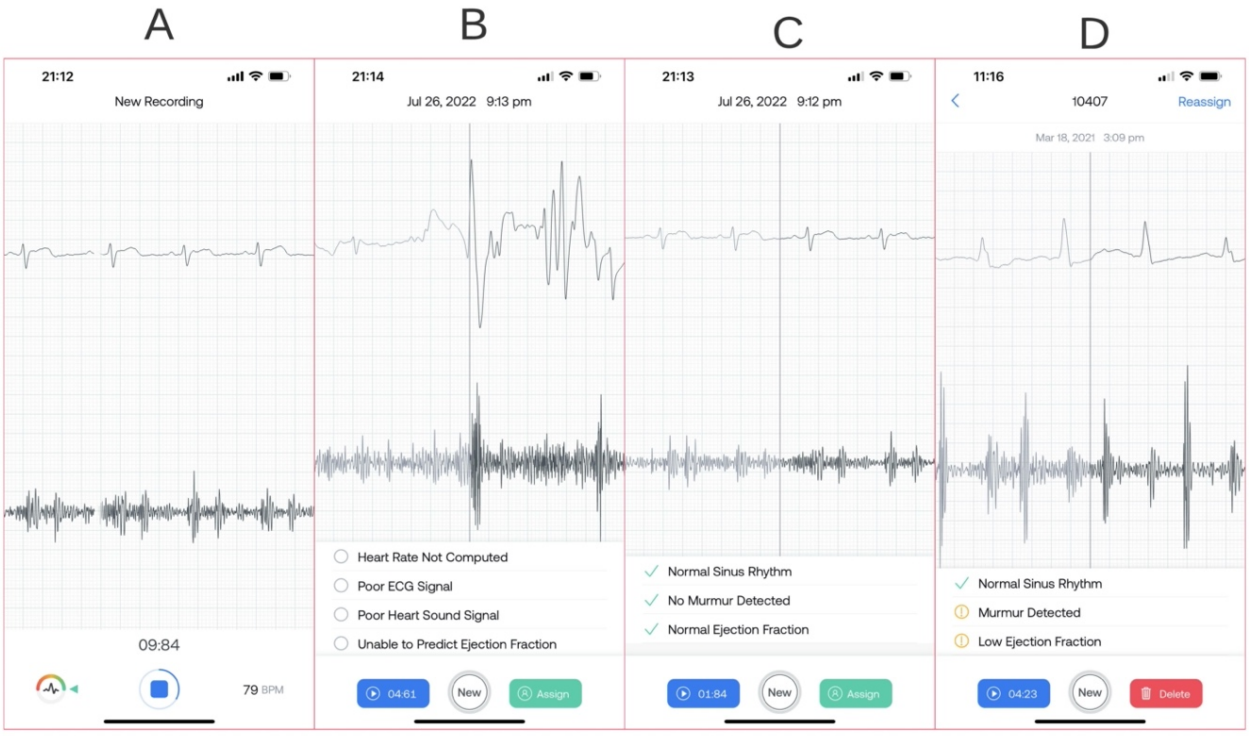


Figure 2.3 Eko app display on smartphone.

Panel A: electrocardiogram (ECG) and phonocardiogram (PCG) waveforms being recorded with a 15 second countdown, bottom left red-amber-green dial indicates signal quality; Panel B: noisy, uninterpretable ECG and PCG waveforms; Panel C: normal ECG and PCG waveforms; Panel D: ECG waveform suggestive of "low ejection fraction" at arbitrary starting threshold of 0.5 (note separate AI algorithm unrelated to this study also suggested presence of murmur).

2.3.8 Eko DUO App, dashboard, and study database

The AI-ECG stethoscope (Eko DUO) transmits single-lead ECG recordings via Bluetooth for visualisation via an Android or iOS smartphone app (Eko App, Eko Health). The app notified the operator when the ECG signal was of adequate or inadequate quality for attempting interpretation by the algorithm, using a red-amber-green dial (figure 2.3, panel A). Only one recording attempt was allowed for each position within the protocol, where maintaining the dial at the "green" level maximised the likelihood of adequate signal quality.

The ECG waveform data were analysed in real-time by AI-ECG via a cloud-based CNN, hosted by the device manufacturer using protocols compliant with the Health Insurance Portability and Accountability Act and General Data Protection Regulation. No information was stored on individual users' smartphones. Overall, the full examination took approximately 2 minutes per patient.

As an arbitrary starting point, the app displayed the result of AI-ECG on the basis of an arbitrary starting threshold of 0.5 – where any AI-ECG prediction above this would display as "low ejection fraction" (figure 2.3, panel C). All operators received the same training and education on the scientific background for AI-ECG and were aware that the results displayed in the app were subject to threshold optimisation and therefore did not confer any clinical utility (and could not be used for this purpose anyway given the still non-regulatory approved nature of AI-ECG). Raw AI-ECG predictions (numbers between 0 and 1) for each single-lead ECG were retrieved from the stethoscope manufacturer's online dashboard (figure 2.4) and combined with a secure, de-identified database containing relevant demographic and clinical variables for each participant. Patients' ethnicity was self-reported from a list of 18 options drawn from the UK Office of National Statistics Census for England.[119]

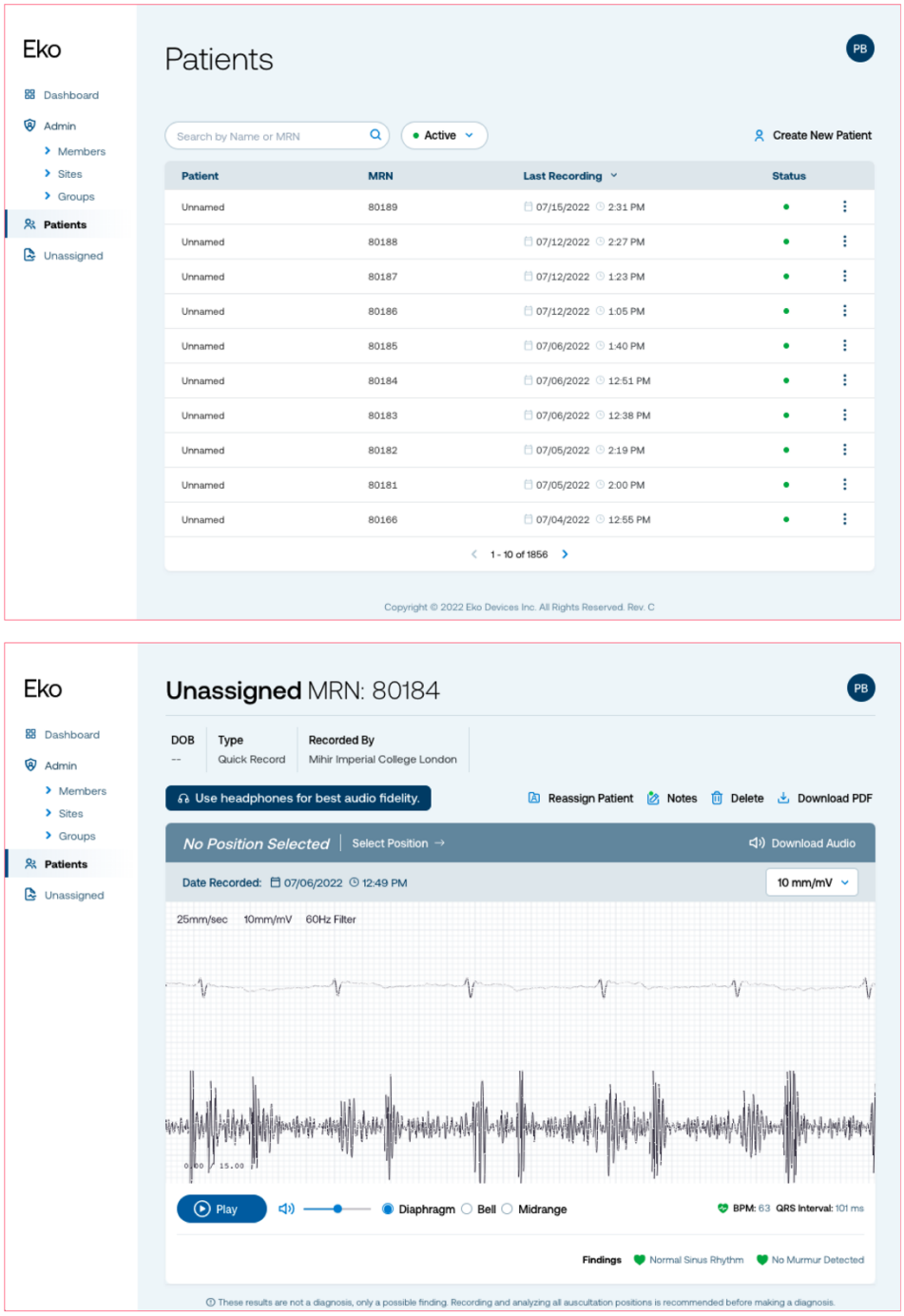


Figure 2.4 Eko web dashboard.

Top: dashboarded list of patients recruited during the study. Bottom: close-up view of an individual patient's ECG and PCG waveform recordings.

2.3.9 Outcomes

The primary outcome was the identification of patients with an LVEF≤40% from single-lead ECG recordings obtained by the ECG-enabled stethoscope. For diagnostic accuracy assessment, the gold standard was percentage LVEF as measured on a 2D transthoracic echocardiogram (TTE) acquired by echocardiographers accredited by the British Society for Echocardiography.[120] LVEF was recorded in line with the same approach taken by the Mayo Clinic for labelling the ground-truth training dataset for AI-ECG. Namely, the first LVEF available from a standard hierarchical sequence: a biplane approach using the Simpson method, a 2D method, or M-mode and, in the absence of any of the preceding, the reported visually estimated LVEF. Where LVEF was reported as a range, the midpoint value was used.

Secondary outcomes included scrutiny for bias by age, sex, and non-White ethnicity. I also aimed to compare the performance of AI-ECG across the three most high-volume recruiting operators. Lastly, I evaluated differences in the consistency of AI-ECG predictions from clinician vs. patient self-administered recordings.

2.3.10 Sample size

This study was not comparing the impact of an invention between groups, therefore a conventional sample size and power calculation was not indicated. However, in order to inform what would be an impactful number of recruited participants, I anchored a sample size calculation with the aim of testing a hypothesis that AI-ECG could detect EF≤40% at or above the clinically meaningful accuracy of previous external validation studies using 12-lead ECG.[95] Specifically, with expectation of a minimum sensitivity of 81% (one-sided 95% CI, given low prevalence of LVEF≤40% and therefore

fewer anticipated 'positive' cases) and specificity of 67% (two-sided 95% CI). Setting the lower limits of confidence intervals for sensitivity and specificity at 70% and 60%, respectively, I calculated the need to recruit 371 patients with EF>40%, and 94 with EF≤40%, to have 80% power to detect a difference at the 0.05 level of significance. Given the unselected nature of the approach to recruitment (pre-test LVEF was unknown) combined with the low prevalence of LVEF≤40%, the study would continue recruiting patients until the requisite number of these patients had been achieved; with the expectation of a large majority sample consisting of patients with LVEF>40%.

2.3.11 Statistical analysis

Demographic and clinical variables were summarised for the overall cohort using means and standard deviations. I compared groups stratified by LVEF (>40 *vs* ≤40%) using Student's *t* tests for continuous variables or Pearson's χ^2 test for categorical variables, as appropriate, with $p<0{\cdot}05$ considered statistically significant. The 18 possible options for ethnicity were grouped into White, Black, Asian, mixed, and other for presentation in the main demographics table.

Using outputs from the AI-ECG model in the interval range of 0–1, performance at classifying LVEF (>40% *vs* ≤40%) was measured for each position by calculating the area under the receiver operating characteristic curve (AUC), using a reference standard of TTE-derived percentage LVEF. I tested the AUC results between the best and second best performing single position using the Delong test for significance.[17] For each position, sensitivity, specificity, negative and positive predictive value, and F1 score was reported at (1) the optimal threshold maximising the sum of sensitivity and specificity (i.e., Youden's index, also usually the point closest to the top left of the corner of the plot), and (2) a restricted threshold that would maximise the sum of sensitivity and specificity, with a minimum

sensitivity of 81% and (where possible) a minimum specificity of 67%. 95% CIs are reported using the latter restriction.

Using the single-best performing position and compared with the overall population, I also report diagnostic ORs stratified by sex for two age bands (age 18–69 years and ≥70 years) and by non-White ethnicity. Diagnostic OR is the ratio of positive likelihood ratio (sensitivity / [1 – specificity]) to the negative likelihood ratio ([1 – sensitivity] / specificity). I applied the Breslow-Day test for homogeneity to test for significant ($p<0·05$) variation in performance.[121] Performance is reported using only ECG recordings of adequate quality to attempt AI-ECG analysis.

Expanding beyond the single-best position alone, performance is also reported when considering the best combination of two positions when using a rule-based approach, where either position predicting LVEF≤40% was considered a positive test result. Using the dataset of 0–1 values for AI-ECG model predictions from each of the two optimally combined positions as inputs, several classification models (including logistic regression) were tested for predicting LVEF≤40%. These models used 60% of the dataset for training and 40% for testing, and consisted of equal proportions of patients with each LVEF status (>40% *vs* ≤40%; randomly allocated). The best model was selected using five-fold cross validation.

Using predictions from the AI-ECG neural network in the interval range 0–1 for each single-lead ECG, receiver operating characteristic curves were plotted to display performance across a full range of thresholds. I generated a receiver operating characteristic curve summarising the single-best position, rule-based optimal combination of two positions, and best overall classification model. Confusion matrices are presented using the restricted threshold.

Lastly, reproducibility of recording results (AI-ECG number 0-1) between clinical researcher and patient self-examination was measured by calculating Intraclass Correlation Coefficients (ICCs)

within each of these users, and between them (informed by three sequential recordings). I considered an ICC of 0.75 to represent 'good' reproducibility.[122] AI-ECG data from patient and clinician recordings were also visualized using a linear distribution plot and a Bland Altman plot. All analyses were done in R (version 3.6.1) and Python (version 3.7.6).

2.3.12 Role of the funding source

This study was funded by the National Institute for Health and Care Research (NIHR) Artificial Intelligence in Health Award (ref. AI_AWARD01849). I was a co-applicant and led on writing the grant proposal and research protocol. The funder had no role in the study design, data collection, data analysis, data interpretation of data, or writing of the report. Eko Health and the Mayo Clinic were not involved in funding the study or in the design of the study protocol or analysis.

2.4 Results

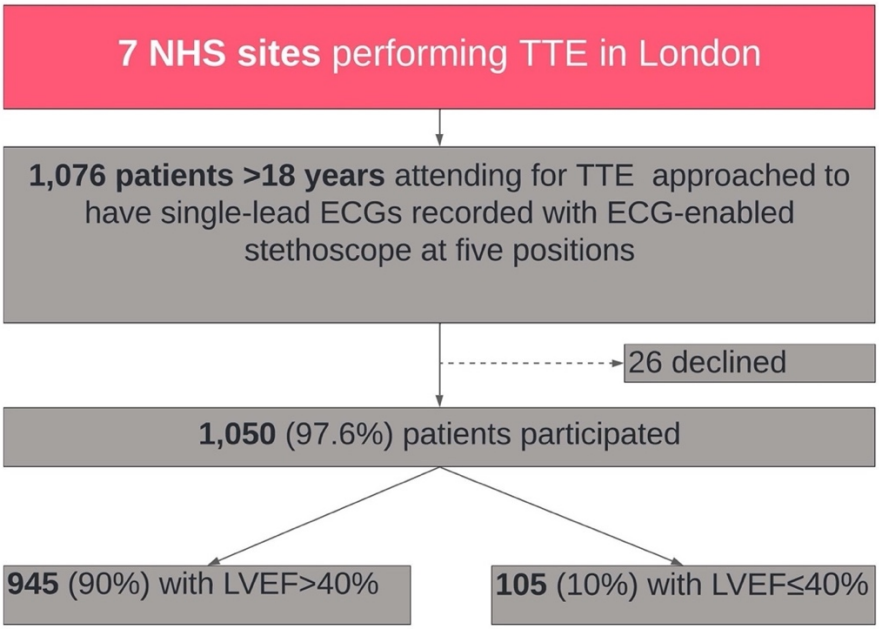


Figure 2.5 Study recruitment flow diagram.

TTE, transthoracic echocardiogram; ECG, electrocardiogram; LVEF, left ventricular ejection fraction. Adapted from Bachtiger et al., 2022.

2.4.1 Participant cohort profile

Between February 6th, 2021, and May 27th, 2021, 1,050 patients were recruited (figure 2.5), of whom 105 (10%) had an LVEF≤40% and 945 (90%) had an LVEF of at least 40% (table 2.2). Overall, the mean age was 62 years (17·4); 535 (51%) patients were male and 432 (41%) were non-White. Full ethnicity breakdown is available in table 2.3. Compared with the normal LVEF group (LVEF >40%), the reduced LVEF group (LVEF ≤40%) was older (mean age 62 years [SD 17·5] *vs* 67 years [15·3]) and had fewer female participants (36 [34%] of 105 *vs* 479 [51%] of 945; table 1). Most comorbidities were more prevalent among the reduced LVEF group.

Characteristic	Total, N = 1,050[1]	LVEF>40%, N = 945[1]	LVEF≤40%, N = 105[1]	p-value[2]
Age, years	62 (17.4)	62 (17.5)	67 (15.3)	<0.001
Age Groups	..	..	..	0.034
18 - 69	636 (61%)	583 (62%)	53 (50%)	..
70+	414 (39%)	362 (38%)	52 (50%)	..
Male, n	535 (51%)	466 (49%)	69 (66%)	0.002
TTE LVEF %	54 (10.3)	57 (5.8)	30 (8.2)	<0.001
Ethnicity, n	..	..	..	0.4
Asian	199 (19%)	176 (19%)	23 (22%)	..
Black	95 (9.0%)	84 (8.9%)	11 (10%)	..
Mixed	22 (2.1%)	18 (1.9%)	<5	..
Other	116 (11%)	102 (11%)	14 (13%)	..
White	618 (59%)	565 (60%)	53 (50%)	..
Hypertension	395 (38%)	338 (36%)	57 (54%)	<0.001
Myocardial Infarction	102 (9.7%)	62 (6.6%)	40 (38%)	<0.001
Atrial Fibrillation	173 (16%)	146 (15%)	27 (26%)	0.011
Pacemaker	59 (5.6%)	43 (4.6%)	16 (15%)	<0.001
Diabetes	224 (21%)	181 (19%)	43 (41%)	<0.001
Stroke/TIA	100 (9.5%)	85 (9.0%)	15 (14%)	0.11
Chronic Kidney Disease	98 (9.3%)	74 (7.8%)	24 (23%)	<0.001
Smoking	148 (14%)	132 (14%)	16 (15%)	0.8
Alcohol excess	26 (2.5%)	25 (2.6%)	<5	0.5
Hypercholesterolaemia	188 (18%)	159 (17%)	29 (28%)	0.009
Pregnancy	21 (2.0%)	21 (2.2%)	0 (0%)	0.2
COPD	57 (5.4%)	48 (5.1%)	9 (8.6%)	0.2
..	[1] Mean (sd); n (%) [2] t test; Pearson's Chi-squared test			

Table 2.2 Baseline characteristics of study participants.

TTE LVEF, transthoracic echo left ventricular ejection fraction; TIA, transient ischaemic attack; COPD, chronic obstructive pulmonary disease; IQR, interquartile range. Numbers below 5 suppressed and displayed as <5.

Race/ethnicity, total n = 1,050	LVEF>40%, N = 945	LVEF≤40%, N = 105	Overall
African	31 (3.3%)	3 (2.9%)	34 (3.2%)
Any other Black/ African/Caribbean background	27 (2.8%)	6 (5.7%)	33 (3.1%)
Any other ethnic group	83 (8.7%)	13 (12%)	96 (9.1%)
Any other Mixed/Multiple ethnic background	7 (0.7%)	2 (1.9%)	9 (0.9%)
Any other White background	118 (12%)	12 (11%)	130 (12%)
Arab	20 (2.1%)	1 (1.0%)	21 (2.0%)
Bangladeshi	3 (0.3%)	0 (0%)	3 (0.3%)
Caribbean	26 (2.7%)	2 (1.9%)	467 (44%)
Chinese	9 (0.9%)	1 (1.0%)	28 (2.7%)
English/Welsh/Scottish/N Irish/ British	429 (45%)	38 (36%)	10 (0.9%)
Indian	59 (6.2%)	11 (10%)	70 (6.6%)
Irish	21 (2.2%)	3 (2.9%)	24 (2.3%)
Other Asian background	94 (9.9%)	7 (6.7%)	101 (9.6%)
Pakistani	11 (1.2%)	4 (3.8%)	15 (1.4%)
White and Asian	2 (0.2%)	1 (1.0%)	3 (0.3%)
White and Black African	6 (0.6%)	0 (0%)	6 (0.6%)
White and Black Caribbean	3 (0.3%)	1 (1.0%)	4 (0.4%)

Table 2.3 Full breakdown of self-reported race/ethnicity according to ONS Census list for England.

2.4.2 Performance of ECG-enabled smart stethoscope

Single-lead ECG recordings were attempted at all precordial positions in 1045 (99·5%) of 1,050 participants. For the handheld position, this was 1006 (95·8%); reasons for not attempting ECG recording included patients being unable to hold the device (e.g., due to grip weakness from previous stroke). Recording of a 15 second ECG of adequate signal quality for attempting AI-ECG interpretation varied across positions, with the aortic (846 [80·6%] of 1,050) and pulmonary (979

[93·2%]) positions performing worst and best, respectively (table 2.4). Taking position 2 as an example, baseline characteristics for age and sex did not differ between those who did and did not have adequate quality recordings (p>0·05).

2.4.3 Performance of AI-ECG

The performance of the AI-ECG algorithm is summarised in table 2.4. ROC curves are displayed in figure 2.6. The single best performing position was over the pulmonary valve, with an AUC of 0·85 (95% CI 0·81–0·89), sensitivity of 84·8% (76·2–91·3), and specificity of 69·5% (66·4–72·6). For this position, table 2.5 shows confusion matrices and table 2.6 shows differences in model performance among the three operators who recruited the most patients. The second-best position was handheld, with an AUC of 0·79 (0.74 - 0.84); p=0·02).

The pulmonary and handheld positions performed best when combined using a rule-based approach: either one or both predicting LVEF≤40% being considered a positive test. For this analysis, 864 (82·3%) of 1050 patients had adequate quality single-lead ECG for attempted AI-ECG prediction at both positions. The resultant AUROC was 0·85 (95% CI 0·81–0·89), with 82·7% (72·7–90·2) sensitivity and 79·9% (77·0-82·6) specificity.

The model with the best performance used weighted logistic regression with l2 regularisation. Data from 864 patients (number of patients with adequate ECG recordings at both pulmonary and handheld positions) was used; 518 (60%) for training and 346 (40%) for testing. Using AI-ECG outputs from these two positions, a weighted logistic regression with l2 regularisation resulted in an AUC of 0·91 (0·88–0·95), sensitivity of 91·9% (78·1–98·3), and specificity of 80·2% (75·5–84·3.

			Maximising sensitivity and specificity equally (Youden index)						Maximising sensitivity and specificity with rule Se>81, Sp>67; or Se>81, maximising Sp					
Pos	**Adequate ECG (n, %)**	**AUC**	**T**	**Se**	**Sp**	**PPV**	**NPV**	**F1**	**T**	**Se**	**Sp**	**PPV**	**NPV**	**F1**
1	846/1,050 (80.6)	0.75	0.37	77.1	60.7	17.3	95.9	0.28	0.35	81.9	53.3	15.8	96.2	0.26
2	**979**/1,050 **(93.2)**	**0.85**	**0.43**	**71.7**	**86.5**	**37.0**	**96.3**	**0.49**	**0.34**	**84.8**	**69.5**	**23.6**	**97.4**	**0.37**
3	946/1,050 (90.1)	0.78	0.49	68.1	77.4	24.7	95.5	0.36	0.28	81.9	55.2	16.6	96.3	0.28
4	968/1,050 (92.2)	0.78	0.42	62.9	80.6	26.2	95.0	0.37	0.31	81.4	58.4	17.7	96.4	0.29
5	916/1,050 (87.2)	0.79	0.42	62.8	83.4	27.7	95.5	0.39	0.30	81.4	60.1	17.5	96.8	0.29
2 + 5	**864/1,050 (82.3)***	**0.85**	**0.45**	**82.7**	**79.9**	**29.9**	**87.8**	**0.44**	**0.45**	**82.7**	**79.9**	**29.9**	**87.8**	**0.44**
2 + 5 lr	**346****	**0.91**	**0.49**	**91.9**	**80.2**	**35.1**	**98.4**	**0.50**	**0.50**	**91.9**	**80.2**	**35.1**	**98.4**	**0.50**

Table 2.4 Performance characteristics of AI-ECG.

Performance shown across all five individual positions, rule-based approach, and logistic regression model. Pos, position; AUC, area under curve; 1, aortic; 2, pulmonary; 3, tricuspid; 4, mitral; 5, hand-held; lr, logistic regression; T, threshold; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; F1, F1 score.

**Number of patients who had adequate recordings at both position 2 and 5, where 'positive' AI-ECG as per threshold was considered a positive test.*

***Representing 40% testing dataset from the original 864 participants with both position 2 and 5 recordings.*

Adapted from Bachtiger et al., 2022.

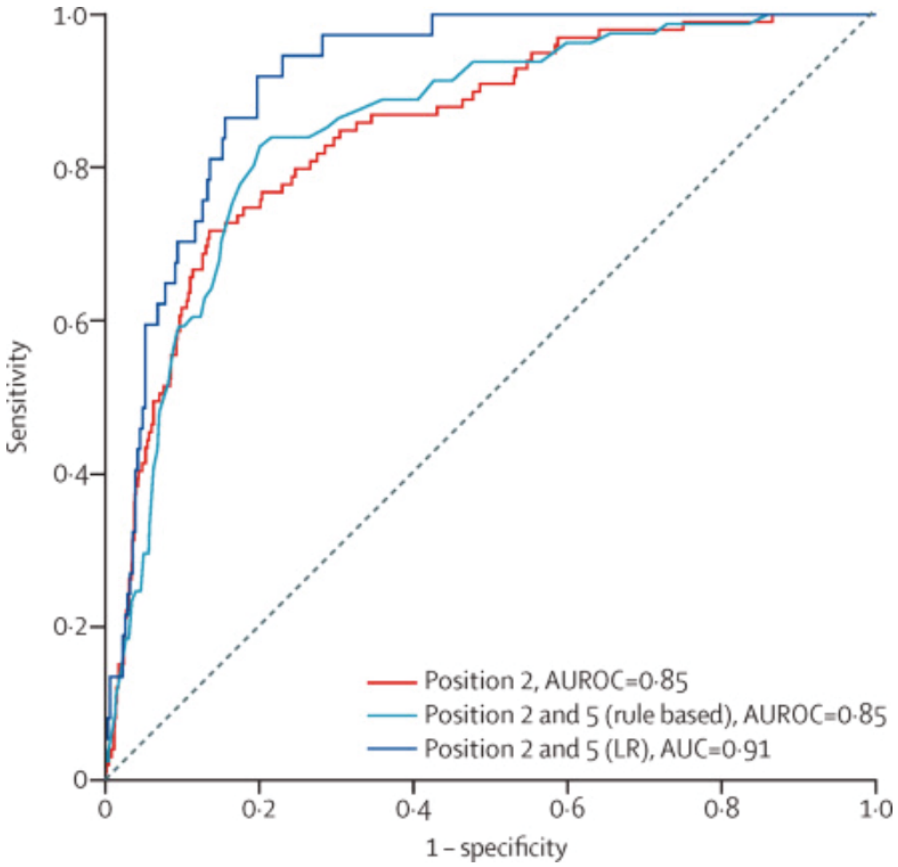


Figure 2.6 ROC curves for performance of AI-ECG.

From Bachtiger et al., 2022. CC BY 4.0

Position 2		
N = 979	LVEF≤40%, n = 99	LVEF>40, n = 880
AI-ECG positive, n = 352	84	268
AI-ECG negative, n = 627	15	612
Positions 2 + 5 (rule-based)		
N = 864	LVEF≤40%, n = 81	LVEF>40, n = 783
AI-ECG positive, n = 224	67	157
AI-ECG negative, n = 640	14	626
Positions 2 + 5 (LR)		
N = 346	LVEF≤40%, n = 37	EF>40, n = 309
AI-ECG positive, n = 95	34	61
AI-ECG negative, n = 251	3	248

Table 2.5 Confusion matrices.

Displayed according to the restricted threshold for maximising sensitivity and specificity, with rule Se>81, Sp>67; or Se>81, maximising Sp. Adapted from Bachtiger et al., 2022.

	Position 2 (pulmonary)		Maximising sensitivity and specificity with rule Se>81, Sp>67; or Se>81, maximising Sp				
Op.	**Adequate recording**	**AUC (CI)**	**Se**	**Sp**	**PPV**	**NPV**	**F1**
1	198/218 (90.7%)	0.80 (0.70-0.88)	77.8 (52.3-96.6)	67.4 (60.0-74.2)	19.4	96.8	0.31
2	192/197 (97.4%)	0.87 (0.77-0.94)	85.0 (62.1-96.7)	75.6 (68.3-82.0)	29.8	97.6	0.44
3	182/204 (89.1%)	0.88 (0.78-0.96)	78.6 (48.2-9.3)	72.9 (65.4-79.4)	19.6	97.6	0.31

Table 2.6 Differences in model performance among the three operators who recruited the most patients.

Adapted from Bachtiger et al., 2022.

2.4.4 False positive results

When considering the restricted threshold for recordings over the pulmonary valve, the number of false positive results was higher in the LVEF 41–50% range (47 [43%] of 109) than in those with a normal LVEF of 50–70% (215 [26·2%] of 820, p=0·01) (table 2.7). Figure 2.7 shows the distribution of AI-ECG results.

Group	LVEF 41-50%	LVEF 50-70%	Total
Positive AI-ECG	47 (43.1%)	215 (26.2%)	262
Negative AI-ECG	62 (56.9%)	605 (73.8%)	667
Total	109	820	929

Table 2.7 Comparison of false positive rates in LVEF 41-50% vs. LVEF 50-70% group.

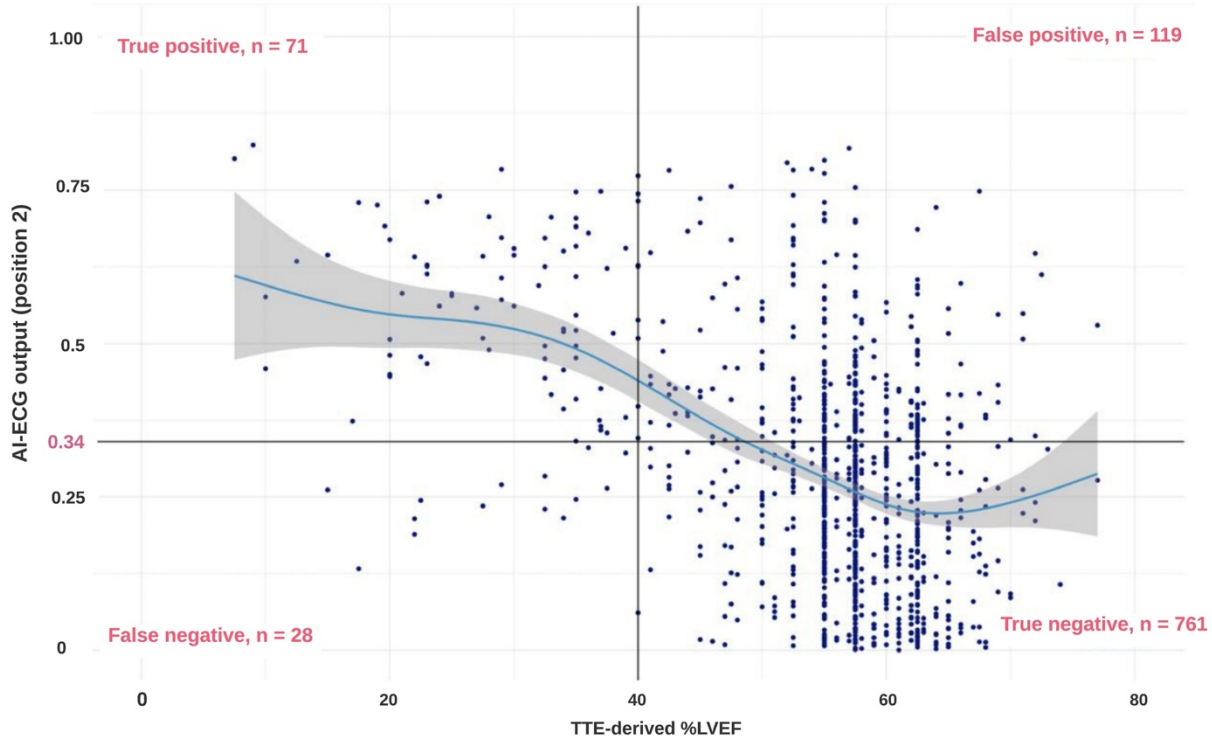


Figure 2.7 Distribution of AI-ECG results.

Plot shows raw AI-ECG outputs (number from 0-1) plotted against %LVEF. AI-ECG results are from pulmonary recording position (position 2). The threshold of 0.34 represents the threshold maximising sensitivity and specificity with rule Se>81, Sp>67; or Se>81, maximising Sp.

2.4.5 AI-ECG bias

The performance of the AI-ECG algorithm stratified by sex and age (18–69 years *vs* ≥70 years), and by non-White ethnicity, is presented in figure 4. Compared with the overall diagnostic OR for the whole study population, no significant differences were seen.

Group	Sensitivity	Specificity	DOR (95% CI)
Female 18-69	87.5 (14/16)	84.0 (242/288)	22.8 (6.2-83.2)
Female 70+	73.3 (11/15)	84.0 (136/162)	10.5 (3.3-33.1)
Male 18-69	88.9 (32/36)	67.9 (188/277)	13.1 (4.9-34.8)
Male 70+	80.6 (25/31)	75.0 (114/152)	10.3 (4.1-25.8)
Non-White	68.0 (34/50)	84.9 (304/358)	10.9 (5.7-21.0)
Overall	**71.7 (84/99)**	**86.4 (612/880)**	**12.38 (7.49-20.45)**

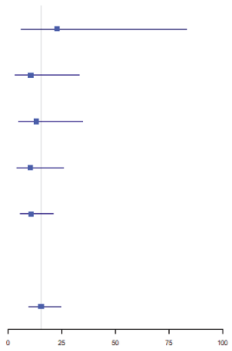


Figure 2.8 AI-ECG bias by age, sex, and non-White ethnicity.

Data are tabulated across a range stratified into age bands by sex, and by non-White ethnicity, using results from the pulmonary position at the threshold maximising the sum of sensitivity and specificity. The diagnostic OR *and associated 95% CI is shown for each group and for the overall study sample. For sensitivity, data presented in brackets represents the number of patients in each group who had a positive result, with the denominator the number of patients with LVEF≤40%. For specificity, this is the number of patients with a negative result, with the denominator patients with LVEF>40%. OR=odds ratio. LVEF=left ventricular ejection fraction. Adapted from Bachtiger et al., 2022.*

2.4.6 Reproducibility of AI-ECG across clinical and patient operators

I found good inter-operator reproducibility between and within and between patient and clinician AI-ECG results (table 2.8). Figure 2.9 shows the distribution of all AI-ECG values from clinician recordings and patient self-recordings. The Bland Altman plot in figure 2.10 displays a mean error of -0.01 (-0.21-0.18), with no obvious systematic error. There was good intra-operator reproducibility for both patients (ICC = 0.82, 95% CI) and clinician (ICC 0.8, 95% CI).

Operator	**Intraclass Correlation Coefficient**	**95% CI**
Patient	0.82	0.72, 0.89
Clinician	0.89	0.83, 0.93
Patient vs. Clinician	0.84	0.77, 0.89

Table 2.8 Intraclass correlation coefficients for AI-ECG recordings within and between patients (n = 50) and clinician.

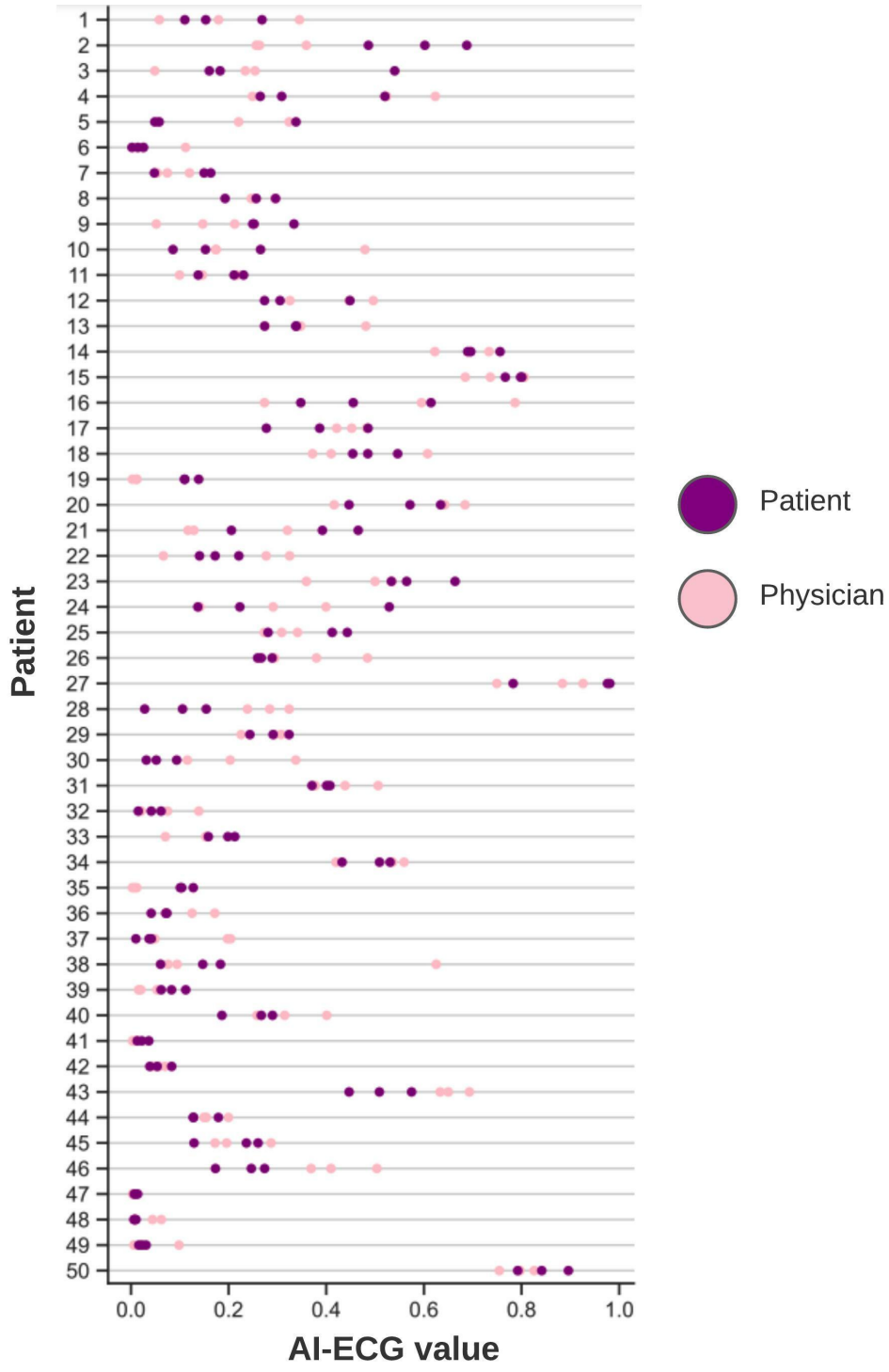


Figure 2.9 AI-ECG reproducibility with three recordings each for physician and patient

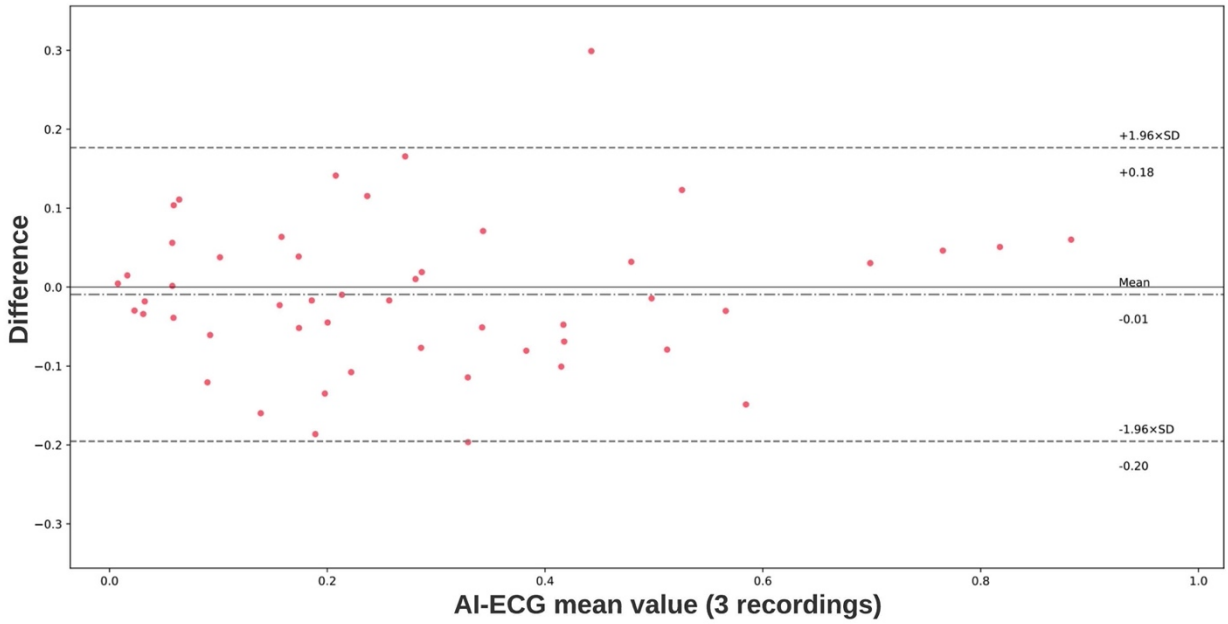


Figure 2.10 Bland Altman plot for difference in clinician recorded vs. patient self-recorded AI-ECG.

Using the mean value of the three recordings taken by clinician and patient. Clinician recording taken as gold standard.

2.5 Discussion

2.5.1 Summary of Results

This independent observational, prospective, multicentre study showed for the first time the performance of AI-ECG for detecting LVEF≤40% using only single-lead ECGs recorded during an ECG-enabled stethoscope examination. My study of 1,050 patients undergoing TTE found that a single-best position, two best positions combined, and an exploratory logistic regression model attained AUCs of 0·85, 0·85, and 0·91, respectively. As well as being accurate, results generation was reliable, operator-independent, and AI-ECG proved unbiased across age, sex, and ethnicity. These results suggest that the (smart) stethoscope examination, a universal component of the clinician–patient interaction, can be used as a screening tool for LVEF≤40% by combining ECG recording and AI at the point of care. Furthermore, the ease with which the examination can be performed extends to the opportunity for patient self-administered testing.

2.5.2 AI-ECG Stethoscope for HFrEF Screening

From a public health perspective, combining AI with an ECG-enabled stethoscope examination for low-cost screening for reduced LVEF fulfils key criteria for a screening programme; including the underlying condition being a public health priority,[123] involving a latent or early symptomatic phase,[124] and for which evidence-based therapies are available. Further evaluation of the potential cost-effectiveness and effects on patient outcomes will be needed, especially in conjunction with established screening tests for HF, such as natriuretic peptide blood tests, which could further improve predictive capability. Easily available clinical tabular variables, such as age, sex, blood pressure, or the presence of comorbid illness, might further improve the model output and aid in the

identification of systolic dysfunction. In the clinic, pre-test probability is likely to be greatest among those with HFrEF symptoms (eg, breathlessness, ankle oedema, and fatigue). However, these are non-specific and can result in a host of other acute or chronic conditions being investigated first. Here, where a stethoscope examination would always be indicated, delays to diagnosis might be avoided by flagging the possibility of HFrEF early. Given the substantial expense of echocardiography and the NHS-wide shortage of echocardiographers[40], the high negative predictive value (97%) could also enable resource prioritisation.

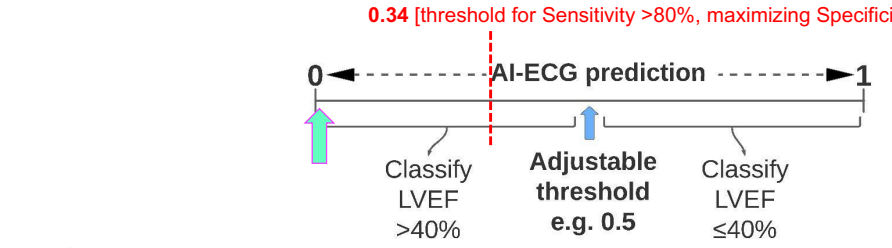
2.5.3 Interpretability of AI-ECG

Successful system-wide adoption of any AI tool will require trust from patients and clinicians, and behavioural change in the latter to both adopt and follow recommendations from algorithms.[21,22] The unknown, 'black box' nature of the neural network means that the specific ECG features that determine individuals' classification of LVEF status are not obvious, although it likely draws on established pathological effects of reduction in LVEF on the ECG.[125-127]

Previous screening programmes and surveys of patients with suspected HF suggest that between 1 and 10% of patients with reduced ejection fraction will have an ostensibly "normal" ECG i.e. nearly all patients with HFrEF will have some ECG abnormalities.[128] One such study analysed data from 6,664 participants who were free of cardiovascular disease at baseline. Authors used a competing risks analysis to compare the association of several baseline ECG predictors with HFrEF and HFpEF detected during a median follow-up of 12.1 years. In a multivariable adjusted model, prolonged QRS duration, left-axis deviation, right-axis deviation, delayed intrinsicoid deflection, prolonged QT interval, abnormal QRS-T axis, ST/T-wave abnormalities, left ventricular hypertrophy, and left

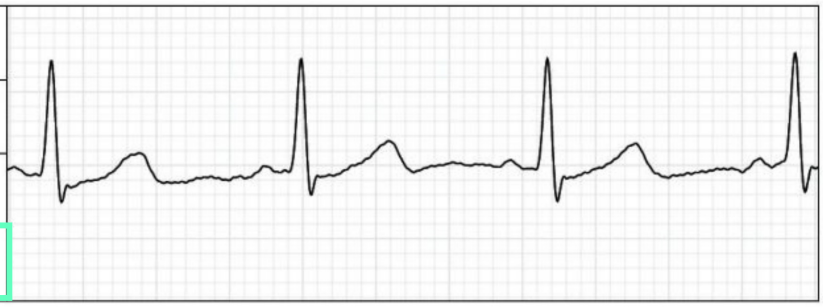
bundle-branch block were associated with HFrEF. In contrast, higher resting heart rate, abnormal P-wave axis, and abnormal QRS-T axis were associated with HFpEF.[126]

These ECG features are all readily detectable by the human eye – but weighing up the ECG in its full richness of temporal and spatial features to accurately predict presence of reduced ejection fraction is beyond human skill. My study therefore highlights use of AI to interrogate a widely available digital biomarker and glean previously inaccessible, clinically useful insight. A similar example might be the application of neural networks for predicting coronary artery disease from retinal imaging.[129] To further illustrate this point, I include examples of single-lead ECGs at the pulmonary position classified by the optimised threshold as true positive, false positive, and false negative, for visual inspection in figure 2.9.

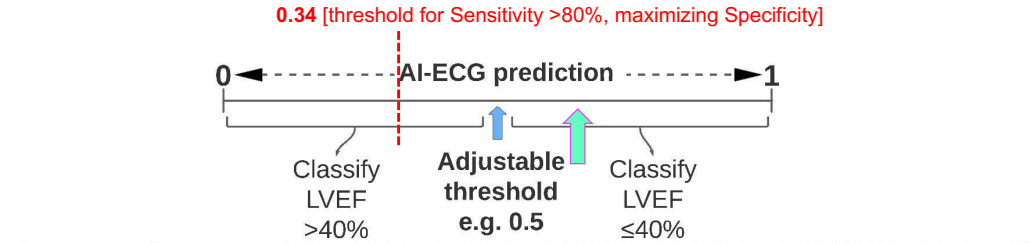


A.

Age	42
Sex	Female
%LVEF	55
AI-ECG	0.02

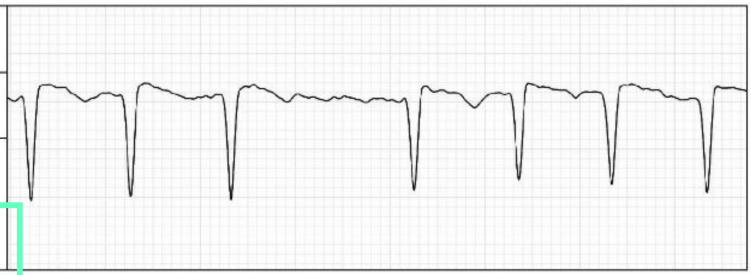


True Negative



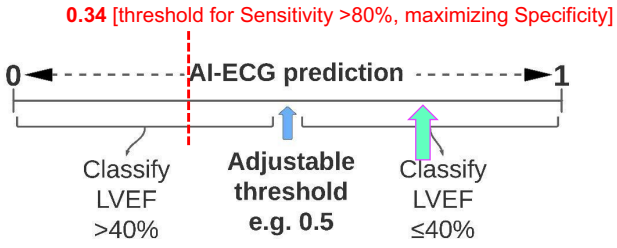
B.

Age	60
Sex	Male
%LVEF	27.5
AI-ECG	0.64

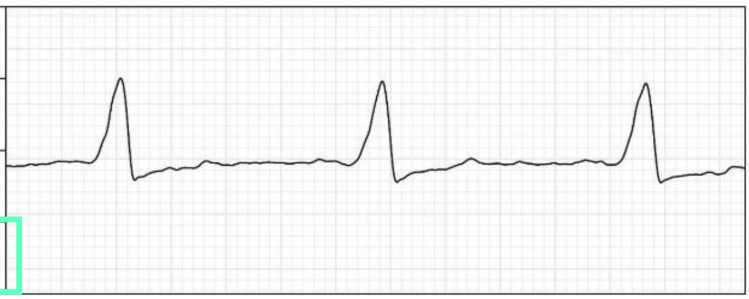


True Positive

C.

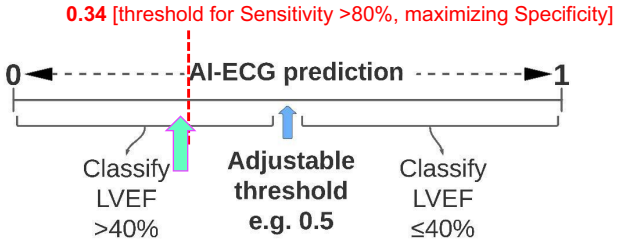


Age	81
Sex	Male
%LVEF	35
AI-ECG	0.75

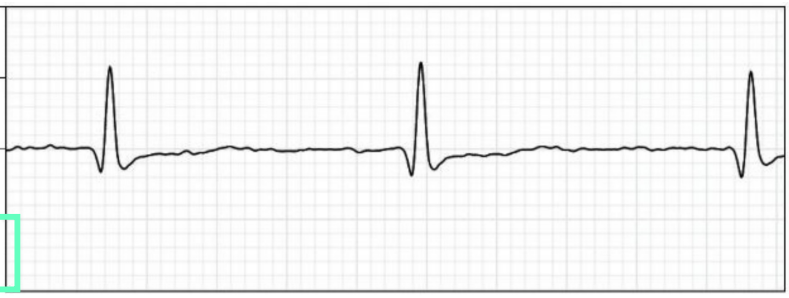


True Positive

D.



Age	81
Sex	Male
%LVEF	41
AI-ECG	0.33



True Negative

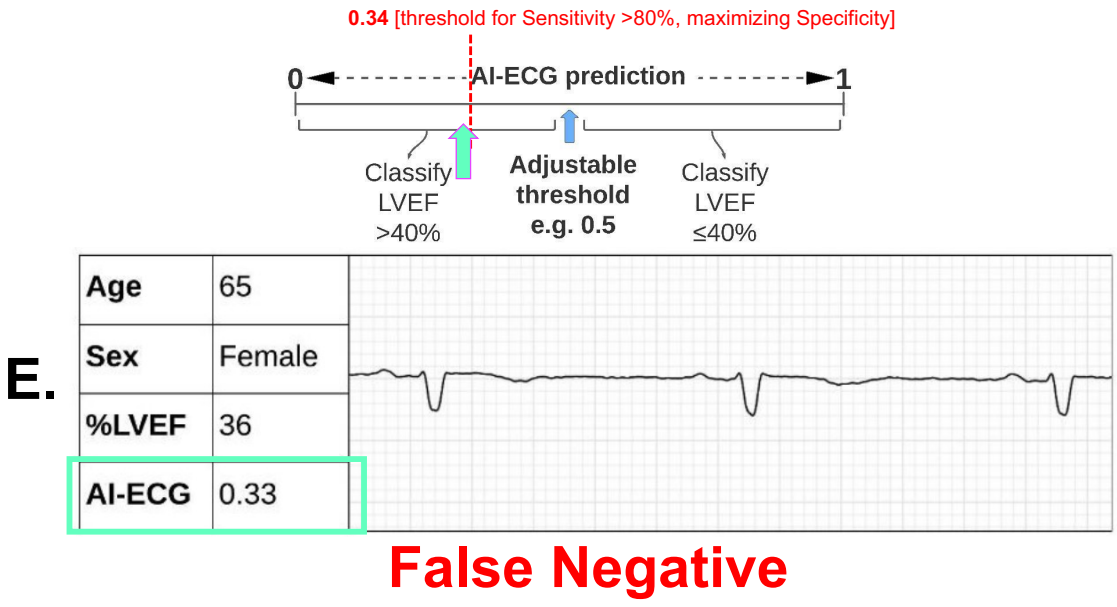


Figure 2.11 Examples of single-lead ECG recordings and associated AI-ECG predictions.

Single-lead ECG recordings from position 2 (pulmonary valve, angled). Recordings were categorised as true positive, false positive, true negative, and false negative based on the raw AI-ECG output (between 0 and 1), relative to an optimum classification threshold of 0.341 (achieving minimum sensitivity and specificity of 81% and 67%, respectively). ECGs displayed at recording calibration of 25mm/second, 10mm/mV.

2.5.4 Performance Compared to 12-lead ECG Studies

An external validation study of AI-ECG using 12-lead inputs reported an AUC of 0·82[95], compared with 0·93 for the original internal validation study.[92] Both of these studies used retrospective data. My study, using only prospectively collected single-lead ECG inputs therefore compares favourably. The low positive predictive value should be interpreted in the context of both the selected study population patients attending for ECG and, therefore, more likely to have abnormal ECG features that risk a false positive result and that false positives occurred most frequently in the LVEF 41–50% range. This is within the diagnostic spectrum of HF where, from a clinical perspective, further investigation would be warranted, entailing minimal-risk natriuretic peptide blood test. Therefore, a key challenge will be how to select and define specific thresholds and cutoff points. For example, more specific cutoffs could optimise against false positive rates, but at the expense of lower sensitivity. Population-specific cutoffs might be necessary to optimise test performance for differing demographic and disease profiles, including different underlying disease prevalence.

2.5.5 Workflow for Performing AI-ECG Stethoscope Examination

Defining the unit of examination (one or multiple positions/recordings) will also be important. I have shown that performance was moderately improved when applying AI-ECG at two positions. Considering the four universal positions for stethoscope examination (plus one handheld) for each patient, the study obtained five raw AI-ECG outputs. When exploring just two of these, inputting AI-ECG values (0 to 1) into a logistic regression model showed substantial improvement in performance. Notably, such an approach should not compromise the tool's ease-of-use by requiring an end-user to

take multiple recordings or manually input data into a further model, which may be avoidable through automation and considered user experience design.

Compared with a positive/negative test output for a particular LVEF cut-off, displaying percentage LVEF predicted as a continuous variable might empower individual clinician choice on the appropriate threshold for ordering further investigations on a case-by-case basis; however, as a decision aid for non-specialists, this feature might be less desirable.

2.5.6 Can AI-ECG predict risk of reduced LVEF?

The original internal validation study of AI-ECG (for 12-lead ECGs) identified a four-times increased risk of developing an LVEF<35% or lower in subsequent years if AI-ECG predicted LVEF<35%, but TTE-derived LVEF was above 35%. This finding highlights the possibility that ECG changes might predate deterioration in LVEF detectable by echocardiography. Accordingly, to test this hypothesis, the Health Research Authority has granted ethical approval for those participants in my study who had a "positive" AI-ECG result for LVEF≤40% but TTE finding of LVEF>40% to be followed up longitudinally. "False positives" in the study cohort may need to be reframed in if AI-ECG shows ability to predict future LVEF≤40%, and in doing so could propose a cohort for surveillance.

2.5.7 Patient self-administered AI-ECG screening

My study took a practical and pragmatic approach to inform whether patients could perform a self-examination with an Eko DUO and obtain AI-ECG predictions comparable to those obtained by a clinician. The high intraclass correlation coefficients (ICCs all >0.8) reassure on the important

question of adequate reproducibility, and in doing so highlights affirms the opportunity for patient self-administered examination. The only abundant existing use case for cardiovascular disease screening using DHTs integrating AI involves detection of AF using medical and consumer technologies, both with a regulatory approved AI algorithm (software as a medical device) for over-read and flagging of an irregular heart rhythm. As consumer technologies with single-lead ECG become cheaper and more accessible, AI-ECG for tracking LVEF may become part of a future suite of algorithms for individual cardiovascular disease monitoring.

2.5.8 Strengths of Study

Given substantial concerns and criticisms of validation studies of health-related AI tools,[130-132] my study design has several strengths. First, data were collected prospectively across multiple real-world settings and by many operators. Second, my research team and I were independent of the groups who developed both the AI-ECG and the AI-ECG stethoscope. Third, my study population was unrelated to the training cohort, and our sample's ethnic diversity (41% non-White) is unmatched by previous, retrospective external validations studies of AI-ECG (which were <10% non-White). Fourth, beyond testing the performance of AI-ECG alone, I evaluated a form factor and workflow for frontline clinical delivery that has several advantages over 12-lead ECGs. Namely, the recording of ECGs during auscultation over the single best (pulmonary) position achieved adequate recordings to attempt prediction in at least 93% of patients (versus 87% for handheld position); taking 15 seconds to complete and requiring minimal training. Fifth, use of AI and an AI-ECG stethoscope upgrades a familiar tool already in daily clinical use. This potentially overcomes barriers, such as maintained use by clinicians, previously identified as a challenge for other devices capable of recording single-lead ECGs.[133]

Such a tool could be particularly impactful in the busy primary care setting, given that, among the 80% of patients diagnosed with HF in hospital, 40% have had a recent primary care encounter with symptoms of HF that would have warranted a stethoscope examination. This approach would also be of value in low-income countries with health systems where access to cardiological care and imaging is scarce.[134]

Finally, beyond the scope of this thesis, the dual acquisition of precordial ECG and phonocardiogram (heart sounds) highlights an opportunity to also screen for further priority cardiovascular diseases, such as valvular heart pathology, using AI-enabled phonocardiography.[135] Similarly, the 15 s single-lead ECG offers an opportunity for the detection of atrial fibrillation, either by visual inspection, or also supported by AI.[136,137] Improvements in accuracy for predicting reduced LVEF might be achievable by combined AI analysis of synchronous ECG and PCG waveforms.

2.5.9 Limitations

The results of this study are best interpreted in the context of its limitations. First, the patient cohort is not fully representative of a screening population, where lower prevalence of LVEF≤40% could influence performance characteristics. Here it is particularly important to consider that PPV and NPV are impacted by the underlying prevalence in the population being screened. Considering my study's sample of – beyond attending for echocardiography – largely unselected, sequential participant pool attending, the prevalence of HFrEF was 10%. At best, my study observed a PPV of 35% i.e. taking 100 patients with positive tests in this population, 35 of them will truly have HFrEF. Consider a different setting, such as the primary care interface, the prevalence of HFrEF will be much lower. Here, the real-world use of an ECG-enabled stethoscope could (should) be as routine as that of a traditional stethoscope. Assuming a 2% underlying prevalence of HF (not further defined by subtype)

prevalence, a hypothetical prediction that 5% of patients examined with a stethoscope in primary care have HFrEF, despite an impressive sensitivity and specificity of 92% and 80% respectively, PPV would drop to 19.5%.; while NPV would increase further to 99.5% Perception of a one in five "hit rate" for referral to echocardiography should be tempered by the transformative impact of a HFrEF diagnosis and the opportunities unlocked by prompt commencement of evidence-based therapies. Ultimately the artificial scarcity of echocardiography in the NHS may require stricter referral criteria that increases PPV and justifies any prioritisation of patients with positive AI-ECG. Further investigation of real-world application of such a tool may also address the current paucity of data describing prevalence of asymptomatic disease.

Second, without comprehensive access to all participant's electronic health records to determine any previously normal LVEF, I am unable to precisely characterise how many participants were flagged as positive by AI-ECG as part of an index diagnosis of HFrEF. Third, there is established inter-operator variability in measurement of LVEF from echocardiography, giving rise to the possibility that some participants close to the LVEF 40% borderline were misclassified. Further studies with a higher number of operators across the wider clinical workforce will be required to determine if the device really is universally easy to use.

2.6 Chapter Conclusion

This chapter described my study that found AI-ECG could identify patients with reduced LVEF (≤40%) from single-lead ECG inputs. Through use of an AI-ECG stethoscope, I highlight an AI algorithm embedded in a familiar clinical tool that fits into routine and universal clinical workflows. Given the frequent clinical encounters of undiagnosed patients before index hospital admission for HF, the stethoscope examination has the potential to be a point-of-care screening opportunity, and through further AI algorithms, to become a tool for comprehensive detection of cardiovascular disease. The feasibility of patient self-examination with this technology highlights and opportunity for expanded access through decentralisation – meeting the screening needs of patients wherever they are.

3 Survival and Health Economic Outcomes in Heart Failure by Place of Index Diagnosis: A Propensity-Matched Analysis

This chapter presents results from a study of HF diagnosis across North West London's 2.5 million patient population. Specifically, taking data from 2015-2020, I present the contemporary reality for routes to index HF diagnosis – community vs. hospital pathways – and use a propensity score matched population to highlight worse early clinical outcomes and increased long-term costs in patients first diagnosed through hospital admission. These findings are framed against the substantial national challenge of HF diagnosis and poses an early health economic model for the impact of a point-of-care screening technology: AI-ECG within an ECG-enabled stethoscope.

3.1 Abstract

3.1.1 Background

This observational cohort study aimed to quantify the clinical and health economic impacts of index HF diagnosis made through hospital admission versus community settings. This is framed against the burden of undiagnosed HF, and collectively informs a health economic model for the potential impact of a point-of-care AI screening tool.

3.1.2 Methods

I examined records from 34,208 patients receiving an index diagnosis of HF between January 2015 and December 2020 across North West London (NWL), UK. A propensity-score-matched (PSM) cohort was identified to adjust for differences in socioeconomic status, cardiovascular risk factors, and pre-diagnosis health resource utilisation cost. Outcomes were stratified by two pathways to index HF diagnosis: a 'hospital pathway' was defined by diagnosis following hospital admission; and a 'community pathway' by diagnosis via a primary care physician or outpatient services. The primary clinical and health economic endpoints were all-cause mortality and cost-consequence differential, respectively. An open-source data repository of predicted vs. expected HF cases was used to measure the burden of underdiagnosis in NWL and nationally. Collectively, these results were used to construct a health economic model for the potential impact of deploying a point-of-care AI screening tool for HF in primary care.

3.1.3 Results

In the overall cohort, 23,273 (68.2%) patients were diagnosed via hospital pathway, compared with 10,885 (31.8%) via community pathway. The PSM cohort consisted of 17,174 patients (8,582 in each group). The ratio of deaths per person-months at 24 months comparing community vs. hospital diagnosis was 0.780 (95% CI 0.722 - 0.841, $p<0.0001$), changing to 0.960 (0.905 -1.020, $p = 0.18$) by 72 months. Diagnosis via hospital pathway incurred an overall extra longitudinal cost of £2,485 per patient. In the context of nearly 50% of HF in NWL being undiagnosed, a point of care screening tool using AI-ECG could offer the sector conservative net savings close to £1.5 million.

3.1.4 Conclusion

Index diagnosis of HF through hospital admission continues to dominate and is associated with a significantly greater short-term risk of mortality and substantially increased long-term costs than if first diagnosed in the community. My study highlights that efforts to increase community diagnosis may provide opportunities for improved clinical and health-economic outcomes, and that in the context of substantial under-diagnosis, an AI-based intervention to increase detection through community pathways may carry a strong health economic justification.

3.2 Introduction

Heart failure (HF) affects 5% of the population aged 75 years or older, with 60,000 new cases annually in the United Kingdom.[27,138] There are multiple evidence-based therapies that improve survival and quality of life [139,140], with early dose optimisation associated with better clinical and health economic outcomes.[141,142]

Detection of HF via primary care is a priority in the United Kingdom NHS Long Term Plan, recognising that between 2011-2013, 80% of all new (index) diagnoses of HF were made via hospital admission.[27] Hospitalisation with chronic HF is associated with increased hazard of death, repeat and prolonged hospitalisation[143], and the unit cost of such a hospital admission can exceed £10,000.[144] Despite several initiatives to improve community-based detection of HF[145], only 4% of eligible patients complete the diagnostic pathway recommended by the National Institute for Health and Care Excellence (NICE) to time and target – with overall minimal change in survival over the last decade.[146] Alarmingly, the gap in underdiagnosis of major cardiovascular risk factors remains persistent and extends such that a substantial portion of HF in the community remains subsequently undetected.[147] Part of the challenge is a lack of screening tools that are quick, reliable, and accurate. Even in the presence of a tool that fulfils these criteria, those in charge of healthcare budgets require a robust health economic justification for any capital outlay on a novel instrument. This requires an up-to-date understanding of the health economic burden of HF to help model any potential cost savings.

Although hospitalisation around the time of HF diagnosis and time-to-diagnosis may adversely affect survival,[148] the underlying assumption that diagnosis through community pathways confers clinical and health economic benefits has not been tested. Testing this assumption poses substantial methodologic challenges. Importantly, given the continually changing healthcare landscape, contemporary estimates of both survival and health economic burdens based on place-of-diagnosis

are essential for shaping health policy interventions. This is now possible through linkage of contemporary, granular, real-world primary and secondary care clinical and cost data. The objective of this study was therefore to measure the combined prognostic and health economic impacts of different routes to index HF diagnosis, and to combine these observations to inform a health economic model for a novel point-of-care screening tool intended to increase detection of HF in primary care.

3.3 Methods

3.3.1 Study Design and Data sources

I used a cohort study design following STROBE[149] and RECORD[150] checklists for reporting observational research in routinely collected health data. The study required interrogation of the Whole Systems Integrated Care dataset within the Discover-NOW Trusted Research Environment. This pools de-identified, contemporary, linked primary and secondary care electronic patient records from over 2.5 million patients in North West London (NWL). In addition to comprehensive demographic and clinical data, the dataset also captures health service utilisation and associated cost.[151]

To describe the burden of undiagnosed HF – or 'detection gap' – across all eight NHS Clinical Commissioning Groups (CCGs) in NWL, I also examined and extracted data from the open-source Heart Failure Dashboard developed by Imperial College Health Partners (ICHP) in collaboration with Novartis. The detection gap is defined as the difference between predicted number of patients with HF in the population and the observed number of HF patients. To estimate the predicted population, the dashboard applies an established HF prevalence model for small populations[152], in this case applying the model to Public Health England's Public Health Profiles. The observed population was defined as all registered patients with HF collected through routine data collection as part of the NHS Quality and Outcomes Framework (QOF 2019-20).[153]

3.3.2 Inclusion and exclusion criteria

I included patients aged over 18 years old, with HF diagnosed between 1st January 2015 and 31st December 2020 (Figure 1). Patients aged 18 or under were excluded; as well as those diagnosed with HF before 1st January 2015; and those who left the North West London area during the study

period. Specific codes for HF were considered, in accordance with previous literature[6,7,9] and by expert clinical consensus. For primary care data, HF labels were selected from an inclusive list of established read codes (table 3.1). This was similarly performed using ICD-10 codes for hospital data (table 3.2). The clinical activity of the cohort was examined by identifying the first coded diagnosis of HF in the study period (index date) for each unique patient, and then mapping the health care resource utilisation of each unique patient in the course of the study period. Two all-encompassing strata for routes to diagnosis were considered:

1. The **'community pathway'** reflected HF diagnoses first coded within *primary care records*. This included patients diagnosed through specialist outpatient settings via primary care referral. We considered this the preferred route to HF diagnosis, in line with National Institute for Health and Care Excellence (NICE) guidance.[22]

2. The **'hospital pathway'** reflected HF diagnoses made via an *inpatient hospital admission*. Such admissions were either non-elective (acute/emergency) or elective (e.g. planned procedure). For both, we included those patients where, upon discharge, HF diagnostic codes were listed as either a primary or secondary diagnosis. This was informed by clinical consensus, since inconsistencies in clinical coding results in predominantly non-hierarchical coding of primary and secondary diagnoses.

Read Code	Read Code Preferred Term	Read Code	Read Code Preferred Term
G58..	Heart failure	G583.	Heart failure with normal ejection fraction
662g.	New York Heart Association classification - class II	G580.	Congestive heart failure
662h.	New York Heart Association classification - class III	SP111	Cardiac insufficiency as a complication of care
G5yy9	Left ventricular systolic dysfunction	G581.	Left ventricular failure
G580.	Congestive heart failure	1O1..	Heart failure confirmed
585f.	Echocardiogram shows left ventricular systolic dysfunction	G1yz1	Rheumatic left ventricular failure
662f.	New York Heart Association classification - class I	G232.	Hypertensive heart & renal dis with (congestive) heart failure
G581.	Left ventricular failure	G234.	Hypertension heart & renal failure
G58z.	Heart failure NOS	G580.	Congestive cardiac failure
G5802	Decompensated cardiac failure	G580.	Right heart failure
G5yyC	Diastolic dysfunction	G580.	Right ventricular failure
G580.	Congestive heart failure	G580.	Biventricular failure
G5yyA	Left ventricular diastolic dysfunction	G581.	Asthma - cardiac
585g.	Echocardiogram shows left ventricular diastolic dysfunction	G581.	Impaired left ventricular function
G5801	Chronic congestive heart failure	G58z.	Cardiac failure NOS
G5800	Acute congestive heart failure	14A6.	H/O: heart failure
G583.	Heart failure with normal ejection fraction	14AM.	H/O: Heart failure in last year
G58..	Cardiac failure	SP111	Heart failure as a complication of care
G581.	Left ventricular failure	662i.	New York Heart Association classification - class IV
G582.	Acute heart failure	G58z.	Weak heart
G580.	Congestive heart failure	G5y4z	Post cardiac operation heart failure NOS
G58..	Heart failure	1J60.	suspected heart failure
G584.	Right ventricular failure	388D.	new york heart assoc classification heart failure symptoms
G581.	Left ventricular failure	G210.	malignant hypertensive heart disease
G5810	Acute left ventricular failure	G2101	malignant hypertensive heart disease with ccf
G5804	Congestive heart failure due to valvular disease	G2111	benign hypertensive heart disease with ccf
21264	Heart failure resolved	G21z1	hypertensive heart disease nos with ccf
G581.	Left ventricular failure	G230.	malignant hypertensive heart and renal disease
G58z.	Heart failure NOS	G41z.	chronic cor pulmonale
G5yyB	Right ventricular diastolic dysfunction	G5540	congestive cardiomyopathy
G5803	Compensated cardiac failure	G5540	congestive obstructive cardiomyopathy

Table 3.1 Read codes from primary care records included for patients considered to be diagnosed with HF.

ICD-10 codes	
I500 - Congestive heart failure	I420 - Dilated cardiomyopathy
I501 - Left ventricular failure	I255 - Ischaemic cardiomyopathy
I509 - Heart failure, unspecified	I110 - Hypertensive heart disease with (congestive) heart failure
I429 - Cardiomyopathy, unspecified	-

Table 3.2 ICD-10 codes for possible HF labels in hospital electronic health records.

3.3.3 Cost data within Discover

Primary care costs described a combination of General Medical Services, Personal Medical services, and Alternative Personal Medical Services contracts commissioned by NHS England and locally commissioned Clinical Commissioning Group (CCG) schemes such as Local Improvement Schemes, Local Enhanced Services, and Out of Hours Services. These costs reflected the actual outturn costs for historic years.

GP practice level costs were apportioned across age groups based on historic analysis of appointment utilisation and then to patients, based on the number of recorded daily contacts that patients had with the practice. The cost allocation assumes that all patient contacts for the specified age group consume the same resource so all contacts will have the same unit price. This reflects the way contracts are commissioned by NHS England and locally-commissioned schemes for GPs and hence recorded for charging purposes.

Hospital costs are based on actual activity and costs (i.e. primarily cost per case) as reported by NHS Trust-issued patient-level service-level agreement monitoring reports. Some contractual adjustments e.g. Emergency Threshold adjustments, re-admission and other contractual penalties were applied retrospectively at the patient level.

3.3.4 Patient characteristics

Age, sex, and ethnicity were extracted for each patient. Comorbidities of interest identified for inclusion included chronic obstructive pulmonary disease (COPD), atrial fibrillation (AF), chronic kidney disease (CKD), ischaemic heart disease (IHD), stroke, type 2 diabetes mellitus (T2DM) and hypertension.

3.3.5 Endpoints

The primary outcome was all-cause mortality. Secondary outcome focused on the cost consequence associated with diagnosis of HF, again by hospital vs. community pathway to diagnosis. Additionally, a HF detection gap was modelled in order to frame and potential cost-savings from a tool such as AI-ECG applied using an ECG-enabled stethoscope.

3.3.6 Data extraction, Propensity Score Matching, and Statistical analysis

In collaboration with fellow researchers at Imperial College Health Partners, I interrogated the Discover dataset through the Discover-NOW Health Data Research Hub for Real World Evidence. For the overall cohort and to facilitate comparisons between patients within each pre-specified strata, continuous variables are expressed as mean ± SD and categorical variables as percentages. I used chi-squared tests to examine differences in baseline clinical characteristics between patient pathways.

Time from diagnosis-to-death was captured for patients who died after the index date. Patient survival curves for mortality were constructed according to the Kaplan-Meier method for each pathway and compared by Fleming-Harrington weighted log-rank test, allowing for early, middle and late

differences through the class of weights, proven to be more efficient when the proportional hazards assumption does not hold.[154]

I *a priori* assumed that there would be important differences and therefore potential confounders between the main analytic groups. I chose to use a propensity score matched (PSM) approach since this would allow independent characterisation of how route to diagnosis impacts the two main outcomes of interest (survival and cost-consequence), while adjusting for confounders between the community and hospital groups. Additionally, this method avoids the less easily interpretable and presentable approach using different multivariate models. Lastly, a PSM approach makes no assumption about the relationships between covariates and an outcome of interest.

Therefore, a propensity score was calculated using a logistic regression model to adjust for baseline differences in patient characteristics. This incorporated available predisposing covariates for HF, including age, male gender, ethnicity, index of multiple deprivation (IMD) rank, hypertension, AF, IHD, CKD and T2DM. Additionally, differences in cost (health service utilisation) before HF diagnosis was also included in the PSM model. This served as a holistic means of accounting for differences between the two groups that would not be captured by individual codes, for example, high-cost, complex conditions such as cancer. I performed a 1:1 comparison between nearest matching neighbours, using a caliper width of 0.2, aligning with similar previous studies.[155,156]

Cost-consequence analysis is highlighted by the UK Government[157] as a preferred tool in the economic evaluation of clinical pathways, highlighting advantages including the output of a simple broken-down (disaggregated) summary of costs and effects that allows policymakers to choose the combination of costs and effects that are most relevant to their context, and apply their own weighting to the effects. For cost-consequence analysis in my study, healthcare utilisation cost was extracted from the index HF diagnosis date to the end of the study period, and included primary care contacts, outpatient appointments, elective and non-elective admissions, and, where relevant, non-elective

readmissions at 30 days. Patient level costs refer to the indicative spend calculated separately for each patient for each healthcare sector. A P-value of < 0.05 was considered statistically significant. Analyses were performed using R Studio (version 1.4.1717).

To represent undiagnosed HF burden, data from the HF Dashboard were converted into a heat map of NHS Clinical Commissioning Groups (CCGs), highlighting both the predicted percentage of undiagnosed HF cases, as well as the predicted absolute number of undiagnosed patients per CCG. The overall trend across all CCGs in England is displayed for comparison.

Lastly, informed by the cost consequence analysis and HF detection gap, I constructed a health economic model to inform on the potential impact of an AI-ECG tool hypothesised to increase the rates of HF detection through community pathways. Given that this model was derived from the core results, I present this as part of the chapter Discussion, rather than a standalone results item.

3.3.7 Ethical approval

My collaborators among the Discover-NOW team at ICHP have secured Health Research Authority approval until 2023 to use the Discover Research Platform for research purposes of studies submitted to the NWL Data Access Committee. Favourable ethical opinion was secured from the NHS Health Research Authority in October 2018. The REC reference is 18/WM/0323 and the IRAS project ID is 253449.

3.4 Results

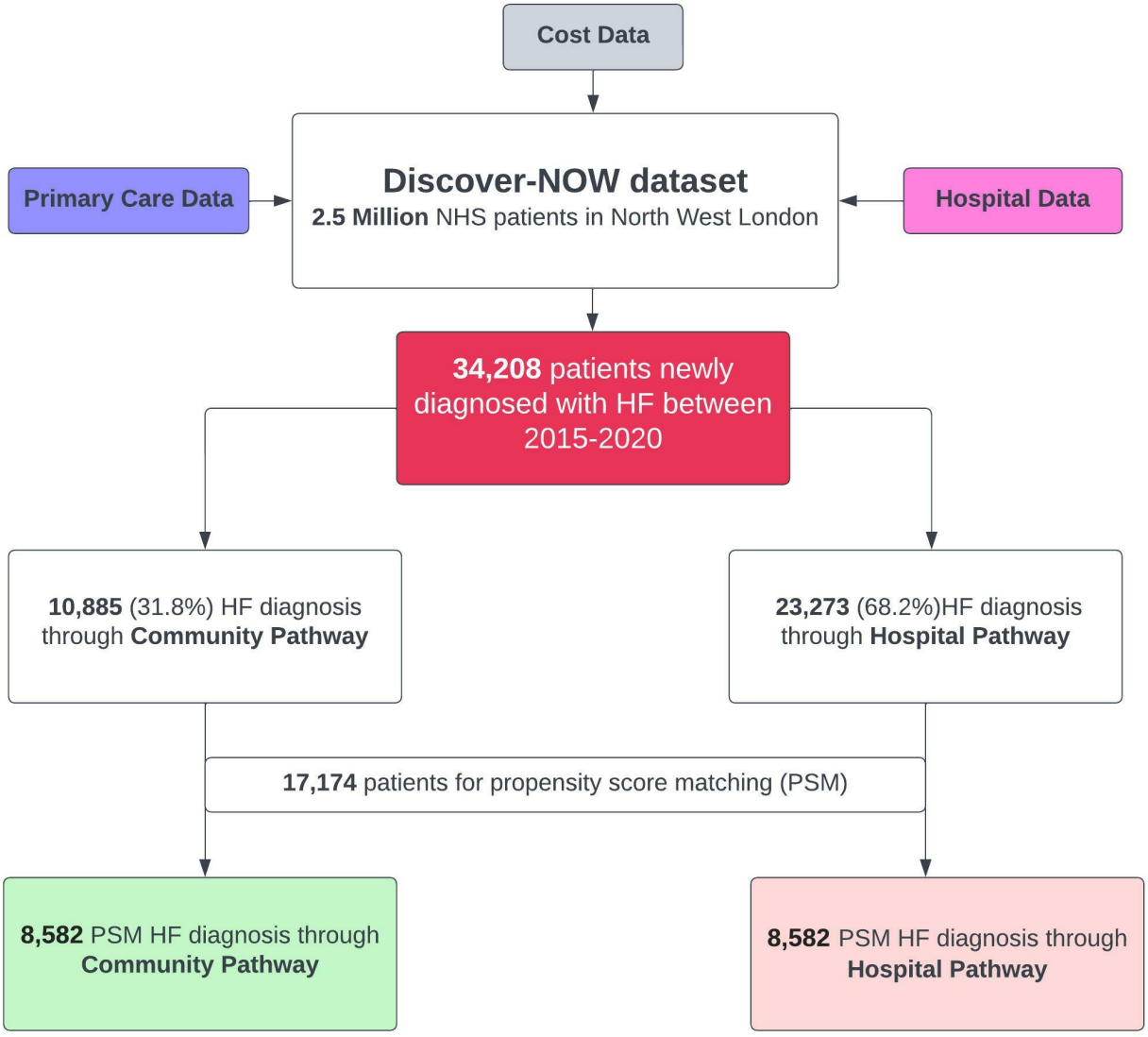


Figure 3.1 Flow diagram for selection of patients with HF for inclusion in the PSM analysis.

3.4.1 Patient Characteristics

Between 1st January 2015 and 31st December 2020, 34,208 patients received a diagnosis of HF. 23,323 (68.2%) had this first recorded during a hospital admission (hospital pathway), and 10,885 (31.8%) in primary care (community pathway). Patient characteristics are summarised in (table 3.2). The cohort diagnosed through hospital admission was older, had a higher representation of male sex, and were more deprived. 31,062 (91.3%) of patients had at least one comorbidity at the time of HF diagnosis. Table 3.3 lists the most frequent HF-associated diagnostic codes in the community pathway; similarly, tables 3.4 shows this for the hospital pathway. In the hospital pathway, 4,686 (20.1%) were recorded as having HF as their primary diagnosis. The remainder recorded HF as a secondary diagnosis (table 3.5). 2,738 (25.2%) of patients in the community pathway had at least one HF symptom recorded in primary care prior to the index date, compared to 4,975 (21.5%) of patients in the hospital pathway ($p<0.05$) (table 3.6).

Propensity Score Matched Cohorts

Prior to matching, the cohorts were unbalanced across most variables. Following PSM, the hospital and community pathway cohorts were well-matched across all variables of interest. The cohort consisted of 17,174 patients (8,582 in each group) (table 3.2).

	Total population (n = 34,208)			Propensity-matched cohort (n = 17,212)		
	Community	**Hospital**	**p**	**Community**	**Hospital**	**p**
Total	10,885 (31.8%)	**23,323 (68.2%)**	**-**	8,582	8,582	
Age (SD)	72.26 ± 13.50	73.84 ± (13.33)	<0.0001	73.53± 12.18	73.61 ± 12.62	**0.66**
Male gender (%)	5,007 (0.46)	11,895 (0.51)	<0.0001	4644 (54.1)	4626 (53.9)	**0.79**
IMD (SD)	5.02 ± 2.30	4.83 ± 2.31	<0.0001	6.63 ± 2.28	6.68 ± 2.31	**0.13**
Ethnicity	-	-	<0.0001	-	-	**0.92**
Asian or Asian British (%)	2994 (27.5)	5913 (25.4)	-	2591 (30.2)	2567 (29.9)	**-**
Black or Black British (%)	948 (8.7)	1961 (8.4)	-	787 (9.2)	807 (9.4)	**-**
Mixed (%)	181 (1.7)	433 (1.9)	-	148 (1.7)	136 (1.6)	**-**
Other ethnic groups (%)	456 (4.2)	912 (3.9)	-	363 (4.2)	367 (4.3)	**-**
White (%)	5,539 (50.9)	12,000 (51.5)	-	4693 (54.7)	4705 (54.8)	**-**
Unknown ethnicity (%)	767 (7.1)	2104 (9.0)	-	-	-	**-**
COPD (%)	1633 (15.0)	4234 (18.2)	< 0.0001	1470 (17.1)	1440 (16.8)	**0.56**
AF (%)	3592 (33.0)	7763 (33.3)	.601851	2498 (29.1)	2571 (30.0)	**0.23**
CKD (%)	2939 (27.0)	6822 (29.3)	<0.0001	2338 (27.2)	2323 (27.1)	**0.81**
IHD (%)	4027 (37.0)	8704 (37.3)	.564231	3862 (45.0)	3935 (45.9)	**0.27**
Stroke (%)	1,013 (9.3)	2,733 (11.7)	<0.0001	899 (10.5)	889 (10.4)	**0.82**
Ventricular arrhythmia (%)	130 (1.2)	281 (1.2)	.933738	104 (1.2)	105 (1.2)	**1.00**
T2DM (%)	3592 (33.0)	8233 (35.3)	<0.0001	3099 (36.1)	3040 (35.4)	**0.36**
Hypertension (%)	7293 (67.0)	15055 (64.6)	<0.0001	7001 (81.6)	6994 (81.5)	**0.91**

Figure 3.2 Demographics and comorbidities of HF population before and after propensity score matching.

ReadCode Term	Number of patients	% of patients
Heart failure	3,487	32%
Congestive heart failure	996	9%
Diastolic dysfunction	679	6%
Left ventricular failure	607	6%
Left ventricular systolic dysfunction	461	4%
Left ventricul systol dysfunc	452	4%
Left ventric diastolic dysfunc	359	3%
Echocardiogram shows left ventricular systolic dys	333	3%
Suspected heart failure	323	3%
Echocardiogram shows LVSDF	301	3%
Heart failure NOS	278	3%
Echocardiogram shows left ventricular diastolic dy	227	2%
Echocardiogram shows LVDDF	220	2%
Biventricular failure	178	2%
Chronic congestive heart failure	151	1%
Heart failure with normal ejection fraction	149	1%
Left ventricular diastolic dysfunction	145	1%
Heart failure annual review	110	1%
Heart failure review completed	103	1%
New York Heart Association classification - class	88	1%
Other	1,238	11%
Grand Total	**10,885**	**100%**

Table 3.3 Distribution of ReadCodes for patients receiving index HF diagnosis through community pathway.

ICD10 diagnosis	Number of patients	% of patients
I500 - Congestive heart failure	2911	62%
I501 - Left ventricular failure	1180	25%
I509 - Heart failure, unspecified	464	10%
I420 - Dilated cardiomyopathy	60	1%
I255 - Ischaemic cardiomyopathy	26	1%
I110 - Hypertensive heart disease with (congestive) heart failure	23	0%
I429 - Cardiomyopathy, unspecified	22	0%
Grand total with HF code listed as primary diagnosis	**4,686 (20.1% total)**	**100%**

Table 3.4 Distribution of ICD-10 codes for patients receiving index HF diagnosis through hospital pathway.

ICD-10 diagnosis	Number of patients	% of patients
J181 - Lobar pneumonia, unspecified	1261	7%
I214 - Acute subendocardial myocardial infarction	759	4%
I251 - Atherosclerotic heart disease	676	4%
J189 - Pneumonia, unspecified	639	3%
I489 - Atrial fibrillation and atrial flutter, unspecified	538	3%
A419 - Sepsis, unspecified	420	2%
N390 - Urinary tract infection, site not specified	335	2%
J440 - Chronic obstructive pulmonary disease with acute lower respiratory infection	310	2%
I249 - Acute ischaemic heart disease, unspecified	301	2%
R074 - Chest pain, unspecified	280	1%
J22X - Unspecified acute lower respiratory infection	277	1%
I210 - Acute transmural myocardial infarction of anterior wall	272	1%
N179 - Acute renal failure, unspecified	252	1%
J690 - Pneumonitis due to food and vomit	196	1%
R296 - Tendency to fall, not elsewhere classified	184	1%
I48X - Atrial fibrillation and flutter	177	1%
L031 - Cellulitis of other parts of limb	175	1%
U071 - Emergency use of U07.1	173	1%
I211 - Acute transmural myocardial infarction of inferior wall	172	1%
R060 - Dyspnoea	163	1%
Other	11,278	60%
Grand total patients with index HF listed as a secondary ICD-10 code	**18,838 (80.1%)**	**100%**

Table 3.5 Primary diagnosis during hospital admission that also made index HF diagnosis (hospital pathways)

Pathway	Number of patients with at least one symptom	Patients with no symptoms recorded	SOB	Ankle swelling	SOB Ankle swelling	Fatigue	SOB Fatigue	Ankle swelling Fatigue	SOB Ankle swelling Fatigue
Community n = 10,873	2,738 (25.2%)	8,135 (74.8%)	1,996 (18.3%)	434 (4.0%)	215 (2.0%)	51 (<1.0%)	25 (<1.0%)	8 (<1.0%)	9 (<1.0%)
Hospital n = 23,272	4,975 (21.5%)	18,170 (78.5%)	4,043 (17.5%	538 (2.3%)	276 (1.2%)	67 (<1%)	44 (<1%)	4 (<1%)	3 (<1%)

Table 3.6 Presence of HF symptoms prior to index diagnosis of HF.

3.4.2 All cause mortality

The median follow up period was 29 months overall; 29 months for the hospital pathway, and 30 months for the community pathway. At 24 months, the event rate for all-cause mortality in the hospital pathway cohort was 0.0094 per person-month, versus 0.0073 in the community pathway cohort (Figure 2). Comparing community vs. hospital diagnosis, the ratio of deaths per person-month at 24 months was 0.780 (95% CI 0.722 - 0.841, p<0.0001). At 72 months, the event rate for all-cause mortality in the hospital pathway cohort was 0.0082 per person-month, versus 0.0079 in the community pathway (inter-pathway event ratio 0.960 [95% CI 0.905 -1.020, p = 0.18] (Table 3.7).

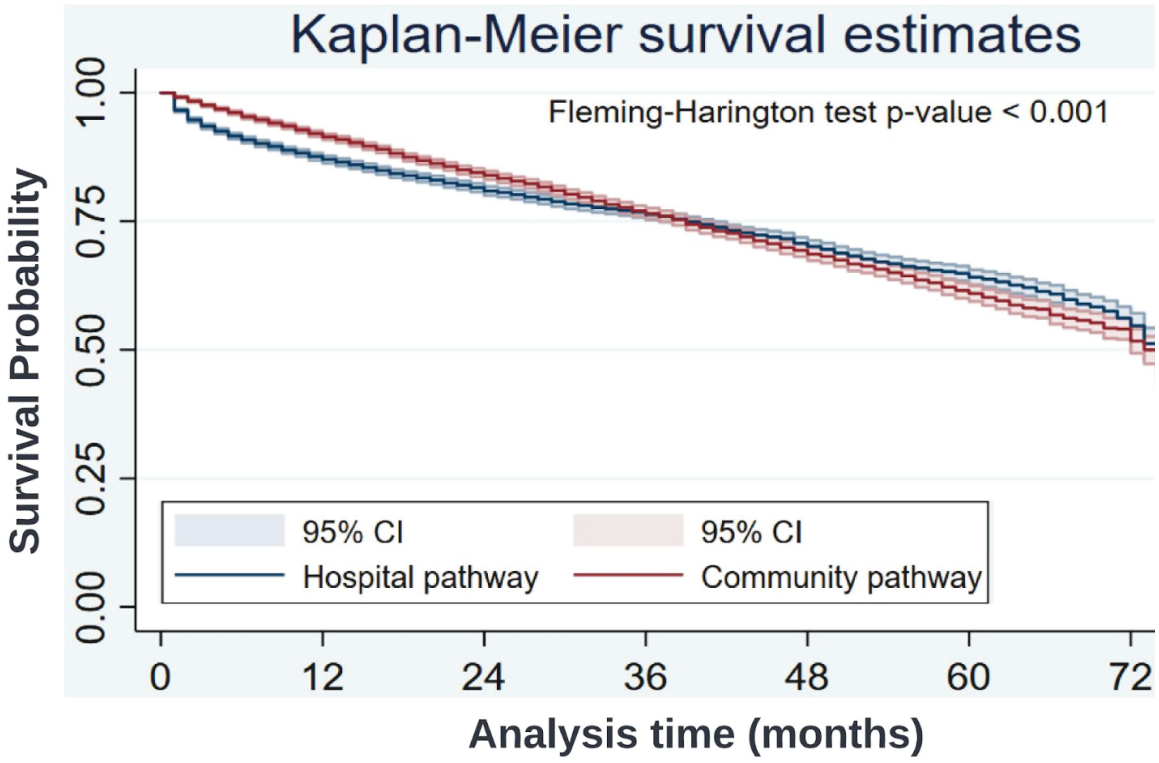


Figure 3.3 Kaplan-Meier Survival curve. All-cause mortality in the hospital pathway vs community pathway for Heart Failure

	Community Pathway		Hospital Pathway		-	
Month	**Patients at risk***	**Death/ person months**	**Patients at risk***	**Death/ person months**	**Rate ratio Community:Hospital, 95% CI**	**P value**
24	5,224	0.0073 (1,215/166,255)	5,040	0.0094 (1,507/160,881)	0.780 (0.722 - 0.841)	<0.0001
48	2,316	0.0076 (1,920/252,467)	2,145	0.0081 (1,978/243,533)	0.936 (0.879 - 0.998)	0.04
72	254	0.0079 (2,220/280,584)	223	0.0082 (2,214/268,890)	0.960 (0.905 -1.020)	0.18

Table 3.7 Differences in death at 24, 48, and 72 months comparing community and hospital pathways.

3.4.3 Cost-consequence analysis

Table 3.8 displays cost-consequence analysis for the propensity matched cohorts. Overall, across all available metrics of health service utilisation, there was a £2,485 longitudinal difference in cost associated with a HF diagnosis made through a community pathway versus hospital admission.

	Costs £ (Mean ± SD) for PSM cohorts						
Category	**Non-elective**	**Elective**	**ED Cost**	**Outpatient**	**Primary Care**	**Total cost**	**Difference**
Community £	8714 ± 16409	3755 ± 11344	419 ± 686	4901 ±7772	9508 ± 20240	27298 ± 29470	-
Hospital £	10804 ± 17719	4407 ± 14628	472 ± 753	5341 ± 8565	8759 ± 20954	29783 ± 32264	**+ 2,485**

Table 3.8 Cost-consequence differential in PSM matched cohort.

Table 1. Demographics and comorbidities for patients across total patient population diagnosed with heart failure in NWL in 2015-2020 (n = 34,208) and PSM cohort (n = 8,582 for each of hospital and community pathway)

3.4.4 Heart failure detection gap in NWL and England

Using data extracted from the ICHP HF Dashboard, figure 3.4 visualises the detection gap in NWL, with patient numbers described in table 3.9. Figure 3.5 expands on this with an impression of the England-wide burden of undetected HF.

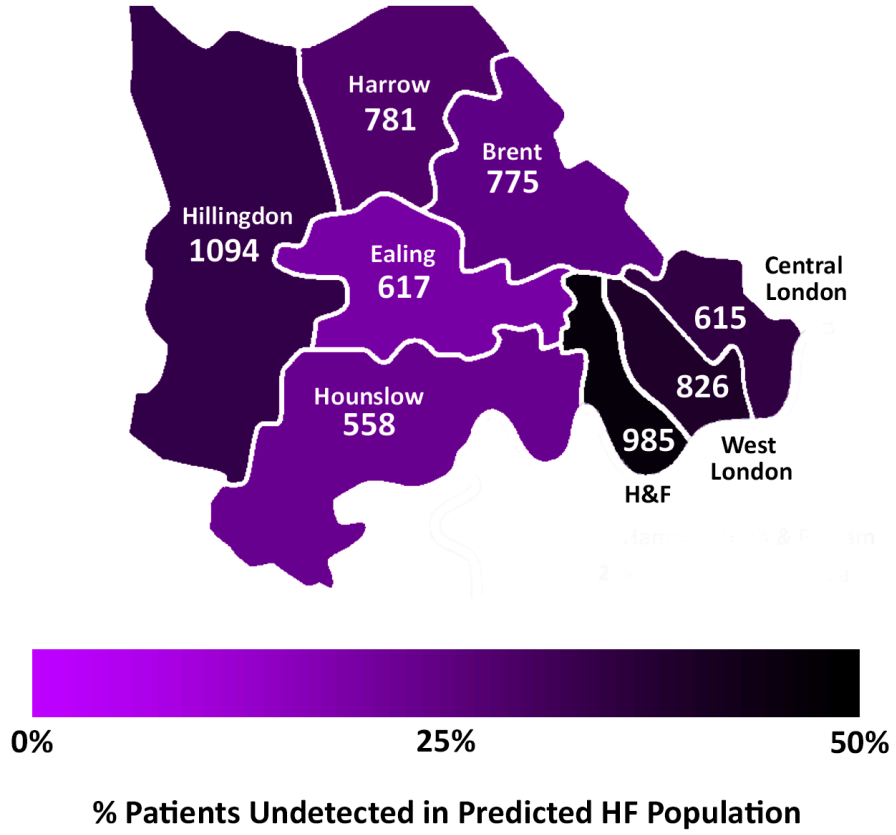


Figure 3.4 Detection gap in the 8 Clinical Commissioning Groups in North West London.

Numbers within each CCG represent predicted number of patients with undiagnosed HF. Image rendered in Photoshop (Adobe, US).

CCG	HF Patient Population	Predicted Undetected Patients	Cumulative Detectable Patients
Brent	3164	775	775
Central London	1685	615	1390
Ealing	3373	617	2007
Hammersmith and Fulham	2054	985	2992
Harrow	2695	781	3773
Hillingdon	3049	1094	4867
Hounslow	2487	558	5425
West London	1970	826	**6251**

Table 3.9 Detection gap for in the 8 Clinical Commissioning Groups in North West London.

Informed by data extracted by Imperial College Healthcare Heart Heart Failure Dashboard.

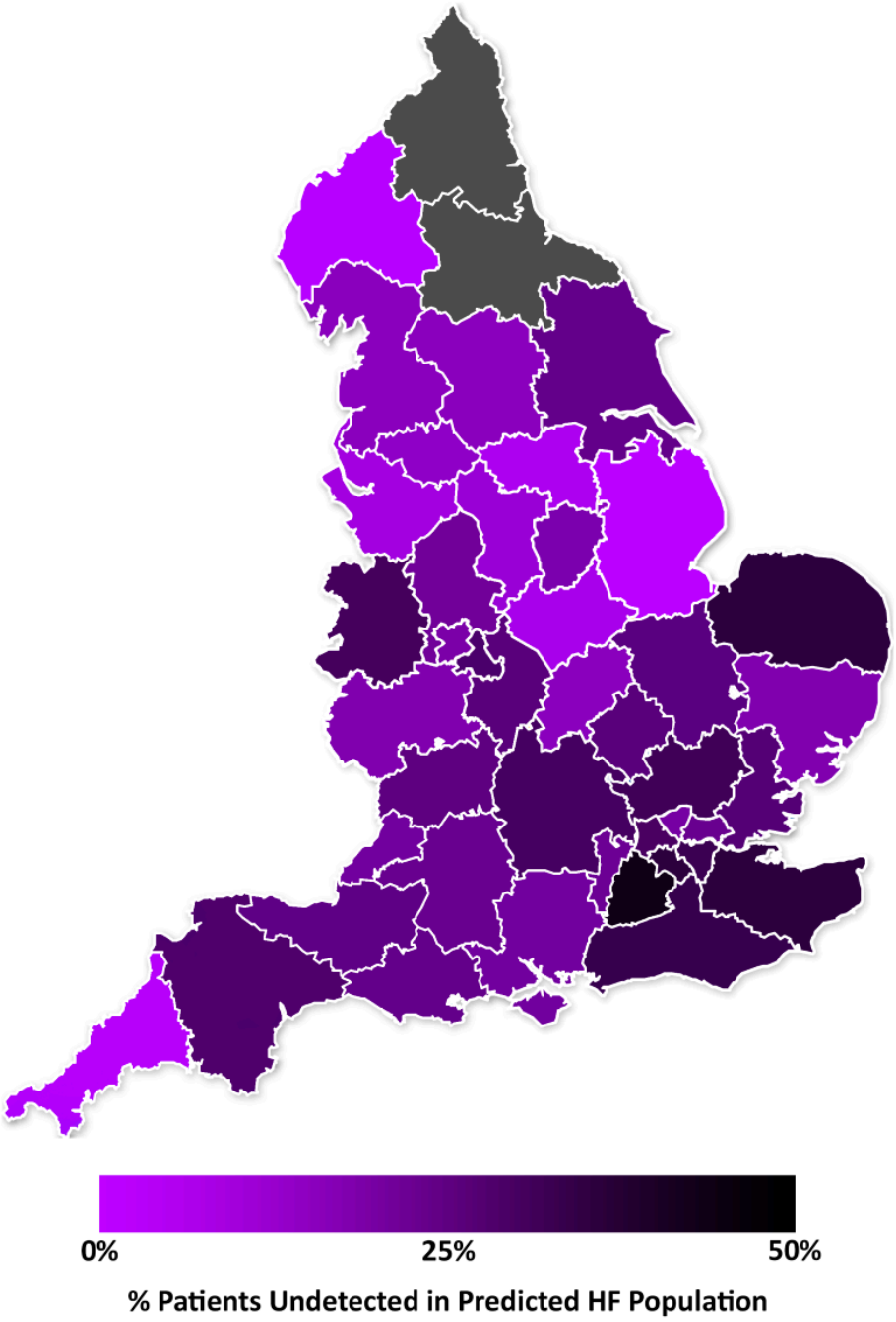


Figure 3.5 Map of detection gap in CCGs across England.

Image rendered in Photoshop (Adobe, US).

3.5 Discussion

3.5.1 Summary of main results

My study of a large propensity-score-matched population of NHS patients with HF demonstrates that across a six-year year period, index diagnosis of HF via hospital versus community pathway was associated with an increased rate of death in the first 24 months, with no difference between groups by 72 months. I found a substantial longitudinal cost saving (£2,500) following index HF diagnosis taking place through community pathways. Increasing rates of community-based HF diagnosis may therefore represent a substantial cost saving opportunity, particularly given the possibility that in some regions nearly 50% of patients with HF are undiagnosed.

3.5.2 Routes to diagnosis

Across the overall population of nearly 35,000 patients diagnosed with HF in NWL from 2015-2020, 70% were first diagnosed via hospital pathways. The overall balance of this differential is consistent with findings reported by Bottle et al. (cited in the NHS Long Term Plan), that between 2010 and 2013, 80% of HF diagnoses were first documented in hospital records.[28] My findings using data from 2015-2020 indicate that intervening efforts to improve community-based detection of HF have, at best, had modest impact.

A substantial portion of HF is precipitated by acute disease e.g. following severe myocardial infarction, requiring urgent hospitalisation.[158] Such cases should not be counted as missed opportunities. However, previous studies found that among those diagnosed in hospital, the vast majority of patients had also seen a GP in the previous year, with 37% having documented symptoms of HF.[28] Notably, in our overall HF population there was no difference in the number of primary care

encounters between the hospital and community pathway cohorts, with similar documented rates of HF symptoms prior to diagnosis. This may represent heterogeneity in awareness of HF within primary care services, but may also highlight the non-specific nature of cardinal HF symptoms, which overlap with other common cardio-respiratory pathologies (e.g. COPD[159]) – a diagnostic challenge discussed in recent international guidelines.[139] This is reflected in the unmatched study population, where multimorbidity (the presence of 2 or more long term conditions) was more pronounced amongst patients in the hospital pathway.

3.5.3 Prognostic association of route to diagnosis

My study observed an early survival advantage associated with community pathway-based diagnosis to 24 months, which was not sustained at 72 months when compared with hospital pathway-based diagnosis. Translating the long-term protective effects observed in clinical trials requires dose optimisation, monitoring, and patient concordance. It is possible that patients diagnosed via the community pathway were earlier in their HF disease course, and therefore more likely to realise the benefits of early initiation of prognostically beneficial therapies.[142] If so, the waning protective effect by 24 months may reflect real-world estimates of adherence to gold standard therapy. These are low, with the most optimistic ranging from 40-60%, and worryingly declining most as HF progresses.[160,161]

The mortality associated with HF is estimated to be between 53-67% five years after diagnosis[146], with hospitalisation a known adverse prognostic marker in established HF. Consequently, the convergence of survival curves by 72 months may take contribution from a 'regression to the mean' effect associated with heterogeneous adherence to gold standard therapy in both cohorts over a sustained period.

My study's findings are consistent with Taylor *et al.*, who examined data from 2000-2017 to identify a patient's first coded instance of a HF diagnosis in primary care records, and reported that those without hospital admission three months before or after diagnosis had better survival.[37] A study reporting on data from 1997-2010 found significantly worse outcomes among patients where HF was only ever coded in hospital records and never registered in the primary care record.[162]

3.5.4 Cost of Heart Failure

Across all health systems, the costs associated with HF are rising. However, detailed contemporary estimates on a per-patient level are lacking. To my knowledge, this is the first study to quantify the health economic opportunity of diagnosis through community pathways vs. hospital pathways.

Hospitalisation is the main driver of cost for HF. [22] As might be expected, I found non-elective admission costs accounted for the majority (84%) of the long-term increased costs of patients in the hospital pathway. However, understanding the health economic burden of HF through units of hospitalisation has substantial limitations. Studies have shown that post-HF diagnosis, there is an average of one hospital admission per year, of which two thirds are attributable to non-cardiovascular comorbidities.[163] However, I have shown that index HF diagnosis through hospital admission is unlikely to be coded hierarchically i.e. the primary diagnosis may be listed as common mimics and exacerbating conditions (e.g. COPD, pneumonia) with new HF listed among the secondary diagnoses. Teasing apart the contribution of HF to the cost of each hospital admission is therefore challenging.

More pragmatically, my study highlights that a community pathway-based diagnosis of HF offers an overall longitudinal cost-saving of £2500 per patient. This offers a compelling variable for

cost modelling and an intelligible, robust metric for policymakers. Realising even a fractional increase in community diagnosis could release substantial cost savings and return on investment.

3.5.5 Strengths

The population of NWL represents a wide spectrum of sociodemographic inequality, and includes the areas of highest ethnic diversity in the UK (as also reflected in results described in previous chapter).[164] To my knowledge, this is the first study to quantify the per-patient cost implications of route to index diagnosis of HF, adding a compelling health economic argument to the more established clinical rationale for investing in community diagnostic services[28]. Notably, this study accounted for a period of 6 years ending in 2020, and thus offers a contemporary picture of HF care, across a large population. This has been enabled by using real-world primary care, secondary care, and cost data within the Discover-NOW dataset, which more broadly takes advantage of now routine systematic sharing of hospital discharge summaries to improve fidelity between primary and secondary care coding.

Not only was I able to extensively match on demographic and comorbidity profiles, but also cost before HF diagnosis. This will have controlled for potential biases introduced by patients with extreme rates of service utilisation or rare/high-cost conditions; specifically mitigating over-estimation of the benefits of community-based diagnosis due to the hospital cohort having higher costs before HF diagnosis.

3.5.6 Comparison with other real-world datasets

Discover is one of the largest linked longitudinal datasets in Europe, but other databases exist where the same protocol could be applied to glean a more national picture. Among the most notable is the Clinical Practice Research Datalink (CPRD), which collects anonymised patient data from a network of GP practices across the UK, encompassing 18 million registered patients. A recent comparison of CPRD with Discover demonstrated that the population of the latter dataset matches the age and sex distribution of the UK population, but is more ethnically diverse than the CPRD dataset. Rates of chronic disease prevalence in Discover are comparable to those reported nationally. Discover also carries the advantage of including the ability to identify care organisations and postcodes, which allow for linkage to healthcare providers as well as social, community and mental health providers. On a practical note, the high-level consideration cost of database access/licensing may favour Discover (e.g. £35k vs £60k for commercially funded studies).[165]

3.5.7 Limitations

The results of my study are best interpreted in the context of its limitations. Despite extensive propensity-score-matching across demographic, clinical, and cost variables, some residual confounding is likely to remain and tempers my conclusions. The examination of real-world data is universally limited by the inconsistency and variable fidelity of medical coding in capturing variables of interest. HF is rarely coded with a granularity that describes preserved, moderately reduced, or reduced ejection fraction. One previous study using data from the ESC HF Long-Term Registry suggested HFrEF, HFmrEF, and HFpEF respectively account for 53%, 18%, and 29% of HF hospital admission – where HFrEF is more severe and has greater in-hospital mortality and post-discharge cardiovascular risk.[166] This may be addressed by future improvements in the coding of echocardiography results. Though this study interrogated granular clinical and cost data, sub-strata

important to understanding a patient's HF management were not available e.g., specific doses of disease modifying drugs that would have allowed inference beyond whether a patient was prescribed a HF-related medication, to whether this was optimised. Similarly, HF-specific quality of life metrics were not available for this cohort. Collectively this means I have likely been unable to fully account for differences in HF severity, where a degree of over-representation of severe disease in the hospital pathway may have skewed my observations. Lastly, though diverse in most other ways, the population of NWL does not encompass any rural/remote communities, whose experience of community versus hospital care may not be represented by our study. Reassuringly, the granularity of clinical coding continues to improve and future studies will likely be able to overcome this limitation.

3.5.8 Opportunities

A recent study by Kahn *et al.* searched primary care EHRs to identify a missed cohort of patients with HF, inviting them to a primary-care-based HF service that enabled optimisation of prognostic medication and an increase in device prescription.[167] Future research could quantify the clinical and health economic impact of invited and/or opportunistic screening of an at-risk population identified through analysis of population-wide linked datasets. Emerging point-of-care testing technologies[168] could underpin a programme comparable to the NHS Diabetes Prevention Programme and the NHS Health Check. Ultimately, despite progress in therapies and evidence for best-practice, the outlook for HF remains bleak, and community pathways may be best positioned to address this if powered by disruptive innovations that leverage integrated data and technology.

3.5.9 Health economic model for AI-ECG

Many patients have primary care encounters before their index HF diagnosis, with the expectation that most would have had a stethoscope examination, prompted by symptoms such as breathlessness. In the context of the findings described in the previous chapter – that AI-ECG applied to single-lead ECG recorded by the Eko DUO poses a quick, accurate, and reliable point-of-care screening test for HFrEF – the findings from this chapter enable the generation of a health economic model to inform whether investment in this technology would be a good use of NHS funding. My choice of a cost-consequence analysis was informed by this already being a UK government recommended tool for such evaluation of DHTs.[157]

For this health economic model, I consider an example cohort of 6,000 patients, close to the number of patients diagnosed with HF every 12 months in NWL. My study would suggest that the following:

- **70% (4,200)** can be expected to be diagnosed through hospital admission;
- **30% (1,800)** will be diagnosed through community pathways.

Next, I assume a conservative price-point for the hardware and AI-ECG license and further considerations around amortization, as follows:

- Eko DUO hardware at full published list price **=£320per unit** (before any bulk discount)
- Access to AI-ECG software at a subscription price of **£600 per license**, per annum, per primary care network (consortia of geographically close GPs working together and sharing resource and facilities).
- There are 400 GP practices in NWL, split into 49 primary care networks (i.e., multiple GP practices can use the same license)
- 3 year lifespan of hardware

Given the above variables, table 3.10 outlines a conservative health economic model anchored in AI-ECG yielding a 10% reduction in index diagnosis being made through hospital admission.

Health economic model for AI-ECG with Eko DUO in context of 6,000 HF cases p/a in NWL 400 (GP practices in NWL) x 2 devices per practice = 800 devices @ £320 per device = £256,000				
Expense year	**Cost**	**70:30 split**	**60:40 split**	**Cost differential**
Year 1	£256,000 (one-off hardware cost) £29,000 (license) **£285,000 total**	4,200 x £2,500 **= £10,500,000**	3,600 x £2,500 **= £9,000,000**	**10,500,000 - 9,000,000 = £1,500,000** - 285,000 (Y1 cost) **= £1,215,000 NET SAVING**
Year 2, 3, 4...	**£29,000 (license)**	4,200 x £2,500 **= £10,500,000**	3,600 x £2,500 **= £9,000,000**	£1,500,000 - £29,000 - **= £1,471,000 NET SAVING**

Table 3.10 Health economic model for deployment of AI-ECG in North West London primary care.

Importantly, though conservative, this model makes several assumptions and has several limitations. First, despite being a putatively quick, accurate, and easy-to-use tool, it remains uncertain how many patients at risk of HF seen in primary care where an Eko DUO is available would receive an examination with this. Second, though impressive, the performance and accuracy of the tool is not perfect. Third, currently the tool is tuned to specifically detect HFrEF; though as discussed, these patients account for a disproportionate amount of the HF cost burden. Finally, there will be unanticipated downstream costs e.g. unnecessary investigation of patients who had a false positive AI-ECG result. However, even when taking an even more conservative approach focused on "breaking even", the health economic case remains compelling. Considering the Year 1 outlay of £285,000 for

hardware, offsetting this cost would require an increase of only 114 patients (in a 2.5 million population) diagnosed through community pathways in 12 months. This represents a shift in the ratio of hospital to community pathway diagnoses from 70:30 to 68:32 i.e. breaking even requires only a 2% increase in the detection of HF cases through community pathways. Overall, and particularly given my study's finding that this pathway also unlocks a significant early survival benefit, the outlined health economic model for AI-ECG with Eko DUO forms the foundation of an argument that this DHT could offer substantial value for money.

3.6 Chapter Conclusion

Index diagnosis of HF through inpatient hospital admission continues to predominate and is associated with a significantly increased short-term risk of mortality and substantially higher long-term cost compared with community pathways. The results from this chapter highlight the need and opportunity for new approaches to increase community-based diagnoses, which may unlock longer, healthier lives for patients while substantially reducing NHS cost burden. AI-ECG with the Eko DUO offers a potentially cost-effective solution towards achieving this.

4 Patient Self-administered Screening for Cardiovascular Disease Using Artificial Intelligence in the Home

Previous chapters described evidence for the clinical and health economic potential of AI-ECG administered by a device such as an ECG-enabled stethoscope. Findings also included strong correlation between AI-ECG results obtained directly by clinical operators and patients during self-examination. This poses an opportunity for a new clinical screening pathway: decentralised, patient self-administered screening for cardiovascular disease using AI. However, such an approach poses unprecedented ethical questions, which are addressed in this chapter through a normative ethical analysis. This work draws on my book chapter edited by Harvard Law School, in press with Cambridge University Press.

4.1 Abstract

The UK government recently committed £250M across the National Health Service (NHS) towards the deployment of technologies that leverage artificial intelligence (AI). Many since funded technologies focused on home-based diagnosis. One particularly compelling exploratory use case involves patient-recorded cardiac waveforms that are interpreted real-time by AI to predict the presence of common, clinically actionable cardiovascular diseases. In this case, both electrocardiograms (ECG) and phonocardiograms (heart sounds) are recorded by a handheld device applied by the patient in a self-administered stethoscope examination, communicating via smartphone, with subsequent AI interpretation of waveforms.

Previous studies suggest the accuracy of this technology is commensurate with or superior to many of the established national screening programmes for other diseases. However, deployment of such a home-based screening programme combining hardware, AI, and a cloud-based digital platform for administration – all anchored in patient self-administration – raises ethical challenges including considerations of ***equity***, ***agency, data rights***, and ultimately ***responsibility*** for safe, effective, and trustworthy implementation of this powerful but novel diagnostic pathway. For example, the NHS cares for patients across disparate geographies and sociodemographic backgrounds, therefore successful deployment depends on universal smartphone access, internet connection, and sufficient digital literacy. Variability across these metrics may exacerbate existing disparities in health care access and outcomes. Moreover, meaningful use of these devices without direct clinician involvement ultimately offloads responsibility for conducting a diagnostic test with potentially life-threatening consequences onto the patient. Use of patients' own smartphones and internet connections should also meet the data security standards expected of NHS activity. Additional complexity arises from rapidly evolving questions around data 'ownership', by European law a term applicable only to the

patient from whom the data originate, when 'controllership' of patient data falls to commercial entities. Clarifying the appropriate consent mechanism – and the data usage to which it extends – requires reconciliation of commercial, patient, and health system rights and obligations.

Oriented to this real-world clinical setting, I evaluate the ethical considerations of extending home-based, self-administered AI diagnostics in the NHS. I discuss the complex field of stakeholders, including patients, academia, and industry, all ultimately beholden to governmental entities. I propose a multi-agency approach to balance permissive regulation and deployment (to align with the speed of innovation) against ethical and statutory obligations to safeguard public health. I further argue that a strong centralised approach to carefully evaluating and integrating home-based AI diagnostics is necessary to balance the ethical considerations outlined above. I conclude with specific policy recommendations applicable to NHS stewardship of this novel diagnostic pathway in a manner that preserves and promotes trust, patient engagement, and public health.

4.2 Introduction

The United Kingdom (UK) government has committed £250M across the National Health Service (NHS) towards the deployment of technologies that leverage artificial intelligence (AI). Many since-funded technologies focus on evidence generation for home-based diagnosis.[169] One compelling potential use case, as detailed in previous chapters, involves patient-recorded cardiac waveforms that are interpreted real-time by AI to predict the presence of common, clinically actionable cardiovascular diseases. In this case, both electrocardiograms (ECG) and phonocardiograms (heart sounds) are recorded by a handheld device applied by the patient in a self-administered stethoscope examination, communicating via smartphone with subsequent AI interpretation of waveforms. Validation studies suggest the accuracy of this technology approaches or exceeds many established national screening programmes for other diseases.[168] More broadly, the combination of a new device (a modified handheld stethoscope), novel AI algorithms, and communication via smartphone coalesce into a technology for delivering a distinct clinical care pathway that may become increasingly prevalent across multiple disease areas.

However, deployment of a home-based screening programme combining hardware, AI, and a cloud-based digital platform for administration – all anchored in patient self-administration – raises distinct ethical challenges for safe, effective, and trustworthy implementation. This chapter approaches these concerns in five parts. **First**, I briefly outline the organisational structure of the NHS and associated regulatory bodies responsible for evaluating efficacy and safety of digital health technologies (DHTs). **Second**, I highlight NHS' plans to prioritise digital health and the specific role of AI in advancing this goal, with a focus on cardiovascular disease. **Third**, I review the clinical imperative for early diagnosis of heart failure (HF) in community settings, and the established clinical evidence supporting the use of a novel AI-ECG based tool to do so. **Fourth**, I examine the ethical concerns

with the AI-ECG diagnostic pathway according to considerations of ***equity***, ***agency, and data rights*** across key stakeholders. **Lastly**, I propose a multi-agency strategy anchored in a purposefully centralised view of this novel diagnostic pathway – with the goal of preserving and promoting trust, patient engagement, and public health.

4.2.1 The UK National Health Service and Responsible Agencies

The UK NHS is an exemplar of a government-funded health system where the payor is also the provider of services. The NHS continues to operate on three founding principles: 1. To meet the needs of everyone, 2. To be free at the point of delivery, 3. To be based on clinical need, not ability to pay. The NHS is funded through general taxation, and every decade since its inception has brought challenges through shifts in policy, clinical priorities, and subsequently, size and distribution of budget. The UK government has devolved responsibility for the NHS to the Scottish Government, Welsh Government, and Northern Irish Assembly; accordingly, NHS England, NHS Wales, NHS Scotland, and Northern Ireland Health and Social Care Services provide health services in the UK. For the purposes of this chapter, I focus on England, where NHS England and NHS Improvement are the responsible government entities for delivery of healthcare.

Similar to health systems globally, the NHS has been slow to catch up with the momentum of technological adoption seen in non-health industries.[170,171] The increasing societal and political pressure to modernise the NHS in this way led to formation of agencies tasked with this specific mandate, each of which plays a key role in evaluating and deploying the technology at issue in this chapter. Since re-integrated into NHS England, "NHSX" was founded as its own entity in 2019, with the aim of setting national policy and developing best practice across technology, digital and data, including data sharing and transparency. Closely related, NHS Digital (established in 2016) is the

national provider of information, data and IT systems for commissioners, analysts and clinicians in health and social care in England. Additionally, the NHS Accelerated Access Collaborative works to bring together industry, government, regulators, patients, and the NHS to remove barriers and accelerate the introduction of innovative new treatments and diagnostics capable of transforming care delivery.

From a regulatory perspective, the Medicines and Healthcare products Regulatory Agency (MHRA) is responsible for ensuring that medicines and medical devices (including software) work and are acceptably safe within the scope of their labelled indications. The MHRA grants authorisation for UK market entry in a similar fashion to the European Medicines Agency for the European Union, and the Food and Drug Administration in the United States. Post-Brexit, the UK has moved away from *Conformitè Europëenne* marking, instead requiring UK Conformity Assessed for new products, though the underlying risk-based classification system remains similar to international counterparts.

In practice, most diagnostic technology (including ECG machines, stethoscopes, and similar) would be considered relatively low-risk devices compared with invasive/implantable technology, or those that are explicitly life-sustaining. One implication of this risk tiering is that unlike a new implanted cardiac device such as a novel pacemaker or coronary stent, diagnostic technology (including AI-ECG) would not be expected to demonstrate safety and effectiveness prior to legal marketing through a large trial with hard clinical endpoints.

Once a medical device receives regulatory authorization from the MHRA, the UK takes additional steps to determine whether and what the NHS should pay for the service at issue. The National Institute for Health and Clinical Excellence (NICE) evaluates the clinical and cost-effectiveness of drugs, health technologies and clinical practice for the NHS. Rather than negotiate prices, NICE makes recommendations for system-wide funding, and therefore, deployment,

principally informed by tools such as quality adjusted life years. In response to the increasing number and complexity of DHTs, NICE partnered with NHS England to develop standards that aim to ensure new digital health technologies are clinically effective and offer economic value. The subsequent evidence standards framework for digital health technologies aims to inform digital health innovators on NHS decision making, support NHS commissioners (executors of local funding) in exacting appropriate evidence, and be dynamic and value driven, with a focus on offering maximal value to patients.[42] This backdrop informed the methodological approach for the health economic modelling of AI-ECG and an ECG-enabled stethoscope outlined in the previous chapter.

Summing the role of the bodies above, as applied to a novel AI-ECG device, I observe the following: Manufacturers seeking marketing authority for new digital health tools primarily focused on *diagnosis* rather than treatment of a specific condition (like HF), must meet the safety and effectiveness standards of the MHRA – but this does not necessarily (or likely) require a dedicated clinical trial illustrating real-world clinical value. By contrast, convincing the NHS to pay for the technology may require more comprehensive evidence sufficient to sway NICE, which is empowered to take a more holistic view of the costs and potential benefits of novel health tools – including digital health tools increasingly supported by NHS sub-agencies tasked with advancing evidence generation and implementation.

4.2.2 Policy Prioritisation Towards a Digital NHS

The NHS Long Term Plan, published in 2019, was drawn up by frontline staff, patient groups, and national experts to define key challenges and set an ambitious but realistic vision for the next ten years of health and social care in the UK.[27] "Digital transformation" is one of the core pillars of the Plan, outlining a broad set of deliverables, including "providing digital services and tools to give people

more control over their own health and the care they receive from the NHS." The Plan singles out AI as a key driver for digital transformation. Specifically, "use of decision support and AI to help clinicians in applying best practice, eliminate unwarranted variation across the whole pathway of care, and support patients in managing their health and condition." Here I already note implicit ethical principles: reducing unjustified variability in care (as a consideration of justice); and promoting patient autonomy by disseminating diagnostic capabilities that otherwise may be accessible only behind layers of clinical or administrative gatekeeping. In this chapter I further discuss whether either of these or other ethical targets are, on balance, advanced by AI-ECG.

Although well-known proponents of AI such as Dr Eric Topol argue that it will "help make healthcare human again"[58], including through liberating staff of tedious administrative tasks, this shift towards increasing use of AI in the NHS may pose challenges to an unprepared workforce, including uncertainty about real-world use and outcomes associated with AI-driven diagnostics.[73] Thus, at its inception in 2019, NHSX allocated £250 million towards the NHS AI Lab, with over half of this funding marked for deployment to support the testing, evaluation, and scaling of promising AI-driven technologies through the AI in Health and Care Award (AI Award). With a mission statement to "drive the digital transformation of the NHS and social care", the AI Award sought to fund technologies specifically addressing public health priorities outlined in the NHS Long Term Plan. To date, only a handful of such AI technologies has achieved both MHRA approval and NICE endorsement, therefore widespread use remains limited to a few examples, with important ethical considerations still lacking exploration through empirical data. Among the most prominent successful deployments of a DHT using AI, HeartFlow's (HeartFlow Inc., California, US) technology uses computed tomography images to create a personalised 3D model of the heart that shows how blood is flowing around it, allowing blockages due to coronary artery disease to be characterised non-invasively – without the need for a comparatively more expensive angiogram procedure.[172]

Within the broad remit to pursue "digital transformation", cardiovascular disease is emphasised throughout the Long Term Plan as a priority target for public health interventions. There is a particular focus on HF. The symptomatic burden and mortality risks of HF remain worse than many common, serious cancers. Among all chronic conditions, HF has the greatest impact on quality of life[173], and costs the National Health Service (NHS) over £625 million per year – 4% of its annual budget[174], with HF-related hospital admissions and related demand for social care expected to increase by 50% by 2040.[4]

It is critical to emphasise that HF is a syndrome made by clinical diagnosis, and not clearly established by any single diagnostic test.[1] While the exact combination of data will vary by context, a clinical diagnosis of HF may include integration of patients' symptoms, physical exam (including traditional stethoscope auscultation of the heart and lungs), measurement of peripheral blood pressure and intracardiac chamber pressure (either invasively or though other techniques), plasma biomarkers, ECG findings, radiographs, and cardiac ultrasound data. Individually, compared with a clinical diagnosis gold standard, the test characteristics of each modality above vary widely, with sensitivity generally higher than specificity.[175,176]

Traditionally, HF has been divided into distinct phenotypes based on the measurement of left ventricular ejection fraction (LVEF), a normally distributed variable. This is observed and typically estimated visually on ultrasound (echocardiography), though other imaging modalities such as nuclear tracer technology, magnetic resonance, direct ventricular angiography, or computed tomography may also be used to calculate LVEF. HF spans the range of LVEF, where LVEF ≤40% is the most clinically severe – while also being the only subtype of heart failure with multiple proven therapies for improving prognosis.[177]

Similar to most chronic diseases in high-income countries, the burden of HF is greatest in those most deprived and tends to have an earlier age of onset in minority ethnic groups, who experience worse HF related outcomes.[178] Therefore HF generally and the LVEF ≤40% subtype in particularly present attractive targets for disseminated technology with the potential to speed up diagnosis and direct patients towards proven therapies, particularly if this mitigates the social determinants of health driving observed disparities in care.

IV. AI-ECG for Community Diagnosis of Heart Failure

The NHS Long Term Plan states that "80% of heart failure is currently diagnosed in hospital, despite 40% of patients having symptoms that should have triggered an earlier assessment." The previous chapter analysed contemporary data and concluded that this imbalance persists. Subsequently the Plan's advocacy for "using a proactive population health approach focused on … earlier detection and intervention to treat undiagnosed disorders, such as heart failure" remains relevant.[27] Given the epidemiology of the problem (including its disparate impact across communities) and imperative for practical screening, a tool supporting community-based diagnosis of HF has the potential to be both clinically impactful and economically attractive. The myriad diagnostics applicable to HF described above, however, variously require phlebotomy, specialty imaging, and clinical interpretation tying together signs and symptoms into a clinical syndrome. AI-supported diagnosis may overcome these limitations.

As broadly defined previously, AI is the ability of computer algorithms to interpret data at human- or super-human levels of performance.[51] Enabled in the last decade through the convergence of increased computing power, massive 'big' datasets, and improved data science techniques for machine learning, this decade continues the trend for accelerated development of AI-driven

technologies with health-related applications. The near ubiquity of ECGs in well-phenotyped cardiology cohorts supports the training and testing of AI algorithms among tens of thousands of patients, with additional opportunities for external validation. This has resulted in both clinical and, increasingly, consumer-facing applications, where AI can interrogate ECGs and accurately identify the presence, for example, of heart rhythm disturbances, demographic details, and – as discovered recently – findings suggestive of HF.[97,137]

To recapitulate, the standard 12-lead ECG waveforms (requiring attachment of multiple electrodes to the patient) have been described as an accurate digital biomarker for the stages of heart failure. Previous research by the Mayo Clinic showed that an AI algorithm (AI-ECG) could detect LVEF of 35% or lower with 86% sensitivity and 86% specificity.[92] This AI-ECG model has since been validated further in cohorts in the United States and Russia.[94,95] More recently (2021), a cluster randomised controlled trial made AI-ECG accessible for 12-lead ECG interpretation in a cohort of Mayo Clinic primary care practices, highlighting an increased rate of detection of HF.[97] However, access to 12-lead ECG machines is limited by factors including cost and expertise.

The emergence of ECG-enabled stethoscopes, capable of recording ***single-lead*** ECGs during contact for routine auscultation (listening), highlighted an opportunity to apply AI-ECG for point-of-care screening. The Eko DUO (Eko Health, Oakland, CA, US) is one example of such an ECG-enabled stethoscope, taking the traditional form factor of earpieces connected to tubing that plugs into a device at the end (the "bell"). Detaching the tubing leaves a small mobile-phone-sized device embedded with sensors for recording both ECG (electrodes) and phonocardiogram (heart sounds, using in-built microphone). Connectivity via Bluetooth allows subsequent live streaming of both ECG and phonocardiographic waveforms to a user's smartphone and the Eko app, and further onwards transmission to any remote recipient via a telemedicine dashboard. Waveforms can also be recorded

and transmitted to cloud-based infrastructure, allowing them to be analysed by cloud-based AI algorithms, such as AI-ECG.

Though single-lead ECG-enabled stethoscopes offer advantages over 12-lead recordings, including reduced cost, remote connectivity, ease-of-use, speed, and adaptation to workflow, the utility of AI depends on single-lead ECG input alone being adequate. Chapter 2 described results of my study that recruited over 1,000 patients attending for echocardiography, where the echo LVEF result was used as the 'ground truth' against which AI-ECG performance was compared. This study demonstrated incremental accuracy of the AI-ECG algorithm when just one 15s ECG recording (sensitivity 85%, specificity 70%) was increased further with two (sensitivity 92%, specificity 80%). Importantly, the study showed there was no observed difference in performance of AI-ECG across 15 operators with varying clinical experience, with no apparent impact of the subtle variation in the exact positioning of the device.[168] Lastly, excellent consistency (reproducibility) between clinical user and patient self-administered recordings was also observed.

While the current programmatic focus of the Long Term Plan is on driving up community HF diagnoses (as supported by results from Chapter 3), AI can in theory also be applied to ECG and phonocardiographic waveforms (heart sounds) to inform on the presence of two further public health priorities within cardiovascular disease: atrial fibrillation (AF), a common irregular heart rhythm; and valvular heart disease, typified by presence of heart murmurs. Both of these conditions can be both causes and effects of HF.[179,180] Therefore, with ECG and PCG recorded in combination, a 15-second examination with an ECG-enabled smart stethoscope may offer a three-in-one screening test for substantial drivers of cardiovascular morbidity and mortality and systemically important drivers of health care costs.

The proposed approach for embarking on the first stage of deploying such a screening pathway would be to anchor in primary care, given the high rates of undiagnosed HF (as outlined in Chapter 3) and further cardiovascular disease (including AF and valvular disease) in communities across England.[181] The early stages of this pathway could build on successful approaches that have interrogated NHS general practitioner electronic health records, with application of search logic to identify those at risk for HF (e.g. risk factors such as hypertension, diabetes, previous myocardial infarction). This would generate a list of patients, who are subsequently contacted to be offered at-home screening. Informed by existing screening programmes in the NHS, those patients agreeing to participate would be posted a small parcel containing an ECG-enabled stethoscope (Eko DUO), a 4G connected smartphone with pre-installed app for AI-ECG (Eko App), and a simple instruction leaflet on how to perform self-recording (similar to supplementary material 10.4). It would also be possible for patients to download the appropriate app directly to their own phone and conduct the recording this way instead, with all data stored on a cloud-based server (the expectation would be that no data is stored on any phone). Results of AI interpretation would be reviewed centrally by a team of clinicians. Patients whose data, as interpreted by AI, suggests the presence of HF, AF, or valvular heart disease would be invited for further investigation in line with established NICE clinical pathways (figure 4.1).

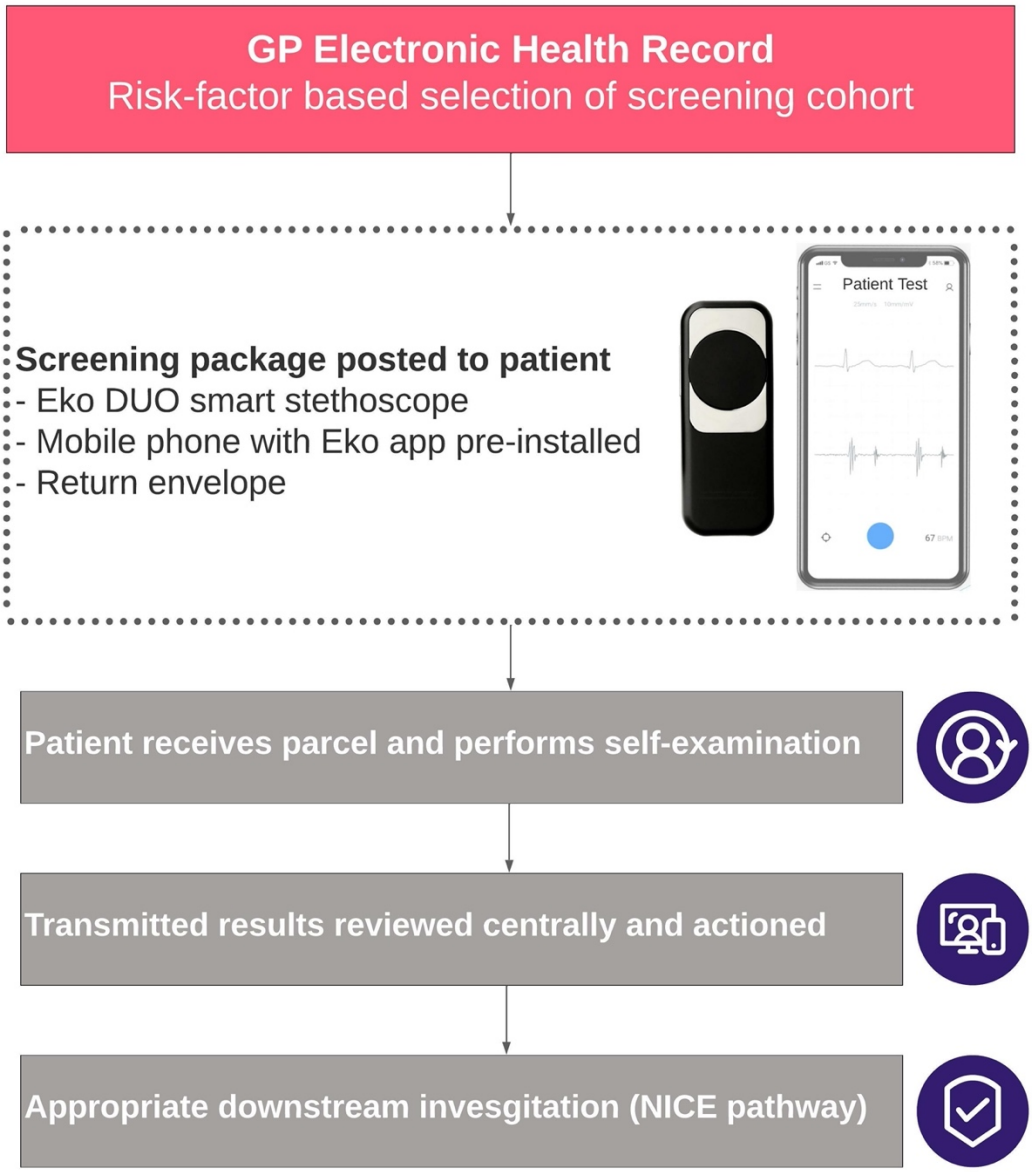


Figure 4.1 Outline for proposed patient self-administered screening pathway using AI-ECG and ECG-enabled smart stethoscope.

This scenario sets the scene for a novel population health intervention that would draw on a technology-driven screening test, initiated in the patient's home, by the patient themselves. The current focus on common and costly cardiovascular conditions serendipitously combines available

technology and clinical expertise and compelling expected benefits; however, more opportunities for similar decentralised, patient-activated screening with digital diagnostics will surely follow if AI-ECG proves tractable. Notably, here I have described what I believe to be the first application of "super-human" AI – accurately inferring the presence of HF from single-lead ECG was previously thought impossible – with the potential for meeting a major unmet need through a clinical pathway that scales access to this potentially transformative diagnostic.

4.3 Ethical Considerations for Self-Administered Cardiovascular Disease Screening at Home

The widespread enthusiasm for community, patient-driven cardiovascular screening reflects not just clinical expediency but recognition that this pathway may support normative public health goals, particularly around equity and patient empowerment. Despite these good-faith expectations, deployment of such a home-based screening programme combining hardware, AI, and a cloud-based digital platform for administration – all hinging on patient self-administration – raises distinct ethical challenges. In this section, I explore the ethical arguments in favour of a patient self-administered AI-ECG programme, as well its potential pitfalls.

4.3.1 Equity

One durable and compelling argument supporting AI-ECG arises from well-known disparities in cardiovascular disease and treatment. Cardiovascular disease follows a social gradient; the most disadvantaged are the most affected, where inequitable access to health services is a substantial driver of poor outcomes. This is particularly pronounced for HF, where under-diagnosis in England is most

severe in the poorest geographies (e.g. Chapter 3, figure 3.5). Consistent with this observation, Chapter 3 highlighted those receiving an index diagnosis of HF through emergency hospital admission rather than through primary care were more likely to be socioeconomically deprived. Some of this can be attributed to the persistence in the NHS of the "inverse care law", where communities most in need of good healthcare are those least likely to receive it e.g. the best hospitals are in the wealthiest cities. As a key social determinant of health, nearly a million people in England (2%) have poor English language skills, related to lower uptake of preventative healthcare and subsequently worse health outcomes.[182] Community deployment of AI-ECG potentially attenuates these disparities in several ways. First, targeted screening based on risk factors (such as high blood pressure and diabetes) will, based on epidemiologic trends, necessarily and fruitfully be enriched for vulnerable patient groups in which these conditions are more prevalent.[167] These same patients may also be less able to access traditional facility-based cardiac testing, which by definition requires ease of transport, schedule flexibility (including forgiving work and childcare arrangements), and availability of translation services – all established barriers to accessing healthcare.[183-185] AI-ECG may overcome these concerns for the patients most in need.

Second, AI-ECG administered as proposed explicitly transfers a key gatekeeping diagnostic screen away from the cognitive biases of traditional bedside medicine. Cross-cultural challenges in subjective diagnosis and treatment escalation are well documented, including in HF across a spectrum of disease severity ranging from outpatient symptoms ascertainment to referral for advanced cardiac therapies and even transplant.[96,186] AI-ECG may overcome the biases embedded even in traditional screening by binarizing complex syndromic diagnoses into an objective, accessible, and clearly documented test finding that is programmatically entwined with subsequent specialist referral.

These supporting arguments grounded in equity may be balanced by equally salient concerns. Even a charitable interpretation of the proposed AI-ECG pathway assumes a relatively savvy, engaged, and motivated patient. The ability to post an AI-ECG screening package widely to homes is just the first step in a series of necessary steps: opening and setting up the screening kit, including the phone and ECG-enabled stethoscope, successfully activating the device, and recording a high-quality tracing that is then processed centrally without data loss. While the early experience using this technology in various settings has been reassuring (including successful training of non-physicians and high reproducibility and correlation with patient's self-recordings), it remains uncertain whether the established digital divide will complicate equitable application of AI-ECG screening. Assuming equal (or even favourably targeted) access to the technology, are patients able to use it, and *do they want to*? The last point is critical: in the UK as well as the United States, trust in health care varies considerably and (broadly speaking) tracks unfortunately and inversely with clinical need.[187,188] Importantly, certain characteristics traditionally associated with the digital divide such as older age should not pre-empt a patient's inclination to participate.[189,190] I have limited data to anchor expectations that all affected patient communities will eagerly pursue this screening option. Some optimism may be drawn from the study described in Chapter 2, where only 26 out of 1,050 – 2.5% – of patients approached and informed of the AI-ECG study declined to participate, in an overall highly ethnically diverse cohort.

Indeed, one well-grounded reason for suspicion recalls another problem for equity-driven enthusiasm for AI-ECG, which is the training and validation of AI algorithms themselves. The "black box" nature of some forms of AI, where the reasons for model prediction cannot easily be inferred, has appropriately led to concerns over insidious algorithmic bias and subsequent reservations around deploying these tools for patient care.[191,192] Even low-tech HF screening confronts this same problem, as, for example, the most widely-used biomarker for HF diagnosis (natriuretic peptide blood test) has well-known performance variability according to age, sex, ethnicity, patient weight, renal function, and

clinical comorbidities.[1] Conversely, results described in Chapter 2 suggest that AI-ECG for HF detection does not exhibit similar biases. It may still be the case that biases exist, but that they require further large-scale deployment to manifest. The proposed clinical pathway would likely accumulate a specific form of "ethical debt", precipitated by the reality that some ethical problems with AI DHTs will only be detected after they are deployed.[76] Who pays for ethical debt in AI? The immediate argument would seem to favour that any agency opting to design, develop, and deploy and AI technology without proactively identifying potential ethical concerns should bear responsibility for untoward consequences. Further problems may arise through the risk of "drift" with AI algorithms. ML algorithms use data describing historical episodes, variables, and data formats to make ahead-of-time predictions of clinical outcomes. However, clinical settings are dynamic environments and the underlying nature of data and data distributions characterising episodes can change with time (data drift), and so can the relationship between episode characteristics and associated clinical outcomes (concept drift). This can in part be summed up by a quote on the Covid-19 pandemic's impact on AI: "Machine-learning models trained on normal human behaviour are now finding that normal has changed, and some are no longer working as they should."[193]

To address these concerns, I propose several programmatic features as essential and intentional for reinforcing the potential of wide-scale screening to promote equity. First, it is imperative on programme managers to prominently collect self-identified race, ethnicity, and other socioeconomic data (e.g., language, education) among all participants at each level of outreach – screened, invited, agreed, successfully tested, identified as "positive", referred for specialist evaluation, and downstream clinical results. Disproportionate representation at each level, and differential drop-out at each step, must be explored but can only begin with high-quality patient-level data to inform analyses and programme refinement. The Discover NOW dataset used for analysis in Chapter 3

exemplifies the type of data infrastructure required – granular, updating real-time, and curated for efficient analyses.

An important overlapping point of data capture might include assessment of digital literacy via survey, establishing the baseline capabilities of each patient, and developing appropriate mitigation strategies. Overall, instruction materials and technology components such as the app should be made available in multiple languages and font sizes (for those with visual impairments), with emphasis on using simple language and preferentially using clear figures/graphics. Conveying simplicity and ease-of-use through such user-centred design assuages a patient's potentially low confidence in self-administration and, beyond uptake alone, helps maximise collection of adequate quality data to inform a patient's screening outcome.[194,195] Trust in AI-ECG may be further buttressed in several ways, recognising the resource limitations available for screening programmes generally. One option may be providing accommodation to sceptical patients in a way that still provides suitable opportunity to participate through alternative means. This could simply be attending for an in-person appointment to have the AI-ECG examination performed by a healthcare professional, or the be given the opportunity to nominate an individual e.g. a family member or friend to apply the screening test to the patient.

The centralised administration of NHS screening programmes by NHS England paired with NHS Digital's repository data on uptake of screening offers granular insights to anticipate and plan for geographies and groups at risk of low uptake. I therefore propose enshrining a data monitoring module into the at-home AI-ECG screening protocol, with prespecified targets for uptake and defined mitigation strategies – importantly, if needed, with available funding to deploy this. This monitoring is feasible due to the inherent connectivity of the technology driving AI-ECG, drawing on dashboarded reports of summary metadata that can highlight disparities in access. However, a more

proactive approach to targeting individuals within a population with certain characteristics needs to be balanced against the risk of stigmatisation[196,197], and ultimately potential loss of trust that may further worsen the cardiovascular outcomes seeking to be improved.

Lastly, equity concerns around algorithmic performance are necessarily empirical questions that will also benefit from patient-level data collection. I acknowledge that moving from research in the form of prospective validation studies (Chapter 2) to deployment for patient care requires judgment in the absence of clear consensus, within the NHS or more globally, around the minimum established scrutiny for bias. To avoid these potentially impactful innovations remaining in the domain of research, and to anticipate wide-reaching implications of a deployment found to exhibit bias retrospectively, one possible solution would be to, by design, prospectively monitor for inconsistent test performance. False positives can be measured through linkage to primary care EHR data for outcomes of downstream NICE investigations for HF, atrial fibrillation, or valvular heart disease – comparing to the "ground truth". For false negatives, this may require a more expansive approach in the form of inviting a small sample of patients with negative AI screening tests for "quality control" next-step investigations (echocardiogram or 12-lead ECG). All of this risks adding complexity and therefore cost to a pathway seeking to simplify and save money. However, given this proposed programme's position at the vanguard of AI deployments for health, a permissive approach balanced with rigorous checkpoints for the pitfalls of AI technology may yield lessons to help blueprint best-practice and build confidence for further applications.

4.3.2 Agency

Another positive argument for AI-ECG screening aligns with trends in promoting *agency*, understood here as patient empowerment, particularly around the use of digital devices to measure,

monitor, and manage one's own health care – particularly in cardiovascular disease. Enthusiastic commercial uptake of fitness wearables, for example, moved quickly past counting steps to incorporate heart rhythm monitoring, leveraging either patient-activated ECG electrodes (as in the Apple Watch and Garmin watch) or alternative approaches that leverage plethysmography to identify irregular heartbeats suggestive of atrial fibrillation.[198] Testing of these distributed technologies has shown mixed results, with the yield of positive cases necessarily depending on the population at issue.[199,200] Recalling the equity concerns above, the use of devices such as the Eko DUO may be more immediately popular notion among younger and healthier patients, in whom true positive diagnoses may be uncommon. Yet targeted and invited screening with AI-ECG may balance these concerns through enriching the population at risk by invitation, while preserving patient activation and engagement. While unproven in this context, it may be that patients who effectively self-diagnose their serious yet treatable conditions transfer that agency into therapeutic engagement[201], which for both HF and AF remain essential and intractable problems.[202,203] Both conditions require sustained attention to longitudinal medical therapy, particularly for HF patients in whom careful dose-escalation of multi-drug regimens are critical for reducing morbidity and mortality. Among patients with atrial fibrillation, only approximately half of eligible patients take guideline-recommended anticoagulation medication for stroke prevention, despite decades of public health outreach.[204] There is some appeal, then, to activating patients in the detection and treatment of conditions that might be variously and unhelpfully viewed by them as *fait accompli*, genetically determined, or simply random.

Realistic concerns about agency extend beyond the previous warnings about digital literacy, access to reliable internet, and language barriers to ask more fundamental questions about whether patients actually *want* to assume this central role in their own health care. A key parallel here is the advent of mandates for shared decision-making in cardiovascular disease, particularly in the United States where federal law now requires patients considering certain procedures for atrial fibrillation and

HF management to incorporate "evidence based shared decision-making tools" in their treatment choices.[205,206] Several observers have identified problems with these mandates[207], including observations that for life-threatening conditions, patients may prefer a stronger clinician role in making recommendations even for choices that necessarily incorporate patients' values and preferences. Considering the AI-ECG context, patients may reasonably ask if other screening options might meet the same goal without shifting such a key role into patients' hands (literally). Relatedly, much of this care pathway depends on phones/apps and the digital stethoscope functioning properly, thereby redistributing responsibility for trouble-shooting either to a device manufacturer, the app developer/manager, or simply the patient. For patients this risks potential feelings of guilt around "doing it wrong" or failing to do it at all. By contrast, if a patient is attending hospital for an echocardiogram to screen for cardiovascular disease, the responsibility for resolving faulty data capture sits firmly with the immediate NHS care provider.

Another potentially insidious consequence of distributed screening via AI-ECG focuses not on putting responsibility on patients, but on taking it away from clinicians. While subtle, shifting the cognitive work of integrating complex signs and symptoms into a syndromic diagnosis like HF, or even sleuthing more obscure manifestations of atrial fibrillation to make that diagnosis, may have unwelcome implications for clinicians' diagnostic skills. Reducing opportunities for bedside physical exam diagnoses of HF, valvular disease, or atrial fibrillation may subtly diminish the "reinforcement learning" commonly understood to be a cornerstone of clinicians' training and lifelong practice.[208] I would emphasise that this is not just whimsical nostalgia for a more Oslerian time in medicine, but a genuine worry about reductionism through algorithmic diagnosis that binarises complex constellations of findings into simple yes/no diagnoses (AI-ECG, strictly speaking, only diagnoses reduced LVEF, which is not clinically equivalent to a diagnosis of HF). Though the likes of Topol advocate for AI as

the catalyst for diagnostic clinical excellence[58], there is equal concern about an over-reliance on AI will progressively deskill and create redundancy in the medical workforce.[209,210]

Resolving these tensions may be possible through seeing the educational opportunity and wider clinical application of the hardware enabling AI-ECG. Careful metrics as described previously will allow concerns about agency to be considered empirically, at least within the categories of patient data collected. Moreover, in the high-income world, the stethoscope has increasingly become an ornamental marker of physician status, supplanted by more readily available investigations that provide a definitive answer to cardiorespiratory differential diagnoses. Considering a familiar physical form factor (stethoscope) that, beyond AI, can now also readily display clinically useful waveforms learned about at medical school but rarely visualised or applied at the point of care. Thus, use of AI-ECG may have the unanticipated benefit of restoring the status of the stethoscope examination as a valuable cornerstone of patient assessment. In a move from science fiction to reality, an AI-enabled technology such as the Eko DUO exemplifies how the modern "stethoscope" may become more like the Tricorder used in Star Trek, as explored by Topol when recently writing in The Lancet.[211]

4.3.3 Data Rights

A central government, NHS-funded public screening programme making use of patients' own smartphones necessarily raises important questions about data rights. Beyond the expected guardrails required by general data protection ruling (GDPR) and other UK-specific health data legislation, AI-ECG introduces additional concerns. One is whether patient participants should be obligated, or enrolled via opt-in or opt-out mechanisms, to contribute their health data towards continuous refinement of the AI-ECG algorithms themselves. I note that while employed in this context by a public agency, the intellectual property for AI-ECG is held by Eko Health, the device manufacturer.

Thus while patients and communities may carry some expectation of potential personal or future benefit from algorithmic refinement, the more obvious rewards accrue to private entities for use elsewhere (but perhaps also the NHS). Ensuring that the NHS realises fair financial value from its data has recently been highlighted as a substantial policy failure.[212] Exchange of NHS data outside of the explicit purpose of supporting direct clinical patient care requires public support. Research shows that views about data sharing are mixed and that substantially more research is needed to understand these attitudes across all subpopulations.[213-215] People have conflicting views about how and under what circumstances data should be shared, and many are uncertain about the idea of commercial companies having any access to their health data.[216]

A related question, particularly for patients who use their own smartphones, involves what other data can be folded into the programme either for programme performance analysis or algorithmic refinement. For example, HF diagnoses may be meaningfully influenced by patient physical activity information, which can be readily gleaned from smartphone statistics even without additional wellness apps.[217] Incorporating actual wellness data from patients' phones, including information on heart rate, sleep metrics, and other connected wearables could sharply improve the accuracy and yield of AI-ECG.

Another potential opportunity, not lost on myself and my collaborators as overseers for the nascent AI-ECG programme, is the possibility that AI-ECG data linked to patients her records might support entirely new diagnostic discovery beyond the core cardiovascular conditions at issue. For example, the Mayo Clinic team responsible for AI-ECG have also evaluated the ability of ECG signals to detect SARS-CoV-2 infection.[218] Other conditions may similarly have subtle manifestations in ECG waveforms, phonocardiography, or their combination – invisible to humans but not AI – that could plausibly emerge from widespread use. Beyond just opt-in or opt-out permissions – known to be

problematic for meaningful engagement for patient consent[219] – what control ought patients to have around the use of their health data in this context?

Lastly, AI-ECG will need to consider data security carefully, including the possibility, however remote, of malicious intent or motivated intruders into the system. Health data is an asset for cyber criminals and can be monetized by trading on underground forums.[220] Cyber threat modelling should be performed by the device manufacturer early in the design phase to identify possible threats and their mitigations.[221] Documentation provided about embedded data security features adds valuable information to patients that may have concerns about the protection of their personal data, and can help in making informed decisions on using AI-ECG. Beyond privacy, threat modelling should also account for patient safety, such as from an intruder with access that allows manipulation of code or data. For example, it could be possible to manipulate results to deprive selected populations of appropriate referral for care. Sabotaging results or causing a denial-of-service situation by flooding the system with incorrect data might also cause damage to the reputation of the system in such a way that patients and clinicians become wary of using it.

Cyber incident response preparedness planning is key in dealing with such events.[222] This also mitigates incidents that cause system malfunctioning not necessarily through malicious intent. Misconfigurations, network outages, and software bugs are often the root cause of incidents that at first display the same characteristics as a cyber-attack. If the healthcare provider already has people and processes in place for responding to technical issues with the devices and apps, adding cyber incident response playbooks that describe escalation paths to crisis management adds additional safeguards.

Software is never perfect; the apps as well as the devices will require regular updates to capitalise on new features and protect against new software vulnerabilities that may be exploited by

an attacker. Software security patches would typically either be pushed out to the devices or require user interaction to download and install these. Updates pushed out in the background without any notification to the user could prove problematic e.g. causing confusion when the app suddenly changed its features, or more importantly, changing the way the app processes patient data in such a significant way that it violates the initial user consent. On the other hand, if critical security updates need user interaction to install, this might cause a significant delay in time that leaves the device vulnerable to cyber-attacks. Put succinctly, anticipating these security and other data rights considerations beyond the relatively superficial means of user agreements remains an unmet challenge for AI-ECG.

4.4 Final recommendations

This chapter outlined a novel clinical pathway to screen for cardiovascular disease using an at home, patient self-administered AI technology that can provide a screening capability beyond human expertise. I set this against a backdrop of, 1.) a diverse ecosystem of stakeholders impacted by and responsible for AI-ECG, spanning patients, NHS clinicians, NHS agencies, and responsible regulatory and health economic bodies; and 2.) a health policy landscape eager to progress "use of decision support and AI" as part of a wider push to decentralise (modernise) care.[27] To underscore the outlined considerations of equity, agency, and data rights, I propose two principal recommendations, framed against but generalisable beyond the pathway example of AI-ECG.

First, I advocate for a multi-agency approach that balances permissive regulation and deployment – to align with the speed of AI innovation – against ethical and statutory obligations to safeguard public health. Bodies such as NHS England, the MHRA, and NICE each have unique mandates and responsibilities, but with cross-cutting implications. The clinical and health economic case for urgent innovation for unmet needs, such as AI-ECG for HF, is obvious and compelling (Chapter 3). Agencies working sequentially delays translating such innovations into clinical practice[223,224], missing opportunities to avert substantial cardiovascular morbidity and mortality. Instead, identification of a potentially transformative technology should trigger a multi-agency approach that works together and in parallel to support timely deployment within clinical pathways to positively impact patient care. This approach holds not only for initial deployment, but also as technology progresses. Here I could consider the challenge of AI algorithms continually iterating (improving); for a given version of AI-ECG, the MHRA grants regulatory approval, NICE endorses commissioning, and NHS England guides implementation. What happens when an inevitable improvement to AI-ECG's algorithm is available? AI algorithms can continually iterate in parallel with

new data, machine learning methodology, and computational power.[225,226] Parodying the ship of Theseus – a thought experiment asking whether an object that has had all of its original components replaced remains the same object – at what point is the algorithm substantially different to the original, and what prospective validation, if any, is needed if the claims remain the same? Multi-agency working can bring together different but complementary expertise to reach a consensus on such questions and ideally avoids unfamiliarity with the lifecycle of AI disrupting delivery of care by reactively resetting when new (improved) versions arrive. Encouragingly, in a potential move towards multi-agency working, the NHS AI Lab (now part of NHS England) recently (2022) commissioned NICE to lead on a consultation for an updated digital health evidence standards framework that aims to aligns with regulators.[227]

Second, both to account for the ethical considerations outlined in this chapter and to balance any faster implementation of promising AI technologies, I recommend a robust centralised approach to carefully deploy and thoroughly evaluate programmes such as AI-ECG. This chapter has covered some of the critical variables to measure that will be unique to using an AI technology for patient self-administered screening at home. Forming a comprehensive list would again be amenable to a multi-agency approach, where NHS England can draw on the playbook for already monitoring existing national screening programmes. An evaluation framework consisting of agreed metrics should be considered not only an intrinsic but a mandatory part of the design, deployment, and ongoing surveillance of AI-ECG. The inherent connectivity and instant data flow of such technology offers, unlike screening programmes to date, the opportunity for real-time monitoring and therefore prompt intervention, not only for clinical indications, but also for any disparities in uptake, execution, algorithm performance, or cyber security. Ultimately this will not only bolster the NHS' position of as a world leader in standards for patient safety, but also as an exemplar system for realising effective AI-driven healthcare interventions.

The NHS demonstrated during the most challenging periods of the Covid-19 pandemic that healthcare can use technology to move fast and fix (sic) things.[228] Looking to the future for AI-ECG, seizing this momentum while also addressing the outlined ethical pitfalls may, in the short term, unlock both clinical and health economic benefits and blueprint best-practice that builds confidence for further applications. In the longer term, I see a convergence of commoditised AI algorithms for cardiovascular and wider disease, and continuous monitoring with increasingly sophisticated wearables, implanted devices, and internet-of-things – all with sensors capable or recording digital biomarkers far beyond ECG and phonocardiography alone.[229,230] In this future, home-based screening may become completely passive. While incrementing toward such a reality could unlock major public health benefits, doing so will depend on bold early use cases such as that described in this chapter to reveal unanticipated ethical challenges and allow these to be resolved. For now, the outlined policy recommendations can serve to underpin stewardship of such novel diagnostic pathways in a matter that preserves and promotes trust, patient engagement, and public health.

4.5 Chapter Conclusion

Patient self-administered screening for cardiovascular disease at home using an AI-powered technology offers substantial potential public health benefits but also poses unique ethical challenges. I recommend a multi-agency approach to the lifecycle of implementation for such an AI-powered pathway, combined with a centrally overseen, mandatory prospective evaluation framework that monitors for equity, agency, and data rights. Assuming the responsibility to proactively address any observed neglect of these considerations instils trust – as the foundation for sustainable and impactful implementation of AI technologies for clinical application within patients' own homes.

5 Summary of results

This chapter draws together a summary of my results chapters and how each of the aims and hypotheses outlined at the beginning of my thesis have been addressed. This is followed by aa chapter outlining a synthesis of this work, future directions underpinned by it, and an overall summary of the wider significance and impact of this body of research.

5.1 Summary of findings by aim and hypotheses

5.1.1 Chapter 2: Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during AI-ECG stethoscope examination: prospective, observational, multicentre, external validation study

In Chapter 2, I addressed the following aim and associated hypothesis:

- **Aim**: To conduct a large-scale independent, prospective, external validation study of AI-ECG applied to single-lead ECG recorded by an ECG-enabled stethoscope by recruiting NHS patients attending for echocardiography – ground truth for comparing screening performance
- **Hypothesis**: Detection of LVEF ≤40% using AI-ECG applied to single-lead ECG recorded by a smart stethoscope is reliable, accurate, operator-independent, unbiased; and feasibly for patient self-administration.

The outlined aim was achieved through the successful recruitment of over 1,000 NHS patients attending echocardiography, all of whom agreed to have an ECG-enabled stethoscope examination performed across multiple recording positions. This generated a substantial dataset of single-lead ECGs for interrogation with AI-ECG, tuned to detect LVEF ≤40%. This study represents one of the first-ever large-scale, prospective, ethnically diverse, multi-site, multi-operator, independent (unfunded and uninfluenced by device or AI software manufacturers), external validations studies of any AI technology. This means my study took a methodological approach that addressed head-on some of the major limitations of translational AI research.[231,232] Overall, as the most important starting point, my results show that the outlined workflow of recording a single-lead ECG with an ECG-

enabled stethoscope is suitable for AI-ECG to perform impressively: up to AUC 0.91, sensitivity 92%, and specificity 80%. This level of accuracy needs to be contextualised to other, reassuring findings: AI-ECG exhibited no bias (by age, gender, or non-White ethnicity), and was operator independent. This last point frames the opportunity for patient self-administered examination, which would appear to be feasible.

The findings from this study informed the interpretation of results in subsequent chapters. A health economic model for large-scale deployment of AI-ECG using the proposed approach of integration with a smart stethoscope examination is entirely dependent, first and foremost, on the test having sufficiently high accuracy. Similarly, embarking on devolving the administration of such a test into the hands of patients themselves requires sufficient confidence of performance, operator independence, and ease of use to justify such a clinical pathway.

5.1.2 Chapter 3: Survival and Health Economic Outcomes in Heart Failure by Place of Index Diagnosis: A Propensity-Matched Analysis

For Chapter 3, my aim and hypothesis were as follows:

- **Aim**: To measure the contemporary clinical and health economic implications of HF by the route to diagnosis (through primary care or hospital admission) and estimate the potential cost savings of increasing rates of diagnosis through point-of-care AI-ECG.
- **Hypothesis**: The mechanism for index diagnosis of HF remains dominated by hospital admission, with worse long-term clinical and health economic outcomes compared to patients first diagnosed in primary care to justify the cost of deploying point-of-care AI-ECG.

The above aim was addressed by interrogating a uniquely granular dataset (Discover-NOW) containing a comprehensive and contemporary record of linked primary and secondary care data as well as health economic cost data. Through a propensity-score-matched analysis of the nearly 35,000 patients diagnosed with HF in NWL between 2015-2020, I identified that index diagnosis of HF through community pathways conferred a short-term survival benefit and a substantial (£2,500) long-term health economic benefit. This was in the wider context of the overall route to index HF diagnosis remaining dominated (70%) by hospital admission. Taking the geography of NWL as an example, I modelled the potential cost savings of a deployment of AI-ECG and Eko DUO, on the principal assumption that this would increase rates of HF detection in the community, patients who would otherwise remain undiagnosed until their symptoms precipitated a hospital admission. I found that shifting the ratio by 10% to achieve 40% community pathway diagnosis (60% remaining as hospital pathway) would unlock net cost savings of well over £1.5 million for NWL alone. Taking an even more conservative approach, breaking even from the initial cost outlay of the technology would require only a 2% upward lift in patients (roughly 116 extra patients in one year) diagnosed through community pathways.

5.1.3 Chapter 4: Patient Self-administered Screening for Cardiovascular Disease Using Artificial Intelligence in the Home

Lastly, in Chapter 4, I addressed the following aim and hypothesis:

- **Aim:** To evaluate the ethical implication of extending AI-ECG technology to patient self-administered screening and specifiy policy recommendations to blueprint best practice.

- **Hypothesis:** Patient self-administration of AI-ECG raises ethical challenges including considerations of equity, agency, data rights, and ultimately responsibility for safe, effective, and trustworthy implementation.

Informed by findings form Chapter 2 and Chapter 3, this chapter applied a normative ethical analysis to address the outlined aim. Framed against a complex landscape of stakeholders including NHS, regulators, NICE, and – most importantly – patients, I interrogated questions around equity, agency, data rights, and responsibility for safe, effective, and trustworthy implementation of this powerful but novel diagnostic pathway. My analysis concluded with recommendations for safe, effective, and trustworthy implementation of any such pathway; that this needs to be anchored in effective multi-agency working – NHS, MHRA, NICE, patients, and beyond – and for any programme to be underpinned by a robust protocol for real-time evaluation across sustained high performance of the technology and equitable uptake across patient groups.

6 Synthesis

"The central idea of early disease detection and treatment is essentially simple. However, the path to its successful achievement (on the one hand, bringing to treatment those with previously undetected disease, and, on the other, avoiding harm to those persons not in need of treatment) is far from simple –though sometimes it may appear deceptively easy." From Wilson and Junger, 1968, [34]

This PhD thesis marks a novel body of work addressing important research questions on the clinical, health economic, and ethical considerations of AI-ECG delivered through an ECG-enabled stethoscope examination. Much of this work has framed application of this DHT in the context of screening, opportunistic and/or invited, both at the individual patient level and through a more programmatic, population-based approach. Despite HF being by a common cause of early morbidity and mortality, currently no national screening programmes exist for this disease area.

The UK National Screening Committee appraises the viability, effectiveness, and appropriateness of any proposed population screening programmes. To date, only five conditions are within the portfolio of NHS screening programmes for young people and adults: abdominal aortic aneurysm, bowel cancer, breast cancer, cervical cancer, and diabetic eye screening.[233] Additionally, though not considered a screening programme per se, the NHS Health Check is a health check-up for adults in England aged 40 to 74.[234] This is designed to detect early signs of stroke, renal disease, type 2 diabetes, dementia – and cardiovascular disease. At present this involves a 30-minute encounter, usually with a nurse or healthcare assistant, who asks questions about lifestyle and family history, measures height and weight, takes a blood pressure recording and draws a blood sample (for cholesterol levels). Adding a 12-lead ECG to this encounter for the purposes of overlaying screening for reduced LVEF using AI-ECG would likely be prohibitive due to demands on time, patient inconvenience (needing to dress down), expertise, and resource required to perform and interpret the result of this investigation. However, the approach investigated in this thesis, using a single-lead ECG acquired by an ECG-enabled stethoscope with automated AI-ECG detection of reduced ejection fraction, taking only 15-seconds, may pose an attractive addition to public health touchpoints such as the NHS Health Check. Here it would be remiss to not also acknowledge the obvious added value proposition of two further conditions screened for within the same technology: atrial fibrillation (using a form of AI-ECG) and valvular heart disease (using AI-phonocardiogram), both priority cardiovascular conditions in their

own right – and causes and consequences of HF. These two conditions could be detected in the same brief encounter using a tool such as the Eko DUO.

Considering the UK National Screening Committee's criteria for an effective screening programme, this PhD thesis has approached many of these, and overall highlights that an ECG-enabled stethoscope with automated AI-ECG detection of reduced ejection fraction may have a case to make for programmatic deployment. In considering a synthesis for the work outlined in these pages, the screening criteria are useful reference for holistically assessing the evidence generated, and for highlighting remaining gaps that would still need to be addressed before applying the outlined technology to clinical pathways and populations.

First, the Screening Committee requires the condition at hand to be an important public health problem as judged by its prevalence and/or severity. This criterion is certainly met by HF, which receives extensive mention in the NHS Long Term Plan and beyond given the substantial clinical and health economic burden it poses. A meta-analysis based on echocardiographic screening studies in the general population – thus also counting previously unrecognized cases – showed that the prevalence of 'all type' HF in high-income countries is around 11.8% in those aged 65 years and over.[235] The criteria also require the epidemiology, incidence, prevalence and natural history of the condition to be well understood, including development from latent to declared disease. A substantial body of evidence exists for this in HF, but there are gaps. This PhD thesis contributes a contemporary analysis on the epidemiology of HF with novel findings instructive for health policy. These include highlighting the continuing, disproportionate trend for index diagnosis taking place through hospital admission (70%), and that this route to diagnosis confers an increased medium-term mortality risk. Before the viability of a screening programme is considered, there is also a requirement to for all the cost-effective preventive interventions to have been implemented as far as practicable. Here I would highlight the disappointing record of programmes targeted at but failing to improve detection of HF in the

community[236], with interventions such as providing guidelines, education, and testing blood natriuretic peptide falling short.[237,238] To that effect, there have been no recent, widely deployed, and cost-effective preventative interventions for HF. There have been some small-scale, proactive but highly targeted attempts to increase the rates of detection using primary care electronic health record searches. One such approach focused on specifically increasing the detection of patients with left ventricular systolic dysfunction (LVEF<35%). This study was sponsored by Medtronic, incentivised by being a manufacturer of implantable cardiac devices (complex device therapy) indicated for improving symptoms and prognosis of patients with LVEF below 35%.[167] The study took a collaborative approach between primary care, cardiologists, and data scientists. By screening the primary care records for patients with either 1.) coded symptoms of HF without a diagnosis, or 2.) historic code for HF, the authors defined a cohort for invitation for cardiology consultation. They found 27% of HF patients identified were eligible for complex device therapy, 45% required medicines optimisation, and 47% of patients audited required diagnosis codes adding to their GP record. These results may be compelling, but the complexity of this approach limits the readiness for national scaling. Furthermore, as exemplified by this study and highlighted in Chapter 3, until the fidelity and quality of clinical coding achieves some standardisation across the NHS, geographic disparities in how effectively patients are detected through this approach would likely manifest, and may surface some of the ethical tensions explored in Chapter 4 – through the inconsistent accuracy of clinical coding precipitating negative socioeconomic and ethnic biases.

Second, the Screening Committee would only consider a national screening programme that applies a "simple, safe, precise and validated screening test". My research has addressed some of these considerations. Though the underlying technology behind AI-ECG is highly complex, applying it through a single-lead ECG enabled stethoscope is relatively straightforward. Based on my study examining over 1,000 patients without concern, with 97.5% of those approached participating,

delivering the test this way is clearly both safe and acceptable. The Screening Committee requires that "the distribution of test values in the target population should be known and a suitable cut-off level defined and agreed". How to weigh the precision of such an approach remains one of the fundamental challenges not reconciled by this thesis; where to set the threshold for AI-ECG, and how to balance sensitivity vs. specificity? This is a universal challenge for DHTs using an AI-based classifier. Taking one recording from position 2 (pulmonary) as an example, two different thresholds for the binarisaion of LVEF prediction into >40% or ≤40% achieve substantially different balance of performance. At a threshold of 0.43, sensitivity and specificity were 71.7% and 86.5%, respectively. At a threshold of 0.34, the balance of sensitivity and specificity reverses to 84.8% and 69.5%, respectively. Ideally (and likely), AI-ECG will go through continued iteration as more training data, refinements in ML methodology, and computing power continue to steadily increment towards developing versions of AI-ECG with thresholds that achieve sensitivity and specificity closer to 100%. Until then, it could be argued that against a background of a very high prevalence of undiagnosed HFrEF (Chapter 3) and the availability of prognostically beneficial therapies, sensitivity should be prioritised over specificity to maximise the rates of early detection. Deployment of a such a high sensitivity tool in primary care could draw on the clinical and health economic impact of index diagnosis in this setting, as observed in Chapter 3, to weigh-up the downstream impact of lower specificity i.e. false positives and the burden of additional investigations e.g. blood tests and echocardiography triggered by these. I note here that my study described in Chapter 2 was not one of an unselected population; studying patients attending for echocardiography will, by definition, enrich for cardiovascular disease and HF specifically. However, encouraging parallel adherence to the existing NICE pathway may buttress against false positives by ensuring patients additionally receive natriuretic peptide testing when faced with a positive AI-ECG result, though the latest ESC guidelines recommend direct referral for echocardiography (bypassing natriuretic peptide testing) if clinical suspicion is sufficiently high; this is at odds with the

NICE pathway continuing to suggest natriuretic peptide testing first to inform a 2- or 6-week triage to echocardiography and specialist review. Here, the availability of new diagnostic technology for HF would require the NICE pathway to be updated, and in so doing would align with criteria from the Screening Committee: "There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals."

Third, there is a requirement that effective interventions exist for patients identified through screening at a pre-symptomatic phase, where interventions lead to better outcomes for the screened individual compared with usual care. Asymptomatic reduction in LVEF is present in 3–6% of the general population, where commencement of therapies can relieve future symptom burden and reduce long-term risk of mortality.[239-241] The benefits of such early detection on family members and carers are highlighted as also meriting consideration. Evidence relating to wider benefits of screening, for example, those relating to family members being able to delay the need to assume caring responsibilities, should be considered. Beyond the scope of this thesis, caring for patients with HF poses a substantial physical and mental burden on informal carers (usually family members of close friends) – with wide reaching societal implications.[242,243]

Fourth, unsurprisingly the Screening Committee requires any national screening approach to be underpinned by robust evidence from high-quality randomised controlled trials that observe a reduction in mortality or morbidity as their primary end points. The process of conducting such large-scale RCTs offers opportunity to also address broader questions around the programme's clinical, social, and ethical acceptability to health professionals and the patient public. The expectation would be that the "benefit gained by individuals from the screening programme should outweigh any harms, for example from overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications." The ethical analysis outlined in Chapter 4 offers a starting point for these discussions and critiques and goes further by also highlighting how implementation of AI-ECG – and

indeed, any self-administered, AI-driven screening – would surface challenges not yet faced by existing national screening programmes. These include the need to monitor a continually iterating (improving, hopefully) technology for new biases, addressing the digital divide, and cybersecurity concerns. One of the benefits of the Covid-19 pandemic may be that the substantial acceleration of the societal uptake of DHTs proves to be sustained. One such population-wide tool involves the continuing use of the NHS App as a form of vaccine passport.[244] Whether this generation's savvy for DHTs will be carried forward as they age remains an assumption, but if this proves to be the case then patient self-examination may be increasingly tractable as the digital divide narrows. The aforementioned randomised controlled trials would also be expected to meet the criteria for informing if the opportunity-cost of a screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) is economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criterion should have regard for evidence from cost-benefit, cost-effectiveness, and/or cost-consequence analyses. Chapter 3 has generated several useful variables for making this health economic case, however further data on downstream costs e.g. impact of false positive results are needed. These can be modelled but ultimately the most reliable evidence requires real world deployment through a medium such as a randomised controlled trial. It is increasingly established that the NHS always "runs hot", at capacity, reducing consideration of programmes/interventions that require several years to manifest a health economic payoff. All patients with positive AI-ECG should ultimately have an echocardiogram. For illustrative purposes, considering a very simplistic approach to a number-needed-to-treat, based on NHS tariff cost of one echocardiogram (£59.00) and a PPV of 20%, the cost of imaging five patients with a positive AI-ECG test (one of whom will have HFrEF) would run to £295.00. This would seem palatable given patients diagnosed through primary care pathways may unlock at £2,500 saving (Chapter 3). However, given both the artificial scarcity of echocardiography services – the only finite

resource being funding – and the financial short-termism of a system under pressure, justifying a business case for increased echocardiography services to cope with more referrals due to positive AI-ECG may prove challenging due to the relatively lower tangibility of future savings. There is also some tension here between the demands of meeting a high evidence standards threshold and accumulating urgent public health need that could be addressed by novel technology that is inherently rapidly iterating. Rather than negotiate prices, NICE makes recommendations for system-wide funding, and therefore, deployment, principally on the basis of using tools such as quality adjusted life years. NICE's remit now also extends to evaluating the value of DHT's. However, there is some reservation about using incumbent tools for assessing the value of DHTs given these have principally been developed for pharmaceutical interventions. The value of DHTs does not only depend on clinical and economic aspects but also on technical features, perceived benefits for clinical staff, willingness to use by end-users, and lastly, the healthcare system's capacity to benefit from the innovation.[245]

Finally, any new programmatic deployment of a screening tool would necessitate a robust plan for communication with patients and managing and monitoring the programme with an agreed set of quality assurance standards. This would also require adequate staffing and facilities for testing, diagnosis, treatment, and programme management prior to commencement. Unlike some of the existing national screening programmes, which variously require as their first step radiological investigations, specialist procedures, invasive procedures, and patient attendance at a facility, the first-step approach with AI-ECG could be completely decentralised and as outlined in Chapter 4, take place in the patient's own home at their own convenience. This is well aligned with policy shifts encouraging "place-based care", aiming to "blur institutional boundaries across a location to provide integrated care for individuals, families and communities" and address the imperative to "meet patients where they are."[246] Further work would be needed to generate an evidence-based approach to informing patients; explaining the purpose of AI-ECG, and potential consequences of screening, to

assist them in making an informed choice. Notably, the Screening Committee suggests that early success ought to anticipate "public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process." My PhD broaches the eligibility of those who should be considered for an AI-ECG test, spanning its opportunistic overlay onto different clinical pathways including every routine stethoscope examination to a more targeted screening approach on the basis of symptoms (e.g. breathlessness) or co-morbidities (Chapter 3 demonstrated higher prevalence of multimorbidity e.g. coronary artery disease, hypertension, diabetes) (figure 6.1). Clearly whether it needs to be a clinician administering the test is also up for debate. The downstream protocol for a positive test will ultimately require gold-standard echocardiography, a relatively limited resource,.

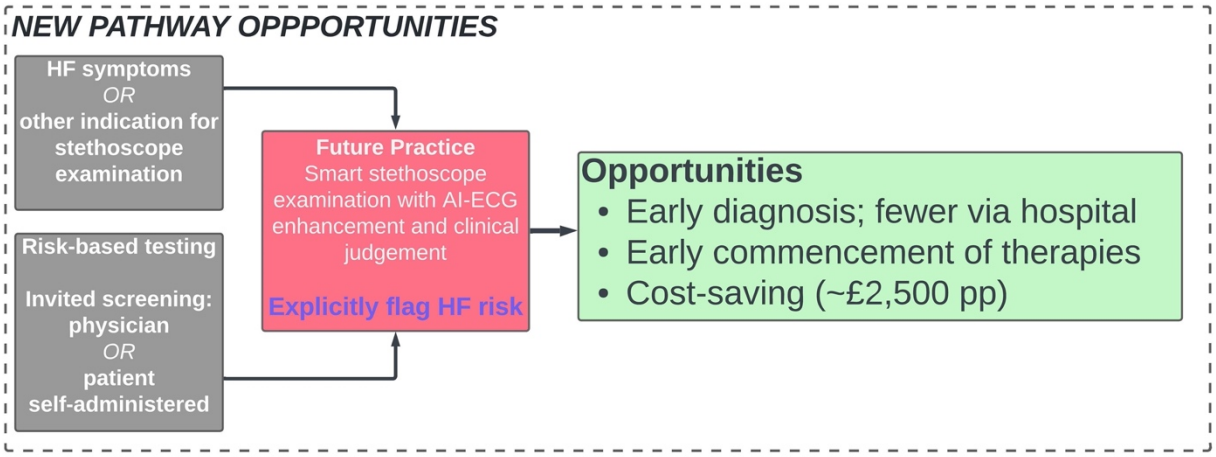


Figure 6.1 New clinical pathway opportunities using AI-ECG

Given that HFrEF somewhat paradoxically has both the worst prognosis and plentiful evidence-based therapies, a case for intelligent triage and prioritisation beyond natriuretic peptide testing could be made. Collectively, the work presented in this PhD thesis lays a foundation of evidence that can be

drawn on to ensure any decisions about these criteria are scientifically, economically, and ethically justifiable to the public.

7 Future work

At the turn of the 21st century the paradigm for healthcare remained anchored to a late industrial age model where paternalistic physicians governed over patient care from large specialist centres.[247] The predicted casting off towards an information age, characterised by a patient-centric philosophy for healthcare delivery, was not being realised. While healthcare remained isolated from the digital and mobile technology revolution, other industries such as banking, travel, and retail redefined how citizens interacted with their offerings. Over twenty years on, healthcare's progress remains slow. The reasons for this are myriad and complex and include the need for DHTs to cross a necessarily high threshold of evidence that goes beyond simply demonstrating that the technology performs a well-circumscribed task with high accuracy. The body of work outlined in this PhD thesis contributes scrutiny of a DHT from multiple, complimentary research perspectives, serving as part of a blueprint for the holistic evaluation of DHTs that will help underpin their widespread uptake and impact. The work presented in these pages has initiated a substantial portfolio of further research that will continue to build the case for DHTs using AI. Built on a foundation of work presented in this thesis, I outline below some of the future directions for this field of research.

7.1 AI-ECG as a consumer technology

For a century, the standard 12-lead ECG has been an essential diagnostic tool limited to clinician use. The ECG remains among the most accessible, cost-effective, and safe means to assess cardiac activity and function. Recent technological and computing advances have progressed this on two fronts: (1) the advent of wearable technologies has shifted the ECG into the consumer space, and (2)

AI interpretation software has expanded the utility ECG for the non-expert. For now, the detection of heart rhythm abnormality, specifically atrial fibrillation, remains the dominant use case.[90,136,137,200]

In Chapter 2, I described my study that took the approach of recording single-lead ECGs from standard precordial positions as well as one hand-held position. The latter somewhat emulates the method and vector for single-lead ECG recordings using some of the most popular consumer devices e.g. Alivecor Kardia and Apple Watch, and I found this position, while still impressive, was most frequently unable to be attempted e.g. due to patient weakness, and had a lower rate of adequate signal quality compared with precordial positions. For ECG acquired by wearables or otherwise, the key determinant of any AI model's performance is the quality of the input data. Automated ECG analysis relies on and assumes optimal signal acquisition. Models will also need to account for biosignal variation as novel means of ECG signal acquisition continue to increase in the consumer market. The single-lead version of AI-ECG referred to in this thesis was trained on all 12 individual ECGs constituting a full 12-lead ECG; and AI-ECG can attempt a prediction using any ECG signal of adequate quality. One approach will be to prospectively or retrospectively obtain the single-lead ECG waveforms recorded by devices other than the Eko DUO e.g. wearables, and evaluate AI-ECG in this context – with the need for synchronous echocardiography as ground truth. Notably, some of these technologies are well suited to repeated single-lead ECG recordings that may enable tracking changes in LVEF over time.

In the event of sustained, clinically meaningful performance, full transition of AI-ECG for HF to direct consumer access seems inevitable. Here, the potential opportunities for improvements in public health would need to be balanced against equally salient and concerns that uses of such consumer AI DHTs cause overutilisation of healthcare resources (through false positives)[248] and risk deepening health inequities if not universally accessible.

7.2 Ensemble AI for detection of heart failure

As might be expected, as shown in Chapter 2, the performance of AI-ECG can be augmented through use of multiple recordings. Performance would likely also be improved through a multimodal model processing additional variables such as age and sex. Considering the use of AI with an ECG-enabled stethoscope to detect HF, the principal waveform instinctively associated with such hardware – the phonocardiography (PCG) or heart sounds – is currently not contributing to informing on the presence of HF. However, PCG has been shown to correlate with changes in LVEF.[249,250] It is widely established that PCG is impacted by structural and mechanical features of the heart, and abnormalities in this manifest in or may be precipitants of HF e.g. valvular heart disease and flow murmurs. Much like AI-ECG for detection of HFrEF, it may be that the PCG waveform can serve as a digital biomarker, containing specific acoustic signatures beyond the capabilities of humans to detect. To capitalise on this additional richness of data, the architecture of neural networks could readily accommodate an ensemble algorithm; taking both ECG and PCG waveforms for combined analysis and likely unlocking an increment in predictive accuracy. This work is underway.

7.3 Do ECG changes precede deterioration in LVEF?

The original publication describing use of AI-ECG for detection of LVEF below 35% from 12-lead ECGs found that patients with an LVEF above 35% on echocardiography but with a positive AI-ECG – ostensibly a false positive result – were at 4 times the risk (hazard ratio, 4.1 [3.3-5.0]) of developing future ventricular dysfunction compared with those with a negative AI-ECG screen.[92] This occurred during a median follow-up period of 3.4 years (IQR, 1.2–6.8). Given this background, I designed the protocol for the study outlined in Chapter 2 to encompass ethical approval from the

Health Research Authority to follow up similarly "false positive" patients longitudinally. Loss to follow up will be a substantial challenge here, however research tools such as the NWL-wide Discover-Now database could mitigate this by, without the need to make contact, highlighting which patients have since gone on to receive a new coded diagnosis of HF. Should future studies discover that AI-ECG has sufficient predictive capability for HF and/or other cardiovascular conditions, the intersection with wearable, consumer technology capable of recording ECGs may offer an opportunity for large-scale application for disease prevention.

7.4 Interrogation of AI-ECG features

The technology powering AI-ECG has created a double-edged sword: the more complex the task we ask it to perform, the less we know how it can do it. Generally, as the field moves forward, we will need to work toward learning what and understanding why AI considers certain features more relevant than others. There are some methodological approaches that could hint at this. For example, saliency maps can highlighted the primacy of the regions in an ECG that are heavily weighted features in the model's decision-making process.[251] However, as model complexity increases, saliency maps become increasingly less … salient, and essentially revert to the black box model that mires trust in complex AI models. There is an argument that a lack of mechanistic understanding should not risk the abandonment of these potentially transformative technologies, but instead we should seize the opportunity and do so safely through systematically and thoroughly scrutinising for bias and shifts in performance, work which can build on the content outlined in all chapters.

7.5 AI-ECG for HFmrEF and HFpEF

In HF with mildly reduced EF (HFmrEF), post-hoc and subgroup analyses suggest that drugs that are effective in HFrEF may also be effective in HFmrEF.[252] Trials in HFmrEF and HFpEF have historically not been positive, but in 2020 and 2021 inhibitors of the sodium–glucose co-transporter 2/1 and 2 were shown to be effective also in HFpEF.[253,254] For now, HFrEF remains both the most common subgroup and the only one with abundant evidence for effective therapies. This will likely change, highlighting the need for tools such as AI-ECG to extend their offering to being able to accurately screen for the full spectrum of LVEF across HF subtypes. This may be achievable if AI-ECG lives up to ML's capacity for continual iteration and improvement as more data become available. Specifically, AI-ECG may transition from being a classification tool – giving an indication of LVEF being above or below a certain threshold – to using regression, thereby displaying a direct prediction of a patient's percentage LVEF.

7.6 Towards a Tricorder

Science fiction has long served as a canvas for exploring the possible futures of health and healthcare. Among the more well-known futuristic medical devices, Star Trek's "tricorder" has set the ambition for the ultimate point of care diagnostic tool.[211] A medical tricorder takes the appearance of a small, grey, hand-held device with a flip-out panel to allow for a larger screen. In the drama posed by an unwell patient, such a tricorder – powered by AI – would be used to externally "scan" the subject to infer a full picture of their physical health.

Considering the DHT investigated by this thesis – a sophisticated convergence of hardware, software, connectivity, and AI – the prospect of a tricorder may already be exiting science fiction and

entering reality. The clinical examination remains an important quantifiable contributor to the diagnosis of HF and disease more generally[255], where the stethoscope is a key tool for gathering data in the process of sleuthing towards a diagnosis. However, for many clinical professionals, the stethoscope has become nothing more than an ornamental marker of physician status, supplanted by ever more readily available, more objective investigations. For example, why listen to the heart when one can perform a point-of-care ultrasound with a probe that plugs into a smartphone?[256] To an extent, widespread use of such technologies remains limited by cost and a requisite for substantial technical expertise. However, the cost of such technology continues to decline, becoming increasingly affordable, and indeed the requirement for technical expertise is being addressed by AI itself. My study has highlighted that some vectors and positions for recording single-lead ECG will maximise the performance of AI-ECG predictions. Here, AI-guided acquisition of the "optimum" recording could steer any user towards completing a high-quality screening test. This parallels efforts in echocardiography to automate the acquisition of optimum images for capturing key variables such as LVEF.[257,258]

Overall, as an ever-increasing number of sensor technologies get smaller, cheaper, more accurate and powerful, and in parallel AI insights begin to surpass human capability, the modern stethoscope may indeed become a relic of late industrial age healthcare, replaced by an information age tricorder-like device that, again through AI, enables patients to apply the technology directly onto themselves. As I approached the end of my PhD, research colleagues have assumed responsibility to test the clinical pathway outlined in Chapter 4, posting an Eko DUO to patients for self-examination and application of AI-ECG to screen for triple cardiovascular disease (HFrEF, AF, and valve disease). This ongoing research may in future prove instructive in the delivery of a democratised and decentralised model of healthcare – the aspiration for this century.

7.7 Real-world, prospective application of AI-ECG

The work presented in this PhD thesis has prompted substantial organic interest from both national health policy leadership as well as more locally from primary care colleagues focusing on cardiovascular population health. In collaboration with these stakeholders and underpinned by the work presented in these pages, one of the major, next steps for this research is to progress to a randomised controlled trial. Though the opportunities for future work in this field all carry scientific merit, prioritisation should be for the translational, prospective application of AI-ECG, in a way that test if such technology remains robust against the challenges of deployment in real-world health system settings. We are chasing the promised clinical, health economic, and patient-centric outcomes of this technology.

Carrying the acronym TRICORDER ("Triple Point-of-Care Cardiovascular Disease Screening With Artificial Intelligence Enabled Stethoscope"), this project will set out to deliver a DHT research study that meets a number of important firsts. Taking the form of a pragmatic, real-world, cluster-randomised controlled trial, TRICORDER will randomise deployment of Eko DUO and a suite of in-built AI tools (for detection of HFrEF, AF, and valve disease) across 250 primary care practices in the NHS (figure 7.1). The ambition will be to deploy across both NWL and North Wales, thereby encompassing geographies with unique sociodemographic, ethnic, and urban/rural diversity.

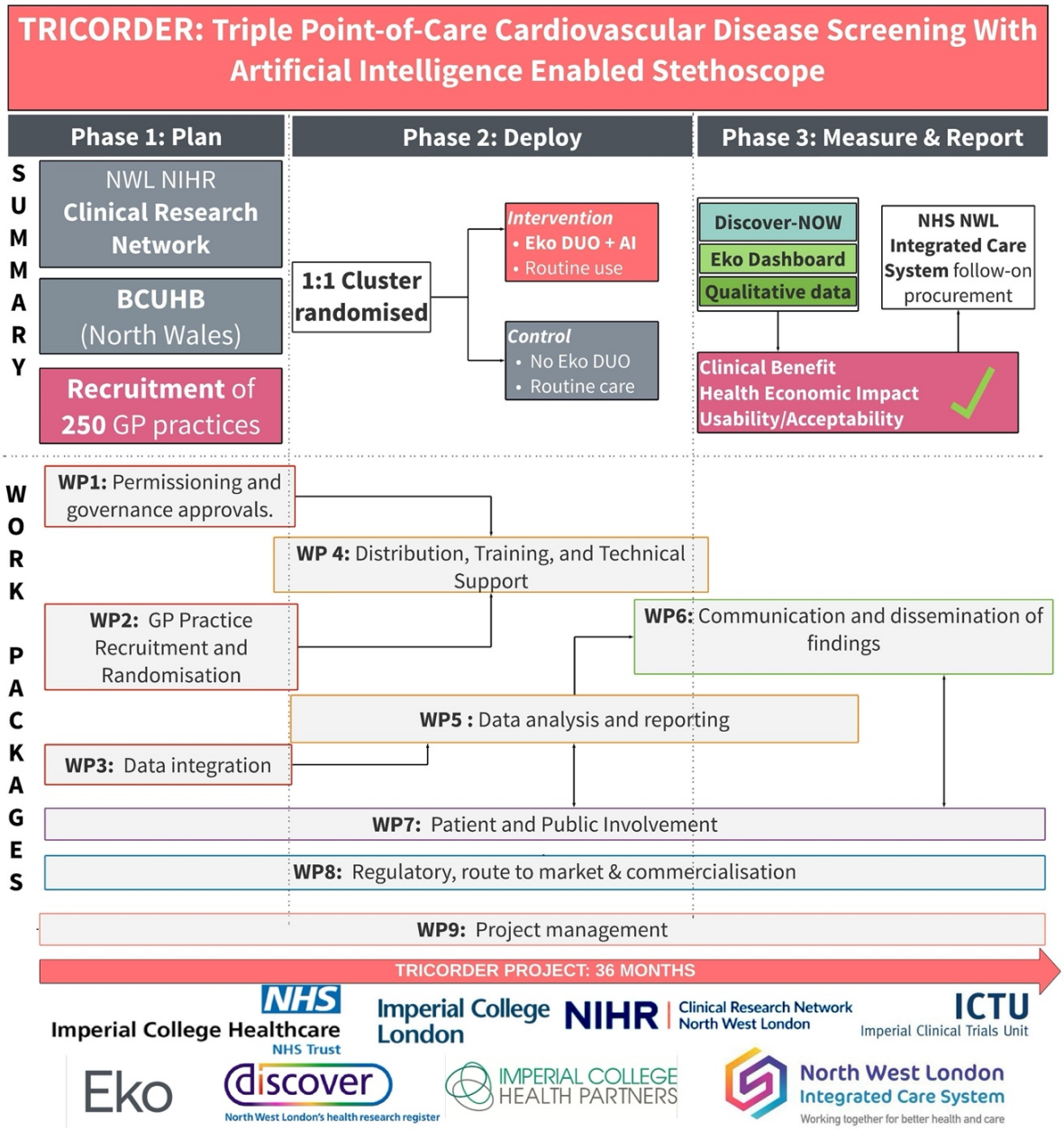


Figure 7.1 Study schematic for TRICORDER cluster randomised controlled trial.

WP: work package; NWL: North West London; NIHR: National Institute for Health and Care Rersearch; BCUHB: Besti Betsi Cadwaladr University Health Board; GP: General Practitioner.

8 Thesis conclusion

As an exemplar of an AI DHT, AI-ECG applied to single-lead ECG recorded by a smart stethoscope could address the public health priority to improve detection of HFrEF, and potentially extend to further common cardiovascular diseases. I have demonstrated that for HFrEF, this tool is reliable, accurate, quick, unbiased, and usable by patients themselves. Faced with the challenge of drastic underdiagnosis of HF in primary care, deployment of this tool would be underpinned by a compelling clinical and health economic argument. Intersecting with this, important ethical considerations have been outlined and framed against potential health policy approaches to balance public risk with public health. This thesis marks novel work that can directly serve patients and health systems; arriving in the wake of substantial momentum to transform healthcare, at the start of a decade where evidence generation can build the clinical and public trust needed to fully realise the potential of DHTs incorporating AI.

9 References

1. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *European heart journal* 2021; **42**(36): 3599-726.

2. Kemp CD, Conte JV. The pathophysiology of heart failure. *Cardiovascular Pathology* 2012; **21**(5): 365-71.

3. Bloom MW, Greenberg B, Jaarsma T, et al. Heart failure with reduced ejection fraction. *Nature reviews Disease primers* 2017; **3**(1): 1-19.

4. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *European journal of heart failure* 2020; **22**(8): 1342-56.

5. Rossignol P, Hernandez AF, Solomon SD, Zannad F. Heart failure drug treatment. *The Lancet* 2019; **393**(10175): 1034-44.

6. Fudim M, Abraham WT, von Bardeleben RS, et al. Device therapy in chronic heart failure: JACC State-of-the-Art Review. *Journal of the American College of Cardiology* 2021; **78**(9): 931-56.

7. Long L, Mordi IR, Bridges C, et al. Exercise-based cardiac rehabilitation for adults with heart failure. *Cochrane Database of Systematic Reviews* 2019; (1).

8. Faris RF, Flather M, Purcell H, Poole-Wilson PA, Coats AJ. Diuretics for heart failure. *Cochrane database of systematic reviews* 2012; (2).

9. Bœuf-Gibot S, Pereira B, Imbert J, et al. Benefits and adverse effects of ACE inhibitors in patients with heart failure with reduced ejection fraction: a systematic review and meta-analysis. *European Journal of Clinical Pharmacology* 2021; **77**(3): 321-9.

10. Gilstrap L, Austin AM, Gladders B, et al. The association between neurohormonal therapy and mortality in older adults with heart failure with reduced ejection fraction. *Journal of the American Geriatrics Society* 2021; **69**(10): 2811-20.

11. Yan Y, Liu B, Du J, et al. SGLT2i versus ARNI in heart failure with reduced ejection fraction: a systematic review and meta-analysis. *ESC heart failure* 2021; **8**(3): 2210-9.

12. Packer M, Anker SD, Butler J, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. *New England Journal of Medicine* 2020; **383**(15): 1413-24.

13. Burnett H, Earley A, Voors AA, et al. Thirty years of evidence on the efficacy of drug treatments for chronic heart failure with reduced ejection fraction: a network meta-analysis. *Circulation: Heart Failure* 2017; **10**(1): e003529.

14. Andronic AA, Mihaila S, Cinteza M. Heart failure with mid-range ejection fraction–a new category of heart failure or still a gray zone. *Maedica* 2016; **11**(4): 320.

15. Lekavich CL, Barksdale DJ, Neelon V, Wu J-R. Heart failure preserved ejection fraction (HFpEF): an integrated and strategic review. *Heart failure reviews* 2015; **20**(6): 643-53.

16. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. *Nature Reviews Cardiology* 2016; **13**(6): 368-78.

17. Roger VL. Epidemiology of heart failure: a contemporary perspective. *Circulation research* 2021; **128**(10): 1421-34.

18. Bottle A, Goudie R, Bell D, Aylin P, Cowie MR. Use of hospital services by age and comorbidity after an index heart failure admission in England: an observational study. *BMJ open* 2016; **6**(6): e010669.

19. Alpert CM, Smith MA, Hummel SL, Hummel EK. Symptom burden in heart failure: assessment, impact on outcomes, and management. *Heart failure reviews* 2017; **22**(1): 25-39.

20. Gerber Y, Weston SA, Redfield MM, et al. A contemporary appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA internal medicine* 2015; **175**(6): 996-1004.

21. Tsao CW, Lyass A, Enserro D, et al. Temporal trends in the incidence of and mortality associated with heart failure with preserved and reduced ejection fraction. *JACC: Heart Failure* 2018; **6**(8): 678-85.

22. NICE. Resource impact report: Chronic heart failure in adults: diagnosis and management (NG106). 2018.

23. Lesyuk W, Kriza C, Kolominsky-Rabas P. Cost-of-illness studies in heart failure: a systematic review 2004–2016. *BMC Cardiovascular Disorders* 2018; **18**(1).

24. Kaplon-Cieślicka A, Benson L, Chioncel O, et al. A comprehensive characterization of acute heart failure with preserved versus mildly reduced versus reduced ejection fraction–insights from the ESC-HFA EORP Heart Failure Long-Term Registry. *European Journal of Heart Failure* 2022; **24**(2): 335-50.

25. Tran RH, Aldemerdash A, Chang P, et al. Guideline-directed medical therapy and survival following hospitalization in patients with heart failure. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 2018; **38**(4): 406-16.

26. Packer M, Metra M. Guideline-directed medical therapy for heart failure does not exist: a non-judgmental framework for describing the level of adherence to evidence-based drug treatments for patients with a reduced ejection fraction. *European Journal of Heart Failure* 2020; **22**(10): 1759-67.

27. NHS. The NHS Long Term Plan. 2019.

28. Bottle A, Kim D, Aylin P, Cowie MR, Majeed A, Hayhoe B. Routes to diagnosis of heart failure: observational study using linked data in England. *Heart* 2018; **104**(7): 600-5.

29. Rørth R, Jhund PS, Yilmaz MB, et al. Comparison of BNP and NT-proBNP in patients with heart failure and reduced ejection fraction. *Circulation: Heart Failure* 2020; **13**(2): e006541.

30. NHS England DoHSC. Evidence for the introduction of Serum NP testing to the diagnostic pathway for Heart failure. 2010.

31. Madamanchi C, Alhosaini H, Sumida A, Runge MS. Obesity and natriuretic peptides, BNP and NT-proBNP: mechanisms and diagnostic implications for heart failure. *International journal of cardiology* 2014; **176**(3): 611-7.

32. Ledwidge M, Gallagher J, Conlon C, et al. Natriuretic peptide–based screening and collaborative care for heart failure: the STOP-HF randomized trial. *Jama* 2013; **310**(1): 66-74.

33. Committee UNS. Guidance: Criteria for a population screening programme. 2022.

34. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization* 2008; **86**(4): 317-9.

35. van Riet EE, Hoes AW, Limburg A, Landman MA, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. *European journal of heart failure* 2014; **16**(7): 772-7.

36. Morgan S, Smith H, Simpson I, et al. Prevalence and clinical characteristics of left ventricular dysfunction among elderly patients in general practice setting: cross sectional survey. *Bmj* 1999; **318**(7180): 368-72.

37. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. *Family practice* 2017; **34**(2): 161-8.

38. Dahlstrom U, Hakansson J, Swedberg K, Waldenstrom A. Adequacy of diagnosis and treatment of chronic heart failure in primary health care in Sweden. *European journal of heart failure* 2009; **11**(1): 92-8.

39. Smeets M, Van Roy S, Aertgeerts B, Vermandere M, Vaes B. Improving care for heart failure patients in primary care, GPs' perceptions: a qualitative evidence synthesis. *BMJ open* 2016; **6**(11): e013459.

40. Cowie MR. The heart failure epidemic: a UK perspective. *Echo research and practice* 2017; **4**(1): R15-R20.

41. Darvishi S, Behnam H, Pouladian M, Samiei N. Measuring left ventricular volumes in two-dimensional echocardiography image sequence using level-set method for automatic detection of end-diastole and end-systole frames. *Research in Cardiovascular Medicine* 2013; **2**(1): 39.

42. National Institute for Health and Care Excellence Evidence standards framework for digital health technologies: National Institute for Health and Care Excellence, 2018.

43. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nature Medicine* 2021; **27**(2): 186-7.

44. Cabitza F, Zeitoun J-D. The proof of the pudding: in praise of a culture of real-world validation for medical artificial intelligence. *Annals of translational medicine* 2019; **7**(8).

45. Uchegbu I, Rodriguez-Guadarrama Y, Rua T, Kartha MR, Abdulkareem M, Petersen SE. Health Economics, Economic Evaluation and Artificial Intelligence Technology. Artificial Intelligence in Cardiothoracic Imaging: Springer; 2022: 535-47.

46. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *Journal of medical Internet research* 2020; **22**(2): e16866.

47. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC medicine* 2019; **17**(1): 1-7.

48. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *The Lancet* 2020; **395**(10236): 1579-86.

49. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations* 2020; **6**(2).

50. Unsworth H, Dillon B, Collinson L, et al. The NICE evidence standards framework for digital health and care technologies–developing and maintaining an innovative evidence framework with global impact. *Digital health* 2021; **7**: 20552076211018617.

51. Bachtiger P, Plymen CM, Pabari PA, et al. Artificial intelligence, data sensors and interconnectivity: future opportunities for heart failure. *Cardiac Failure Review* 2020; **6**.

52. Komkov HB. Reservoir Computing with Boolean Logic Network Circuits: University of Maryland, College Park; 2021.

53. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology* 2019; **103**(2): 167-75.

54. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology* 2018; **29**(8): 1836-42.

55. Saba L, Biswas M, Kuppili V, et al. The present and future of deep learning in radiology. *European journal of radiology* 2019; **114**: 14-24.

56. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine* 2018; **24**(11): 1716-20.

57. Edgcomb JB, Zima B. Machine Learning, Natural Language Processing, and the Electronic Health Record: Innovations in Mental Health Services Research. *Psychiatric Services* 2019; **70**(4): 346-9.

58. Topol E. Deep medicine: how artificial intelligence can make healthcare human again: Hachette UK; 2019.

59. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. *The Lancet Digital Health* 2022; **4**(5): e384-e97.

60. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine* 2019; **25**(9): 1337-40.

61. Schulam P, Saria S. Can You Trust This Prediction? Auditing Pointwise Reliability After Learning. In: Kamalika C, Masashi S, editors. Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research: PMLR; 2019. p. 1022--31.

62. Gehl Sampath P. Governing Artificial Intelligence in an Age of Inequality. *Global Policy* 2021.

63. Brian D. Smedley AYS, Alan R. Nelson. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care: National Academies Press; 2003.

64. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. *The Lancet Digital Health* 2022; **4**(6): e406-e14.

65. Cabellos-García AC, Castro-Sánchez E, Martínez-Sabater A, et al. Relationship between Determinants of Health, Equity, and Dimensions of Health Literacy in Patients with Cardiovascular Disease. *International Journal of Environmental Research and Public Health* 2020; **17**(6): 2082.

66. Kreatsoulas C, Anand SS. The impact of social determinants on cardiovascular disease. *Canadian Journal of Cardiology* 2010; **26**: 8C-13C.

67. O'Dowd A. Major global cyber-attack hits NHS and delays treatment. *BMJ* 2017; **357**: j2357.

68. Ghafur S, Grass E, Jennings NR, Darzi A. The challenges of cybersecurity in health care: the UK National Health Service as a case study. *The Lancet Digital Health* 2019; **1**(1): e10-e2.

69. Martin G, Ghafur S, Kinross J, Hankin C, Darzi A. WannaCry a year on. *BMJ* 2018: k2381.

70. Almohri H, Cheng L, Yao D, Alemzadeh H. On threat modeling and mitigation of medical cyber-physical systems. 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2017: IEEE; 2017. p. 114-9.

71. Spanakis EG, Bonomi S, Sfakianakis S, et al. Cyber-attacks and threats for healthcare–a multi-layer thread analysis. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2020: IEEE; 2020. p. 5705-8.

72. Burki T. The dangers of the digital age. *The Lancet Digital Health* 2019; **1**(2): e61-e2.

73. Topol E. The Topol Review. *Preparing the healthcare workforce to deliver the digital future* 2019: 1-48.

74. Makri A. Bridging the digital divide in health care. *The Lancet Digital Health* 2019; **1**(5): e204-e5.

75. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. *Journal of general internal medicine* 2015; **30**(3): 284-9.

76. Petrozzino C. Who pays for ethical debt in AI? *AI and Ethics* 2021; **1**(3): 205-8.

77. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ digital medicine* 2020; **3**(1): 1-4.

78. Duckworth C, Chmiel FP, Burns DK, et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. *Scientific Reports* 2021; **11**(1).

79. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *The Lancet Digital Health* 2021; **3**(3): e195-e203.

80. Panch T, Duralde E, Mattie H, et al. A distributed approach to the regulation of clinical AI. *PLOS Digital Health* 2022; **1**(5): e0000040.

81. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine* 2019; **25**(1): 30-6.

82. Westerlund AM, Hawe JS, Heinig M, Schunkert H. Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence. *International Journal of Molecular Sciences* 2021; **22**(19): 10291.

83. Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nature Reviews Cardiology* 2021; **18**(8): 600-9.

84. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology* 2021; **18**(7): 465-78.

85. Seetharam K, Shrestha S, Sengupta PP. Artificial Intelligence in Cardiovascular Medicine. *Current Treatment Options in Cardiovascular Medicine* 2019; **21**(5).

86. Wright J, Regele O, Kourtis L, et al. Evolution of the digital biomarker ecosystem. *Digital Medicine* 2017; **3**(4): 154-63.

87. Tooley JE, Turakhia MP. Is it time for a consumerized or home-based 12-lead electrocardiogram? *EP Europace* 2022; **24**(3): 357-8.

88. Nygårds M-E, Hulting J. An automated system for ECG monitoring. *Computers and Biomedical Research* 1979; **12**(2): 181-202.

89. Sajeev JK, Koshy AN, Teh AW. Wearable devices for cardiac arrhythmia detection: a new contender? *Internal medicine journal* 2019; **49**(5): 570-3.

90. Sehrawat O, Kashou AH, Noseworthy PA. Artificial intelligence and atrial fibrillation. *Journal of cardiovascular electrophysiology* 2022.

91. Ko W-Y, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *Journal of the American College of Cardiology* 2020; **75**(7): 722-33.

92. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine* 2019; **25**(1): 70-4.

93. Isaksen JL, Baumert M, Hermans ANL, Maleckar M, Linz D. Artificial intelligence for the detection, prediction, and management of atrial fibrillation. *Herzschrittmachertherapie + Elektrophysiologie* 2022; **33**(1): 34-41.

94. 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. *Circulation: Arrhythmia and Electrophysiology* 2020.

95. Attia IZ, Tseng AS, Benavente ED, et al. External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *International Journal of Cardiology* 2021; **329**: 130-5.

96. Chouairi F, Fuery M, Clark KA, et al. Evaluation of racial and ethnic disparities in cardiac transplantation. *Journal of the American Heart Association* 2021; **10**(17): e021067.

97. 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. *Nature Medicine* 2021.

98. Roman Y. The United States 2020 Census data: implications for precision medicine and the research landscape. *Personalized Medicine* 2022; **19**(1): 5-8.

99. Jingi AM, Noubiap JJN, Ewane Onana A, et al. Access to Diagnostic Tests and Essential Medicines for Cardiovascular Diseases and Diabetes Care: Cost, Availability and Affordability in the West Region of Cameroon. *PLoS ONE* 2014; **9**(11): e111812.

100. Flaxman P. The 10-minute appointment. *British Journal of General Practice* 2015; **65**(640): 573-4.

101. Wegner FK, Kochhäuser S, Ellermann C, et al. Prospective blinded Evaluation of the smartphone-based AliveCor Kardia ECG monitor for Atrial Fibrillation detection: The PEAK-AF study. *European Journal of Internal Medicine* 2020; **73**: 72-5.

102. Stavrakis S, Garabelli PJ, Smith L, Albert D, Po SS. Clinical validation of a smartphone based, 6-lead ECG device. *Circulation* 2017; **136**(suppl_1): A15576-A.

103. Duarte R, Stainthorpe A, Mahon J, et al. Lead-I ECG for detecting atrial fibrillation in patients attending primary care with an irregular pulse using single-time point testing: A systematic review and economic evaluation. *PLOS ONE* 2019; **14**(12): e0226671.

104. Network WAHS. Full Report: Independent Evaluation of the AHSN Network mobile ECG roll-out programm. August 2019.

105. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of Medical Internet Research* 2017; **19**(11): e367.

106. Organization WH. Digital Implementation Investment Guide (DIIG): integrating digital interventions into health programmes. 2020.

107. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. *NPJ digital medicine* 2019; **2**(1): 1-9.

108. Roguin A. Rene Theophile Hyacinthe Laennec (1781-1826): The Man Behind the Stethoscope. *Clinical Medicine & Research* 2006; **4**(3): 230-5.

109. Parati G, Stergiou GS, Asmar R, et al. European Society of Hypertension Practice Guidelines for home blood pressure monitoring. *Journal of Human Hypertension* 2010; **24**(12): 779-85.

110. Rigby MJ. Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics* 2019; **21**(2): 121-4.

111. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine* 2020; **127**: 104065.

112. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circulation: Heart Failure* 2013; **6**(3): 606-19.

113. Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *Journal of the American College of Cardiology* 2014; **63**(12): 1123-33.

114. Bozkurt B, Coats AJ, Tsutsui H, et al. Universal definition and classification of heart failure: a report of the heart failure society of America, heart failure association of the European society of cardiology, Japanese heart failure society and writing committee of the universal definition of heart failure. *Journal of cardiac failure* 2021; **27**(4): 387-413.

115. van Rossum G. Python Tutorial, Technical Report CS-R9526. 1995.

116. Sugrue A, Noseworthy PA, Kremen V, et al. Identification of Concealed and Manifest Long QT Syndrome Using a Novel T Wave Analysis Program. *Circulation: Arrhythmia and Electrophysiology* 2016; **9**(7): e003830.

117. Gholam-Hosseini H, Nazeran H. Detection and extraction of the ECG signal parameters. IEEE.

118. Couderc J-P, Xia X, Peterson DR, et al. T-wave morphology abnormalities in benign, potent, and arrhythmogenic Ikr inhibition. *Heart Rhythm* 2011; **8**(7): 1036-43.

119. Statistics OfN. Ethnic group, national identity and religion. Measuing equality: a guide for the collection and classification of ethnic group, national identity and religion data in the UK. . 2016.

120. Echocardiography BSo. Protocols and guidelines. 2021.

121. Penfield RD. Applying the Breslow-Day test of trend in odds ratio heterogeneity to the analysis of nonuniform DIF. *Alberta Journal of Educational Research* 2003; **49**(3).

122. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 2016; **15**(2): 155-63.

123. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. *Cardiac Failure Review* 2017; **03**(01): 7.

124. De Couto G, Ouzounian M, Liu PP. Early detection of myocardial dysfunction and heart failure. *Nature Reviews Cardiology* 2010; **7**(6): 334-44.

125. Hendry PB, Krisdinarti L, Erika M. Scoring System Based on Electrocardiogram Features to Predict the Type of Heart Failure in Patients With Chronic Heart Failure. *Cardiology Research* 2016; **7**(3): 110-6.

126. O'Neal WT, Mazur M, Bertoni AG, et al. Electrocardiographic Predictors of Heart Failure With Reduced Versus Preserved Ejection Fraction: The Multi-Ethnic Study of Atherosclerosis. *Journal of the American Heart Association* 2017; **6**(6): e006023.

127. Alhamaydeh M, Gregg R, Ahmad A, Faramand Z, Saba S, Al-Zaiti S. Identifying the most important ECG predictors of reduced ejection fraction in patients with suspected acute coronary syndrome. *Journal of Electrocardiology* 2020; **61**: 81-5.

128. Khan NK, Goode KM, Cleland JGF, et al. Prevalence of ECG abnormalities in an international survey of patients with suspected or confirmed heart failure at death or discharge. *European Journal of Heart Failure* 2007; **9**(5): 491-501.

129. Rim TH, Lee CJ, Tham Y-C, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *The Lancet Digital Health* 2021; **3**(5): e306-e16.

130. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* 2019; **25**(1): 44-56.

131. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. *Science* 2019; **363**(6429): 810-2.

132. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020: m689.

133. Orchard J, Li J, Gallagher R, Freedman B, Lowres N, Neubeck L. Uptake of a primary care atrial fibrillation screening program (AF-SMART): a realist evaluation of implementation in metropolitan and rural general practice. *BMC Family Practice* 2019; **20**(1).

134. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Global Health* 2018; **3**(4): e000798.

135. Chorba JS, Shapiro AM, Le L, et al. Deep Learning Algorithm for Automated Cardiac Murmur Detection via a Digital Stethoscope Platform. *Journal of the American Heart Association* 2021; **10**(9).

136. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet* 2019; **394**(10201): 861-7.

137. Hagiwara Y, Fujita H, Oh SL, et al. Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review. *Information Sciences* 2018; **467**: 99-114.

138. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *The Lancet* 2018; **391**(10120): 572-80.

139. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2021; **42**(36): 3599-726.

140. Tran RH, Aldemerdash A, Chang P, et al. Guideline-Directed Medical Therapy and Survival Following Hospitalization in Patients with Heart Failure. *Pharmacotherapy* 2018; **38**(4): 406-16.

141. Ouwerkerk W, Voors AA, Anker SD, et al. Determinants and clinical outcome of uptitration of ACE-inhibitors and beta-blockers in patients with heart failure: a prospective European study. *European Heart Journal* 2017; **38**(24): 1883-90.

142. Abdin A, Anker SD, Butler J, et al. 'Time is prognosis' in heart failure: time-to-treatment initiation as a modifiable risk factor. *ESC Heart Fail* 2021; **8**(6): 4444-53.

143. Solomon SD, Dobson J, Pocock S, et al. Influence of Nonfatal Hospitalization for Heart Failure on Subsequent Mortality in Patients With Chronic Heart Failure. *Circulation* 2007; **116**(13): 1482-7.

144. NICE. Resource impact report: Chronic heart failure in adults, 2018.

145. Hancock HC, Close H, Fuat A, Murphy JJ, Hungin APS, Mason JM. Barriers to accurate diagnosis and effective management of heart failure have not changed in the past 10 years: a qualitative study and national survey. *BMJ Open* 2014; **4**(3): e003866.

146. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, et al. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. *BMJ* 2019; **364**: l223.

147. Chang KC, Vamos EP, Palladino R, Majeed A, Lee JT, Millett C. Impact of the NHS Health Check on inequalities in cardiovascular disease risk: a difference-in-differences matching analysis. *J Epidemiol Community Health* 2019; **73**(1): 11-8.

148. Karlström P, Johansson P, Dahlström U, Boman K, Alehagen U. The impact of time to heart failure diagnosis on outcomes in patients tailored for heart failure treatment by use of natriuretic peptides. Results from the UPSTEP study. *International Journal of Cardiology* 2017; **236**: 315-20.

149. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; **335**(7626).

150. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine* 2015; **12**(10): e1001885.

151. Bottle A, Cohen C, Lucas A, et al. How an electronic health record became a real-world research resource: comparison between London's Whole Systems Integrated Care database and the Clinical Practice Research Datalink. *BMC Medical Informatics and Decision Making* 2020; **20**(1): 71.

152. Martin Cowie SM, Jessica Morris, Roger Newson, Michael Soljak, Bowen Su. Heart failure prevalence models for small populations: Technical Document produced for Public Health England . School of Public Health, Imperial College London; 2016.

153. Partners ICH. Heart Failure Dashboard. 2020. https://ichpanalytics.imperialcollegehealthpartners.com/#/signin?redirect=%2Fsite%2FICHP%2Fviews%2FHeartFailureHIITv1%2FHOME.

154. Garès V, Andrieu S, Dupuy J-F, Savy N. A comparison of the constant piecewise weighted logrank and Fleming-Harrington tests. *Electronic journal of statistics* 2014; **8**(1): 841-60.

155. Vamvakidou A, Danylenko O, Pradhan J, et al. Relative clinical value of coronary computed tomography and stress echocardiography-guided management of stable chest pain patients: a propensity-matched analysis. *European Heart Journal-Cardiovascular Imaging* 2021; **22**(12): 1473-81.

156. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *European heart journal* 2011; **32**(14): 1704-8.

157. Agency UHS. Guidance: Cost consequence analysis: health economic studies. July 2020.

158. Gerber Y, Weston SA, Enriquez-Sarano M, et al. Mortality associated with heart failure after myocardial infarction: a contemporary community perspective. *Circulation: Heart Failure* 2016; **9**(1): e002460.

159. de Miguel Díez J, Morgan JC, García RJ. The association between COPD and heart failure risk: a review. *International journal of chronic obstructive pulmonary disease* 2013; **8**: 305.

160. Oosterom-Calo R, van Ballegooijen AJ, Terwee CB, et al. Determinants of adherence to heart failure medication: a systematic literature review. *Heart Fail Rev* 2013; **18**(4): 409-27.

161. Fitzgerald AA, Powers JD, Ho PM, et al. Impact of medication nonadherence on hospitalizations and mortality in heart failure. *J Card Fail* 2011; **17**(8): 664-9.

162. Koudstaal S, Pujades-Rodriguez M, Denaxas S, et al. Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people. *European journal of heart failure* 2017; **19**(9): 1119-27.

163. Dunlay SM, Redfield MM, Weston SA, et al. Hospitalizations after heart failure diagnosis a community perspective. *J Am Coll Cardiol* 2009; **54**(18): 1695-702.

164. Statistics OfN. Regional Ethnic Diversity. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest; 2018.

165. Bottle A, Cohen C, Lucas A, et al. How an electronic health record became a real-world research resource: comparison between London's Whole Systems Integrated Care database and the Clinical Practice Research Datalink. *BMC medical informatics and decision making* 2020; **20**(1): 1-11.

166. Kaplon-Cieślicka A, Benson L, Chioncel O, et al. A comprehensive characterization of acute heart failure with preserved versus mildly reduced versus reduced ejection fraction – insights from the <scp>ESC-HFA EORP</scp> Heart Failure Long-Term Registry. *European Journal of Heart Failure* 2022; **24**(2): 335-50.

167. Kahn M, Grayson AD, Chaggar PS, et al. Primary care heart failure service identifies a missed cohort of heart failure patients with reduced ejection fraction. *European Heart Journal* 2022; **43**(5): 405-12.

168. Bachtiger P, Petri CF, Scott FE, et al. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre study. *The Lancet Digital Health* 2022.

169. The Rt Hon Matt Hancock MP DoHaSC. Health Secretary announces £250 million investment in artificial intelligence. 2019.

170. Ghafur S, Schneider E. Why are health care organizations slow to adopt patient-facing digital technologies. *Health Affairs* 2019.

171. Llewellyn S, Procter R, Harvey G, Maniatopoulos G, Boyd A. Facilitating technology adoption in the NHS: negotiating the organisational and policy context – a qualitative study. *Health Services and Delivery Research* 2014; **2**(23): 1-132.

172. Excellence NIfHaC. Medical technologies guidance [MTG32] HeartFlow FFRCT for estimating fractional flow reserve from coronary CT angiography. 2021.

173. Mamas MA, Sperrin M, Watson MC, et al. Do patients have worse outcomes in heart failure than in cancer? A primary care-based cohort study with 10-year follow-up in Scotland. *European journal of heart failure* 2017; **19**(9): 1095-104.

174. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *The Lancet* 2018; **391**(10120): 572-80.

175. Roberts E, Ludman AJ, Dworzynski K, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. *Bmj* 2015; **350**.

176. Fonseca C, Mota T, Morais H, et al. The value of the electrocardiogram and chest x-ray for confirming or refuting a suspected diagnosis of heart failure in the community. *European journal of heart failure* 2004; **6**(6): 807-12.

177. Tromp J, Ouwerkerk W, van Veldhuisen DJ, et al. A systematic review and network meta-analysis of pharmacological treatment of heart failure with reduced ejection fraction. *Heart Failure* 2022; **10**(2): 73-84.

178. Lawson CA, Zaccardi F, Squire I, et al. Risk factors for heart failure: 20-year population-based trends by sex, socioeconomic status, and ethnicity. *Circulation: Heart Failure* 2020; **13**(2): e006472.

179. Tsigkas G, Apostolos A, Despotopoulos S, et al. Heart failure and atrial fibrillation: New concepts in pathophysiology, management, and future directions. *Heart Failure Reviews* 2021: 1-10.

180. Bates ER. Treatment options in severe aortic stenosis. *Circulation* 2011; **124**(3): 355-9.

181. Soljak M, Samarasundera E, Indulkar T, Walford H, Majeed A. Variations in cardiovascular disease under-diagnosis in England: national cross-sectional spatial analysis. *BMC cardiovascular disorders* 2011; **11**(1): 1-11.

182. Improving health literacy to reduce health inequalities. In: England PH, editor.; 2015.

183. Ahmed SM, Lemkau JP, Nealeigh N, Mann B. Barriers to healthcare access in a non-elderly urban poor American population. *Health & social care in the community* 2001; **9**(6): 445-53.

184. Wolfe MK, McDonald NC, Holmes GM. Transportation barriers to health care in the United States: findings from the national health interview survey, 1997–2017. *American journal of public health* 2020; **110**(6): 815-22.

185. Flores G. Language barriers to health care in the United States. *N Engl J Med* 2006; **355**(3): 229-31.

186. Lewsey SC, Breathett K. Racial and ethnic disparities in heart failure: current state and future directions. *Current opinion in cardiology* 2021; **36**(3): 320-8.

187. Lord K, Ibrahim K, Kumar S, Rudd N, Mitchell A, Symonds P. Measuring trust in healthcare professionals a study of ethnically diverse UK cancer patients. *Clinical Oncology* 2012; **24**(1): 13-21.

188. Delgadillo J, Asaria M, Ali S, Gilbody S. On poverty, politics and psychology: the socioeconomic gradient of mental healthcare utilisation and outcomes. *British Journal of Psychiatry* 2016; **209**(5): 429-30.

189. Fox G, Connolly R. Mobile health technology adoption across generations: Narrowing the digital divide. *Information Systems Journal* 2018; **28**(6): 995-1019.

190. Morris A, Brading H. E-literacy and the grey digital divide: A review with recommendations. *Journal of Information Literacy* 2007; **1**(3): 13.

191. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics* 2021; **47**(5): 329-35.

192. DeCamp M, Tilburt JC. Why we cannot trust artificial intelligence in medicine. *The Lancet Digital Health* 2019; **1**(8): e390.

193. Heaven WD. Our weird behavior during the pandemic is messing with AI models. *MIT Technology Review) Available online at https:// www technologyreview com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/, checked on* 2020; **6**(15): 2020.

194. Hölzel LP, Ries Z, Kriston L, et al. Effects of culture-sensitive adaptation of patient information material on usefulness in migrants: a multicentre, blinded randomised controlled trial. *BMJ Open* 2016; **6**(11): e012008.

195. Wright P. Criteria and ingredients for successful patient information. *Journal of Audiovisual Media in Medicine* 2003; **26**(1): 6-10.

196. Foster PH. Use of stigma, fear, and denial in development of a framework for prevention of HIV/AIDS in rural African American communities. *Family & Community Health* 2007; **30**(4): 318-27.

197. Zizzo N, Di Pietro N, Green C, Reynolds J, Bell E, Racine E. Comments and reflections on ethics in screening for biomarkers of prenatal alcohol exposure. *Alcoholism: Clinical and Experimental Research* 2013; **37**(9): 1451-5.

198. Duncker D, Ding WY, Etheridge S, et al. Smart wearables for cardiac monitoring real-world use beyond atrial fibrillation. *Sensors* 2021; **21**(7): 2539.

199. Lubitz SA, Atlas SJ, Ashburner JM, et al. Screening for Atrial Fibrillation in Older Adults at Primary Care Visits: VITAL-AF Randomized Controlled Trial. *Circulation* 2022; **145**(13): 946-54.

200. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine* 2019; **381**(20): 1909-17.

201. Nunes F, Andersen T, Fitzpatrick G. The agency of patients and carers in medical care and self-care technologies for interacting with doctors. *Health Informatics Journal* 2019; **25**(2): 330-49.

202. Brown SB, Shannon RP. Improving Medication Compliance in Patients With Heart Failure. *The American Journal of Cardiology* 2008; **101**(2): 274-7.

203. Nerini E, Grip L, Camm AJ, Giugliano RP. Atrial fibrillation and the 'other drug problem': reducing non-adherence with technology. *European Heart Journal* 2013; **34**(27): 2031-3.

204. Gebreyohannes EA, Salter S, Chalmers L, Bereznicki L, Lee K. Non-adherence to thromboprophylaxis guidelines in atrial fibrillation: a narrative review of the extent of and factors in guideline non-adherence. *American Journal of Cardiovascular Drugs* 2021; **21**(4): 419-33.

205. Chung MK, Fagerlin A, Wang PJ, et al. Shared Decision Making in Cardiac Electrophysiology Procedures and Arrhythmia Management. *Circulation: Arrhythmia and Electrophysiology* 2021; **14**(12): e007958.

206. Knoepke CE, Allen LA, Kramer DB, Matlock DD. Medicare mandates for shared decision making in cardiovascular device placement. *Circulation: Cardiovascular Quality and Outcomes* 2019; **12**(7): e004899.

207. Kuehn BM. States Explore Shared Decision Making. *JAMA* 2009; **301**(24): 2539-41.

208. Anderson SJ, Krigolson OE, Jamniczky HA, Hecker KG. Learning Anatomical Structures: a Reinforcement-Based Learning Approach. *Medical Science Educator* 2016; **26**(1): 123-8.

209. Diprose W, Buist N. Artificial intelligence in medicine: humans need not apply? *The New Zealand Medical Journal (Online)* 2016; **129**(1434): 73.

210. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *Jama* 2018; **319**(1): 19-20.

211. Chan J, Raju SC, Topol E. Towards a tricorder for diagnosing paediatric conditions. *The Lancet* 2019; **394**(10202): 907.

212. Fontana G, Ghafur S, Torne L, Goodman J, Darzi A. Ensuring that the NHS realises fair financial value from its data. *The Lancet Digital Health* 2020; **2**(1): e10-e2.

213. West B, Cumella A. Data sharing and technology: exploring the attitudes of people with asthma. 2018.

214. England H. How Do People Feel About Their Data Being Shared by the NHS. *URL: https://www healthwatch co uk/news/2018-05-17/how-do-people-feel-about-their-data-being-shared-nhs [accessed 2019-02-07]* 2018.

215. Thinking C. Understanding public expectations of the use of health and care data. *East Sussex, England: Curved Thinking* 2019.

216. Tully MP, Bozentko K, Clement S, et al. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. *Journal of medical Internet research* 2018; **20**(3): e7763.

217. Cakmak AS, Reinertsen E, Taylor HA, Shah AJ, Clifford GD. Personalized heart failure severity estimates using passive smartphone data. 2018 IEEE International Conference on Big Data (Big Data); 2018: IEEE; 2018. p. 1569-74.

218. Attia ZI, Kapa S, Dugan J, et al. Rapid Exclusion of COVID Infection With the Artificial Intelligence Electrocardiogram. Mayo Clinic Proceedings; 2021: Elsevier; 2021. p. 2081-94.

219. Speer SA, Stokoe E. Ethics in action: Consent-gaining interactions and implications for research practice. *British Journal of Social Psychology* 2014; **53**(1): 54-73.

220. Ibarra J, Jahankhani H, Kendzierskyj S. Cyber-Physical Attacks and the Value of Healthcare Data: Facing an Era of Cyber Extortion and Organised Crime. Springer International Publishing; 2019: 115-37.

221. The MITRE Corporation MDIC. Playbook for Threat Modelling Medical Devices. 2021.

222. Martin G, Martin P, Hankin C, Darzi A, Kinross J. Cybersecurity and healthcare: how safe are we? *BMJ* 2017: j3179.

223. Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. Mass Medical Soc; 2012. p. 848-55.

224. Bergsland J, Elle OJ, Fosse E. Barriers to medical device innovation. *Medical Devices: Evidence and Research* 2014: 205.

225. Pedersen M, Verspoor K, Jenkinson M, Law M, Abbott DF, Jackson GD. Artificial intelligence for clinical decision support in neurology. *Brain Communications* 2020; **2**(2).

226. Ayling J, Chapman A. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* 2021.

227. Excellence NIfHaC. Evidence standards framework (ESF) for digital health technologies update - consultation. 2022.

228. Hutchings R. The impact of Covid-19 on the use of digital technology in the NHS. *Nuffield Trust* 2020: 2020-08.

229. Wang J, Spicher N, Warnecke JM, Haghi M, Schwartze J, Deserno TM. Unobtrusive Health Monitoring in Private Spaces: The Smart Home. *Sensors* 2021; **21**(3): 864.

230. Philip NY, Rodrigues JJPC, Wang H, Fong SJ, Chen J. Internet of Things for In-Home Health Monitoring Systems: Current Advances, Challenges and Future Directions. *IEEE Journal on Selected Areas in Communications* 2021; **39**(2): 300-10.

231. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine* 2019; **17**(1).

232. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. *The Lancet Digital Health* 2019; **1**(1): e13-e4.

233. England PH. Guidance: NHS screening programmes: information for GPs and practice staff. 2018.

234. Government U. NHS Health Check. 2022. https://www.nhs.uk/conditions/nhs-health-check/.

235. Van Riet EES, Hoes AW, Wagenaar KP, Limburg A, Landman MAJ, Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. *European Journal of Heart Failure* 2016; **18**(3): 242-52.

236. Fuat A. Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study. *BMJ* 2003; **326**(7382): 196-.

237. Zaphiriou A, Robb S, Murray-Thomas T, et al. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with suspected heart failure: Results of the UK natriuretic peptide study. *European Journal of Heart Failure* 2005; **7**(4): 537-41.

238. Kelder JC, Cowie MR, McDonagh TA, et al. Quantifying the added value of BNP in suspected heart failure in general practice: an individual patient data meta-analysis. *Heart* 2011; **97**(12): 959-63.

239. McDonagh TA, McDonald K, Maisel AS. Screening for Asymptomatic Left Ventricular Dysfunction Using B-Type Natriuretic Peptide. *Congestive Heart Failure* 2008; **14**: 5-8.

240. Investigators C. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *The Lancet* 2001; **357**(9266): 1385-90.

241. Members ATF, Priori SG, Blomström-Lundqvist C, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC) Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC). *Ep Europace* 2015; **17**(11): 1601-87.

242. Whittingham K, Barnes S, Gardiner C. Tools to measure quality of life and carer burden in informal carers of heart failure patients: A narrative review. *Palliative Medicine* 2013; **27**(7): 596-607.

243. Small N, Barnes S, Gott M, et al. Dying, death and bereavement: a qualitative study of the views of carers of people with heart failure in the UK. *BMC Palliative Care* 2009; **8**(1): 6.

244. Hodes S, Majeed A. Using the NHS App as a covid-19 vaccine passport. *bmj* 2021; **373**.

245. Kolasa K, Kozinski G. How to Value Digital Health Interventions? A Systematic Literature Review. *International Journal of Environmental Research and Public Health* 2020; **17**(6): 2119.

246. Meeting Patients Where They Are. *AMA Journal of Ethics* 2013; **15**(4): 293-8.

247. Sharma A, Harrington RA, McClellan MB, et al. Using digital health technology to better generate evidence and deliver evidence-based care. *Journal of the American College of Cardiology* 2018; **71**(23): 2680-90.

248. Wyatt KD, Poole LR, Mullan AF, Kopecky SL, Heaton HA. Clinical evaluation and diagnostic yield following evaluation of abnormal pulse detected using Apple Watch. *Journal of the American Medical Informatics Association* 2020; **27**(9): 1359-63.

249. Saraf K, Baek CI, Wasko MH, et al. Assessment of Left Ventricular Diastolic Function using Phonocardiogram Signals: A Comparison with Echocardiography. 2020: IEEE.

250. Zeng Y, Yang S, Yu X, et al. A multimodal parallel method for left ventricular dysfunction identification based on phonocardiogram and electrocardiogram signals synchronous analysis. *Mathematical Biosciences and Engineering* 2022; **19**(9): 9612-35.

251. Cho J, Lee B, Kwon J-M, et al. Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography. *ASAIO Journal* 2021; **67**(3): 314-21.

252. Savarese G, Stolfo D, Sinagra G, Lund LH. Heart failure with mid-range or mildly reduced ejection fraction. *Nature Reviews Cardiology* 2022; **19**(2): 100-16.

253. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in heart failure with a preserved ejection fraction. *New England Journal of Medicine* 2021; **385**(16): 1451-61.

254. Bhatt DL, Szarek M, Pitt B, et al. Sotagliflozin in patients with diabetes and chronic kidney disease. *New England Journal of Medicine* 2021; **384**(2): 129-39.

255. Kelder JC, Cramer MJ, Van Wijngaarden J, et al. The Diagnostic Value of Physical Examination and Additional Testing in Primary Care Patients With Suspected Heart Failure. *Circulation* 2011; **124**(25): 2865-73.

256. Bacon DR, Cowles K, Thapa D, et al. Creating an Ultrasound Scholarly Concentration Program for Medical Students. *Advances in Medical Education and Practice* 2021; **Volume 12**: 1103-10.

257. Stewart JE, Goudie A, Mukherjee A, Dwivedi G. Artificial intelligence-enhanced echocardiography in the emergency department. *Emergency Medicine Australasia* 2021; **33**(6): 1117-20.

258. Schneider M, Bartko P, Geller W, et al. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. *The International Journal of Cardiovascular Imaging* 2021; **37**(2): 577-86.

10 Supplementary Appendix

10.1 Patient information sheet – summary

From: Professor Nicholas Peters Chief Investigator, DUO-EF-19. Imperial College London

Imperial College London

SUMMARY PATIENT INFORMATION SHEET:

DUO-EF-19: Diagnosing Heart Failure in the Post-COVID-19 Clinic, Primary Care and Hospital Setting Setting Using a Digital Stethoscope with AI ECG

"A study using a digital stethoscope to evaluate if artificial intelligence algorithm applied to electrocardiogram (ECG) can accurately diagnose reduced pumping function of the heart"

Thank you for showing an interest in this study. The first page of this information sheet sets out a brief summary. The rest of the document provides more detail. You can ask us for more information at any point -- you are not obliged to participate and can withdraw at any time.

What will the study involve?

As applicable, during your visit to the **echo or cardiac MRI department/GP practice/Covid-19 follow up clinic** your clinical team will use a safe digital stethoscope to take **five** ECG readings, each lasting 15 seconds (<2 minutes in total). This can take place at the same time as routine examination and will not interfere with your care; this is completely painless and poses no harm. **The table below tells if you about the criteria to take part:**

You must meet these criteria:	**You must not have:**

Age > 18 (no upper limit); able to give informed consent
 	
Any of the following:
- Attending for echo or cardiac MRI
  - Inpatient or outpatient
 	
- Attending with symptoms suggestive of HF
  - In the GP setting
  - In Covid-19 follow up clinic
	- Any chest wound, skin pathology or other feature that would prohibit routine stethoscope examination

Condition Studied	**Study Length**	**Number of Visits**
Heart failure (HF)	5-10 minutes	No additional visits required; the study can take place at the same time as your appointment (none additional). If you are someone who has a positive result from the algorithm but a negative echo/MRI we will call you in 24 months to check-in.

- There are no direct benefits to you beyond knowing you are helping with research
- There is no risk of harm (the device is safe)
- Participation will not change your treatment

This study has indemnity and insurance coverage from Imperial College London. Your data will be treated in full confidentiality, details of which are outlined in the participant info sheet. Thank you.

Study contact details: patrik.bachtiger@nhs.net Signed: Professor Nicholas Peters 04/02/2021

10.2 Patient consent form for ECG-enabled stethoscope examination

Imperial College Healthcare NHS Trust | Imperial College London

STUDY ID:

Study Protocol Number: IRAS 285417 v2.0 Name of Principal Investigator: **PROFESSOR NICHOLAS PETERS**

INFORMED CONSENT FORM FOR SUBJECTS ABLE TO GIVE CONSENT

Full Title of Project: **DUO-EF-19: Diagnosing Heart Failure in the Post-COVID-19 Clinic, Primary Care and Hospital Setting Setting Using a Digital Stethoscope with AI ECG**

	Initial
1.) I confirm that I have read and understand the subject information sheet dated ..... version ..... for the above study and have had the opportunity to ask questions which have been answered fully.	.....
2.) I understand that my participation is voluntary and I am free to withdraw at any time, without giving any reason and without my medical care or legal rights being affected.	.....
3.) I understand that sections of any of my medical notes may be looked at by responsible individuals from IMPERIAL COLLEGE LONDON or from NHS Trust or from regulatory authorities where it is relevant to my taking part in this research.	.....
4.) I give permission for these individuals to access my records that are relevant to this research.	.....
5.) I give/do not give (delete as applicable) consent for information collected about me to be used to support other research in the future, including those outside of the EEA.	.....
6.) I give consent for the study team to notify my care team of any significant incidental findings	.....
7.) I consent to take part in the above study.	.....
8.) I give/do not give (delete as applicable) consent to being contacted to potentially taking part in other research studies.	.....
9.) If indicated, I give consent for the research team to contact me in 24 months to inquire about my health status	.....

Name of Participant	Signature	Date
Name of Person taking consent	Signature	Date

SOP Ref: JRCO/SOP/016 1 copy for subject; 1 copy for Principal Investigator; 1copy to be kept with hospital notes

V16.0 09 Mar 2020 Page 1 of 1

DUO-EF-19 CONSENT FORM V2.0 04/02/2021

10.3 Health Research Authority Research Ethics Committee Approval

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Professor Nicholas Peters
Hammersmith Hospital
Du Cane Road
London
W12 0HS

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

05 February 2021

Dear Professor Peters

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Diagnosis of Heart Failure in the Post-COVID-19 Clinic, Primary Care and Hospital Setting Using a Digital Stethoscope with AI ECG**
IRAS project ID:	**285417**
Protocol number:	**20HH6360**
REC reference:	**21/LO/0051**
Sponsor	**Imperial College London**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

10.4 Patient self-examination support istructions

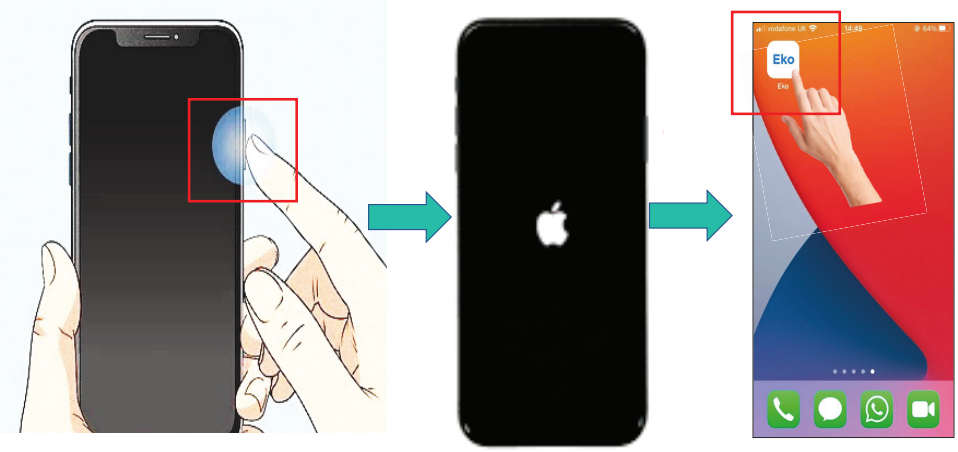
To be done on DUO

STEP 1: Turning on Eko DUO

1. Hold button for 5 seconds and release finger from button when light flashes

To be done on iPhone

STEP 2: Turn on iPhone and enter app



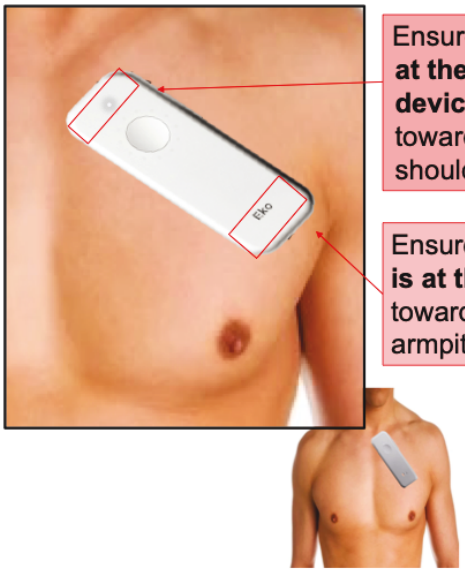
2a. Hold button on right side for 5 seconds until apple logo appears

2b. Apple logo appears release finger from side

2c. Tap Eko app and place iPhone on flat surface - Screen facing up

To be done with DUO

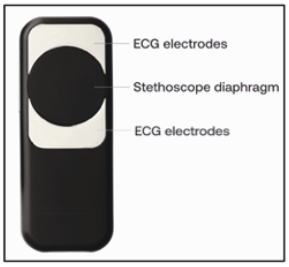
STEP 3: Placing the DUO on your chest



Ensure **light is at the top of device** pointing towards opposite shoulder

Ensure **Eko logo is at the bottom** towards the Left armpit

BACK OF DUO



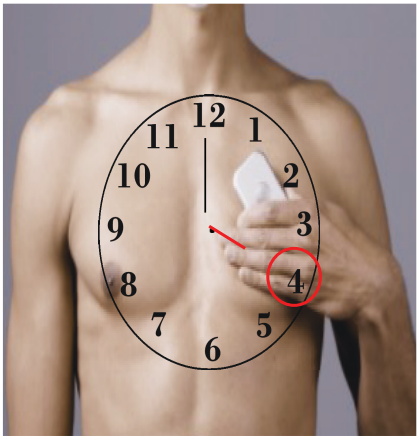
Ensure both grey electrodes are touching, flat to the skin

To be done with DUO

STEP 4 – Positioning DUO on Skin

3a. Place Eko DUO **directly, flat** onto the skin

3b. Position exactly as demonstrated in image



Ensure device is on the upper left side of chest **angled at 4 o'clock**

To be done on iPhone

How to improve the quality of your reading:

IF DIAL ON RED

REPOSITION DUO TO ENSURE POSITION MATCHES STEP 4

IF DIAL ON YELLOW:

ENSURE DUO FLAT TO SKIN AND HOLD DUO FIRMLY AND STILL IN PALM

IF DIAL ON GREEN:

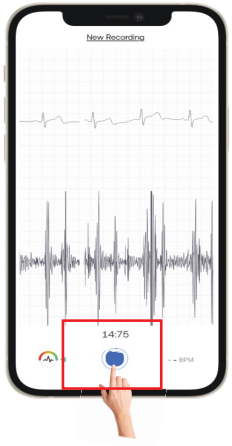
DO NOT MOVE HAND, MOVE TO STEP 4

To be done on iPhone

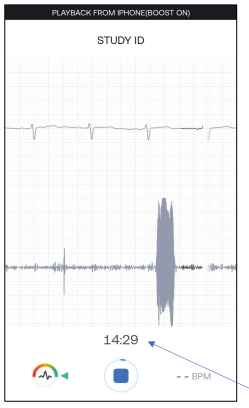
STEP 5: 15 seconds for recording

KEEP DUO STILL AND AVOID MOVING OR SPEAKING DURING RECORDING

When Dial on green 
Tap record Button will turn from  to 



Device will automatically stop and save after 15 seconds



Time remaining

To be done on Duo

STEP 6: Repeat to complete 3x recordings

- Lift the DUO off your chest and place back in same position as step 4
- Repeat recording IN STEP 5
- Complete 3 recordings this way

Thank you! A member of the research team will be with you to help you with any of the steps along the way.

END