Artificial intelligence in cardiology: the debate continues

Folkert W. Asselbergs 🄳 [1,2,3] and Alan G. Fraser 🄳 [4,5]*

[1]Division Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, Netherlands; [2]Institute of Health Informatics and Institute of Cardiovascular Science, University College London, 222 Euston Rd, London NW1 2DA, UK; [3]NIHR BRC Clinical Research Informatics Unit, University College London Hospital, London, UK; [4]School of Medicine, Cardiff University, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK; and [5]Cardiovascular Imaging and Dynamics, Katholieke Universiteit Leuven, UZ Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium

In 1955, when John McCarthy and his colleagues proposed their first study of artificial intelligence, they suggested that 'every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it'. Whether that might ever be possible would depend on how we define intelligence, but what is indisputable is that new methods are needed to analyse and interpret the copious information provided by digital medical images, genomic databases, and biobanks. Technological advances have enabled applications of artificial intelligence (AI) including machine learning (ML) to be implemented into clinical practice, and their related scientific literature is exploding. Advocates argue enthusiastically that AI will transform many aspects of clinical cardiovascular medicine, while sceptics stress the importance of caution and the need for more evidence. This report summarizes the main opposing arguments that were presented in a debate at the 2021 Congress of the European Society of Cardiology. Artificial intelligence is an advanced analytical technique that should be considered when conventional statistical methods are insufficient, but testing a hypothesis or solving a clinical problem—not finding another application for AI—remains the most important objective. Artificial intelligence and ML methods should be transparent and interpretable, if they are to be approved by regulators and trusted to provide support for clinical decisions. Physicians need to understand AI methods and collaborate with engineers. Few applications have yet been shown to have a positive impact on clinical outcomes, so investment in research is essential.

Graphical Abstract
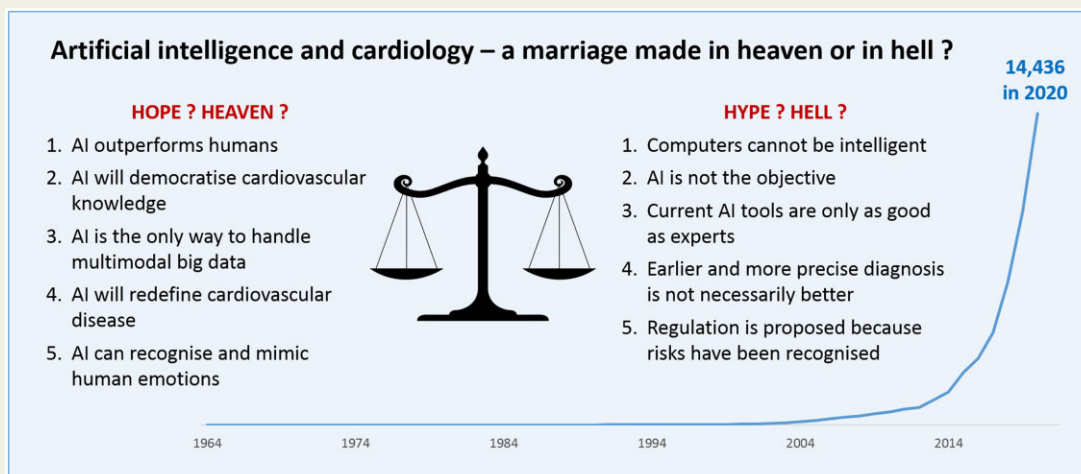
Summary of the main arguments presented in the debate, set against the exponential growth of papers listed on Pubmed relating to machine learning (ML), since the terms artificial intelligence and ML were first included.

Introduction

Artificial intelligence (AI) and machine learning (ML) are already integrated into aspects of routine cardiological practice, and they may become ubiquitous—but is our community just following fashion or will these tools contribute to genuine benefits for patients?

At the European Society of Cardiology (ESC) Congress 2021, a 'Great Debate' considered the question 'Artificial Intelligence in Cardiology: a Marriage Made in Heaven or Hell?'. Its title had clearly been intended by the programme committee to goad the contributors into arguing from extreme positions, but thinking about important issues in that way can serve a useful purpose. Despite initial hopes[1] many uncertainties remain. Here, we recapitulate the major arguments (see *Graphical Abstract*), highlight outstanding questions, and concur on the need for further research.

'AI and cardiology—a marriage made in heaven' (Folkert Asselbergs)

Artificial intelligence and cardiology are already heavily intertwined and this relationship will only intensify in the next years into a solid marriage.

Artificial intelligence has many applications that will benefit the cardiovascular community at large. Keeping up-to-date with the scientific literature or even guidelines is almost impossible, considering their volume; AI could help to find and analyse the wealth of data available in the public domain, thereby supporting researchers and healthcare professionals to provide the best care according to the latest evidence.[2] Artificial intelligence can optimize logistics and operations in a hospital, increase efficiency, and reduce the administrative burden on healthcare professionals and physicians,[3] for example by automatic labelling using natural language processing[4] or by scheduling patients according to their forecasted attendance.[5] In the near future, these applications will be extended with conversational AI that will reduce the time spent on electronic health records by automation of clinical notes and ordering.

Innovative technologies are needed to cope with the trends of an ageing population, increased healthcare utilization, and limited human resources. This does not apply solely to physicians or nurses but also to paramedical personnel like sonographers; in future, routine echocardiograms may be performed by untrained personnel guided by AI algorithms.[6] Innovative applications may ease daily work and increase job satisfaction.

Discussions among healthcare professionals about AI more often concern possible applications in clinical decision-making. They refer mostly to fancy algorithms applied in certain disease groups, which predict certain outcomes using a diverse range of data sources. In this pro-con debate, we should consider the following points.

Artificial intelligence outperforms humans

Humans are prone to error. When tired, distracted, or sick, their ability to make clinical decisions will be impaired—whereas AI works consistently at any time or season. Of course, AI algorithms should be trained and validated according to strict guidelines, but once that has been done, they can be applied repeatedly and easily. Each human observer must be trained separately, which limits the scalability of new methods. An exemplar is measuring left ventricular wall thickness in hypertrophic cardiomyopathy, which can be done more precisely by machines with a clear impact on decision-making by identifying patients who will benefit from implantation of a defibrillator.[7]

Artificial intelligence will democratize cardiovascular knowledge

A large part of the daily work of cardiologists is taken up with answering the questions of colleagues, such as how to interpret a patient's electrocardiogram. Artificial intelligence empowers primary care physicians and non-cardiologists by providing automated electrocardiographic (ECG) diagnoses that can guide decisions whether to treat or to refer for specialist cardiological care.[8] Algorithms can detect not only ischaemia or arrhythmias but also ECG signs of diminished ejection fraction, heart valve disease, or risk for atrial fibrillation.[9] In future, this knowledge will not be limited to healthcare professionals but extended to individuals using smartphone applications.[10]

Artificial intelligence is the only way to handle multimodal big data

Nowadays, numerous types of data from different sources are available to physicians. Increasingly, various omics data, imaging, ECG recordings, unstructured free text, and outputs from sensors and monitoring are collected, all of which needs to be interpreted in order to reach a diagnosis and plan treatment. Artificial intelligence will enable precision medicine by integrating and analysing all these different data sources, creating a digital twin of each individual patient that will provide information on diagnosis, prognosis, and treatments.[11]

Artificial intelligence will redefine cardiovascular disease

Cardiovascular diseases such as heart failure are heterogeneous, with individual patients varying in their responses to treatment. Reclassification based on more precise phenotyping is needed to improve outcomes.[12] A shift from a 'one-size fits all' to a more data-driven approach will identify those patients who will benefit most from particular therapies. For example, ML demonstrated heterogeneity in responses to beta-blockade in patients with heart failure.[13]

Artificial intelligence can recognize and mimic human emotions

Often the argument is made that AI will never replace doctors as computers lack empathy and communicative skills. However, computers do not have any conflict of interest, they are unbiased, and they will increasingly recognize emotions in some detail.[14,15] Patients will be able to interact with computers through chatbots or video within their own living environment, in their own language, together with their caregivers, family and relatives, with unlimited access and with no constraints on time. Even medical students can nowadays be trained in communication by virtual humans.[16] Of course, trust needs to be built between humans and computers to ensure that patients provide the correct information and are willing to adhere to advice about treatment. It is premature to consider such tools as supportive for end-of-life decisions and palliative care, but conversational AI and artificial empathy are developing rapidly and will have a place in healthcare in the future.

'AI and cardiology—a marriage made in hell' (Alan Fraser)

Adopting the role of professional iconoclast when considering AI means being Devil's advocate and uncovering any character flaws or misrepresentation—which is not so difficult because those are rife. We need to question our assumptions.

Computers cannot be intelligent

General-purpose AI, that might replicate capabilities of the human brain, is still a distant dream (or nightmare). Artificial intelligence algorithms can be exceptionally capable but they are fundamentally stupid. The really smart intelligence comes from the engineers who design the architecture and write the software for running neural networks. It is a disservice to use anthropomorphic language that endows AI with human characteristics to which it can never aspire.[17]

All computers do—even when handling petabytes of data—is to process binary codes. Computer vision software is vulnerable to adversarial challenges and highly prone to error in recognizing outlying cases. Machine learning can identify patterns even within random data. In medicine, we must always remember that AI identifies associations rather than causation.

Artificial intelligence is not the objective

The ultimate goal of clinical research is to develop more effective treatment—not another application for AI. There has been exponential growth in the AI medical literature but many studies appear to have been done without a prior hypothesis or clear clinical target. We should think of AI and ML as sophisticated tools that we need for analysing big datasets when conventional statistical methods can no longer cope. The methodology is subservient to the research question, not vice versa. Clinicians should identify important problems for engineers, instead of leaving them to develop tools which then seek an application.

Current artificial intelligence tools are only as good as experts

The integration of AI and ML into clinical practice is most advanced in diagnostic imaging. More than 100 products have already been approved by regulators (CE-marked), although scientific evidence establishing their utility has been published for only one third.[18] Individual trials report better performance by algorithms for specific tasks than by clinicians, but systematic reviews have concluded only that their performance is equivalent.[19,20] When retested on *re-acquired* images, the reproducibility of AI is not always better than expert human analysis.[21] And so far, <10% of the approved tools have been evaluated for their impact on clinical outcomes.[18]

Clinical diagnosis was more accurate when performed by doctors than by computer algorithms called symptom checkers (with correct first diagnoses in 72% vs. 34% respectively, when provided with the same vignettes to interpret).[22] Nor is there convincing evidence that clinician diagnostic performance is improved by using ML-based decision support systems; 46% of results were unchanged.[23] The performance of ML for clinical prediction models is not better than logistic regression, with 68% of ML studies being judged to have potential bias in their validation procedures.[24]

Earlier and more precise diagnosis is not necessarily better

Often AI algorithms have high sensitivity but rather low specificity, which implies a risk of overdiagnosis[25] and excess downstream testing. Unnecessary investigation or treatment of individuals whose subclinical changes would never have developed into significant disease may have psychological and social consequences as well as side-effects.

Deep learning is limited by the labels assigned to each case within the dataset that is used to train the algorithm, which may be problematic if diagnostic categories are suboptimal because a disease is poorly understood. Repeated cross-validation within the same training dataset tends to overestimate reproducibility.[26] Unless all relevant data are used as inputs, then phenotypic characterization by unsupervised ML may be uninformative.[27] Tools developed by comparing highly selected normal and diseased subjects (for example to interpret the electrocardiogram) will work much less well in unselected populations including people with a wide range of pretest probabilities and many comorbidities.[28] Algorithmic bias is a major concern (and potential danger) unless test and independent validation cohorts are representative of all populations in which the method will be applied.[29]

Table I   Professional and regulatory standards for medical AI

Professional quality standards

For clinical trials of interventions involving AI:

- **SPIRIT-AI extension:**
  Guidelines for clinical trial protocols for interventions involving AI.
  BMJ 2020;370:m321. https://doi.org/10.1136/bmj.m3210.

- **CONSORT-AI extension:**
  Reporting guidelines for clinical trial reports for interventions involving AI.
  BMJ 2020;370:m3164. https://doi.org/10.1136/bmj.m3164.

For diagnostic and prognostic prediction model studies based on AI:

- **TRIPOD–AI** Reporting guideline (in preparation)

- **PROBAST–AI** Risk of bias tool (in preparation)
  Protocol published by Collins GS, *et al. BMJ Open* 2021;11:e048008.

For the development-to-implementation gap in clinical AI:

- **DECIDE–AI** Human factors, early clinical evaluation (in preparation)
  Protocol published by DECIDE-AI-Steering Group. *Nat Med.* 2021;27:186–7.

For diagnostic studies including AI:

- **STARD-AI**: Diagnostic test accuracy studies (in preparation)
  Protocol published by Sounderajah V, *et al. BMJ Open* 2021;11:e047709.

- **PRIME**: Cardiovascular imaging-related machine learning evaluation.
  Sengupta PP, *et al. J Am Coll Cardiol Imaging.* 2020;13:2017–35.

Regulatory guidance

International Medical Device Regulators Forum:

- Software as a Medical Device (SaMD): Clinical Evaluation. IMDRF/SaMD WG/N41FINAL: 2017. http://www.imdrf.org/documents/documents.asp.

European Commission: Independent High-Level Expert Group on Artificial Intelligence:

- Ethics Guidelines for Trustworthy AI. 2019.
  https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

US Food & Drug Administration:

- Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device. Action Plan.
  January 2021. https://www.fda.gov/media/145022/download

European Commission:

- Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence. COM/2021/206 final. 21 April 2021.
  https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Extensions to previous consensus recommendations, already published or in preparation (as of September 2021). Documents are available through the EQUATOR website: https://www.ndorms.ox.ac.uk/research/research-groups/equator-network.
The list of regulatory documents is provisional since many jurisdictions are developing new guidance. All URLs accessed on 27 September 2021.

Regulation is proposed because risks have been recognized

There are no legal rules for performing a logistic regression—so why has software been included in the new European definition of a medical device, why have ethical guidelines been published, why is a new EU law on AI being debated, and why have many professional standards been proposed (*Table 1*)? To return to the analogy used to frame this debate, numerous reservations about implementing AI mean that a trusting relationship now would be premature.

Discussion and conclusions

The opposing arguments distil into some key questions. For example, do AI algorithms need to be transparent and explicable, before they can be used in clinical practice? Are we ready to trust a result obtained using AI as the basis for making a diagnosis or recommending a particular treatment?

Of course, there is not one 'AI' or 'ML' but a variety of methods for performing supervised and unsupervised analyses that are more or less transparent in their operations, and so statements (including some used in this debate!) should avoid overgeneralizations. It is important to emphasize that the problem or question for which an AI method may offer the most efficient or accurate solution—whether for managing administrative tasks, addressing research questions, measuring images, or analysing data to make diagnostic predictions or provide recommendations for clinical decisions—should always come first. During the COVID-19 pandemic, almost 2500 papers have been published describing some application of AI or ML, but the crucial clinical advances came instead from large simple randomized clinical trials such as RECOVERY.[30]

One of the working groups recently convened by the European Commission to plan the European Health Data Space, was charged to consider 'Which concrete solutions do we identify and actions could we take, to promote cross-border uptake of AI for health care?'—but surely that is the wrong question. Arguably, we need more research to address unanswered questions in cardiovascular medicine—but we should use AI and ML only if they offer the best way to explore particular hypotheses and answer those questions. Uptake of AI is not the primary objective, and innovation is useful only if it is needed and effective.

It seems clear that AI and ML will be good for circumscribed tasks, but they are unlikely to replace either the expert radiologist or the clinical cardiologist. Healthcare professionals who use these new tools need to learn how they can be integrated safely and appropriately, with their methods transparent, their limitations explicit and their outcomes interpretable (*Figure 1*). That is exemplified by an algorithm that was widely used to identify patients with complex health needs, but which was found to have a large racial bias because it estimated needs according to historical health care utilization.[31] For the most significant tasks, namely selecting and prescribing treatment, there are dangers of relying on AI and still many hurdles to overcome. When risks are highest, randomized trials are needed to prove that an algorithm works without bias in each population where it will be applied. Some AI algorithms will need regulatory approval, depending on their risk and application.
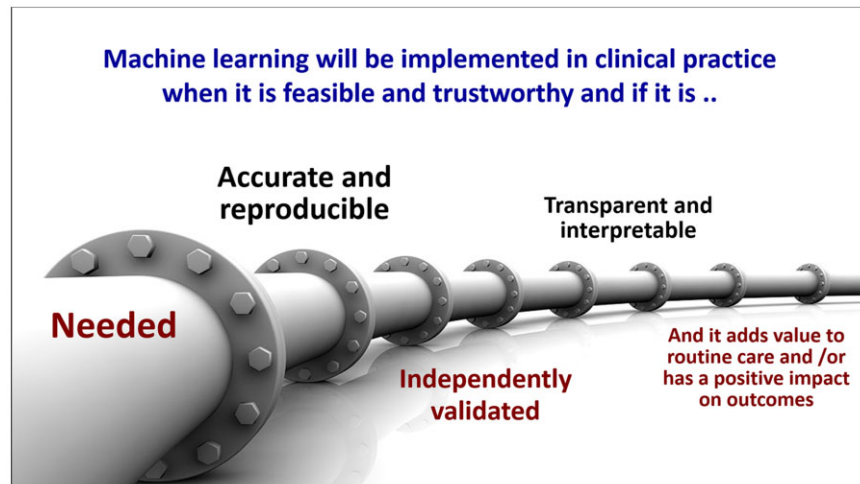
Figure 1 The pipeline of evidence for machine learning tools in medicine. These are the most important factors that need to be addressed when establishing clinical evidence for a new application.

Critics or sceptics are not alone. A report for the National Academy of Medicine in the USA concluded that 'The challenges are unrealistic expectations, biased and non-representative data, inadequate prioritization of equity and inclusion, the risk of exacerbating health care disparities, low levels of trust, uncertain regulatory and tort environments, and inadequate evaluation before scaling narrow AI'.[32] Proponents and enthusiasts should not inflate expectations but ensure that research addresses the right questions. The CORE–MD project (Coordinating Research and Evidence for Medical Devices) which is led by the ESC will develop recommendations for European regulators concerning the approval of AI algorithms as medical devices.[33]

So, for this debate, what is the appropriate analogy? Artificial intelligence and cardiology are already engaged and may even be considering marriage, but now is the time for negotiating a thoughtful prenuptial agreement.

Lead author biography

Alan G. Fraser is Emeritus Professor of Cardiology at the Wales Heart Research Institute, Cardiff University, and Visiting Professor in Cardiovascular Imaging and Dynamics at the University of Leuven. He is a Past-President of the European Association of Cardiovascular Imaging, and now Scientific Coordinator of the EU Horizon 2020 CORE–MD project (Coordinating Research and Evidence for Medical Devices). His research interests include cardiac imaging, heart valve disease, heart muscle disease, and the pathophysiology and diagnosis of heart failure.

Data availability

There are no new data associated with this article.

References

1. McCarthy J, Minsky M, Rochester N, Shannon CE. A proposal for the Dartmouth Summer research project on artificial intelligence. August 1955. http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf (20 October 2021, date last accessed).
2. van de Schoot R, de Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell* 2021; **3**:125–133.
3. Overhage JM, McCallie D Jr. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Ann Intern Med* 2020;**172**: 169–174.
4. Sammani A, Bagheri A, van der Heijden PGM, et al. Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks. *NPJ Digit Med* 2021;**4**:37.
5. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. *NPJ Digit Med* 2019;**2**:26.
6. Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol* 2021;**6**:624–632.
7. Augusto JB, Davies RH, Bhuva AN, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digit Health* 2021;**3**:e20–e28.
8. Schwab K, Nguyen D, Ungab G, et al. Artificial intelligence MacHIne learning for the detection and treatment of atrial fibrillation guidelines in the emergency department setting (AIM HIGHER): Assessing a machine learning clinical decision

support tool to detect and treat non-valvular atrial fibrillation in the emergency department. *J Am Coll Emerg Physicians Open* 2021;**2**:e12534.

9. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* 2021;**18**:465–478.

10. Giudicessi JR, Schram M, Bos JM, et al. Artificial intelligence-enabled assessment of the heart rate corrected QT interval using a mobile electrocardiogram device. *Circulation* 2021;**143**:1274–1286.

11. Corral-Acero J, Margara F, Marciniak M, et al. The 'Digital Twin' to enable the vision of precision cardiology. *Eur Heart J* 2020;**41**:4556–4564.

12. Uijl A, Savarese G, Vaartjes I, et al. Identification of distinct phenotypic clusters in heart failure with preserved ejection fraction. *Eur J Heart Fail* 2021;**23**:973–982.

13. Karwath A, Bunting KV, Gill SK, et al.; card Alc Group and the Beta-blockers in Heart Failure Collaborative Group. Redefining β-blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. *Lancet* 2021;**398**:1427–1435.

14. Park S, Lee SW, Whang M. The analysis of emotion authenticity based on facial micromovements. *Sensors (Basel)* 2021;**21**:4616.

15. Xie B, Sidulova M, Park CH. Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. *Sensors (Basel)* 2021;**21**:4913.

16. Kron FW, Fetters MD, Scerbo MW, et al. Using a computer simulation for teaching communication skills: a blinded multisite mixed methods randomized controlled trial. *Patient Educ Couns* 2017;**100**:748–759.

17. Bishop JM. Artificial intelligence is stupid and causal reasoning will not fix it. *Front Psychol* 2020;**11**:513474.

18. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol* 2021;**31**:3797–3804.

19. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;**1**:e271–e297.

20. 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;**368**:m689.

21. Bhuva AN, Bai W, Lau C, et al. A multicenter, scan-rescan, human and machine learning CMR study to test generalizability and precision in imaging biomarker analysis. *Circ Cardiovasc Imaging* 2019;**12**:e009214.

22. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. *JAMA Intern Med* 2016;**176**:1860–1861.

23. Vasey B, Ursprung S, Beddoe B, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. *JAMA Netw Open* 2021;**4**:e211276.

24. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;**110**:12–22.

25. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health* 2020;**2**:e486–e488.

26. Little MA, Varoquaux G, Saeb S, et al. The need to approximate the use-case in clinical machine learning. *Gigascience* 2017;**6**:1–9.

27. Fraser AG, Tschöpe C, de Boer RA. Diagnostic recommendations and phenotyping for heart failure with preserved ejection fraction—knowing more and understanding less? *Eur J Heart Fail* 2021;**23**:964–972.

28. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;**28**:231–237.

29. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Front Artif Intell* 2020;**3**:561802.

30. Group RC, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med* 2021;**384**:693–704.

31. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;**366**:447–453.

32. Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. *JAMA* 2020;**323**:509–510.

33. Fraser AG, Nelissen RGHH, Kjærsgaard-Andersen P, Szymański P, Melvin T, Piscoi P. Improved clinical investigation and evaluation of high-risk medical devices: the rationale and objectives of CORE–MD (Coordinating Research and Evidence for Medical Devices). *Eur Heart J Qual Care Clin Outcomes* 2021;doi:10.1093/ehjqcco/qcab059.