Article

Consumer wearable devices for evaluation of heart rate control using digoxin versus beta-blockers: the RATE-AF randomized trial

A list of authors and their affiliations appears at the end of the paper

Consumer-grade wearable technology has the potential to support clinical research and patient management. Here, we report results from the RATE-AF trial wearables study, which was designed to compare heart rate in older, multimorbid patients with permanent atrial fibrillation and heart failure who were randomized to treatment with either digoxin or beta-blockers. Heart rate ($n = 143,379,796$) and physical activity ($n = 23,704,307$) intervals were obtained from 53 participants (mean age 75.6 years (s.d. 8.4), 40% women) using a wrist-worn wearable linked to a smartphone for 20 weeks. Heart rates in participants treated with digoxin versus beta-blockers were not significantly different (regression coefficient 1.22 (95% confidence interval (CI) −2.82 to 5.27; $P = 0.55$); adjusted 0.66 (95% CI −3.45 to 4.77; $P = 0.75$)). No difference in heart rate was observed between the two groups of patients after accounting for physical activity ($P = 0.74$) or patients with high activity levels (≥30,000 steps per week; $P = 0.97$). Using a convolutional neural network designed to account for missing data, we found that wearable device data could predict New York Heart Association functional class 5 months after baseline assessment similarly to standard clinical measures of electrocardiographic heart rate and 6-minute walk test (F1 score 0.56 (95% CI 0.41 to 0.70) versus 0.55 (95% CI 0.41 to 0.68); $P = 0.88$ for comparison). The results of this study indicate that digoxin and beta-blockers have equivalent effects on heart rate in atrial fibrillation at rest and on exertion, and suggest that dynamic monitoring of individuals with arrhythmia using wearable technology could be an alternative to in-person assessment. ClinicalTrials.gov identifier: NCT02391337.

The effectiveness and safety of therapeutic interventions are traditionally evaluated with periodic, single time-point assessments. In patients with cardiovascular disease, this often requires multiple clinical visits to obtain tests such as electrocardiograms (ECG) for assessment of heart rate, or 6-minute walk (6MW) tests to appraise physical capacity. These in-hospital measurements are time-consuming, costly and not dynamic, providing a limited 'snapshot' of that person's functional status[1,2]. Wearable technology can provide continuous measurement of physiological parameters[3–5]. However the large volume of data acquired may need advanced analytics, taking account of the lower quality compared with medical devices or frequent missing values[6].

A prime example of the potential clinical value of wearable sensor data could be therapy choice and dose adjustment for heart rate control in patients with atrial fibrillation (AF), an increasingly common rhythm disorder. There is a limited evidence base on this topic; for example, digoxin has typically been considered a poorly effective drug for controlling heart rate, particularly on exertion, although this is based on acute studies only[7]. Wearables offer an opportunity to assess each
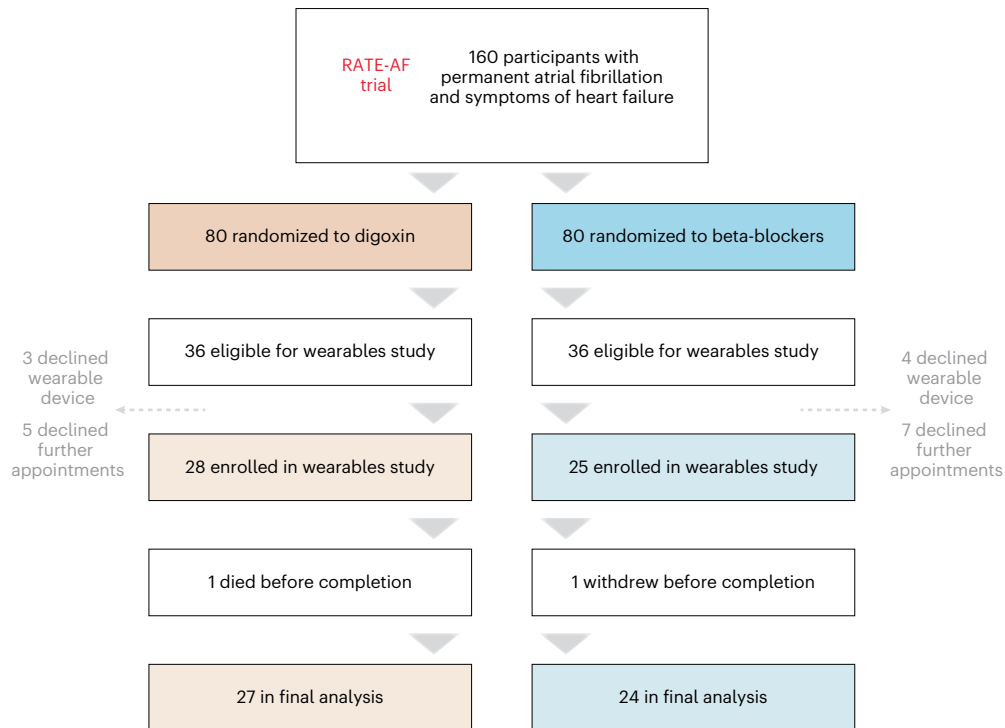
e-mail: d.kotecha@bham.ac.uk

Fig. 1 | RATE-AF wearables substudy flowchart. Flowchart for the wearables study enrollment.

patient in their own environment, with longer-term evaluation better reflecting the extended time taken to achieve therapeutic benefit from digoxin[8], and properly account for physical activity.

Although there is considerable potential for consumer-wearable devices to contribute to cardiovascular disease management, there are few robust studies independent of manufacturers that can highlight opportunities as well as limitations in older, multimorbid patients[9]. Embedded in a randomized controlled trial to mitigate the effects of unmeasured or unknown confounders[10], we hypothesized that a wrist-worn wearable could: (1) address whether digoxin is inferior to beta-blockers for longer-term heart rate control in patients with AF at rest and on exertion; (2) adjust for differences in individual physical activity; and (3) explore whether wearable sensor data are comparable with conventional measurements for the prediction of clinical progress, and ultimately, aid longer-term patient management.

Results

One hundred and sixty patients were randomized to digoxin or beta-blocker therapy in the RAte control Therapy Evaluation in permanent Atrial Fibrillation (RATE-AF) trial, of whom 72 (36 from each treatment arm) were eligible to participate in a substudy that provided a wrist-worn wearable and a connected smartphone. The design and deployment of the study were aided by a patient and public involvement team[11]. The remaining participants had completed the main trial or did not have sufficient time left for data collection (Fig. 1). Of those eligible, 8 in the digoxin group and 11 in the beta-blocker group declined to participate, principally because they did not want to use a wearable device or attend further trial appointments. The characteristics of those who declined were similar to the participating cohort, apart from more women declining and with a lower 6MW distance (Extended Data Table 1).

Fifty-three participants were enrolled in the substudy (Extended Data Fig. 1), with mean age at randomization of 75.6 years (s.d. 8.4; range 61 to 90 years) and 40% women. Twenty-eight participants (53%) had been randomized to digoxin and 25 (47%) to beta-blockers a mean of 30 weeks before their entry into the wearables substudy

(s.d. 8 weeks; range 12 to 46 weeks). Both groups were well balanced with respect to demographics and clinical measurements (Table 1), with the most common comorbidities being hypertension (74%) and heart failure (45%). There were more patients with a formal diagnosis of heart failure in the digoxin group, but ECG-derived left-ventricular ejection fraction (LVEF) measured in a blinded fashion at the start of the trial was similar for both groups: 55.7% (s.d. 8.8%) for digoxin versus 55.4% (s.d. 9.0%) for beta-blockers. One patient from each arm was lost to follow-up.

Heart rate at the start of the substudy was 79.4 beats per minute (bpm; s.d. 12.2) in those randomized to digoxin and 73.1 bpm (s.d. 14.3) for beta-blockers, with one patient in each group presenting at that visit in sinus rhythm. The capacity for physical activity was similar in both groups at baseline, with a median 6MW distance of 351 m for digoxin (i.q.r. 120–454) and 357 m for beta-blockers (i.q.r. 295–411). Symptoms of AF and heart failure were reported in all patients (100%), with modified European Heart Rhythm Association (mEHRA) class 2a or above, and New York Heart Association (NYHA) class II or above. At the trial mid-point, the mean dose of digoxin used was 157 μg (s.d. 10 μg) with a serum digoxin level of 0.79 ng ml^{-1} (s.d. 0.29); the mean dose in the beta-blocker group was 3.2 mg of bisoprolol (s.d. 1.9). The duration of the wearable substudy was a median of 23 weeks (i.q.r. 4–38). There was consistent use of the wrist-worn wearable (more so than the smartphone), with 90.4% of participants using this every day for the last 7 days before the interim review (Extended Data Table 2).

Wearable-acquired heart rate and physical activity

Per patient, the mean duration of ambulatory sensor data collected was 20 weeks (s.d. 7), with an average of 2,623,951 heart rate data points for each patient treated with digoxin (s.d. 907,697) and 2,796,367 for each patient treated with beta-blocker (s.d. 811,956). Across all patients there were 143,379,796 data intervals collected for heart rate and 23,704,307 for corresponding physical activity (Table 2). Figure 2 highlights the considerable variability within and across individual patients despite appropriate rate control therapy in terms of heart rate, response to exertion and the correlation between heart rate and physical activity.

Table 1 | Baseline characteristics

Characteristic	Overall (*N*=53)	Randomized to digoxin (*N*=28)	Randomized to beta-blockers (*N*=25)
Demographics			
Age at randomization, mean years (s.d.)	75.6 (8.4)	74.2 (8.4)	77.2. (8.3)
Gender, women *n* (%)	21 (39.6)	11 (39.3)	10 (40.0)
Baseline comorbidities			
Hypertension, *n* (%)	39 (73.6)	22 (78.6)	17 (68.0)
Diabetes mellitus, *n* (%)	11 (20.8)	5 (17.9)	6 (24.0)
Previous stroke or transient ischemic attack, *n* (%)	8 (15.1)	5 (17.9)	3 (12.0)
Treatment with inhalers for COPD or asthma, *n* (%)	14 (26.4)	8 (28.6)	6 (24.0)
Diagnosed with heart failure, *n* (%)	24 (45.3)	16 (57.1)	8 (32.0)
LVEF on echocardiogram, mean % (s.d.)	55.6 (8.8)	55.7 (8.8)	55.4 (9.0)
Echocardiogram LVEF <50%, *n* (%)	18 (34.0)	11 (39.3)	7 (28.0)
Clinical measurements at baseline trial visit			
ECG heart rate, mean bpm (s.d.)	97.7 (20.4)	95.7 (19.8)	99.6 (20.5)
Systolic blood pressure, mean mmHg (s.d.)	137.9 (17.2)	136.2 (15.2)	139.7 (19.3)
6MW distance, median meters (i.q.r.)	384 (207–437)	372 (150–434.5)	384 (229–438)
NT-proBNP, median pg ml⁻¹ (i.q.r.)	1099 (770–1725)	1112 (766–1831)	1057 (829–1717)
NYHA class, *n* (%)			
I (no limitation of activity)	0 (0)	0 (0)	0 (0)
II (slight limitation of activity)	38 (71.7)	22 (78.6)	16 (64.0)
III (marked limitation of activity)	14 (26.4)	5 (17.9)	9 (36.0)
IV (symptoms of heart failure at rest)	1 (1.9)	1 (3.6)	0 (0)
mEHRA class, *n* (%)			
1 (no symptoms)	0 (0)	0 (0)	0 (0)
2a (mild symptoms)	3 (5.7)	1 (3.6)	2 (8.0)
2b (moderate symptoms)	26 (49.1)	16 (57.1)	10 (40.0)
3 (severe symptoms)	22 (41.5)	11 (39.3)	11 (44.0)
4 (disabling symptoms)	2 (3.8)	0 (0)	2 (8.0)

COPD, chronic obstructive pulmonary disease.

Digoxin versus beta-blockers

Weekly averages of heart rate were no different when comparing patients randomized to digoxin or beta-blockers (Extended Data Table 3). Accounting for all repeated measurements over time, there was no significant difference in heart rate comparing digoxin and beta-blocker therapy using the wearable sensors (Fig. 3). There was no interaction seen according to gender ($P_{interaction}$ = 0.39). The unadjusted regression coefficient for digoxin versus beta-blockers was 1.22 (95% CI −2.82 to 5.27; P = 0.55), and 0.66 when adjusted for age, gender, diagnosis of heart failure and N-terminal pro-hormone B-natriuretic peptide

Table 2 | Sensor data from the wearable device

Data collection	Randomized to digoxin (*N*=28)	Randomized to beta-blocker (*N*=25)
Total number of data points for heart rate	73,470,613	69,909,183
Total number of data points for step count	12,210,254	11,494,053
Number of combined data points for heart rate and step count	4,746,169	4,683,959
Mean number of data points for heart rate per patient (s.d.)	2,623,951 (907,697)	2,796,367 (811,956)
Mean number of data points for step count per patient (s.d.)	436,081 (122,493)	459,762 (110,443)
Mean number of combined data points for heart rate and step count per patient (s.d.)	169,506 (71,217)	187,358 (57,784)
Mean timespan for data collection per patient, days (s.d.)	153 (53)	160 (46)

The data presented are for nonmissing time points.

(NT-proBNP) (95% CI −3.45 to 4.77; P = 0.75). There remained no difference in heart rate between the digoxin and beta-blocker groups after accounting for physical activity (P = 0.74). Post-hoc adjusted subgroup analysis according to activity levels found no difference in heart rate between digoxin and beta-blockers in those with low weekly-averaged activity (<15,000 steps per week; 298 weeks from 44 patients; P = 0.48), minimum recommended activity (15,000–30,000 steps per week; 316 weeks from 37 patients; P = 0.47) or recommended activity (≥30,000 steps per week; 417 weeks from 33 patients; P = 0.97).

Exploratory analysis of wearable data to predict NYHA class

A convolutional neural network (CNN) model was trained on heart rate and step count data using the wearable sensor output from 41 patients who had sufficient time windows for analysis. F1 scores (combining precision and recall) were used to compare the CNN model for prediction of NYHA class at the end of the trial (Extended Data Table 4), with chance returning an F1 score of 0.35. The wearables CNN yielded an F1 score of 0.56 (95% CI 0.41 to 0.70); Fig. 4. This was equivalent to a model generated from conventional trial parameters (ECG heart rate and 6MW test results), which returned an F1 score of 0.55 (95% CI 0.41 to 0.68); P = 0.88 for patient-level comparison with wearables CNN. The wearable data appeared independent of clinical factors such as age, gender and body mass index, with similar F1 score when combining wearable data with clinical factors (0.58; 95% CI 0.45 to 0.73). The corresponding areas under the receiver operator characteristic curves were 0.73 for ECG heart rate and 6MW test, 0.77 for the wearables CNN and 0.78 for wearables plus clinical factors.

Discussion

The RATE-AF wearables study demonstrated a potential use of nonmedical-grade wearable devices in clinical research where they were used to monitor clinical progress and the response to a change in therapy. Embedded in a randomized trial, the study provides robust information on the value, but also the limitations of using these devices. The commonly held view originating from acute trials that digoxin is inferior to beta-blockers for heart rate control in AF was not seen in this longer-term study. No difference in heart rate was seen when considering more than 140 million data points at rest and on exertion during a 20-week period. The equivalence of digoxin to beta-blockers held after accounting for individual differences in day-to-day physical activity, and no difference in heart rate was evident even in periods with high activity. Although limited by the number of patients, the wearable sensor data appeared to be comparable with conventional trial outcomes
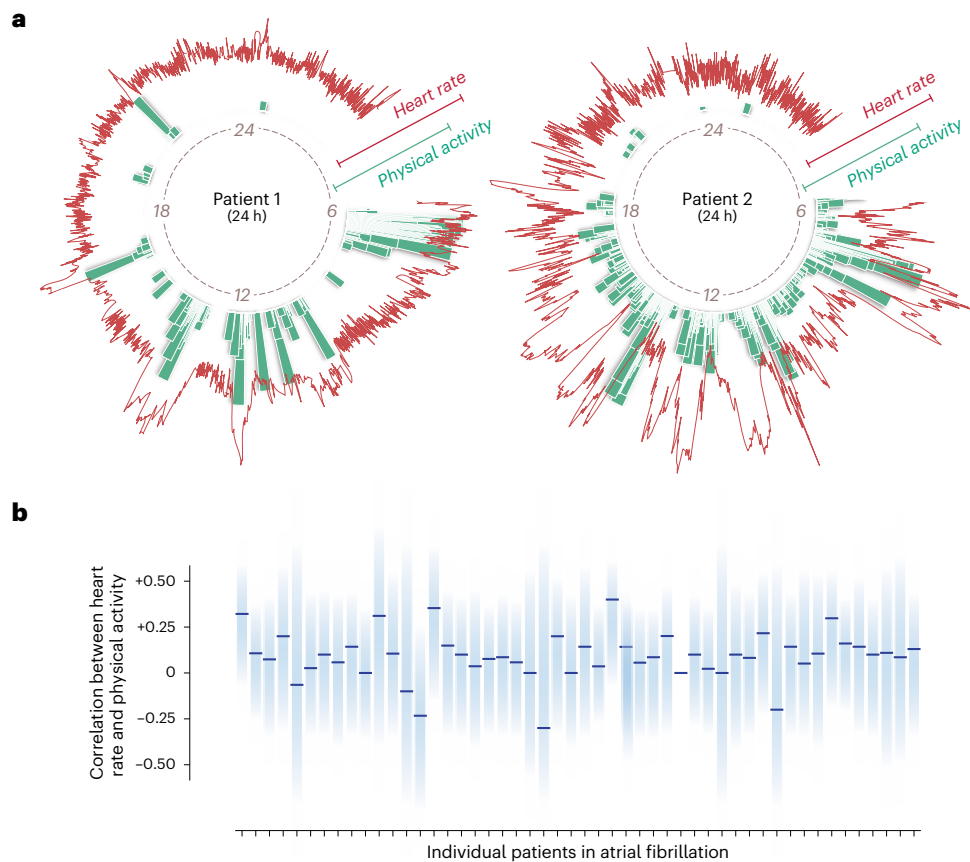
Fig. 2 | Wearable device data for measurement of heart rate and physical activity. a, Examples of data capture for heart rate (red lines) and step count (green bars) using a wrist-worn wearable and smartphone over a single 24-h period for two individual patients with AF and heart failure. **b**, Correlations between daytime 10-s intervals of heart rate and physical activity for 50 patients who remained in AF at each visit. Light blue columns indicate the range of positive and negative correlations between heart rate and physical activity, with medians indicated by dark blue bars (correlation <0.19, very weak; 0.20–0.59, weak to moderate).

for prediction of a clinical outcome (in this case, functional class), but required the development of a neural network pipeline for appropriate analytics. Standardization of these approaches could lead to wearable devices contributing to, or even replacing resource-intensive clinical tests and visits in the future.

The majority of clinical trials in cardiovascular research utilize in-person periodic testing that not only requires patient and staff attendance, but is time-constrained in an environment unlike the patient's own surroundings. Wearables offer an exciting possibility for patient-directed data collection that better reflects real-life day-to-day variations in heart rate and physical activity[12]. In this study, a neural network that was self-training was designed to provide value from the vast amount of information collected by wearables, and address key issues such as inconsistent and missing data. The performance of the wearable neural network was not dependent on clinical factors, and for prediction of future NYHA class was equivalent to (but not better than) conventional parameters such as a 12-lead ECG and 6MW testing. F1 scores >0.5 indicate an acceptable balance between precision and recall, and the wearable model score of 0.56 was significantly better than chance (in this case, 0.35). Further development, testing and validation of these approaches is clearly required. Of note, the trial recruited older patients (mean age 76 years, range up to 90 years) who indicated the use of wearables was appealing to monitor and improve their own functional capacity.

Several studies have utilized wearable devices in cardiovascular research. The Apple smartwatch and Fitbit studies used photoplethysmography to identify new AF[13,14], and there are many different devices being used across heart failure to quantify patient physiology[15]. Early optimism has already started to fade somewhat, and one retrospective matched study in the US found that patients using wearable devices accessed more healthcare but with no difference in heart rate[16]. It remains to be seen how these consumer-bought, nonmedical devices will be integrated into routine care, with or without the acceptance of healthcare professionals[1,17]. Further randomized trials are needed to understand whether continuous monitoring can provide a personalized assessment of treatment response, or help to identify subphenotypes of disease. Similarly, approaches to transparent and validated artificial intelligence (AI) remain in their infancy[18]. AI has demonstrated an ability to go beyond our current linear understanding of disease trajectory and interactions, with the ability to personalize diagnostic and therapeutic strategies even in multimorbid conditions[19,20]. Despite the potential benefits of AI, broader use in clinical practice requires an approach guided by strong methodological principles[6].

AF is a key public health issue, predicted to double in prevalence over the next few decades[21]. Recent evidence suggests that cognitive decline and vascular dementia should be added to the list of adverse events suffered by patients with AF[22,23]. New technologies have rapidly increased our ability to detect AF, but how these should be implemented in clinical practice remains unclear[9]. Despite permanent AF being the most common 'type' of AF[24], there is a remarkable lack of research to support decision-making and improve patient quality of life[7,25–27]. Treatment choices for heart rate control are often made based on evidence from heart failure trials; however, the reduction in mortality from beta-blockers was not evident in double-blind trials in
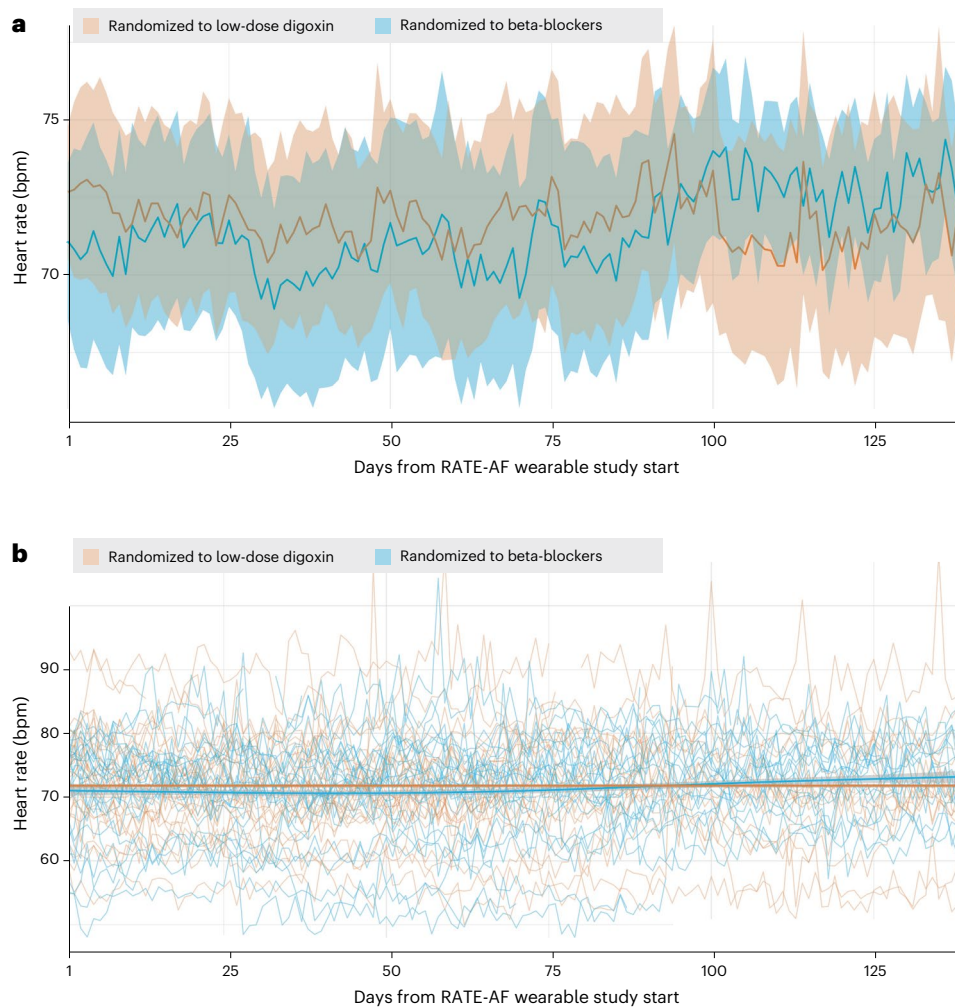
Fig. 3 | Heart rate data from the wearable device. The start of the wearables study was preceded by a period of dose adjustment and stabilization (mean of 30 weeks from randomization). **a**, Mean (solid line) and s.d. (shaded area) in heart rate over the 20-week period of follow-up in patients randomized to treatment with digoxin (brown) or beta-blockers (blue). **b**, Individual patient heart rate trajectories over the 20-week period of follow-up in patients randomized to treatment with digoxin (brown) and beta-blockers (blue). Bold lines indicate the fitted generalized linear model curves with corresponding 95% CI (shaded). No significant difference was demonstrated between the two groups using a generalized linear model with random-effects to account for repeated measurements (unadjusted $P = 0.55$; adjusted $P = 0.75$; after accounting for physical activity $P = 0.74$).

patients with heart failure and concomitant AF[28]. Digoxin has traditionally been reserved for sicker patients or as a second-line agent. Similar to beta-blockers, there is no apparent mortality impact from digoxin in patients with AF and heart failure, and observational analyses are wholly inappropriate to evaluate outcomes with digoxin because of systematic prescription biases[29,30]. The RATE-AF trial was a head-to-head randomized trial of digoxin and beta-blockers, showing that digoxin had a similar impact on patient-report quality life as beta-blockers yet substantially improved functional class. Use of digoxin led to a significant reduction in natriuretic peptides, and fewer than half the number of primary care and unplanned hospital visits[10]. This study now adds a further dimension, showing that low-dose digoxin can be an effective rate control agent. The effect on heart rate and physical activity was consistent with the longer-term activity of digoxin, including pro-parasympathetic effects on cellular, electrophysiological and neurohormonal pathways[8].

Although the number of patients was limited and did not encompass all participants in RATE-AF, this study included a large volume of wearable data and was able to benefit from being embedded in a randomized trial to limit extraneous bias. The wearables were implemented post randomization and hence there is a risk of residual confounding; however, there was no crossover use of digoxin[10], analysis was by

intention-to-treat and reasonable balance in clinical characteristics between groups was maintained for those patients joining the wearable study. As expected, actual usage of wearable devices among participants had considerable variation. There is a known motivational and behavioral impact from wearables, although when asked at final follow-up, the same proportion in each randomized group indicated improved motivation for physical activity: 16 patients (59.3%) in the digoxin group and 13 (54.2%) in the beta-blocker group (P for comparison = 0.71). Use in an older population brought challenges, with many participants not previously owning a smartphone; however, good compliance and data quality were achieved. Missing data points were frequent, but their impact was minimized by using an innovative approach to embed machine learning on the significance of missing data, rather than ignore or impute it. Although subgroup analysis according to activity levels found no difference in heart rate between digoxin and beta-blockers, the use of digoxin in high-activity settings remains largely untested. Diversity in ethnicity was limited (white British or Irish ethnicity accounted for 94% of participants), and so these results cannot be extrapolated to other ethnicity groups. Although technology advances can improve human health, it remains important to ensure that deployment of such devices does not exacerbate health inequalities.
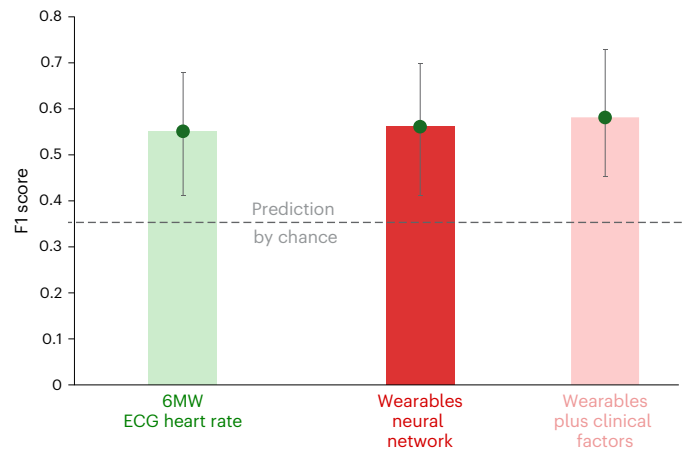
Fig. 4 | Prediction of clinical progress with a neural network based on wearable data. F1 scores combining precision and recall of each model are presented along with 95% CI for the prediction of NYHA functional class at the end of follow-up (mean 5 months); an F1 score of 0.35 (dashed line) is equivalent to chance. Derived from wearable sensor data from $n = 41$ individual patients.

In summary, a wrist-worn consumer-grade wearable device and smartphone were successfully deployed within a randomized controlled trial of older, multimorbid patients to evaluate continuous, ambulatory heart rate and physical activity. Including an average of two to three million data points per patient, including at rest and exertion, digoxin and beta-blocker therapy had similar effects on heart rate measured over a 20-week period. A neural network model of wearable sensor data showed similar performance for predicting future health status as conventional measures used in clinical trials.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-024-03094-4.

References

1. Izmailova, E. S., Wagner, J. A. & Perakslis, E. D. Wearable devices in clinical trials: hype and hypothesis. *Clin. Pharmacol. Ther.* **104**, 42–52 (2018).
2. Bakobaki, J. M. et al. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. *Clin. Trials* **9**, 257–264 (2012).
3. Perry, B. et al. Use of mobile devices to measure outcomes in clinical research, 2010–2016: a systematic literature review. *Digit. Biomark.* **2**, 11–30 (2018).
4. Pevnick, J. M., Birkeland, K., Zimmer, R., Elad, Y. & Kedan, I. Wearable technology for cardiology: an update and framework for the future. *Trends Cardiovasc. Med.* **28**, 144–150 (2018).
5. Redfield, M. M. et al. Isosorbide mononitrate in heart failure with preserved ejection fraction. *N. Engl. J. Med.* **373**, 2314–2324 (2015).
6. Gill, S. K. et al. Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *Eur. Heart J.* **44**, 713–725 (2023).
7. Kotecha, D. et al. A review of rate control in atrial fibrillation, and the rationale and protocol for the RATE-AF trial. *BMJ Open* **7**, e015099 (2017).
8. Ziff, O. J. & Kotecha, D. Digoxin: the good and the bad. *Trends Cardiovasc. Med.* **26**, 585–595 (2016).
9. Gill, S. et al. Smartphone detection of atrial fibrillation using photoplethysmography: a systematic review and meta-analysis. *Heart* **108**, 1600–1607 (2022).
10. Kotecha, D. et al. Effect of digoxin vs bisoprolol for heart rate control in atrial fibrillation on patient-reported quality of life: the RATE-AF randomized clinical trial. *JAMA* **324**, 2497–2508 (2020).
11. Bunting, K. V., Stanbury, M., Tica, O. & Kotecha, D. Transforming clinical research by involving and empowering patients—the RATE-AF randomized trial. *Eur. Heart J.* **42**, 2411–2414 (2021).
12. Bayoumy, K. et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nat. Rev. Cardiol.* **18**, 581–599 (2021).
13. Perez, M. V. et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N. Engl. J. Med.* **381**, 1909–1917 (2019).
14. Lubitz, S. A. et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit Heart Study. *Circulation* **146**, 1415–1424 (2022).
15. DeVore, A. D., Wosik, J. & Hernandez, A. F. The future of wearables in heart failure patients. *JACC Heart Fail.* **7**, 922–932 (2019).
16. Wang, L. et al. Association of wearable device use with pulse rate and health care use in adults with atrial fibrillation. *JAMA Netw. Open* **4**, e215821 (2021).
17. Sana, F. et al. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **75**, 1582–1592 (2020).
18. Liu, X. 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* **1**, e271–e297 (2019).
19. Yao, X. et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat. Med.* **27**, 815–819 (2021).
20. Karwath, A. et al. Redefining beta-blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. *Lancet* **398**, 1427–1435 (2021).
21. Lane, D. A., Skjoth, F., Lip, G. Y. H., Larsen, T. B. & Kotecha, D. Temporal trends in incidence, prevalence, and mortality of atrial fibrillation in primary care. *J. Am. Heart Assoc.* **6**, e005155 (2017).
22. Mobley, A. R. et al. Thromboembolic events and vascular dementia in contemporary patients with atrial fibrillation and low apparent stroke risk. *Nat. Med.* https://doi.org/10.1038/s41591-024-03049-9 (2024).
23. Schnabel, R. B. et al. Early diagnosis and better rhythm management to improve outcomes in patients with atrial fibrillation: the 8th AFNET/EHRA consensus conference. *Europace* **25**, 6–27 (2023).
24. Chiang, C. E. et al. Distribution and risk profile of paroxysmal, persistent, and permanent atrial fibrillation in routine clinical practice: insight from the real-life global survey evaluating patients with atrial fibrillation international registry. *Circ. Arrhythm. Electrophysiol.* **5**, 632–639 (2012).
25. Steg, P. G. et al. Symptoms, functional status and quality of life in patients with controlled and uncontrolled atrial fibrillation: data from the RealiseAF cross-sectional international registry. *Heart* **98**, 195–201 (2012).
26. Nabauer, M. et al. The Registry of the German Competence NETwork on Atrial Fibrillation: patient characteristics and initial management. *Europace* **11**, 423–434 (2009).
27. Jones, J. et al. Importance and assessment of quality of life in symptomatic permanent atrial fibrillation: patient focus groups from the RATE-AF Trial. *Cardiology* **145**, 666–675 (2020).
28. Kotecha, D. et al. Efficacy of β blockers in patients with heart failure plus atrial fibrillation: an individual-patient data meta-analysis. *Lancet* **384**, 2235–2243 (2014).

29. Ziff, O. J. et al. Safety and efficacy of digoxin: systematic review and meta-analysis of observational and controlled trial data. *Br. Med. J.* **351**, h4451 (2015).

30. Champsi, A. et al. Digoxin in patients with heart failure and/or atrial fibrillation: a systematic review and meta-analysis of 5.9 million patient years of follow-up. Preprint at *SSRN* https://doi.org/10.2139/ssrn.4544769 (2023).

Simrat K. Gill[1], Andrey Barsky[2], Xin Guan[2], Karina V. Bunting[1,3,4], Andreas Karwath [2,3,4], Otilia Tica [1], Mary Stanbury[5], Sandra Haynes[5], Amos Folarin [6,7], Richard Dobson [6,7], Julia Kurps[8], Folkert W. Asselbergs[9,10], Diederick E. Grobbee[11], A. John Camm [12], Marinus J. C. Eijkemans[10], Georgios V. Gkoutos [2,4], Dipak Kotecha [1,3,4,9] ✉, BigData@Heart Consortium*, the cardAIc group* & the RATE-AF trial team*

[1]Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, UK. [2]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. [3]West Midlands NHS Secure Data Environment, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [4]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [5]Patient and Public Involvement Team, Birmingham, UK. [6]Department of Biostatistics & Health Informatics, King's College London, London, UK. [7]Health Data Research UK, University College London, London, UK. [8]Real World Data team, The Hyve, Utrecht, the Netherlands. [9]Department of Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. [10]Amsterdam University Medical Center, Department of Cardiology, University of Amsterdam, Amsterdam, the Netherlands. [11]Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands. [12]Cardiology Clinical Academic Group, St George's University of London, London, UK. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: d.kotecha@bham.ac.uk

BigData@Heart Consortium

Simrat K. Gill[1], Amos Folarin[6,7], Richard Dobson[6,7], Julia Kurps[8], Folkert W. Asselbergs[9,10], Diederick E. Grobbee[11], Marinus J. C. Eijkemans[10] & Dipak Kotecha[1,3,4,9]

the cardAIc group

Andrey Barsky[2], Xin Guan[2], Karina V. Bunting[1,3,4], Andreas Karwath[2,3,4], Otilia Tica[1], Mary Stanbury[5], Sandra Haynes[5], Georgios V. Gkoutos[2,4] & Dipak Kotecha[1,3,4,9]

the RATE-AF trial team

Simrat K. Gill[1], Karina V. Bunting[1,3,4], Mary Stanbury[5], Sandra Haynes[5], A. John Camm[12] & Dipak Kotecha[1,3,4,9]

Methods

The RATE-AF trial

The RATE-AF trial was a prospective, randomized, open-label, blinded end-point trial that compared the use of low-dose digoxin versus beta-blockers for long-term heart rate control[7]. Recruitment took place across primary care sites and three hospitals in the West Midlands region of England between 2016 and 2018. Inclusion criteria were: (1) age 60 years or older; (2) permanent AF in need of rate control; and (3) symptoms of heart failure, with breathlessness equivalent to NYHA class II or above. Exclusion criteria were limited so that the trial population reflected routine clinical practice (see published protocol for full list of selection criteria)[10].

Ethics and inclusion statement

The trial was co-designed by a patient and public involvement (PPI) team, with the aim of improving quality of life for patients with AF[11,27]. Ethical approval was obtained from the East Midlands–Derby Research Ethics Committee (16/EM/0178), the Health Research Authority (IRAS 191437) and the Medicines and Healthcare Products Regulatory Agency. The trial was publicly funded by the UK National Institute for Health and Care Research (CDF-2015-08-074) and registered with clinicaltrials.gov (NCT02391337) and clinicaltrialsregister.eu (2015-005043-13).

Randomization and trial process

Each participant was randomized to either digoxin 62.5–250 µg or bisoprolol 1.25–10 mg once daily in a 1:1 ratio at their baseline visit. Randomization was completed using a computer-generated minimization algorithm to ensure treatment arms were balanced for gender and AF symptoms, based on the mEHRA classification.

The trial was embedded into usual care within the National Health Service (NHS), with participants attending formal visits at baseline, 6 months and 12 months. Endpoints acquired were patient-reported quality of life, NT-proBNP, symptoms and functional capacity using mEHRA and NYHA class, 6MW distance and time, heart rate (pulse examination), 12-lead ECG, LVEF using cardiac ultrasound and assessment of adverse events.

Wearables substudy

Funding for the wearables substudy was obtained after the main trial had commenced from the European Union Innovative Medicines Initiative BigData@Heart program (grant no. 116074). The study was supported by the Application of Artificial Intelligence to Routine Healthcare Data to Benefit Patients with Cardiovascular Disease (cardAIc) team at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust. An amendment was made to the trial protocol and subsequently approved by the Research Ethics Committee. One of the original stated aims of the substudy was to correlate wearable sensor data with patient quality of life using the Short Form (36) Health Survey (SF-36). Following work led by the PPI team that showed SF-36 to be as suboptimal measure of assessment[27], this was subsequently changed to NYHA class in the statistical analysis plan completed before data analysis (Supplementary Note).

Participants with at least 2 months remaining in the RATE-AF trial were considered eligible for inclusion in the substudy. All participants were provided with a specific patient information leaflet written by the PPI team, and were asked to sign an optional form to indicate informed consent. As an exploratory analysis, no sample size calculation was performed in advance of recruitment. Heart rate sensor data from the first 5 weeks in the first ten participants was used to estimate the minimum number of participants needed. The average weekly heart rate, s.d. and correlation of repeated measures from this data indicated that a sample size of 40 participants would provide 90% power over 20 weeks to detect a 1/3 s.d. difference in heart rate (2 bpm) between digoxin and beta-blockers (control 72 bpm; s.d. 6 bpm; repeated measures correlation 0.91; two-sided alpha 0.05). A minimum target of 50

enrolled participants would account for death and loss to follow-up during the substudy.

Consenting individuals were given a Samsung A6 Android smartphone (with a prepaid mobile data contract) and wrist-worn Fitbit Charge 2 wearable device for passive monitoring. There were no exclusions related to age or previous proficiency with information technology. Applications were preinstalled and set up for remote data collection, providing active monitoring and an educational resource for patients, including the European Society of Cardiology smartphone application (app) specifically designed for patients with AF[31]. Participants were shown how to charge and carry out basic functions on each device, and how to use the apps. They were instructed to carry the phone throughout the day and to wear the wrist device continuously, only removing it for showering, bathing, swimming or charging. After the set-up appointment, in-person or telephone follow-up was provided after the first week, after 4 weeks and ad hoc to maintain engagement and address any concerns or technical issues raised by participants.

Data collection and storage

Data collected via the device and smartphone was encrypted and uploaded to a secure server, temporarily cached on the smartphone until an appropriate Wi-Fi or mobile data connection was available. The collection of wearable data streams was automated using the RADAR-base platform, funded by the European Union Innovative Medicines Initiative RADAR-CNS (grant no. 115902)[32]. This platform allows for secure streaming of data from wearables, apps and devices to a central location. For this study, the RADAR-base platform was installed on a virtual machine hosted by Amazon Web Services in the Europe (London) region and was maintained by the Hyve (IT company, Netherlands). By applying for Fitbit developer application, the RADAR-base platform automatically collected data from registered participants, who were also able to see their own heart rate and step counts. For clinical data storage, secure electronic case report forms were generated using the Research Electronic Data Capture (REDCap) system hosted by the University of Birmingham, and the main trial case report forms hosted by the Birmingham Clinical Trials Unit.

Statistical analysis

Data were analyzed by intention-to-treat according to the randomized allocation (digoxin versus beta-blockers), with no imputation for any missing data. Continuous measurements of heart rate and step count were pooled at 1-min intervals to form time-series data (heart rate averaged over each minute; step counts summed over each minute), with the primary analysis over a prespecified period of the first 20 weeks of device use. The results were summarized and presented as a number, percentage, mean and s.d. or standard error of the mean, or median with i.q.r. The Kruskal–Wallis nonparametric test or a t-test were used to determine differences between the two treatment arms depending on normality, and Spearman's test was used to quantify correlations. To account for multiple repeated measurements of heart rate over time in individual participants, generalized linear models were generated using a random-effects estimator and exchangeable correlation matrix. A post-hoc subgroup analysis according to activity levels was based on US Centers for Disease Control recommended activity levels (minimum 150 min per week aerobic activity equivalent to 15,000 steps per week, and health benefits goal of 300 min per week aerobic activity equivalent to 30,000 steps per week). Statistical analyses were performed using Stata v.17 (StataCorp LP), with a two-tailed P-value <0.05 denoting statistical significance.

Neural network

Machine learning algorithms were generated to explore whether continuous sensor data were comparable with conventional periodic trial measurements at the closest RATE-AF trial appointment, developed

according to our previously published AI framework[6]. Unlabeled wearable sensor data from staggered 4-h periods were used to develop a self-supervising CNN (Extended Data Fig. 2). The self-supervising model was motivated by the principle that important information is carried not only in the heart rate and step count channels, but also in the temporal interaction between those channels. To learn this interaction, an auxiliary dataset was synthesized from a training set of the original sensor data where the heart rates and step counts of each sample were scrambled across patients and dissociated. For example, a multichannel sample might include the heart rate time-series of patient A, but the step count time-series of patient B. The auxiliary dataset was combined with the original data to create a classification problem: to discriminate whether a given sample came from the original or scrambled data. Because the original data have temporal interdependencies between the channels, and the scrambled data do not, it is believed that learning this objective is equivalent to learning the relationship between those sensor channels.

Heart rate measures were standardized to z-scores; step counts were normalized to the range [0,1] because of frequent and meaningful measurements of zero (inactivity). Both were defined with respect to each participant's individual statistics—a heart rate z-score of 0 indicates the mean average heart rate for that patient. Small amounts of missing data were present throughout the recordings: these may have been short periods in which participants were not wearing their devices, or where data was not received because of connectivity issues. This missing data was neither dropped nor imputed, but used as a third time-series channel alongside heart rate and step count. This allowed the model to learn the significance of missing data instead of making assumptions about its distribution.

Multichannel time-series data were the input for a one-dimensional convolutional layer of 8 filters and a kernel size of 21 (minutes), followed by a one-dimensional max pooling layer of size and stride 2. After pooling are two further convolutional layers with 20 and 32 filters, each with kernel size 21. Finally, one-dimensional global average pooling was performed to reduce the data representation to a vector of length 32. During training, dropout (with probability = 0.5) was applied to this layer to improve regularization. Finally, the prediction is made by a fully connected layer comprising a single sigmoid unit. Every layer but the last used rectified linear unit activation, and the network was trained to minimize binary cross-entropy using an Adam optimizer with a learning rate of 3×10^{-4} and L2 regularization with weight 1×10^{-8} applied to each nonbias parameter. After training the network to convergence, the dropout and classification layers were removed from the model, and the output of the final convolutional layer was used as a 32-dimensional embedding vector representing the time-series data used as input.

Because the objective of this model was to predict the patient's future NYHA class, it was evaluated by embedding data from each patient's first week, and using that embedding to predict their NYHA class at the end of trial as the outcome of interest. This self-supervised model was trained using all data for each patient other than in this first week, while also holding out a subset of patients as a validation set to monitor under- or overfitting during model training. The hold-out set comprised 20% of the patient group, repeated across five iterations with k-fold cross-validation. For this exploratory analysis, participants were only included if they had available nonmissing time windows in the first and subsequent 19 weeks, and reached the final follow-up assessment for NYHA class. Models were compared for prediction of NYHA class at the end of the trial (5 months later): a conventional logistic regression model including ECG heart rate and 6MW test results (distance traveled, time taken and participant speed), and the wearable sensor model using CNN latent time-series embeddings from wearable sensor data as input features with L2 norm regularization (as used in ridge regression).

We prespecified evaluation of models using the F1 score—a metric combining precision and recall that ranges from 0 to 1, with 1 indicating perfect accuracy for classification. For each model, label smoothing was used over the NYHA class targets as a further method of regularization. The 95% CI was estimated by bootstrap resampling. During the peer review process, a post-hoc analysis was added to calculate the area under the receiver operator characteristic curve for each model, which provides an aggregate measure of classification performance with values ranging from 0 to 1 (higher indicates better performance). Machine learning analyses were performed using Python (Python Software Foundation) with scikit-learn, and TensorFlow (Google Brain).

Role of the funding sources
None of the organizations providing funding had any role in the design or conduct of the study (including collection, analysis and interpretation of the data) or any involvement in preparation, review or approval of the manuscript.

Reporting frameworks
The study is reported according to the Minimum Information about Clinical Artificial Intelligence Modeling (MI-CLAIM) checklist[33].

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Summary anonymized wearable sensor data are available for noncommercial purposes on request to the corresponding author (d.kotecha@bham.ac.uk; 60 days response time for decisions). Because of the risk of patient reidentification, access to any individual-level data will require an appropriate ethical committee approval and review by the RATE-AF trial oversight committee, which includes patient and public representatives (applications to D. Kotecha, d.kotecha@bham.ac.uk; 180 days response time for decisions). Anonymized RATE-AF main trial and substudy datasets will be made available in an open-access repository after completion of secondary manuscripts.

Code availability
The machine learning framework for embedding multichannel time-series data is made freely available at: https://github.com/gkoutos-group/wearable_data_embedding.

References
31. Kotecha, D. et al. European Society of Cardiology smartphone and tablet applications for patients with atrial fibrillation and their health care providers. *Europace* **20**, 225–233 (2018).
32. Ranjan, Y. et al. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. *JMIR Mhealth Uhealth* **7**, e11734 (2019).
33. Norgeot, B. et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat. Med.* **26**, 1320–1324 (2020).

Author contributions

Competing interests

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-024-03094-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03094-4.

Correspondence and requests for materials should be addressed to Dipak Kotecha.

Peer review information *Nature Medicine* thanks Laurent Billot, Isabelle van Gelder, Xiaoxi Yao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Michael Basson, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.
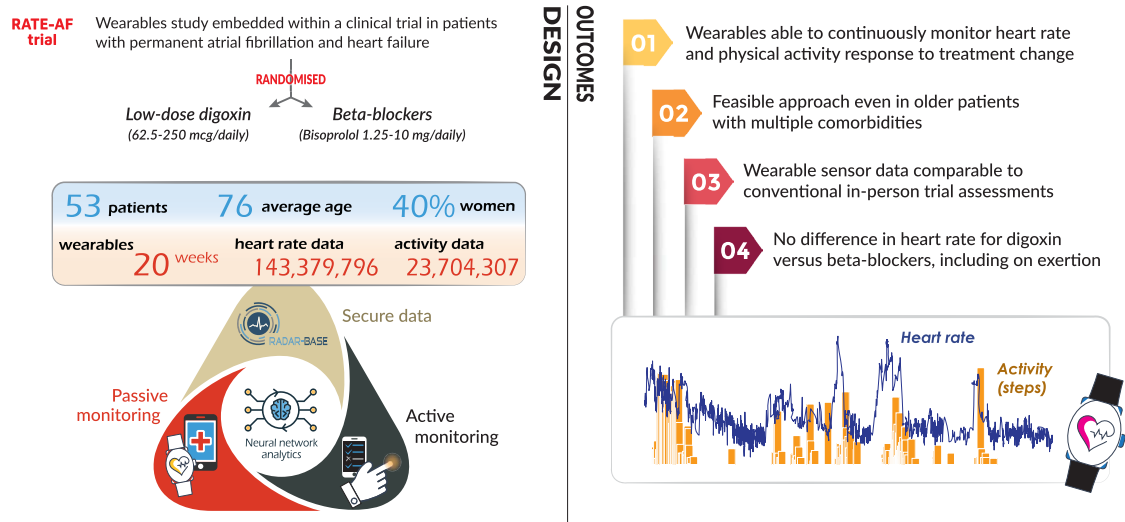
Extended Data Fig. 1 | Overview of the RATE-AF randomized trial wearable study. mcg = micrograms; mg = milligrams; RATE-AF = RAte control Therapy Evaluation in permanent Atrial Fibrillation.
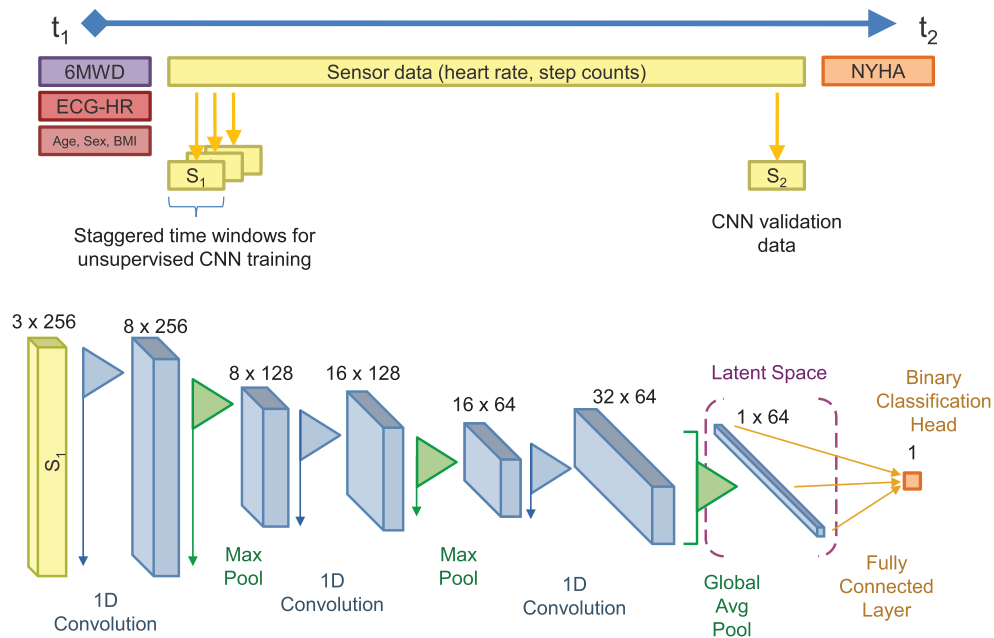
Extended Data Fig. 2 | Neural network architecture. Top panel: At time-point 1 (t1), 6-minute walk distance (6MWD), electrocardiogram (ECG) heart rate (HR), age, sex and body mass index are taken at the closest trial appointment to the start of the wearables sub-study. S1 denotes start of sensor data collection. S2 denotes sensor data used for validation. Time-point 2 (t2) is the future for New York Heart Association (NYHA) functional class prediction. Lower panel: The convolutional neural network (CNN) architecture uses dilated convolution and max-pool layers to capture long-term structure. Global average pooling is used to reduce spatial extent to fixed-length vector representation. Dropout and L2 penalty are applied to avoid overfitting to training set and label smoothing is used for faster learning. The final layer is binary sigmoid for data discrimination task. The penultimate layer of 32 units is used as latent space features for regression modeling.

Extended Data Table 1 | Baseline characteristics of patients that declined participation

Characteristic	Included in the wearables sub-study (n=53)	Declined the wearables sub-study (n=19)
Age at randomisation, mean years (SD)	75.6 (8.4)	76.4 (7.1)
Gender, women n (%)	21 (39.6%)	12 (63.2%)
Hypertension, n (%)	39 (73.6%)	16 (84.2%)
Diabetes mellitus, n (%)	11 (20.8%)	6 (31.6%)
Previous stroke or transient ischemic attack, n (%)	8 (15.1%)	5 (26.3%)
Treatment with inhalers for COPD or asthma, n (%)	14 (26.4%)	10 (52.6%)
Diagnosed with heart failure, n (%)	24 (45.3%)	12 (63.2%)
LVEF on echocardiogram, mean % (SD)	55.6 (8.8)	63.2 (5.8)
NT-proBNP, median pg/mL (IQR)	1099 (770-1725)	913 (584-1691)
NYHA class, n (%)		
I	0 (0%)	0 (0%)
II	38 (71.7%)	9 (47.4%)
III	14 (26.4%)	10 (52.6%)
IV	1 (1.9%)	0 (0%)
mEHRA class, n (%)		
1	0 (0%)	1 (5.3%)
2a	3 (5.7%)	7 (36.8%)
2b	26 (49.1%)	8 (42.1%)
3	22 (41.5%)	3 (15.8%)
4	2 (3.8%)	0 (0%)
ECG heart rate, mean beats/min (SD)	97.7 (20.4)	100.8 (18.0)
Systolic blood pressure, mean mmHg (SD)	137.9 (17.2)	136.1 (19.0)
6-minute walk distance, median metres (IQR)	384 (207-437)	266 (44-330)

ECG=electrocardiogram; IQR=interquartile range; LVEF=left ventricular ejection fraction; mEHRA=modified European Heart Rhythm Association; NT-proBNP=N-terminal pro-hormone B-natriuretic peptide; NYHA=New York Heart Association; SD=standard deviation.

Extended Data Table 2 | Usage of devices at interim review

Characteristic	Digoxin N = 28	Beta-blocker N = 24*	Total N = 52
Carrying smartphone regularly in the last 7 days, n (%)	22 (78.6%)	17 (70.8%)	39 (75.0%)
Number of days wearable device worn in last 7 days, n			
0 days	1 (3.6%)	1 (4.2%)	2 (3.9%)
1 day	0 (0%)	0 (0%)	0 (0%)
2 days	0 (0%)	1 (4.2%)	1 (1.9%)
3 days	1 (3.6%)	0 (0%)	1 (1.9%)
4 days	0 (0%)	0 (0%)	0 (0%)
5 days	0 (0%)	1 (4.2%)	1 (1.9%)
6 days	0 (0%)	0 (0%)	0 (0%)
7 days	26 (92.9%)	21 (87.5%)	47 (90.4%)

*1 patient did not attend the interim review.

Extended Data Table 3 | Wearable sensor parameters over sequential weeks

Wearable sensor output	Weeks 1-5			Weeks 6-10			Weeks 11-15			Weeks 16-20		
	Digoxin	Beta-blockers	p-value	Digoxin	Beta-blockers	p-value	Digoxin	Beta-blockers	p-value	Digoxin	Beta-blockers	p-value
Average heart rate, beats per minute (SEM)	72.1 (0.6)	70.8 (0.7)	0.54	71.6 (0.7)	70.5 (0.8)	0.97	71.9 (0.9)	71.5 (0.7)	0.96	71.2 (1.0)	72.7 (0.6)	0.25
Average total step count (SEM)	32,519 (2480)	28,287 (2052)	0.53	31,461 (2065)	30,114 (2163)	0.53	30,045 (2102)	25,856 (2251)	0.025	33,539 (2499)	29,105 (2694)	0.07

Statistical comparison between patients randomised to digoxin and beta-blockers for each time period uses a one-way analysis of variance by ranks (non-parametric Kruskal-Wallis test). SEM=standard error of the mean.

Extended Data Table 4 | Change in NYHA classification

NYHA class	Randomised to digoxin		Randomised to beta-blockers	
	Baseline N = 28	Final N=27	Baseline N = 25	Final N=24
I: No limitation of physical activity, n (%)	0 (0%)	17 (63.0%)	0 (0%)	6 (25.0%)
II: Slight limitation of physical activity, n (%)	22 (78.6%)	8 (29.6%)	16 (64.0%)	18 (75.0%)
III: Marked limitation of physical activity, n (%)	5 (17.9%)	2 (7.4%)	9 (36.0%)	0 (0%)
IV: Symptoms of heart failure at rest, n (%)	1 (3.6%)	0 (0%)	0 (0%)	0 (0%)

NYHA=New York Heart Association.

nature portfolio

Corresponding author(s): Dipak Kotecha

Last updated by author(s): May 13, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection was automated using the RADAR-base platform, funded by the European Union Innovative Medicines Initiative RADAR-CNS (115902). This platform allows for secure streaming of data from wearables, apps and devices.
Data analysis	Machine learning analyses were performed with custom coding developed using Python (Python Software Foundation, Delaware; version: continually updated) with scikit-learn, and TensorFlow (Google Brain, California). The machine learning framework for embedding multi-channel time-series data is made freely available at: https://github.com/gkoutos-group/wearable_data_embedding. Statistical analyses were performed using Stata version 17 (StataCorp LP, Texas), with a two-tailed p-value <0.05 denoting statistical significance.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Summary anonymised wearable sensor data are available for non-commercial purposes on request to the corresponding author (Prof Dipak Kotecha; d.kotecha@bham.ac.uk; 60-days response time for decisions). Due to the risk of patient re-identification, access to any individual-level data will require an appropriate ethical committee approval and review by the RATE-AF trial oversight committee which includes patient and public representatives (applications to Prof Dipak Kotecha; d.kotecha@bham.ac.uk; 180-days response time for decisions). Anonymised RATE-AF main trial and substudy datasets will be made available in an open-access repository after completion of secondary manuscripts.

Research involving human participants, their data, or biological material

Reporting on sex and gender	Gender used as this was self-identified by participants on enrollment. Gender reported in main results and table 1. Gender accounted for in adjusted analyses. Added gender interaction for main result.
Reporting on race, ethnicity, or other socially relevant groupings	Ethnicity was self-reported by participants on enrollment, as required by UK policies on diversity and equality for research within the National Health Service. The UK census categories were used for ethnicity reporting. Ethnicity is reported in the limitations as this was primarily a White British/Irish population. Full data on ethnicity have already been published (JAMA 2020;324:2497-2508).
Population characteristics	Fifty-three participants were enrolled in the sub-study, with mean age at randomisation of 75.6 years (SD 8.4; range 61 to 90 years) and 40% women. 28 (53%) had been randomised to digoxin and 25 (47%) to beta-blockers a mean of 30 weeks prior to their entry into the wearables sub-study (SD 8 weeks; range 12 to 46 weeks). Both groups were well balanced with respect to demographics and clinical measurements, with the commonest comorbidities being hypertension (74%) and heart failure (45%).
Recruitment	Recruitment for the RATE-AF trial occurred from 3 hospitals and primary care sites in the West Midlands, England, between 2016 and 2018. Inclusion criteria were: (1) age 60 years or older; (2) permanent AF in need of rate-control, with no plans to restore sinus rhythm; and (3) symptoms of possible heart failure, with breathlessness equivalent to New York Heart Association (NYHA) class II or above. Exclusion criteria were limited so the population enrolled reflected clinical practice. Patients that met these selection criteria were invited to join the trial (see JAMA 2020;324:2497-2508 for full details). Funding for the wearables sub-study was obtained after the main trial had commenced from the European Union Innovative Medicines Initiative BigData@Heart programme (116074). Participants with at least 2 months remaining in the RATE-AF trial were considered eligible for inclusion in the sub-study. All eligible participants were invited to join the wearable substudy regardless of age, gender, clinical status or familiarity with technology. 53 patients joined the study and 19 declined (see Extended Data Table 1 for characteristics).
Ethics oversight	Ethical approval was obtained from the East Midlands–Derby Research Ethics Committee (16/EM/ 0178), the Health Research Authority (IRAS 191437), and the Medicines and Healthcare Products Regulatory Agency.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	As an exploratory analysis, no sample size calculation was performed in advance of recruitment. Sensor data from the first 5 weeks of heart rate data in the first 10 participants provided average weekly heart rate, SD and correlation of repeated measures, indicating that a sample size of 40 participants would provide 90% power over 20 weeks (2-sided alpha 0.05) to detect a 1/3 SD difference in heart rate (2 beats/min) between digoxin and beta-blockers (values inserted from the data extraction on first 10 participants: control 72 beats/min; SD 6 beats/min; repeated measures correlation 0.9). A minimum target of 50 enrolled participants was targeted to account for death and loss to follow-up during the substudy.

Data exclusions	One patient from each arm was lost to follow up, with 51 participants included in the final analysis.
Replication	A post-hoc subgroup analysis according to activity levels was based on US Centers for Disease Control recommended activity levels (minimum 150 minutes/week aerobic activity equivalent to 15,000 steps/week, and health benefits goal of 300 minutes/week aerobic activity equivalent to 30,000 steps/week). CNN F1 scores used bootstrapping. See methods for full details on evaluation of machine learning approaches. The wearable sensor data neural network was retrained by a different data scientist with generation of new code to test the validity of our original model. The results for the wearables neural network prediction were extremely stable and demonstrated successful replication: Scientist#1: F1 score for NYHA class prediction=0.55 (95% CI 0.40 - 0.70) Scientist#2: F1 score for NYHA class prediction=0.56 (95% CI 0.41 - 0.70).
Randomization	As part of the main RATE-AF trial, each participant was randomised to either digoxin 62.5-250µg or bisoprolol 1.25-10mg once daily (or alternative beta-blockers) in a 1:1 ratio. Randomisation was completed using a computer-generated minimisation algorithm to ensure treatment arms were balanced for gender and AF symptoms, based on the modified European Heart Rhythm Association classification score.
Blinding	The RAte control Therapy Evaluation in permanent Atrial Fibrillation (RATE-AF) trial was a prospective, randomised, open-label trial embedded in usual care.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	registered with clinicaltrials.gov (NCT02391337) and clinicaltrialsregister.eu (2015-005043-13)
Study protocol	available at JAMA 2020;324:2497-2508
Data collection	Recruitment occurred from 3 hospitals and primary care sites in the West Midlands, England, between 2016 and 2018. Substudy sensor data collection between 2018 and 2019.
Outcomes	Endpoints acquired in the main trial were patient-reported quality of life, N-terminal pro-hormone B-natriuretic peptide, symptoms and functional capacity using mEHRA and NYHA class, 6-minute walk distance and time, heart rate (pulse examination), 12-lead ECG, left ventricular ejection fraction using cardiac ultrasound, and assessment of adverse events. The primary and secondary outcomes were predefined in the protocol (see JAMA 2020;324:2497-2508). Substudy endpoints for the wearable sensor data were focused on heart rate and step count (as detailed in the pre-specified addendum to the RATE-AF trial statistical analysis plan) to test the null hypothesis of no difference between sensor-derived continuous variables and single time-point assessment using 6-minute walk distance and electrocardiograph heart rate for predicting change in New York Heart Association functional class from substudy start to trial end. Wearable sensor data were pooled at 1-minute intervals to form time-series data, with an additional channel to denote missingness.

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A