nature medicine

Article

https://doi.org/10.1038/s41591-024-02961-4

# AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial

Received: 16 April 2023

Accepted: 29 March 2024

Published online: 29 April 2024

Check for updates

Chin-Sheng Lin[1,2], Wei-Ting Liu[1], Dung-Jang Tsai[2,3,4], Yu-Sheng Lou[3], Chiao-Hsiang Chang[1], Chiao-Chin Lee[1], Wen-Hui Fang[3,5], Chih-Chia Wang[5], Yen-Yuan Chen[6], Wei-Shiang Lin[1], Cheng-Chung Cheng[1], Chia-Cheng Lee[7,8], Chih-Hung Wang[9,10], Chien-Sung Tsai[11], Shih-Hua Lin[12] & Chin Lin[2,3,13,14]

The early identification of vulnerable patients has the potential to improve outcomes but poses a substantial challenge in clinical practice. This study evaluated the ability of an artificial intelligence (AI)-enabled electrocardiogram (ECG) to identify hospitalized patients with a high risk of mortality in a multisite randomized controlled trial involving 39 physicians and 15,965 patients. The AI-ECG alert intervention included an AI report and warning messages delivered to the physicians, flagging patients predicted to be at high risk of mortality. The trial met its primary outcome, finding that implementation of the AI-ECG alert was associated with a significant reduction in all-cause mortality within 90 days: 3.6% patients in the intervention group died within 90 days, compared to 4.3% in the control group (4.3%) (hazard ratio (HR) = 0.83, 95% confidence interval (CI) = 0.70–0.99). A prespecified analysis showed that reduction in all-cause mortality associated with the AI-ECG alert was observed primarily in patients with high-risk ECGs (HR = 0.69, 95% CI = 0.53–0.90). In analyses of secondary outcomes, patients in the intervention group with high-risk ECGs received increased levels of intensive care compared to the control group; for the high-risk ECG group of patients, implementation of the AI-ECG alert was associated with a significant reduction in the risk of cardiac death (0.2% in the intervention arm versus 2.4% in the control arm, HR = 0.07, 95% CI = 0.01–0.56). While the precise means by which implementation of the AI-ECG alert led to decreased mortality are to be fully elucidated, these results indicate that such implementation assists in the detection of high-risk patients, prompting timely clinical care and reducing mortality. ClinicalTrials.gov registration: NCT05118035.

The overall burden of critical illness is greater than commonly recognized and is expected to escalate with an aging population[1]. It is widely acknowledged that providing intensive care to critically ill patients reduces mortality[2]. Delays in providing intensive care for critically ill patients results in catastrophic outcomes[3]. Most in-hospital cardiac arrests are potentially preventable; however, the early signs of deterioration might be difficult to identify[4]. Rapid response systems (RRS), including afferent and efferent limbs, have been introduced in hospitals

A full list of affiliations appears at the end of the paper. e-mail: xup6fup@mail.ndmctsgh.edu.tw

https://doi.org/10.1038/s41591-024-02961-4

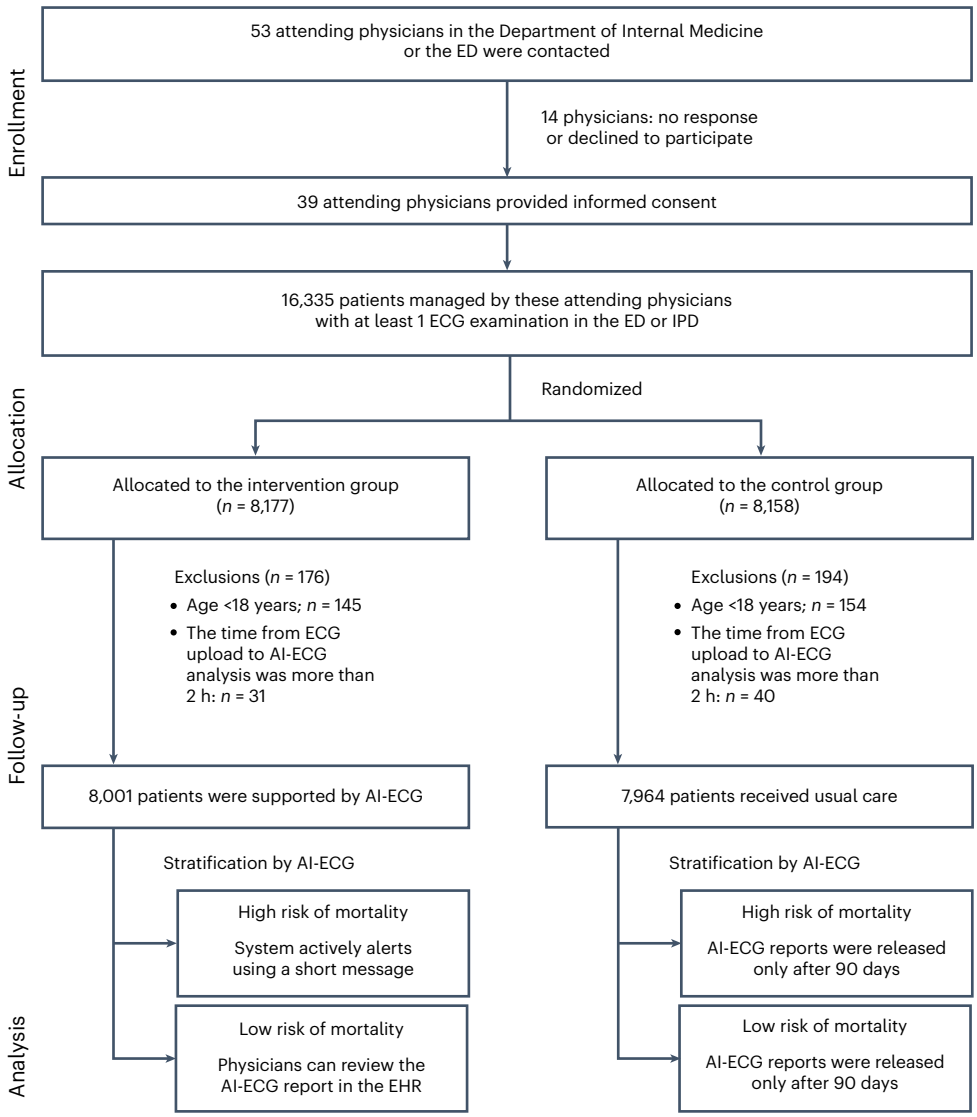



**Fig. 1 | CONSORT-AI flow diagram.** This flowchart illustrates a randomized control trial investigating the effectiveness of AI-enabled electrocardiogram (AI-ECG) analysis in emergency and inpatient departments. The study enrolled 39 physicians and analyzed 16,335 patients who cared by participants, divided into intervention and control groups, to compare outcomes with and without AI-ECG support. Exclusion criteria included patients under 18 years and delays over two hours in AI-ECG analysis post-ECG upload. The intervention involved real-time AI-ECG alerts for high mortality risk, while control group reports were delayed.

to manage clinical deterioration[5]. The afferent limb, also known as the track-and-trigger system (TTS), is the key to activating RRS. With the popularization of electronic health records (EHRs), these TTS may be embedded into the hospital to generate real-time alerts, which could potentially improve the quality of critical care[6].

Previous TTS could be categorized into two approaches. The first approach relies on single medical data, which is applicable but lacks accuracy, resulting in alert fatigue[7]. The second approach involves aggregated weighting scores, such as the National Early Warning Score (NEWS)[8], modified early warning score (MEWS)[9] or artificial intelligence (AI) models of physiological data[10]. Although offering improved accuracy[11], potential missing data without proper protocolized procedures are the Achilles' heel of this approach[12]. Notably, existing aggregated weighting scores only integrate established information, which might be critical to elucidate the minimal impact of this particular RRS on the reduction of mortality during randomized controlled trials (RCTs)[13]. To facilitate comprehensive care by physicians, an effective TTS may necessitate three key components: utmost accuracy; simplicity; and the inclusion of additional concealed information.

The 12-lead electrocardiogram (ECG) is a commonly used diagnostic tool in clinical practice, particularly in emergency departments (EDs) and inpatient departments (IPDs), where it is used on a large proportion of patients. We developed an AI-enabled ECG (AI-ECG) for mortality risk stratification and to predict all-cause mortality[14]. In an external validation, the performance of the AI-ECG achieved an area under the curve (AUC) of 0.858 on 1-year mortality prediction[14], which is compatible with an AUC of 0.855 (ref. 15) from previous research, which is much better than the predictions by a single parameter[7] or aggregated weighting scores[11]. Importantly, we found that our AI-ECG exhibited better performance in predicting 30-day mortality with an AUC of 0.906 compared to 1-year mortality[14], emphasizing its strength for critically ill patients. Furthermore, this AI-ECG score demonstrated the associations with subtle changes of several underlying cardiovascular diseases, exhibiting its potential capabilities beyond physicians[14]. Although the AI-ECG serves as an effective TTS, there is no pertinent RCT to date[16].

We sought to apply the AI-ECG to the TTS to identify deteriorating patients whose clinical conditions are possibly reversible to evaluate

**Table 1 | Patient characteristics stratified according to randomization**

| | All patients | | High risk as identified by the AI-ECG | | Low risk as identified by the AI-ECG | |
|---|---|---|---|---|---|---|
| | Intervention group (*n*=8,001) | Controls(*n*=7,964) | Intervention group(*n*=709) | Controls(*n*=688) | Intervention group(*n*=7,292) | Controls(*n*=7,276) |
| Stratification using the AI-ECG | | | | | | |
| High risk of mortality | 709 (8.9) | 688 (8.6) | | | | |
| Low risk of mortality | 7,292 (91.1) | 7,276 (91.4) | | | | |
| Hospitalization type | | | | | | |
| Academic medical center | 7,404 (92.5) | 7,390 (92.8) | 678 (95.6) | 655 (95.2) | 6,726 (92.2) | 6,735 (92.6) |
| Community hospital | 597 (7.5) | 574 (7.2) | 31 (4.4) | 33 (4.8) | 566 (7.8) | 541 (7.4) |
| Patient source | | | | | | |
| ED | 4,610 (57.6) | 4,569 (57.4) | 386 (54.4) | 372 (54.1) | 4,224 (57.9) | 4,197 (57.7) |
| IPD | 3,391 (42.4) | 3,395 (42.6) | 323 (45.6) | 316 (45.9) | 3,068 (42.1) | 3,079 (42.3) |
| Demographics and comorbidities | | | | | | |
| Sex (male) | 4,076 (50.9) | 4,168 (52.3) | 385 (54.3) | 385 (56.0) | 3,691 (50.6) | 3,783 (52.0) |
| Age (mean±s.d.) | 60.9±18.5 | 61.5±18.2 | 71.6±16.0 | 71.4±16.2 | 59.9±18.4 | 60.6±18.1 |
| Age group | | | | | | |
| <65 years | 4,369 (54.6) | 4,224 (53.0) | 224 (31.6) | 223 (32.4) | 4,145 (56.8) | 4,001 (55.0) |
| 65–74 years | 1,773 (22.2) | 1,853 (23.3) | 167 (23.6) | 171 (24.9) | 1,606 (22.0) | 1,682 (23.1) |
| ≥75 years | 1,859 (23.2) | 1,887 (23.7) | 318 (44.9) | 294 (42.7) | 1,541 (21.1) | 1,593 (21.9) |
| Coronary artery disease | 2,098 (26.2) | 2,089 (26.2) | 267 (37.7) | 240 (34.9) | 1,831 (25.1) | 1,849 (25.4) |
| Heart failure | 884 (11.0) | 903 (11.3) | 213 (30.0) | 166 (24.1) | 671 (9.2) | 737 (10.1) |
| Atrial fibrillation | 592 (7.4) | 580 (7.3) | 143 (20.2) | 119 (17.3) | 449 (6.2) | 461 (6.3) |
| Diabetes mellitus | 2,074 (25.9) | 2,146 (26.9) | 305 (43.0) | 299 (43.5) | 1,769 (24.3) | 1,847 (25.4) |
| Hypertension | 3,442 (43.0) | 3,434 (43.1) | 420 (59.2) | 412 (59.9) | 3,022 (41.4) | 3,022 (41.5) |
| MEWS[a] | | | | | | |
| Score (mean±s.d.) | 1.5±1.0 | 1.5±1.1 | 2.5±1.7 | 2.6±1.9 | 1.4±0.9 | 1.4±0.9 |
| MEWS group | | | | | | |
| 0–2 (low risk) | 7,009 (87.6) | 6,988 (87.7) | 280 (39.5) | 280 (40.7) | 6,591 (90.4) | 6,570 (90.3) |
| 3 and 4 (medium risk) | 797 (10.0) | 764 (9.6) | 346 (48.8) | 312 (45.3) | 589 (8.1) | 590 (8.1) |
| ≥5 (high risk) | 195 (2.4) | 212 (2.7) | 83 (11.7) | 96 (14.0) | 112 (1.5) | 116 (1.6) |

All values shown are *n* (%) unless otherwise stated. [a]A total of 1,079 (6.8%) individuals had missing data for at least one component of MEWS. To account for these missing values, a score of zero was imputed in the analysis.

the potential benefits of the simple TTS without conducting additional ECG and altering routine daily practice. The aim of this multisite single-blind RCT was to evaluate whether a warning message to the treating physician generated by the AI-ECG, identifying patients at high risk for all-cause mortality based on a single ECG in an ED or IPD, would prompt management decisions leading to a decrease in deaths.

## Results

### Participant and patient characteristics

The primary clinical hypothesis of this RCT was that the AI-ECG would identify human-unrecognized deteriorating signs and prompt early intervention to reduce mortality. A total of 39 attending physicians participated in this trial and their characteristics are shown in Supplementary Table 1. Mean age was 45.6 ± 6.2 years and 97.4% were male. As shown in Fig. 1, the final analysis included 8,001 and 7,964 patients in the intervention and control groups, respectively. Among the intervention and control groups, the AI-ECG identified 709 (8.9%) and 688 (8.6%) patients with a high risk of mortality, respectively; participating physicians received AI-ECG alerts for these patients in the intervention group. Participating physicians carefully assessed the patient's current condition on receiving the AI-ECG alert and arranged appropriate intensive monitoring or intensive care accordingly. Moreover, given the AI-ECG's demonstrated capability in identifying and predicting heart failure[17] and paroxysmal atrial fibrillation[18] beyond human knowledge, additional cardiac tests were considered for patients without overt cardiac abnormalities. Mean age was 60.9 ± 18.5 and 61.5 ± 18.2 years; 50.9% and 52.3% were male in the intervention and control groups, respectively (Table 1). Because of the differences in intervention approaches and the definition of follow-up periods between the high-risk and low-risk subgroups, our prespecified statistical analysis plan also encompassed stratified analyses according to AI-ECG predictions. Consequently, we have presented the patient characteristics for these subgroups in Table 1. Furthermore, Supplementary Tables 2 and 3 provide detailed information on the MEWS components and the ECG features.

### Performance and components of the AI-ECG risk score

Supplementary Tables 4 and 5 show the patient characteristics and ECG features based on AI-ECG stratification. Extended Data Fig. 1 presents the overt risk features that were most correlated with the AI-ECG risk score. We found that age had the highest correlation, but this was mainly driven by further stratification of low-risk patients, which may have limited clinical significance. In comparison, MEWS and heart rate showed stronger associations with the AI-ECG risk score in medium to high-risk patients. Extended Data Fig. 2 compares the AI-ECG alerts with

https://doi.org/10.1038/s41591-024-02961-4

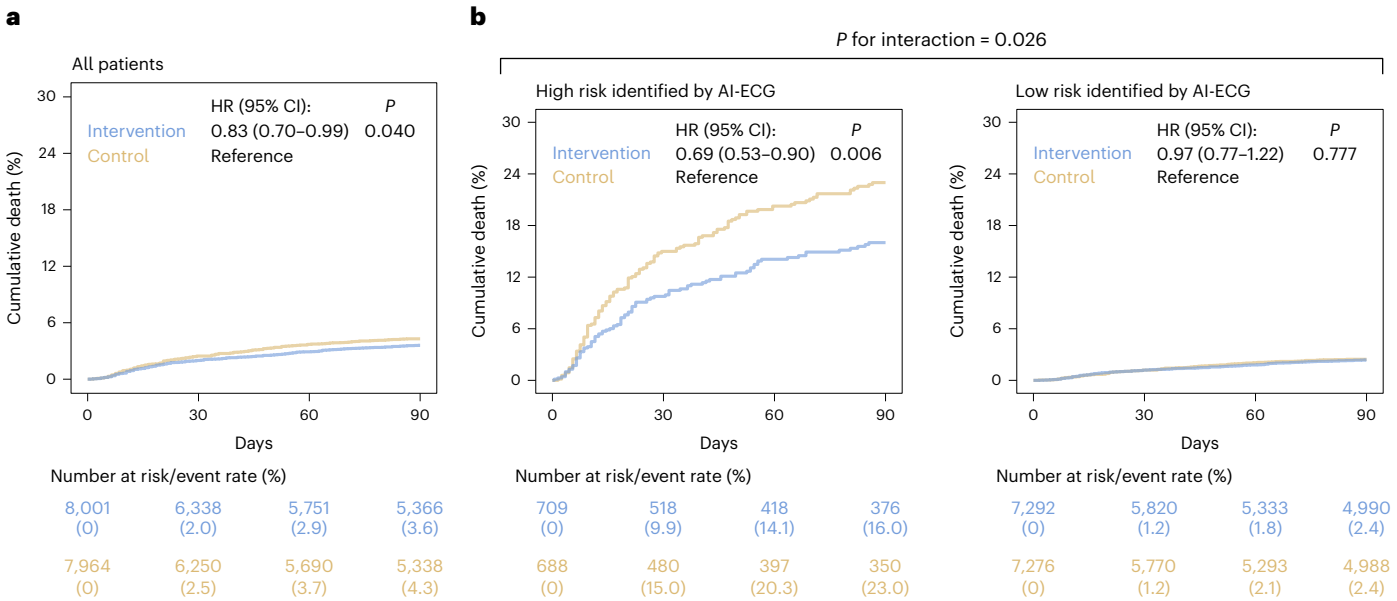



**Fig. 2 | Prespecified primary analysis. a**, Overall effect of the AI-ECG intervention on mortality. **b**, Effect of the AI-ECG intervention stratified according to AI-ECG prediction (high or low risk). For a patient identified as high risk by the AI-ECG in the intervention group, the corresponding physician received a short message as an alert. For a patient identified as low risk by the AI-ECG in the intervention group, an AI-ECG report was made available only for review. Cox proportional hazards, mixed-effect models without covariates were used for the statistical test, which was two-sided, with no adjustment for multiple comparisons. The blue and brown lines represent the intervention and control groups, respectively. The tables show the at-risk population and cumulative risk for the given time intervals in each group.

all baseline characteristics. Even applying all available characteristics, the performance of the model to predict AI-ECG alerts had an AUC of 0.844; the baseline logistical model using only patient characteristics achieved an AUC of 0.722. Notably, the AI-ECG outperformed the combination of all baseline characteristics in predicting the 90-day mortality risk, with an AUC of 0.886, indicating its capability beyond baseline characteristics.

Extended Data Fig. 3 shows the ability of the AI-ECG to perform risk stratification as stratified by each cause of death. After adjusting for sex and age, the high-risk group had a hazard ratio (HR) of 7.53 (95% confidence interval (CI) = 5.94–9.54) on all-cause mortality. We first distinguished cause of death into cardiac and noncardiac deaths; we found that its predictive ability (HR = 10.78 and 95% CI = 4.59–25.32) for cardiac death was higher when compared to noncardiac death (HR = 7.33 and 95% CI = 5.73–9.38). This interesting finding revealed that the AI-ECG might identify more information on cardiac-related deaths, although it was trained using the label of all-cause mortality in this study. For a more detailed cause-of-death analysis, the AI-ECG presented the highest predictive ability for death due to arrhythmia, followed by death due to myocardial infarction and death due to cancer. Particularly, while no future information regarding tachycardia, atrial fibrillation and heart failure was provided during model training, Extended Data Fig. 4 illustrates that risk stratification based on this model exhibited predictive capabilities for these diseases. This demonstrates its utility in detecting subtle changes of several underlying cardiovascular diseases.

## Primary outcomes

In the overall population, Fig. 2a shows that the cumulative proportion of death within 90 days was significantly different between the groups (3.6% in the intervention group versus 4.3% in the control group, HR = 0.83 and 95% CI = 0.70–0.99, $P = 0.040$). As shown in Fig. 2b, this mortality risk reduction originated from the high-risk cases identified by the AI-ECG ($P$ for the intervention group × risk group interaction = 0.026). An active warning message in the intervention group significantly reduced mortality risk from 23.0% to 16.0% (HR = 0.69 and 95% CI = 0.53–0.90, $P = 0.006$). An opportunity to review the AI-ECG reports only provided limited help in the AI-defined low-risk population (HR = 0.97 and 95% CI = 0.77–1.22, $P = 0.777$).

## Secondary outcomes

We initially categorized causes of death into cardiac and noncardiac. Figure 3a shows a significant reduction in the overall risk of cardiac death with the aid of the AI-ECG (HR = 0.27 and 95% CI = 0.11–0.67, $P = 0.005$). Subsequently, in Fig. 3b, stratification according to AI-ECG prediction further demonstrated a significant decrease in the risk of cardiac death in the high-risk subgroup (HR = 0.07 and 95% CI = 0.01–0.56, $P = 0.012$). Figure 3c shows that the AI-ECG exhibited an overall HR of 0.88 for noncardiac death (95% CI = 0.74–1.05, $P = 0.170$). In Fig. 3d, further stratified analysis showed a significant reduction in noncardiac death risk in the high-risk subgroup in the intervention arm compared to the control group (HR = 0.76 and 95% CI = 0.58–0.99, $P = 0.045$). However, because of the small number of events, the interaction tests for intervention group × risk group did not show significant results regarding either cardiac or noncardiac deaths. Extended Data Fig. 5 presents the impact of the AI-ECG on predefined detailed causes of death. The only significant result emerged from the interaction tests in deaths due to cancer ($P$ for intervention group × risk group interaction = 0.044) before correction for multiple comparisons. However, because of the limited sample size, there was no significant risk reduction observed even in the high-risk subgroup (HR = 0.62 and 95% CI = 0.37–1.05, $P = 0.078$).

Table 2 shows the details of the prespecified subsequent tests. A significantly higher proportion of patients in the intervention group were admitted to the intensive care unit (ICU) (HR = 1.40 and 95% CI = 1.06–1.85, $P = 0.016$), received amiodarone treatment (HR = 1.58 and 95% CI = 1.19–2.10, $P = 0.002$) and underwent echocardiography (HR = 1.36 and 95% CI = 1.15–1.61, $P < 0.001$), N-terminal pro-brain natriuretic peptide (NT-proBNP) (HR = 1.26 and 95% CI = 1.08–1.46, $P = 0.002$), free calcium (HR = 1.22 and 95% CI = 1.05–1.42, $P = 0.016$) and

https://doi.org/10.1038/s41591-024-02961-4

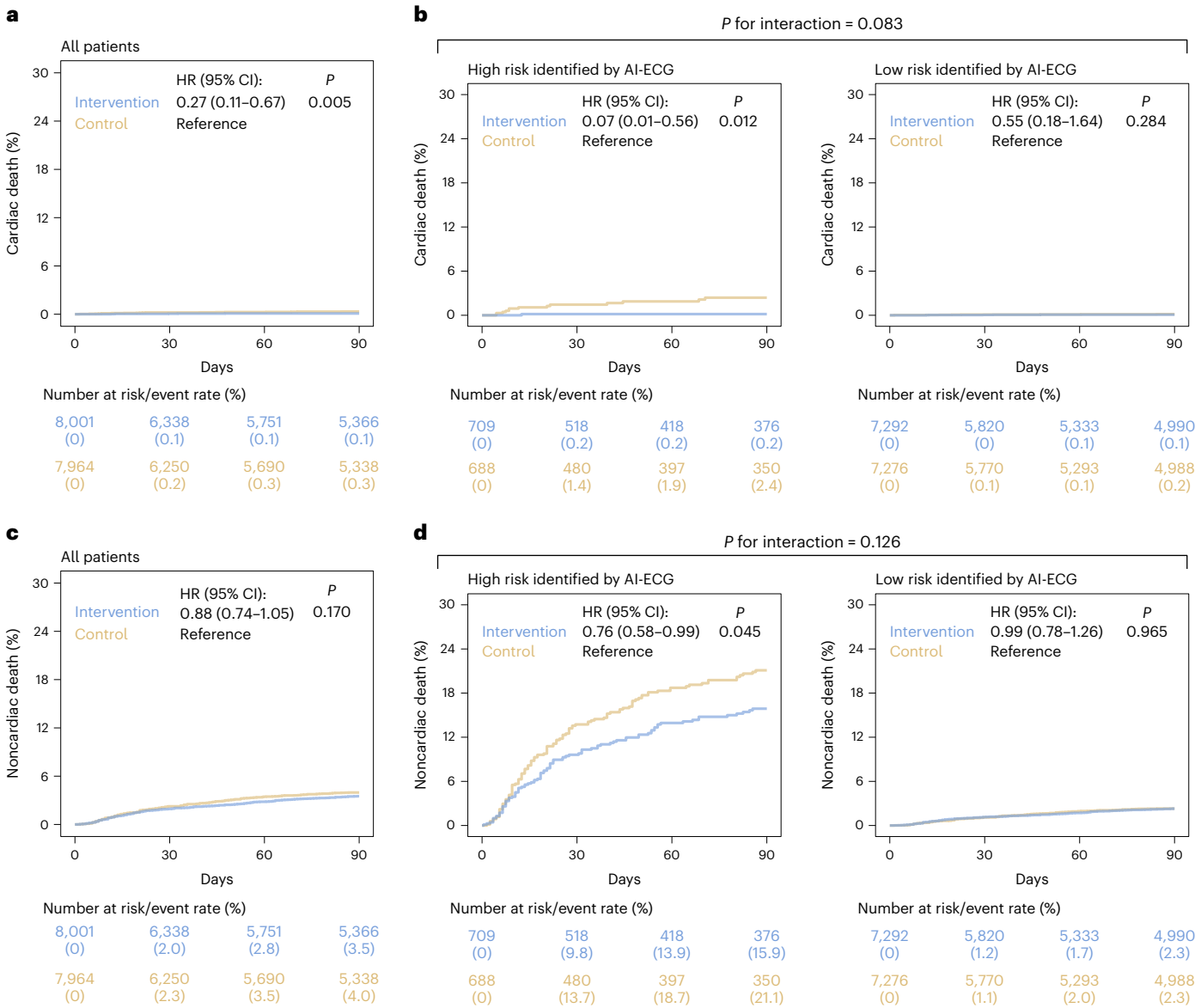



**Fig. 3 | Prespecified analysis of cause of death. a**, Overall effect of the AI-ECG intervention on cardiac deaths. **b**, Effect of the AI-ECG intervention on cardiac deaths stratified according to AI-ECG prediction. **c**, Overall effect of the AI-ECG intervention on noncardiac deaths. **d**, Effect of the AI-ECG intervention on noncardiac deaths stratified according to the AI-ECG prediction. Cox proportional hazards, mixed-effect models without covariates were used for the statistical test, which was two-sided, with no adjustment for multiple comparisons. The blue and brown lines represent the intervention and control groups, respectively. The tables show the at-risk population and the cumulative risk for the given time intervals in each group.

magnesium tests (HR = 1.19 and 95% CI = 1.03–1.38, $P = 0.020$). These additional medical interventions were not significantly increased in low-risk cases recognized by the AI-ECG ($P$ for intervention group × risk group interaction < 0.05 for all).

With increased echocardiographic and NT-proBNP tests in the high-risk subgroup, more new echocardiographic findings (HR = 1.64 and 95% CI = 1.02–2.64, $P = 0.042$) and NT-proBNP abnormalities (HR = 1.35 and 95% CI = 1.13–1.60, $P = 0.001$) were identified in the intervention than in the control group. Detailed analyses of specific echocardiographic findings identified merely a significant increase on mitral regurgitation (MR) (HR = 2.45 and 95% CI = 1.08–5.57, $P = 0.032$). However, electrolyte abnormalities, including free calcium and magnesium, were similar between the two groups, despite more tests being arranged.

## Prespecified exploratory outcomes

We performed stratified analyses to investigate how the intervention would affect low-risk and high-risk subgroups. For the primary endpoint, we found a significant effect on mortality in the high-risk but not the low-risk subgroup (Fig. 2). However, in an analysis of specific causes of death, the high-risk subgroup did not benefit from the AI-ECG more than the low-risk subgroup, possibly because of the limited sample size (Fig. 3). In the case of the secondary endpoints, we observed significant effects for the high-risk but not the low-risk subgroup on ICU admission, amiodarone use, echocardiogram, NT-proBNP test, free calcium test, magnesium test, new echocardiographic findings and MR (Table 2). Notably, high-sensitivity troponin I (TnI) was the only secondary endpoint for which a significant difference between the intervention and control arms was observed for the low-risk group. These results, indicating greater effectiveness of the AI-ECG for high-risk compared to low-risk patients, are consistent with the expected effects of an RRS. In a stratified analysis based on baseline characteristics, the impact of the intervention on mortality risk reduction was largely consistent across subgroups (all $P$ for intervention group × characteristic group interaction > 0.05) (Fig. 4).

https://doi.org/10.1038/s41591-024-02961-4

**Table 2 | Prespecified secondary endpoints after ECG testing**

| Event | High-risk individuals identified by the AI-ECG | | | Low-risk individuals identified by the AI-ECG | | | *P* for interaction |
|---|---|---|---|---|---|---|---|
| | Intervention group(*n*=709), number ofevents (%) | Controls(*n*=688), number ofevents (%) | HR (95% CI)and *P* | Intervention group(*n*=7,292), numberof events (%) | Controls(*n*=7,276), numberof events (%) | HR (95% CI)and *P* | |
| Intervention or test | | | | | | | |
| ICU admission (3 days) | 120 (16.9) | 86 (12.5) | 1.40 (1.06–1.85), *P*=0.016 | 167 (2.3) | 188 (2.6) | 0.89 (0.72–1.09), *P*=0.256 | 0.009 |
| Amiodarone use (3 days) | 123 (17.3) | 78 (11.3) | 1.58 (1.19–2.10), *P*=0.002 | 79 (1.1) | 92 (1.3) | 0.86 (0.63–1.16), *P*=0.310 | 0.004 |
| Digoxin use (3 days) | 63 (8.9) | 43 (6.2) | 1.45 (0.99–2.14), *P*=0.059 | 86 (1.2) | 95 (1.3) | 0.90 (0.67–1.21), *P*=0.491 | 0.054 |
| Diltiazem use (3 days) | 49 (6.9) | 45 (6.5) | 1.07 (0.71–1.61), *P*=0.739 | 130 (1.8) | 135 (1.9) | 0.96 (0.75–1.22), *P*=0.742 | 0.648 |
| Lidocaine use (3 days) | 61 (8.6) | 50 (7.3) | 1.20 (0.82–1.74), *P*=0.345 | 344 (4.7) | 350 (4.8) | 0.98 (0.84–1.14), *P*=0.772 | 0.323 |
| Defibrillation (3 days) | 19 (2.7) | 10 (1.5) | 1.85 (0.86–3.98), *P*=0.114 | 9 (0.1) | 6 (0.1) | 1.50 (0.53–4.21), *P*=0.444 | 0.746 |
| Echocardiogram (7 days) | 316 (44.6) | 240 (34.9) | 1.36 (1.15–1.61), *P*<0.001 | 1,347 (18.5) | 1,357 (18.7) | 0.99 (0.92–1.07), *P*=0.802 | 0.001 |
| NT-proBNP test (3 days) | 378 (53.3) | 320 (46.5) | 1.26 (1.08–1.46), *P*=0.002 | 1,273 (17.5) | 1,301 (17.9) | 0.97 (0.90–1.05), *P*=0.468 | 0.007 |
| TnI test (3 days) | 486 (68.5) | 452 (65.7) | 1.09 (0.96–1.23), *P*=0.207 | 2,677 (36.7) | 2,714 (37.3) | 0.98 (0.93–1.03), *P*=0.468 | 0.282 |
| Coronary angiography (3 days) | 41 (5.8) | 26 (3.8) | 1.55 (0.95–2.54), *P*=0.080 | 259 (3.6) | 284 (3.9) | 0.91 (0.77–1.07), *P*=0.255 | 0.042 |
| Free calcium test (3 days) | 360 (50.8) | 304 (44.2) | 1.22 (1.05–1.42), *P*=0.010 | 984 (13.5) | 971 (13.3) | 1.01 (0.93–1.10), *P*=0.806 | 0.047 |
| Magnesium test (3 days) | 386 (54.4) | 338 (49.1) | 1.19 (1.03–1.38), *P*=0.020 | 1,074 (14.7) | 1,092 (15.0) | 0.98 (0.90–1.06), *P*=0.621 | 0.033 |
| Potassium test (3 days) | 562 (79.3) | 537 (78.1) | 1.03 (0.92–1.16), *P*=0.581 | 4,729 (64.9) | 4,694 (64.5) | 1.00 (0.96–1.04), *P*=0.871 | 0.732 |
| Sodium test (3 days) | 554 (78.1) | 530 (77.0) | 1.03 (0.92–1.16), *P*=0.591 | 4,694 (64.4) | 4,655 (64.0) | 1.00 (0.96–1.05), *P*=0.853 | 0.821 |
| Chloride test (3 days) | 351 (49.5) | 340 (49.4) | 1.02 (0.88–1.18), *P*=0.823 | 2,008 (27.5) | 1,921 (26.4) | 1.05 (0.99–1.12), *P*=0.117 | 0.634 |
| Test results | | | | | | | |
| New echocardiography findings (7 days) | 45 (6.3) | 27 (3.9) | 1.64 (1.02–2.64), *P*=0.042 | 97 (1.3%) | 79 (1.1) | 1.20 (0.89–1.61), *P*=0.231 | 0.314 |
| *New-onset low EF (7 days)* | 9 (1.3) | 5 (0.7) | 1.76 (0.59–5.25), *P*=0.311 | 9 (0.1) | 12 (0.2) | 0.75 (0.32–1.78), *P*=0.513 | 0.230 |
| *New-onset pulmonary arterial hypertension (7 days)* | 4 (0.6) | 10 (1.5) | 0.39 (0.12–1.23), *P*=0.109 | 35 (0.5) | 34 (0.5) | 1.03 (0.64–1.65), *P*=0.907 | 0.126 |
| *New-onset pericardial effusion (7 days)* | 6 (0.8) | 1 (0.1) | 5.87 (0.71–48.78), *P*=0.101 | 3 (0) | 3 (0) | 1.00 (0.20–4.95), *P*=0.998 | 0.192 |

https://doi.org/10.1038/s41591-024-02961-4

**Table 2 (continued) | Prespecified secondary endpoints after ECG testing**

| Event | High-risk individuals identified by the AI-ECG | | | Low-risk individuals identified by the AI-ECG | | | *P* for interaction |
|---|---|---|---|---|---|---|---|
| | Intervention group(*n*=709), number ofevents (%) | Controls(*n*=688), number ofevents (%) | HR (95% CI)and *P* | Intervention group(*n*=7,292), numberof events (%) | Controls(*n*=7,276), numberof events (%) | HR (95% CI)and *P* | |
| *New aortic stenosis or progression (7days)* | 6 (0.8) | 1 (0.1) | 5.87 (0.71–48.78), *P*=0.101 | 8 (0.1) | 5 (0.1) | 1.60 (0.52–4.89), *P*=0.411 | 0.288 |
| *New aortic regurgitation or progression (7days)* | 6 (0.8) | 6 (0.9) | 0.97 (0.31–3.02), *P*=0.964 | 15 (0.2) | 9 (0.1) | 1.67 (0.73–3.81), *P*=0.226 | 0.454 |
| *New MR or progression (7days)* | 20 (2.8) | 8 (1.2) | 2.45 (1.08–5.57), *P*=0.032 | 34 (0.5) | 37 (0.5) | 0.89 (0.56–1.42), *P*=0.633 | 0.041 |
| *New pulmonary regurgitation or progression (7days)* | 4 (0.6) | 6 (0.9) | 0.65 (0.18–2.30), *P*=0.503 | 13 (0.2) | 8 (0.1) | 1.62 (0.67–3.92), *P*=0.280 | 0.243 |
| *New tricuspid regurgitation or progression (7days)* | 20 (2.8) | 13 (1.9) | 1.50 (0.75–3.03), *P*=0.251 | 46 (0.6) | 28 (0.4) | 1.53 (0.97–2.43), *P*=0.068 | 0.840 |
| NT-proBNP abnormality (3days) | 285 (40.2) | 224 (32.6) | 1.35 (1.13–1.60), *P*=0.001 | 476 (6.5) | 443 (6.1) | 1.07 (0.94–1.22), *P*=0.282 | 0.052 |
| TnI abnormality (3days) | 147 (20.7) | 131 (19.0) | 1.11 (0.87–1.40), *P*=0.403 | 157 (2.2) | 197 (2.7) | 0.79 (0.64–0.98), *P*=0.030 | 0.039 |
| Percutaneous coronary intervention (3days) | 28 (3.9) | 17 (2.5) | 1.61 (0.88–2.95), *P*=0.120 | 113 (1.5) | 137 (1.9) | 0.82 (0.64–1.05), *P*=0.121 | 0.042 |
| Free calcium abnormality (3days) | 225 (31.7) | 214 (31.1) | 1.04 (0.86–1.25), *P*=0.707 | 614 (8.4) | 603 (8.3) | 1.02 (0.91–1.14), *P*=0.780 | 0.872 |
| Magnesium abnormality (3days) | 160 (22.6) | 145 (21.1) | 1.09 (0.87–1.36), *P*=0.470 | 283 (3.9) | 308 (4.2) | 0.92 (0.78–1.08), *P*=0.290 | 0.232 |
| Potassium abnormality (3days) | 194 (27.4) | 180 (26.2) | 1.05 (0.86–1.29), *P*=0.616 | 737 (10.1) | 718 (9.9) | 1.02 (0.92–1.13), *P*=0.675 | 0.810 |
| Sodium abnormality (3days) | 241 (34.0) | 231 (33.6) | 1.03 (0.86–1.23), *P*=0.779 | 928 (12.7) | 927 (12.7) | 1.00 (0.91–1.09), *P*=0.950 | 0.831 |
| Chloride abnormality (3days) | 170 (24.0) | 170 (24.7) | 0.98 (0.80–1.22), *P*=0.889 | 558 (7.7) | 522 (7.2) | 1.07 (0.95–1.20), *P*=0.275 | 0.506 |

*P* values for the interaction are two-sided, with no adjustment for multiple comparisons. Detailed echocardiography findings are shown in italics.

## Discussion

Although many RCTs report the limited effectiveness of RRS[13], the heterogeneous nature of RRS and its dependency on user factors suggest the need for case-by-case discussions on intervention and design in each RCT[19]. This study evaluates the efficacy of AI models in RRS and demonstrates that the use of an AI-ECG led to a significant reduction in mortality. Importantly, mortality reduction was largely attributed to the high-risk subgroup, as our RRS effectively and timely communicated warning messages to physicians. Our results highlight that the deployment of real-time AI-ECG analysis in TTS provides an opportunity to reduce mortality in the ED and IPD.

Alert fatigue may be the most important limitation of RRS. It has been reported that 49–96% of clinical alerts by TTS were overridden[20]. Most alerts were ignored by frontline clinicians, which led to failed improvement in quality of care by RRS deployment[21]. Therefore, the success of the proposed RRS using an AI-ECG may be attributed to the increased attention of the enrolled physicians. The proportion of high risk in mortality cases and the cumulative mortality rate were 49.5% and 23.0%, respectively in the control group, which was consistent with previous studies[14,15] and much improved sensitivities and positive predictive values of most RRS when using aggregated weighting scores[11]. Compared to the other AI models developed to detect patient deterioration[10], we suggest that the accuracy of our AI-ECG was not inferior. Moreover, alert fatigue may happen more commonly, leading to a cycle of increased alert fatigue[22]. This high mortality risk (23.0%) is enough to interest physicians in investing more time in patient care and break the cycle of alert fatigue. Moreover, one potential explanation for the reduction of alert fatigue in this study may be linked to the novelty of the AI-ECG. In this study, 39 enrolled physicians received only 709 alerts in 4.5 months, which enhanced their willingness to order further intensive care. In the future, when the AI-ECG results and alerts are accessible in real time for all patients, we anticipate an average of fewer than ten alerts per month for each physician. This burden remains within an acceptable range and is conducive to maintaining quality of care.

The efferent limb of the RRS is triggered by the identified 'calling criteria'. Many hospitals have a rapid response team, including interdisciplinary members usually led by an attending physician, for providing

https://doi.org/10.1038/s41591-024-02961-4

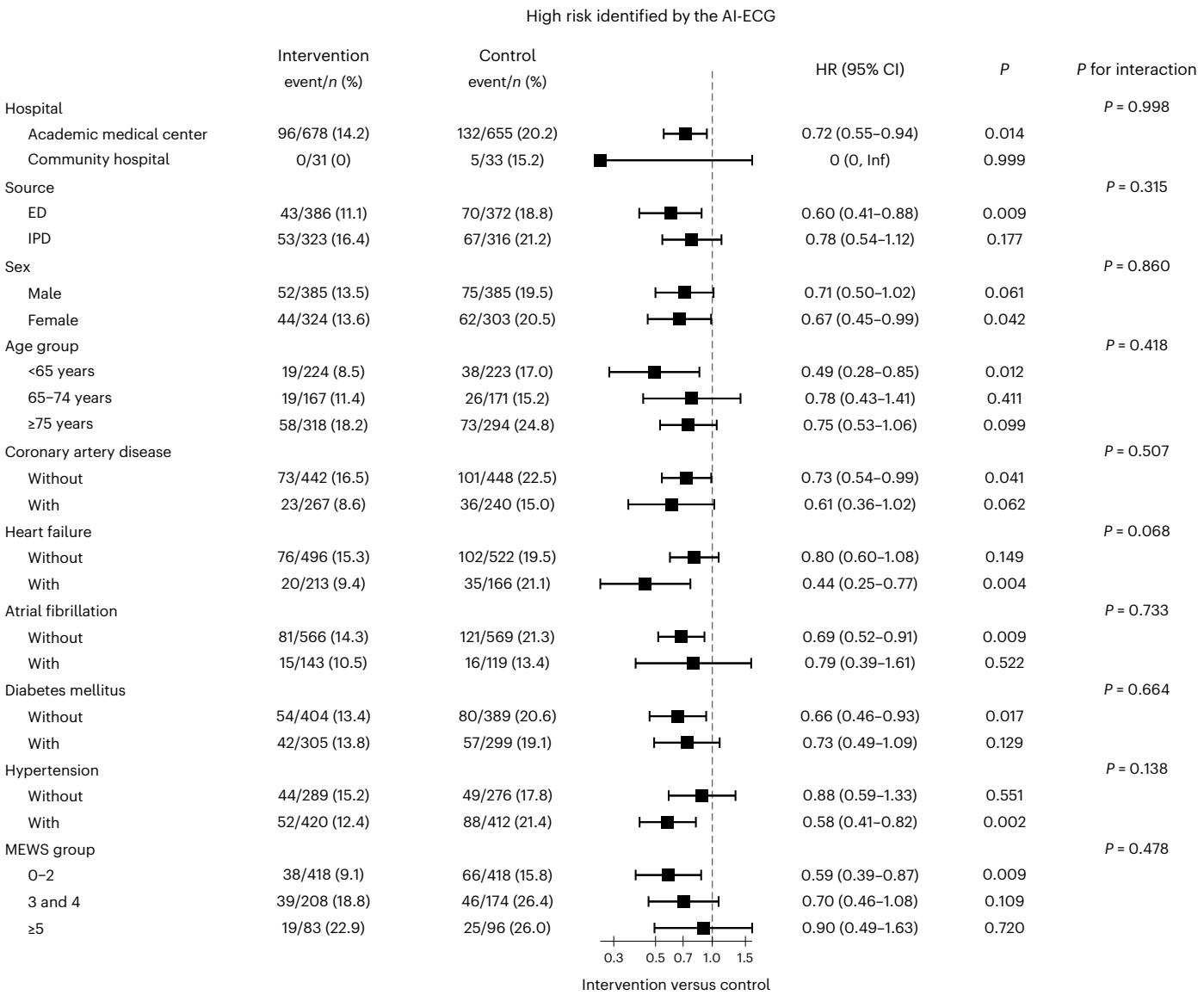



**Fig. 4 | Prespecified exploratory outcomes in patients identified as high risk by the AI-ECG for all-cause mortality within 90 days, stratified using baseline characteristics.** The black squares represent the point estimates of the HR, while the error bars indicate the 95% CI. Cox proportional hazards, mixed-effect models without covariates were used for the statistical test, which was two-sided, with no adjustment for multiple comparisons.

high levels of care[5]. In this study, there was no fixed team to handle the AI-ECG alerts. The enrolled physicians might apply appropriate medical resources to respond to the activation of the RRS. While we did not implement a system to capture the responses of each individual physician, the intervention group indeed received more diagnostic testing and medical interventions than the control group in the high-risk subgroup, supporting the activation of RRS by the AI-ECG alerts. Note that subsequent intensive care had no specific form; it is fully dependent on the condition of the individual patient, ranging from adding diagnostic testing and medical support to intensive monitoring[23]. Because of the diverse care for each individual critically ill patient, it is challenging to comprehensively list and analyze the full spectrum of medical interventions. Therefore, the list of prespecified secondary endpoints only included a limited number of anticipated items before the trial. However, we observed that the AI-ECG alerts prompted a significant number of prespecified treatments or tests. As expected, more new echocardiogram findings and NT-proBNP abnormalities were uncovered because of increased testing, resulting in early diagnoses of cardiac diseases, which provide beneficial effects for disease management[24,25]. An interesting finding in this study was that despite AI-ECG alerts triggering more electrolyte tests, electrolyte abnormalities were not increased. This observation indicates that current electrolyte management may be adequate. Collectively, the efferent limb of our AI-ECG-based RRS could effectively conduct its task; participating physicians paid more attention to the patients identified as high risk by the AI-ECG.

It is interesting that a simple AI-ECG intervention has led to a dramatic reduction in mortality, even though the precise mechanism is not fully understood. There are two potential explanations. First, the AI-ECG exhibits exceptional risk stratification capabilities, allowing physicians to pay more attention to patients with a higher mortality risk (23.0%) rather than those with a lower risk (2.4%). Previous meta-analyses comparing the outcomes between high-intensity and low-intensity ICU proposed a 29% reduction in mortality for critically ill patients[2], which is consistent with our finding of 31%. Second, the AI-ECG identified subtle changes in underlying cardiac conditions from unknown ECG features, particularly in patients with low ejection fraction (EF)[17] and paroxysmal atrial fibrillation[18]. Although previous

RCTs based on NEWS or MEWS for RRS indicated that automatic alerts provided only marginal assistance in enhancing patient outcomes[13], this might be attributed to the fact that significant vital sign changes themselves prompted intensive care, even in the absence of RRS assistance[26]. The correlation between AI-ECG and MEWS in our study was relatively low, with less than 15% of patients in the AI-ECG high-risk subgroup having a MEWS score of five or higher. This emphasizes AI's ability to identify high-risk patients who may not be recognized by human experts, potentially leading to timely treatments and reduced mortality. In our prespecified exploratory stratified analysis, we also observed a nonsignificant trend with more favorable impacts on younger patients or those with lower MEWS scores, who were supported by our AI-ECG system. However, whether the mechanisms underlying the reduction in mortality rates from AI-ECG alerts are attributed to the identification of deteriorating signs by the AI-ECG in these patients with subclinical characteristics require further validation in future large-scale, multicenter RCTs.

In addition to treatment changes and comprehensive tests immediately after the AI-ECG alerts, we suggest another potential mechanism of intensive monitoring. Because of the limitations of our EHRs, detailed records of bedside monitoring are lacking in our hospital. However, before the trial, we designed several prespecified secondary endpoints related to arrhythmia treatment as an alternative indicator of intensive cardiac monitoring. Interestingly, the analytical results suggested higher use of amiodarone in the AI-ECG high-risk intervention group compared to the control group. Increased use of amiodarone might be the result of more intensive monitoring, providing critical evidence of increased intensive cardiac monitoring after AI-ECG alerts. A study reported that 62% of inpatients who experienced cardiac arrest exhibited physiological instability for more than 6 h before the event, yet only 22% had explicit documentation[27], highlighting the importance of intensive monitoring. Although physicians were unaware of the required immediate actions due to the 'black box' characteristics of the AI-ECG, increased attention and detailed assessment might reduce mortality. As mentioned earlier, further large-scale, multicenter RCTs are necessary to confirm the effects of intensive monitoring for the AI-ECG high-risk subgroup regarding mortality reduction. Just as the AI-ECG was successfully validated by an RCT for the early diagnosis of patients with low EFs[28], this study revealed another excellent application of the AI-ECG for the identification of deteriorating signs, resulting in significant reduction in mortality.

Some limitations should be addressed. First, during the study period, physicians enrolled in the study may have increased medical attention after receiving an AI-ECG alert, which suggests that the effect of the AI-ECG on reducing mortality may be lower in real-world settings because of the Hawthorne effect. Nevertheless, our study emphasizes the significant impact of physician awareness on reducing mortality. Second, participating physicians demonstrated increased medical attention in the control group during the study period compared to before the RCT. The Hawthorne effect may underestimate the effectiveness of the AI-ECG on mortality reduction. Third, the effect of the RRS depends on the cutoff point selected to trigger the AI-ECG alert. In this RCT, we selected a cutoff point to identify the top 10% high-risk individuals using preliminary data, which was the consensus regarding clinical loading among the enrolled physicians before the trial. Fourth, the ECG tests were triggered by clinical issues but they were not conducted for all patients at regular intervals. Moreover, the heterogeneity of the RRS may result in variations in the efferent arm. The effectiveness of the AI-ECG may be highly dependent on local testing policy and physician decisions. Fifth, all study data were collected from EHRs; thus, incomplete information may lead to an underestimation of absolute values. However, this should have a consistent impact on both the intervention and control groups, making the effect estimation of the AI-ECG intervention still reliable. Lastly, the study was conducted in a country with almost 100% insurance coverage. However, this study offers valuable insights for authorities to consider whether reimbursement of subsequent medical costs after an AI-ECG alert is worthy of investment.

This trial establishes a pragmatic framework for the further development of RRS, which is similar to A/B testing[29,30], providing an easily importable system for continuous improvement of RRS by AI models[31]. Physicians who received an alert from the AI-ECG paid more attention to high-risk patients, leading to timely medical interventions and reduced mortality. As the ECG is a low-cost test frequently performed for several purposes, the AI model run using existing ECGs may be applied to most patients in EDs and IPDs. Future research may involve the improvement of the RRS by understanding better how the AI-ECG alerts result in lowered mortality, which may substantially improve the quality of care for critically ill patients.

## 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/s41591-024-02961-4.

## References


1. Adhikari, N. K., Fowler, R. A., Bhagwanjee, S. & Rubenfeld, G. D. Critical care and the global burden of critical illness in adults. *Lancet* **376**, 1339–1346 (2010).
2. Pronovost, P. J. et al. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. *JAMA* **288**, 2151–2162 (2002).
3. Chalfin, D. B., Trzeciak, S., Likourezos, A., Baumann, B. M. & Dellinger, R. P. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Crit. Care Med.* **35**, 1477–1483 (2007).
4. Hodgetts, T. J. et al. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation* **54**, 115–123 (2002).
5. Jones, D. A., DeVita, M. A. & Bellomo, R. Rapid-response teams. *N. Engl. J. Med.* **365**, 139–146 (2011).
6. Goldstein, B. A., Navar, A. M., Pencina, M. J. & Ioannidis, J. P. A. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J. Am. Med. Inform. Assoc.* **24**, 198–208 (2017).
7. Smith, G. B., Prytherch, D. R., Schmidt, P. E., Featherstone, P. I. & Higgins, B. A review, and performance evaluation, of single-parameter 'track and trigger' systems. *Resuscitation* **79**, 11–21 (2008).
8. Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E. & Featherstone, P. I. The ability of the national early warning score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* **84**, 465–470 (2013).
9. Subbe, C. P., Kruger, M., Rutherford, P. & Gemmel, L. Validation of a modified Early Warning Score in medical admissions. *QJM* **94**, 521–526 (2001).
10. Mann, K. D. et al. Predicting patient deterioration: a review of tools in the digital hospital setting. *J. Med. Internet Res.* **23**, e28209 (2021).
11. Romero-Brufau, S. et al. Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation* **85**, 549–552 (2014).
12. Ludikhuize, J. et al. Standardized measurement of the Modified Early Warning Score results in enhanced implementation of a rapid response system: a quasi-experimental study. *Resuscitation* **85**, 676–682 (2014).
13. Chan, P. S., Jain, R., Nallmothu, B. K., Berg, R. A. & Sasson, C. Rapid response teams: a systematic review and meta-analysis. *Arch. Intern. Med.* **170**, 18–26 (2010).

14. Tsai, D.-J. et al. Mortality risk prediction of the electrocardiogram as an informative indicator of cardiovascular diseases. *Digit. Health* **9**, 20552076231187247 (2023).
15. Raghunath, S. et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat. Med.* **26**, 886–891 (2020).
16. Plana, D. et al. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw. Open* **5**, e2233946 (2022).
17. Attia, Z. I. et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat. Med.* **25**, 70–74 (2019).
18. Attia, Z. I. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* **394**, 861–867 (2019).
19. Downey, C. L., Tahir, W., Randell, R., Brown, J. M. & Jayne, D. G. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. *Int. J. Nurs. Stud.* **76**, 106–119 (2017).
20. van der Sijs, H., Aarts, J., Vulto, A. & Berg, M. Overriding of drug safety alerts in computerized physician order entry. *J. Am. Med. Inform. Assoc.* **13**, 138–147 (2006).
21. Bedoya, A. D. et al. Minimal impact of implemented Early Warning Score and best practice alert for patient deterioration. *Crit. Care Med.* **47**, 49–55 (2019).
22. Embi, P. J. & Leonard, A. C. Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. *J. Am. Med. Inform. Assoc.* **19**, e145–e148 (2012).
23. Lyons, P. G., Edelson, D. P. & Churpek, M. M. Rapid response systems. *Resuscitation* **128**, 191–197 (2018).
24. Iung, B. & Vahanian, A. Epidemiology of valvular heart disease in the adult. *Nat. Rev. Cardiol.* **8**, 162–172 (2011).
25. de Lemos, J. A., McGuire, D. K. & Drazner, M. H. B-type natriuretic peptide in cardiovascular disease. *Lancet* **362**, 316–322 (2003).
26. Kollef, M. H. et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J. Hosp. Med.* **9**, 424–429 (2014).
27. Perkins, G. D., Temple, R. M. & George, R. Time to intervene: lessons from the NCEPOD report. *Resuscitation* **83**, 1305–1306 (2012).
28. 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).
29. Austrian, J. et al. Applying A/B testing to clinical decision support: rapid randomized controlled trials. *J. Med. Internet Res.* **23**, e16651 (2021).
30. Horwitz, L. I., Kuznetsova, M. & Jones, S. A. Creating a learning health system through rapid-cycle, randomized testing. *N. Engl. J. Med.* **381**, 1175–1179 (2019).
31. Simon, G. E., Platt, R. & Hernandez, A. F. Evidence from pragmatic trials during routine care—slouching toward a learning health system. *N. Engl. J. Med.* **382**, 1488–1491 (2020).



[1]Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [2]Medical Technology Education Center, School of Medicine, National Defense Medical Center, Taipei, Taiwan, Republic of China. [3]Department of Artificial Intelligence and Internet of Things, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [4]Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan, Republic of China. [5]Department of Family and Community Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [6]Department and Graduate Institute of Medical Education and Bioethics, National Taiwan University College of Medicine, Taipei, Taiwan, Republic of China. [7]Department of Medical Informatics, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [8]Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [9]Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [10]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan, Republic of China. [11]Division of Cardiovascular Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [12]Division of Nephrology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, Republic of China. [13]School of Public Health, National Defense Medical Center, Taipei, Taiwan, Republic of China. [14]Graduate Institute of Aerospace and Undersea Medicine, National Defense Medical Center, Taipei, Taiwan, Republic of China. ✉e-mail: xup6fup@mail.ndmctsgh.edu.tw

## Methods

### Trial design and ethics statement

This trial was registered before initiation (ClinicalTrials.gov registration: NCT05118035); the trial protocol is shown in Supplementary Note 1. We followed the guidelines outlined in the CONSORT-AI extension checklist[32]. This work was approved by the institutional review board of the Tri-Service General Hospital, Taipei, Taiwan (A202105120). As the research team did not have direct contact with the patients and relied on EHRs to collect patient data, informed consent was obtained from the attending physicians before commencing the trial; patient consent was waived.

We conducted this RCT at both an academic medical center and a community hospital in Taiwan. Thirty-nine eligible attending physicians in the Department of Internal Medicine or EDs provided informed consent. Physicians who did not provide informed consent were excluded from the AI-ECG report system. Although patients were not considered 'participants' in this study, we analyzed patient-level data collected from EHRs to explore the effect of the AI-ECG intervention. Patient data were included in the analysis if patients in the EDs and IPDs received at least one ECG for any indication between 15 December 2021 and 30 April 2022. At our hospital, ECG indications align with international healthcare systems that are primarily based on routine ward protocols and relevant clinical criteria, as follows[33]: (1) patients with middle-to-advanced age who require surgery or hospitalization received routine ECGs; (2) patients with symptoms related to disturbances of heart rhythm (for example, syncope, palpitation and intermittent shortness of breath) and chest pain or suspected acute coronary syndrome; (3) patients with existing or suspected cardiovascular diseases (for example, congestive heart failure, hypertension, pericarditis, valvular heart disease and cardiomyopathies); (4) evaluation of bradycardia or tachycardia; (5) evaluation of electrolyte imbalances or drug toxicity; and (6) monitoring patients with implanted cardiac devices. Patients under 18 years old or with a time delay of more than 2 h between their ECG and the AI-ECG analysis were excluded. We only analyzed patients who were cared for by attending physicians who had provided informed consent.

### Randomization

Randomization was conducted according to all possible medical numbers (Supplementary Fig. 1); this is a seven-digit-long serial number at our hospital, with a total of $10^7$ combinations. The patient-level randomization process was implemented to ensure longer patient follow-up, avoiding potential loss of follow-up that could occur with physician-level assignment. This approach enhances the study's reliability by maintaining consistency in the care received by patients throughout the 90-day period. Before the start of the trial, we completed the randomization process using simple random sampling; half of possible medical numbers were allocated to the intervention group. In other words, the randomization process may have occurred before the creation of the medical record number.

### The AI-ECG system

The AI-ECG system used in this study is a convolutional neural network; its technical details have been described in a previous publication[14]. It was trained on more than 450,000 ECGs, with the primary label being all-cause mortality, using survival data with censored events. The model's original output was a value ranging from negative to positive infinity, representing the relative risk of death for patients, which was not easily interpretable. Therefore, we transformed the model's output into a percentile score based on a hospital-based population, representing the patient's risk of death higher than a specific percentage of individuals. Generally, patients with a percentile score below 75 were considered to have an extremely low risk of death. As this study was conducted in EDs and IPDs, these patients had higher percentile scores compared to the entire hospital population. Thus, we prespecified a high-risk threshold consisting of the 95th percentile, which approximated 10% of patients in EDs and IPDs. In this study, we categorized patients as 'high risk' or 'low risk' based on this prespecified threshold and implemented our TTS accordingly. It is important to emphasize that the percentile score had broad predictive ability for several future cardiovascular diseases[14]. Moreover, it was more accurate for short-term than long-term predictions; its accuracy surpasses that of physicians and existing clinical tools. Therefore, before the start of the trial, participating physicians were aware that patients with high percentile scores might require comprehensive cardiovascular disease assessment.

We conducted a series of analyses to demonstrate the performance of our AI-ECG's percentile scores in this trial. First, we compared the differences in patient characteristics and ECG features between the AI-ECG high-risk and low-risk subgroups, using the Student's *t*-test and chi-squared test as appropriate. Subsequently, the three most significant variables correlated with the AI-ECG percentile scores were selected for Spearman correlation coefficient analysis. We then used XGBoost to model the AI-ECG predictions using all variables and ranked their importance. The results of these machine learning models, including a baseline linear model of logistic regression, were compared with the AI-ECG scores for all variables. Finally, we used Cox proportional hazards models to analyze the relationship between AI-ECG risk stratification and each cause of death. Additionally, we explored the AI-ECG's predictive capability for future tachycardia, atrial fibrillation and heart failure based on the previous literature.

### AI-ECG intervention and blinding

ECG recordings were collected using a Philips 12-lead ECG machine (PageWriter TC30 and PageWriter TC50); the ECG signal was recorded in digital format. The sampling frequency was 500 Hz, with 10 s recorded by each lead. At our hospital, ECGs are conducted by nurses. After the recording of a 10-s ECG signal, the nurse starts the process by clicking the 'upload' button. Subsequently, all ECGs are transmitted in real time to our private cloud server, where our AI-ECG platform operates on an NVIDIA DGX-1 for mortality risk predictions. As shown in Supplementary Fig. 2, physicians can review the AI-ECG reports in EHRs.

Once the AI-ECG indicated a high risk of mortality, a warning message was immediately triggered and sent to the corresponding attending physician. The threshold for sending the AI-ECG alerts was predetermined. The decision to alert approximately 10% of patients in the trial was made to avoid alert fatigue. Notifications appeared in the recipient's smartphone message system for prompt attention. The message notified the physician that: 'an ECG was received for patient X, which indicates high risk of mortality. Please tend to your patient's conditions. If you require further information, the following link connects to the ECG and the result of AI-ECG prediction'. The short message was only sent once for the earliest high-risk ECG for each patient. Before the trial, we encouraged physicians to conduct a comprehensive assessment of patients on receiving the AI-ECG alert and provide appropriate tests and medical interventions based on known clinical conditions. Physicians may order additional tests based on previous evidence that the AI-ECG identified subtle changes in underlying cardiac diseases[34], particularly low EF[17] and paroxysmal atrial fibrillation[18]. However, the criteria for treatment, ICU admission and surgery would not be adjusted based on AI-ECG alerts because the predictions of the AI-ECG were not incorporated into any medical guidelines before the trial. To ensure that additional medical interventions aligned with known clinical conditions, we manually reviewed the medical records of these patients. Therefore, we anticipated that patients might undergo additional tests and treatments as a result of identifying potential conditions through the AI-ECG alerts. Notably, although we actively sent a warning message to high-risk cases, the AI-ECG report for low-risk cases still presented a degree of risk. Physicians could check the relative severity by accessing the EHRs of patients in the intervention group.

In summary, physicians were able to review the AI-ECG report of patients in the intervention group, while patients in the control group received conventional care. This was a single-blind study as the AI-ECG report system presented different information for patients in the intervention and control groups. Importantly, regardless of whether patients were assigned to the intervention or control group, physicians were responsible for reviewing and interpreting the ECGs. Our AI-ECG was deemed an auxiliary tool for interpretation.

### Data collection

Because the warning message may significantly adjust the medical pathway in the intervention group, we stratified our patients into high-risk and low-risk subgroups. For patients who had at least one high-risk ECG, whether in the intervention or control group, we used the first high-risk ECG identified by the AI-ECG as the start of the follow-up time, although patients in the control group were not covered by the active warning message service. For patients without a high-risk ECG, the start of the follow-up time was defined as the initial ECG assessment. The reason for selecting the first high-risk ECG as the index time is because it is the most likely time point at which medical interventions may be initiated, leading to improved patient outcomes. However, as there were no high-risk ECGs in the low-risk group, the first ECG was chosen as the index time.

Patient characteristics were obtained from our hospital's information system, including sex and age; disease history before AI-ECG analysis was identified using the corresponding International Classification of Diseases, Ninth (ICD-9) and Tenth Revision (ICD-10) codes, including coronary artery disease (ICD-9 codes 410.x to 414.x, and 429.2, and ICD-10 codes I20.x to I25.x), heart failure (ICD-9 codes 428.x, 398.91 and 402.x1, and ICD-10 codes I50.x), atrial fibrillation (ICD-9 code 427.31 and ICD-10 codes I48.x), diabetes mellitus (ICD-9 codes 250.x and ICD-10 codes E11.x) and hypertension (ICD-9 codes 401.x to 404.x and ICD-10 codes I10.x to I16.x).

We also collected traditional ECG features using the Philips system. The Philips system provided an automatic analysis for each ECG; 35 ECG patterns and eight ECG measurements were extracted from the XML documents. The eight ECG measurements included heart rate, percentile interval, QRS complex duration, QT interval, correct QT interval, and P, RS and T wave axes. The data for the eight variables were 91–100% complete. Missing values were imputed using multiple imputations by chained equations using the R package mice v.3.5.0. The 35 ECG patterns were collected using statements from the Philips system, including sinus rhythm, sinus arrhythmia, sinus pause, ectopic atrial rhythm, junctional rhythm, pacemaker rhythm, early precordial R/S transition, ST elevation, ST depression, abnormal T wave, abnormal Q wave, RSR wave, low voltage, left and right axis deviation, left and right ventricular hypertrophy, left and right atrial enlargement, nonspecific intraventricular conduction delay, left fascicular block, right and left bundle branch block, first-degree and second-degree atrioventricular block, complete degree atrioventricular block, atrial fibrillation, atrial flutter, supraventricular tachycardia, Wolff–Parkinson–White syndrome, ventricular tachycardia, atrial premature complex and ventricular premature complex.

### Primary and secondary endpoints

The primary endpoint was all-cause mortality. Previous research suggested determination of all-cause mortality because it is less susceptible to clinician bias[26], and long-term mortality was considered a marker beyond hospital mortality[23]. Patient status (dead or alive) was acquired via the EHRs. Although it is possible that some patients may have died outside our hospital and therefore were not recorded by the EHR, we ensured that censored patients had survived until their last recorded hospital encounter. The endpoint of this study was set as 31 July 2022 because the follow-up period was 90 days for patients under the study as of 30 April 2022. The data for non-event visits were censored at the patient's last known hospital encounter to limit bias from incomplete records. Moreover, according to a previous study of readmissions in Taiwan based on the National Health Insurance Research Database, only 0.16% of readmissions occurred at a different hospital[35]; therefore, we are confident that no unknown deaths were included in our trial.

The prespecified secondary endpoints included a cause-of-death analysis, in which we performed a manual review and classified cases as either cardiac (including myocardial infarction, arrhythmia and other conditions) or noncardiac (including cancer, sepsis and other conditions) deaths. As the AI model that identified a high risk of mortality is based entirely on ECG information, we assumed that medical adjustments in cardiac drugs and associated tests may be present. The prespecified secondary endpoints also included follow-up tests and medical treatment after the ECG, including ICU admission, arrhythmia-related interventions (amiodarone, digoxin, diltiazem, lidocaine and defibrillation), heart function tests (echocardiogram, NT-proBNP, TnI and coronary angiography) and electrolyte tests (free calcium, and magnesium, potassium, sodium and chloride). In addition to arrhythmia-related interventions and ICU admission, the importance of other tests was to identify abnormal findings that may lead to further treatment affecting the risk of mortality. Therefore, we defined the abnormality of each test as follows: (1) echocardiogram: new-onset low EF (≤40%), new-onset pulmonary arterial hypertension (≥40 mmHg), new-onset pericardial effusion (≥10 mm) and new-onset moderate-to-severe valvular diseases (aortic stenosis, aortic regurgitation, MR, pulmonary regurgitation and tricuspid regurgitation), and progression of valvular diseases from moderate to severe; (2) NT-proBNP (≥1,000 pg ml$^{-1}$); (3) TnI (≥200 pg ml$^{-1}$); (4) coronary angiography: percutaneous coronary intervention; (5) free calcium (<4.5 or >5.3 mg dl$^{-1}$); (6) magnesium (<1.7 or >2.2 mg dl$^{-1}$); (7) potassium (<3.5 or >5.1 mmol l$^{-1}$); (8) sodium (<136 or >145 mmol l$^{-1}$); and (9) chloride (<98 or >107 mmol l$^{-1}$). Patients without examination within the follow-up time were considered normal in this analysis. Before the trial, we assumed that the effect of the AI-ECG intervention, most probably due to the warning message, lasted only 3 days. Additionally, because of the extended scheduling and execution time required for echocardiography at our hospital, which was known before the start of this trial, we defined a prespecified follow-up period of 7 days for echocardiography.

### Sample size

A previous study reported that intensive care was associated with a lower mortality rate compared to conventional care (relative risk = 0.71, 95% CI = 0.62–0.82)[2]. We performed sample size estimation using a two-sided significance level of 0.05, a statistical power of 0.80, a sample size ratio in the intervention and control groups of 1.0, a hypothetical ratio of 1.0 in patients from the ED and IPD, and a hypothetical proportion of controls with a primary endpoint of 0.17 (ref. 14) and a relative risk of 0.71; the minimum number in the intervention and control groups was the same, that is, 801 per arm. Because we predesigned the study to identify approximately 10% of ECGs as high-risk to reduce alert fatigue, the final sample size of this trial was approximately 16,020. The trial started on 15 December 2021 when the AI-ECG support was turned on for patients in the intervention group and ended on 30 April 2022 when the number of patients per arm approximated 8,000 patients per arm.

### Prespecified statistical analysis

The statistical analysis was performed using R v.3.4.4, with a two-sided significance level of $P < 0.05$. Patient characteristics and ECG features are presented as means and standard deviations or percentages where appropriate. A Cox proportional hazards, mixed-effect model was used to compare the intervention and control groups regarding primary and secondary endpoints with the enrolled physicians as the random effect; this was performed using the R package coxme v.2.2-18.1. HRs and 95% CIs were used as effect indicators; Kaplan–Meier curves were

used to visualize and calculate the cumulative incidence of events. Furthermore, the AI-ECG high-risk and low-risk groups differed in their intervention approaches and index time. Therefore, our prespecified analysis included testing the effects between the intervention and control groups regarding the differences of effects between high-risk and low-risk subgroups. To conduct this analysis, we used a Cox proportional hazards model and included an additional interaction term.

Before the start of the trial, we formulated a prespecified exploratory analysis strategy aimed at understanding whether the effectiveness of the AI-ECG intervention primarily came from the warning message. Assuming this to be the case, three validation hypotheses emerged: (1) no difference between the intervention and control groups in AI-ECG-defined low-risk cases; (2) a disparity between the intervention and control groups in AI-ECG-defined high-risk cases; and (3) a statistically significant interaction between AI-ECG high and low risk and the intervention and control groups. This prespecified exploratory analysis includes all primary and secondary endpoints.

The prespecified exploratory analysis also includes subgroup analyses based on age, sex and baseline comorbidities. Subgroup effects, specifically whether the HRs were consistent across subgroups, were assessed by examining the significance of the interaction term incorporated into the Cox proportional hazards model. This told us which patients might benefit most from the AI-ECG intervention.

### Reporting summary

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

## Data availability

Patient data cannot be made publicly available due to privacy concerns. De-identified tabular data can be obtained from the corresponding author on approval from the ethics committee of the Tri-Service General Hospital. Approval from this committee can be requested from the Tri-Service General Hospital's Clinical Trial Management System (https://tsgh.cims.tw/wiPtms/index.html), with an expected review period of approximately 2–3 months. After approval, researchers will be granted VPN access to perform analyses, ensuring data security and confidentiality (summary data can be exported), with measures in place to prevent any breach of personal information.

## Code availability

The model weights of the AI algorithm used in this study cannot be made publicly available due to the proprietary nature of the algorithm. However, the computer code for training is available from GitHub: https://github.com/Imshepherd/ECGSurvNet. This code can be used to train a survival deep learning model using public ECG databases with survival information, such as SaMi-Trop (https://doi.org/10.5281/zenodo.4905617)[36] and CODE-15% (https://doi.org/10.5281/zenodo.4916205)[37], both available from Zenodo.

## References


32. Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J. & Denniston, A. K. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat. Med.* **26**, 1364–1374 (2020).
33. Crawford, M. H. et al. ACC/AHA Guidelines for Ambulatory Electrocardiography. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the Guidelines for Ambulatory Electrocardiography). Developed in collaboration with the North American Society for Pacing and Electrophysiology. *J. Am. Coll. Cardiol.* **34**, 912–948 (1999).
34. Attia, Z. I., Harmon, D. M., Behr, E. R. & Friedman, P. A. Application of artificial intelligence to the electrocardiogram. *Eur. Heart J.* **42**, 4717–4730 (2021).
35. Huang, P.-F., Kung, P.-T., Chou, W.-Y. & Tsai, W.-C. Characteristics and related factors of emergency department visits, readmission, and hospital transfers of inpatients under a DRG-based payment system: a nationwide cohort study. *PLoS ONE* **15**, e0243373 (2020).
36. Ribeiro, A. L. P. et al. Sami-Trop: 12-lead ECG traces with age and mortality annotations. *Zenodo* https://doi.org/10.5281/zenodo.4905617 (2021).
37. Ribeiro, A. H. et al. CODE-15%: a large scale annotated dataset of 12-lead ECGs. *Zenodo* https://doi.org/10.5281/zenodo.4916205 (2021).


## Acknowledgements

This study was supported by funding from the National Science and Technology Council, Taiwan (NSTC110-2314-B-016-010-MY3 to C.L.; NSTC112-2321-B-016-003 to S.-H.L.), the Cheng Hsin General Hospital, Taiwan (CHNDMC-113-01 to C.-S.L.; CHNDMC-113-11205 to C.L.) and the Medical Affairs Bureau, Taiwan (MND-MAB-C07-113022, MND-MAB-C13-112051, MND-MAB-C08-111032 and MND-MAB-C08-11032 to C.-S.L.; MND-MAB-110-113, MND-MAB-D-111045, MND-MAB-C13-112050 and MND-MAB-C07-113021 to C.L.). We thank Y.-C. Wong for English editing and final proofreading of the manuscript.

## Author contributions

C.-S.L. and C.L. conceived and designed the study. D.-J.T. and Y.-S.L. acquired the data. C.-S.L. and C.L. analyzed the data. C.-S.L., W.-T.L. and C.L. interpreted the data. W.-T.L., C.-H.C., W.-S.L., C.-C.C. and Chiao-Chin Lee reviewed the medical records. C.-C.W. and Y.-Y.C. provided expert opinion on medical ethics when designing the study. D.-J.T., Chia-Cheng Lee and C.L. provided a deep learning model to stratify a high risk of mortality. W.-H.F., Chia-Cheng Lee, C.-H.W., C.-S.T. and S.-H.L. integrated the deep learning model with the hospital information system. C.-S.L. and C.L. drafted the initial manuscript. W.-T.L., W.-H.F. and S.-H.L. revised the manuscript for important intellectual content. C.L. takes final responsibility for the manuscript and provided final approval for the version to be published.

## Competing interests

The National Defense Medical Center has granted Quanta Computer, originally a personal computer and cloud server manufacturer, a license for its AI-ECG algorithm as part of Quanta's shift toward investing in smart hospital information systems. The National Defense Medical Center will not receive any financial gains from deploying the AI-ECG technology in patient care across Taiwan's military hospitals. Financial benefits from using the AI-ECG outside these military hospitals may accrue to C.-S.L., W.-T.L., Y.-S.L., C.-H.C., Chiao-Chin Lee, W.-H.F., W.-S.L., C.-C.C., Chia-Cheng Lee, C.-H.W., C.-S.T., S.-H.L. and C.L. The remaining authors declare no competing interests.

## Additional information

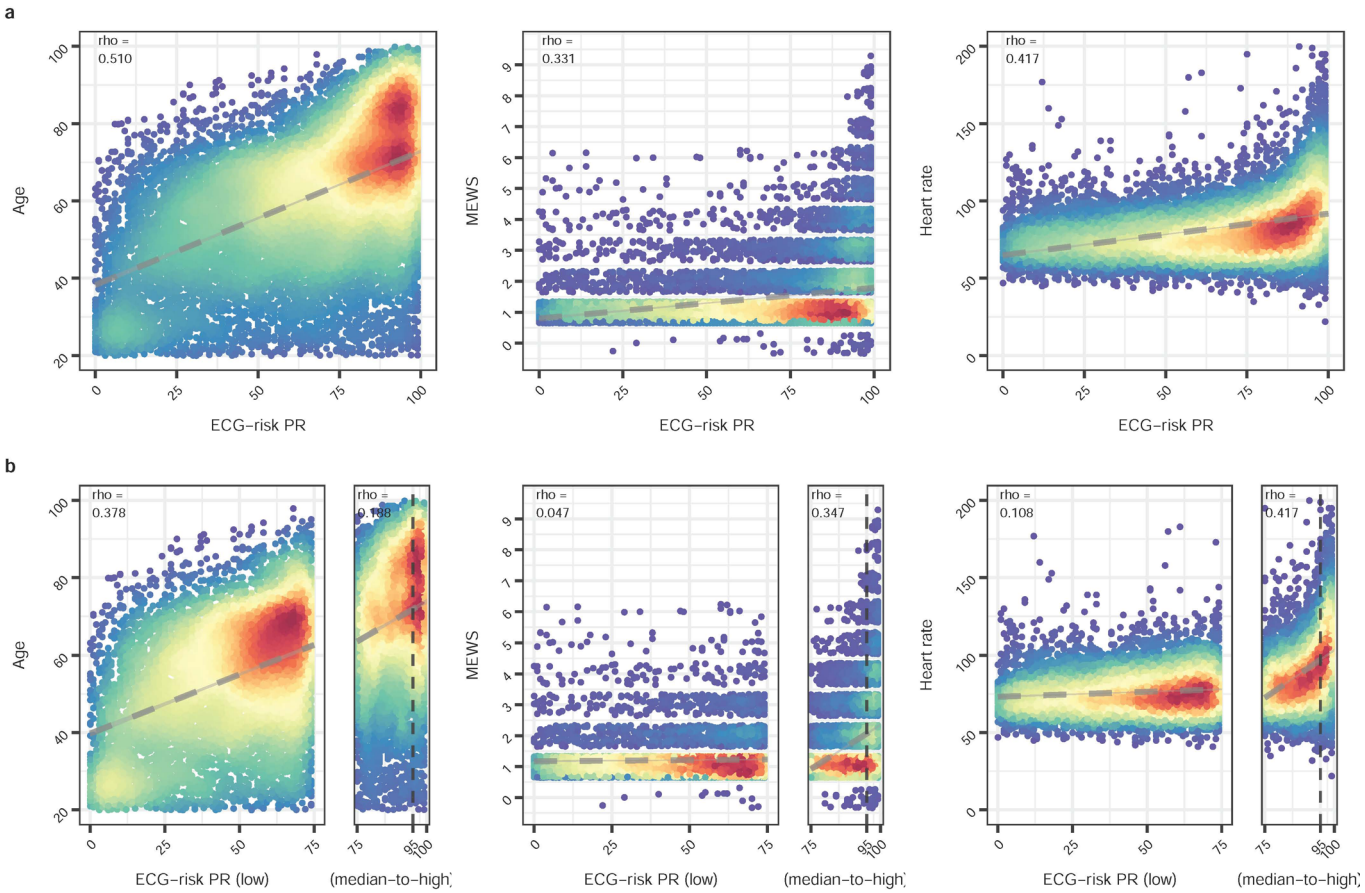
**Extended data** is available for this paper at https://doi.org/10.1038/s41591-024-02961-4.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-02961-4.

**Correspondence and requests for materials** should be addressed to Chin Lin.

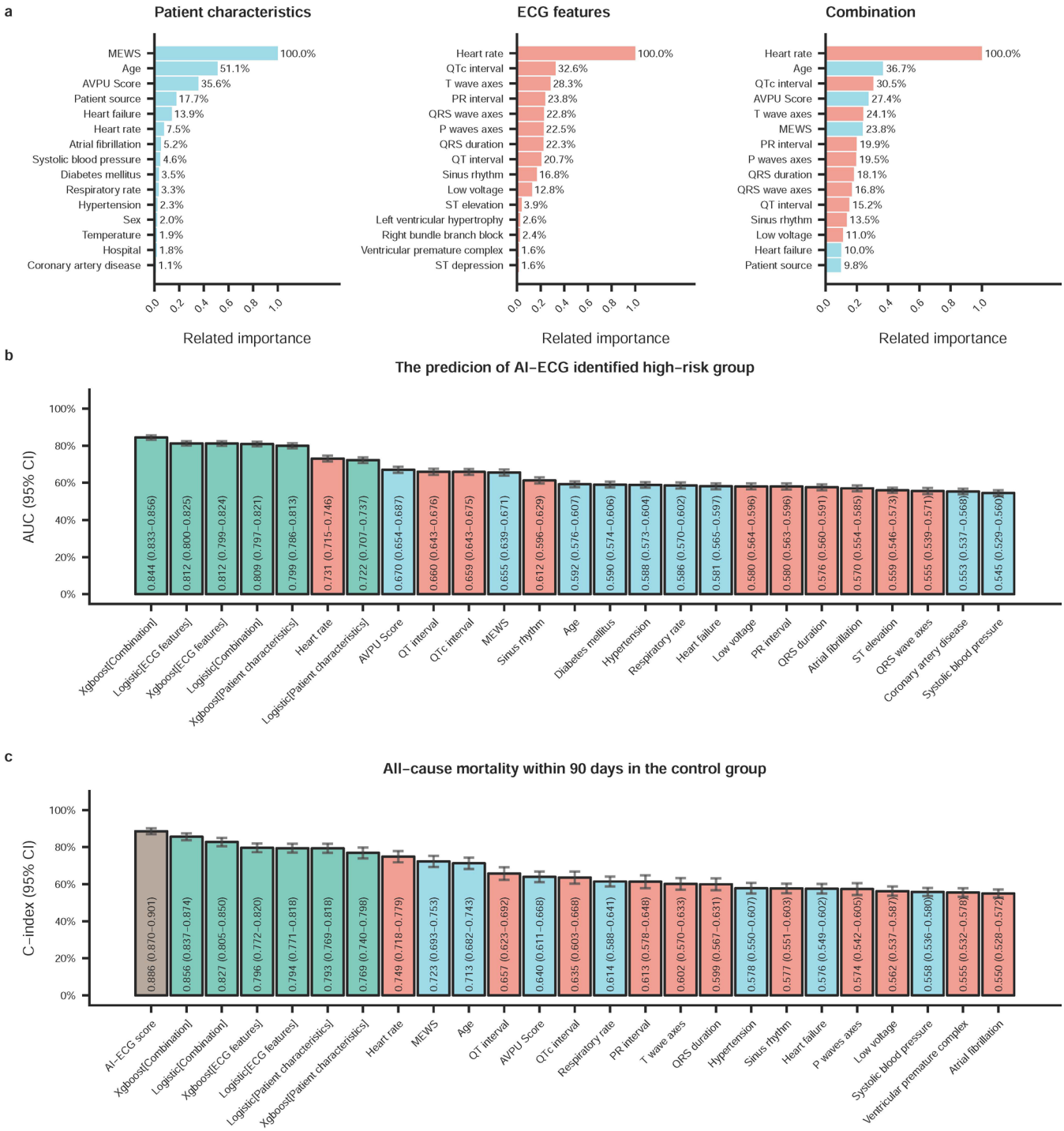
**Peer review information** *Nature Medicine* thanks Zachi Attia, Rahul Deo, Fu Siong Ng, Jill Waalen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Michael Basson, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-024-02961-4



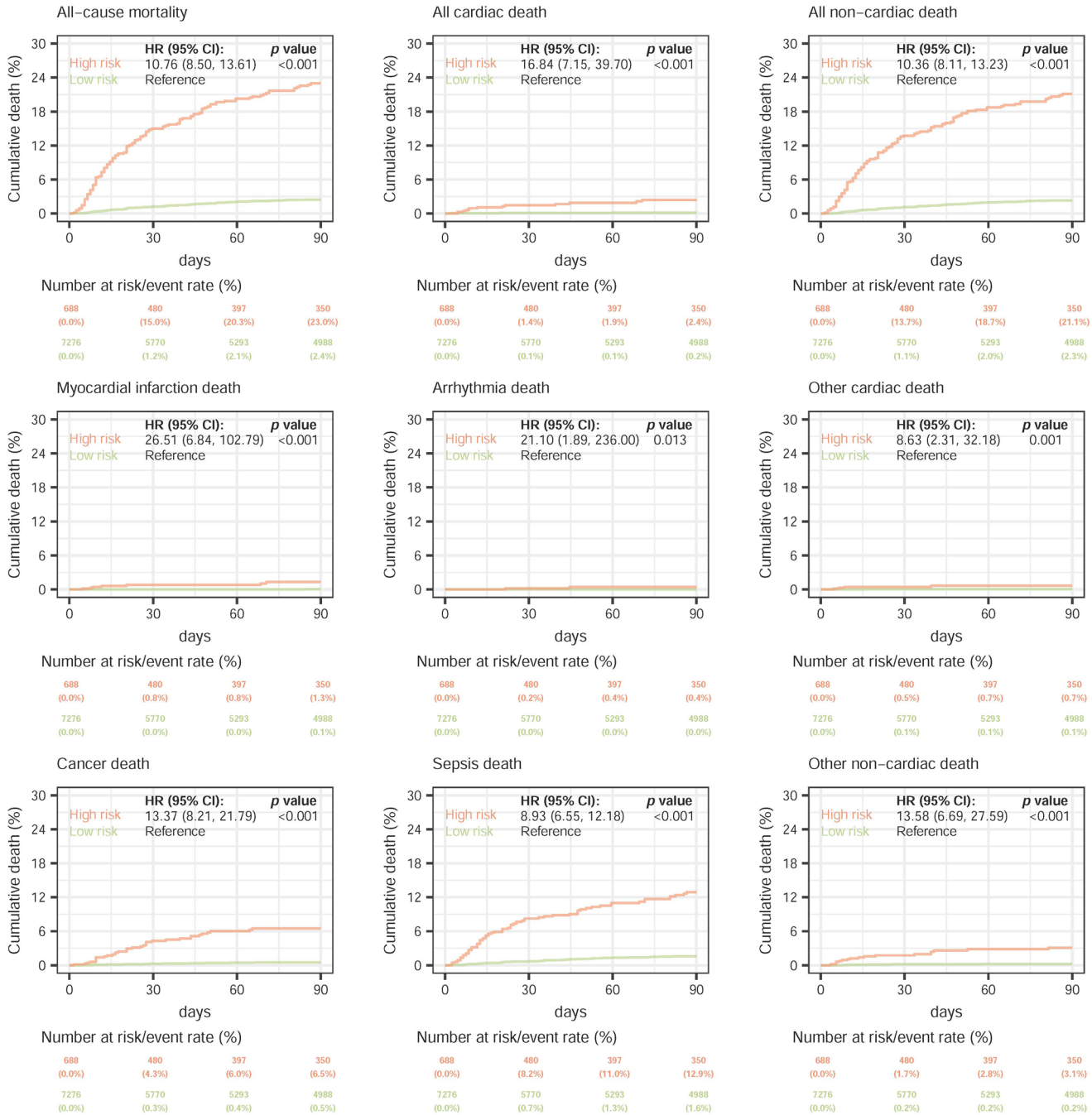

**Extended Data Fig. 1 | The association between the AI-ECG predicted all-cause mortality risk score and traditional risk features.** (**a**) Scatter plot and Spearman's rho coefficient were used to examine the relationship. In this analysis, the ECG-risk score was transformed into percentiles (PR), with a PR of 95 indicating that the patient's ECG-risk score is higher than 95% of individuals, which was the threshold to send AI-ECG alert in this trial. These PR values were determined based on previous research based on all population. Therefore, there were approximately 10% patients in emergency department (ED) and inpatient department (IPD) with an ECG-risk score higher than 95 because their conditions were collectively worse than the overall population. Age, MEWS (Modified Early Warning Score), and heart rate (HR) were chosen as they exhibited the highest correlation. The scatter plot color-codes the data points, with red indicating the highest density, followed by yellow, green, light blue, and dark blue. To better present MEWS on the scatter plot, we assigned a random number (without changing the rank) to all values during plotting. (**b**) A segmented scatter plot. As the mortality risk is extremely low for patients with a PR < 75 reported by a previous study, we performed a stratified analysis for this group (even if PR75 is not the cutoff point to send alarm message in this study, it still distinguishes between patients with low risk and median-to-high risk among these features). It can be observed that the correlation between age and ECG-risk score mainly occurs among relatively low-risk patients, while MEWS and HR are associated with relatively high-risk patients.

https://doi.org/10.1038/s41591-024-02961-4



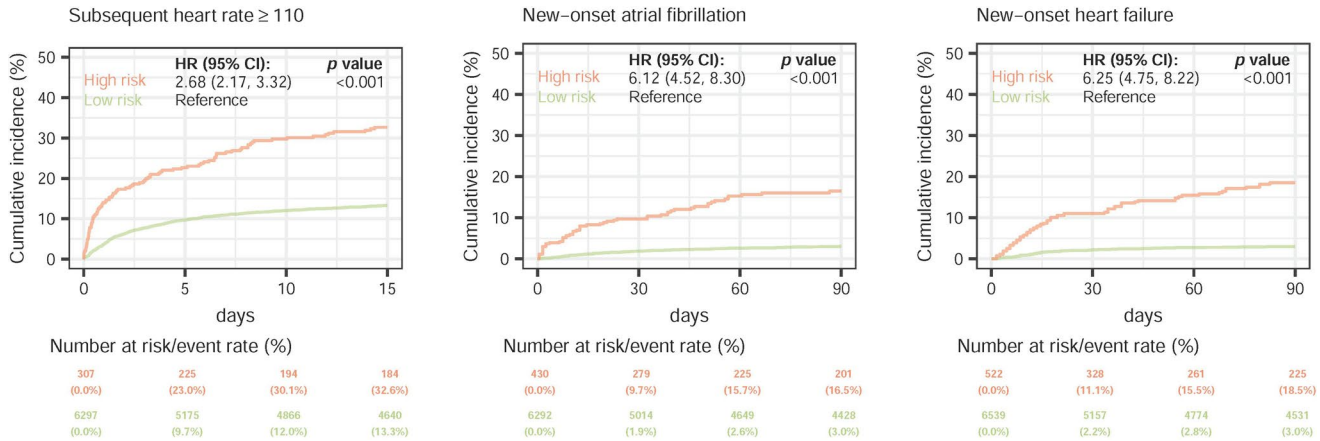

**Extended Data Fig. 2 | The comparison between AI-ECG and patient data.** (**a**) The components of AI-ECG identified high-risk group. We trained three xgboost models using patient characteristics, ECG features, and combination of them to predict AI-ECG results. The bars are the related importance of the components to predict AI-ECG. (**b**) The prediction abilities of all patient data on the AI-ECG results (AUC). The error bars are the 95% confidence intervals (CI) of each AUC. **c**) The prediction abilities of all patient data on all-cause mortality within 90 days (C-index). The error bars are the 95% confidence intervals (CI) of each AUC. The sky-blue and light-red bars represent the results of prediction using individual patient characteristics and ECG features, respectively. The blue-green bars represent predictions integrating features from xgboost and logistic regression. We used the continuous value of AI-ECG in this analysis and presented it in the gray bar. All analyses were based on the entire population data in this trial (n = 15,965).

https://doi.org/10.1038/s41591-024-02961-4



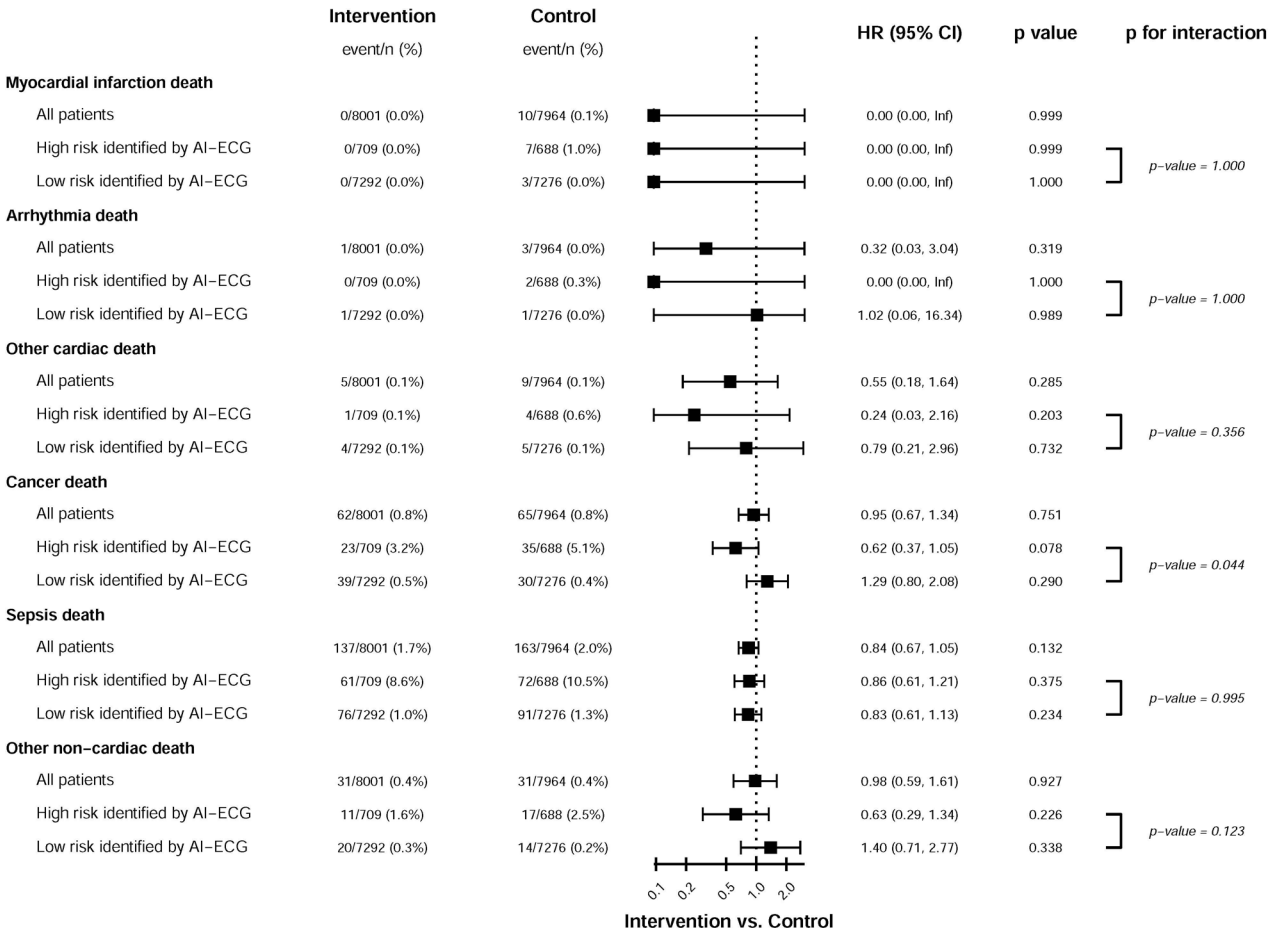

**Extended Data Fig. 3 | Performance of AI-ECG on risk groups stratified by causes of death in the control group.** Cox proportional hazard models were used for the statistical test; this was two-sided, with no adjustment for multiple comparison. The hazard ratios (HRs) were adjusted by age and sex. Red line and green line represent high risk [indicating an AI-ECG prediction greater than the operational cutoff (PR95)] and low risk [indicating an AI-ECG prediction less than the operational cutoff (PR95)] groups, respectively. The table shows the at-risk population and cumulative risk for the given time intervals in each risk group. In the analysis of causes of death, the patients died due to other cause were considered as censored data. The exact $p$ values were $2.5 \times 10^{-87}$ (all-cause mortality), $1.1 \times 10^{-10}$ (all cardiac mortality), $4.4 \times 10^{-78}$ (all non-cardiac mortality), $2.1 \times 10^{-6}$ (myocardial infarction death), $2.3 \times 10^{-25}$ (cancer death), $1.8 \times 10^{-43}$ (sepsis death), and $5.3 \times 10^{-13}$ (other non-cardiac death), respectively.

https://doi.org/10.1038/s41591-024-02961-4




**Extended Data Fig. 4 | AI-ECG risk groups on subsequent heart rate of ≥110 bpm, new-onset atrial fibrillation, and new-onset heart failure in the control group.** Cox proportional hazard models were used for the statistical test; this was two-sided, with no adjustment for multiple comparison. The hazard ratios (HRs) were adjusted by age and sex. Red line and green line represent high risk [indicating an AI-ECG prediction greater than the operational cutoff (PR95)] and low risk [indicating an AI-ECG prediction less than the operational cutoff (PR95)] groups, respectively. The table shows the at-risk population and cumulative risk for the given time intervals in each risk group. For subsequent heart rate of ≥110 bpm, we only included patients with heart rate of <100 bpm on index ECG. For new-onset atrial fibrillation, we only included patients with sinus rhythm on index ECG and without history of atrial fibrillation. For new-onset heart failure, we only included patients without history of heart failure. The exact $p$ values were $6.2 \times 10^{-20}$ (subsequent heart rate ≥110), $1.4 \times 10^{-31}$ (new-onset atrial fibrillation), and $3.9 \times 10^{-39}$ (new-onset heart failure), respectively.

https://doi.org/10.1038/s41591-024-02961-4




**Extended Data Fig. 5 | The pre-specified secondary analysis for effectiveness of AI-ECG for detailed causes of death.** The black square represents the point estimate of the HR, while the error bars indicate the 95% confidence intervals (CI). Cox proportional hazard mixed effect models were used for the statistical test; this was two-sided, with no adjustment for multiple comparison.

nature portfolio

Corresponding author(s): Chin Lin

Last updated by author(s): Mar 22, 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<br>*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<br>*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 | All data utilized in this study were electronically retrieved. R version 3.4.4, FileZilla Portable version 3.66.5, SQLite version 3.34.0, and our hospital information system were utilized. |
| Data analysis | Data analysis were conducted by R version 3.4.4 with packages "survival" version 2.43-3, "coxme" version 2.2-18.1, "xgboost" version 0.71.2, and "pROC" version 1.18.5; Model training and inference were conducted by R version 3.4.4 and package "mxnet" version 1.3.0. |

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

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The patient data cannot be made publicly available due to privacy concerns. The de-identified tabular data can be obtained from the corresponding author upon approval from the ethical committee of Tri-Service General Hospital. The ethical approval can be applied through the Tri-Service General Hospital's Clinical Trial

Management System (https://tsgh.cims.tw/wiPtms/index.html), which can expect a review period of approximately 2-3 months. Following approval, data requesters will be granted VPN access to perform analyses, ensuring data security and confidentiality (summary data can be exported), with measures in place to prevent any breach of personal information.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | The findings of this study are unrelated to gender. The gender information used in this study were collected from electronic health records, which is based on biological definition. |
| Reporting on race, ethnicity, or other socially relevant groupings | This study was conducted in Taipei, Taiwan and unrelated to race, ethnicity, or other socially relevant groupings. Taipei has a population composed of over 99% Han ethnicity, with less than 1% being indigenous peoples who, due to regulatory constraints, may not be recorded in electronic medical records. Furthermore, as this study was based on randomization process, the results were unlikely to be influenced by factors such as race, ethnicity, or other socially relevant groupings. |
| Population characteristics | A total of 39 attending physicians participated in this trial and their characteristics were shown in Extended Table 1. The mean age was 45.6±6.2 years old and 97.4% were male. As shown in Figure 1, the final analysis included 8,001 and 7,964 patients in the intervention and control groups, respectively. Among the intervention and control groups, the AI-ECG identified 709 (8.9%) and 688 (8.6%) patients with a high risk of mortality, respectively. The mean age was 60.9±18.5/61.5 ±18.2 years old, and 50.9%/52.3% were male in the intervention/control groups, respectively (Table 1). |
| Recruitment | We introduce the clinical trial protocol at hospital monthly meeting. All eligible attending physicians were asked to participate this trial, and those willing to participate will provide their inform consents before the start of this trial. A total of 39 eligible attending physicians in the department of internal medicine or ED provided informed consent. Physicians (n = 14) who did not provide informed consent were excluded in the AI-ECG report system. |
| Ethics oversight | This work was ethically approved by the institutional review board (IRB) at Tri-Service General Hospital, Taipei, Taiwan (A202105120). |

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

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

| | |
|---|---|
| Sample size | A previous study reported that intensive care was associated with a lower mortality rate compared to conventional care (relative risk: 0.71, 95% CI: 0.62-0.82).[JAMA 288, 2151-2162 (2002)] We performed sample size estimation using a two-sided significance level of 0.05, a statistical power of 0.80, a sample size ratio in intervention and control groups of 1.0, a hypothetical ratio of 1.0 in patients from the emergency departments and inpatient departments, a hypothetical proportion of controls with a primary endpoint of 0.17,[Digital health 9, 20552076231187247 (2023)] and a relative risk of 0.71, and the minimum number in intervention and control group were both 801 per arm. Since we pre-designed the study to identify approximately 10% of ECGs as high-risk to reduce alert fatigue, the final sample size of this trial was approximately 16,020. The trial started on December 15th, 2021 when the AI-ECG support was turned on for patients in the intervention group and ended on April 30th, 2022 when the number of patients per arm approximated 8,000 patients per arm. |
| Data exclusions | We conducted this RCT at both an academic medical center and a community hospital in Taiwan. A total of 39 eligible attending physicians in the department of internal medicine or emergency departments provided informed consent. Physicians who did not provide informed consent were excluded in the AI-ECG report system. Although the patients were not considered "participants" in this study, we analyzed patient-level data collected from electronic health records to explore the effect of AI-ECG intervention. Patient data were included in the analysis if the patient in the emergency departments and inpatient departments received at least one ECG for any indication between December 15, 2021 and April 30, 2022. In our hospital, ECG indications align with international healthcare systems which is primarily based on routine ward protocols and relevant clinical criteria:[Journal of the American College of Cardiology 34, 912-948 (1999)] (1) patients with middle to advanced age who require operation or hospitalization received routine ECG examinations, (2) patients with symptoms related to disturbances of heart rhythm (e.g., syncope, palpitation, and intermittent shortness of breath) and chest pain or suspected acute coronary syndrome, (3) patients with existing or suspected cardiovascular diseases (e.g., congestive heart failure, hypertension, pericarditis, valvular heart disease, and cardiomyopathies), (4) evaluation of bradycardia or tachycardia, (5) evaluation of electrolyte imbalances or drug toxicity, and (6) monitoring patients with implanted cardiac device. Patients under 18 years old or had a time delay of more than 2 hours between their ECG and AI-ECG analysis were excluded. We only analyzed patients who were cared for by attending physicians who had provided informed consent. |
| Replication | There were no replication attempts for this prospective randomized trial due to cost and time requirements. |
| Randomization | Randomization was conducted according to all possible medical numbers (Extended Data Figure 6), which is 7 digits long as a serial number in |

| | |
|---|---|
| Randomization | our hospital with a total of 10^7 combinations. The patient-level randomization process was implemented to ensure longer patient follow-up, avoiding potential loss of follow-up which could occur with physician-level assignment. This approach enhances the study's reliability by maintaining consistency in the care received by patients throughout the 90-day period. Prior to the start of this trial, we have completed the randomization process using simple sampling at random, and a half of possible medical numbers were allocated to the intervention group. In other words, the randomization process may have occurred before the creation of the medical record number. |
| Blinding | Blinding is not feasible for participating physicians, as they can easily confirm through the electronic health record whether the patients under their care belong to the intervention group (able to use the AI-ECG module) or the control group (unable to use the AI-ECG module). For researchers, maintaining blinding during data collection was ensured because the generation of electronic medical records is independent of the research team, and during preliminary data analysis, blinding was also maintained as patient identifiers are de-identified. |

# 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 | This trial was registered before initiation (ClinicalTrials.gov number NCT05118035). |
| Study protocol | Full study protocol has been provided in supplementary note. |
| Data collection | This trial was conducted in Tri-Service General Hospital System, including an academic medical center located in Neihu district, Taipei City, and a community hospital located in Zhongzheng district, Taipei City. Physicians' data were self-reported via the questionnaire to be filled in providing informed consent, and the patient-level data were collected from electronic health records. Physicians' consents were collected between Nov 1, 2021 and Nov 30, 2021. Patient data were collected between Dec 15, 2021 and Apr 30, 2022. Outcome data were collected within 90 days of the index ECG. |
| Outcomes | The primary outcome in this study was all-cause mortality. Patient status (dead/alive) was acquired through the EHR. Although it is possible that some patients may have died outside of our hospital and therefore were not captured by our electronic medical record, we ensured that censored patients were known to have survived until their last recorded hospital encounter. The data for non-event visits were censored at the patient's last known hospital live encounter to limit bias from incomplete records. The end point of this study was set as July 31, 2022 since we needed to follow-up at 90 days for patients under study as of April 30, 2022.<br>The secondary outcomes included follow-up examination and medical treatment after ECG exam, including intensive care unit (ICU) admission, arrhythmia-related intervention, heart function examination, and electrolyte examination. Moreover, to determine the cause of death within 90 days, we conducted a manual review and categorized the cases as either cardiac (including myocardial infarction, arrhythmia, and others) or noncardiac (including cancer, sepsis, and others). |

## Plants

| | |
|---|---|
| Seed stocks | Not applicable |
| Novel plant genotypes | Not applicable |
| Authentication | Not applicable |