Artificial intelligence-based electrocardiographic biomarker for outcome prediction in patients with acute heart failure

Youngjin Cho, Minjae Yoon, Joonghee Kim, Ji Hyun Lee, Il-Young Oh, Chan Joo Lee, Seok-Min Kang, Dong-Ju Choi

Submitted to: Journal of Medical Internet Research
on: August 27, 2023

Table of Contents

Artificial intelligence-based electrocardiographic biomarker for outcome prediction in patients with acute heart failure

Youngjin Cho[1*] MD, PhD; Minjae Yoon[1*] MD; Joonghee Kim[1] MD; Ji Hyun Lee[1] MD, PhD; Il-Young Oh[1] MD, PhD; Chan Joo Lee[2] MD, PhD; Seok-Min Kang[2*] MD, PhD; Dong-Ju Choi[1*] MD, PhD

[1]Seoul National University Bundang Hospital, Seoul National University College of Medicine Division of Cardiology, Department of Internal Medicine Seongnam, Gyeonggi-do KR
[2]Severance Hospital, Yonsei University College of Medicine Division of Cardiology, Department of Internal Medicine Seoul KR
[*]these authors contributed equally

Corresponding Author:
Dong-Ju Choi MD, PhD
Seoul National University Bundang Hospital, Seoul National University College of Medicine
Division of Cardiology, Department of Internal Medicine
82, Gumi-ro 173 Beon-gil, Bundang-gu
Seongnam, Gyeonggi-do
KR

Abstract

Background: Although several biomarkers exist for patients with heart failure (HF), their use in routine clinical practice is often constrained by their cost and limited availability.

Objective: We examined the utility of an artificial intelligence (AI) algorithm that analyses printed electrocardiograms (ECGs) for outcome prediction in patients with acute HF.

Methods: We retrospectively analysed prospectively collected data of patients with acute HF in two tertiary centres. Baseline ECGs were analysed using a deep learning system called Quantitative ECG (QCG™) trained to detect several urgent clinical conditions, including shock, cardiac arrest, and reduced left ventricular ejection fraction (LVEF).

Results: Among the 1,254 patients enrolled, in-hospital cardiac death (IHCD) occurred in 53 (4.2%) patients, and the QCG score for critical events (QCG-Critical) was significantly higher in these patients than in survivors (0.57±0.23 vs. 0.29±0.20, P<.001). QCG-Critical score was an independent predictor of IHCD after adjustment for age, sex, comorbidities, HF aetiology/type, atrial fibrillation, and QRS widening (adjusted odds ratio [OR], 1.68; 95% confidence interval [CI], 1.47–1.92; P<.001, per 0.1 increase), and even after additional adjustments for echocardiographic LVEF and N-terminal pro-B-type natriuretic peptide (adjusted OR, 1.59; 95% CI, 1.36–1.87; P<.001, per 0.1 increase). During long-term follow-up, patients with higher QCG-Critical scores (>0.5) had higher mortality rates than those with low QCG-Critical scores (<0.25) (adjusted hazard ratio, 2.69; 95% CI, 2.14–3.38; P<.001).

Conclusions: Predicting outcomes in patients with acute HF using the QCG-Critical score is feasible, indicating that the AI-based ECG score may be a novel biomarker for these patients. Clinical Trial: The study design has been registered in ClinicalTrial.gov NCT01389843.

(JMIR Preprints 27/08/2023:52139)
DOI: https://doi.org/10.2196/preprints.52139

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
✓ **Only make the preprint title and abstract visible.**
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Artificial intelligence-based electrocardiographic biomarker for outcome prediction in patients with acute heart failure

Youngjin Cho[1*], Minjae Yoon[1*], Joonghee Kim[2], Ji Hyun Lee[1], Il-Young Oh[1], Chan Joo Lee[3], Seok-Min Kang[3**], Dong-Ju Choi[1**]

[1]Division of Cardiology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
[2]Department of Emergency Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
[3]Division of Cardiology, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]These authors contributed equally to this work.

Running title: Artificial intelligence-based electrocardiography for acute heart failure
Word count: 3,598

[**]**Co-Correspondence:**
Professor Dong-Ju Choi, MD, PhD
Cardiovascular Center, Department of Internal Medicine,
Seoul National University Bundang Hospital, Seoul National University College of Medicine,
82, Gumi-ro 173 Beon-gil, Bundang-gu, Seongnam, Gyeonggi-do, 13620, Republic of Korea
Telephone: 82-31-787-7007; Fax: 82-31-787-7041
E-mail: djchoi.snu@gmail.com

Professor Seok-Min Kang, MD, PhD
Division of Cardiology, Department of Internal Medicine,
Severance Hospital, Yonsei University College of Medicine,
50-1, Yonsei-Ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
Telephone: 82-2-2228-8450; Fax: 82-2-2227-7732
E-mail: smkang@yuhs.ac

ABSTRACT

Background

Although several biomarkers exist for patients with heart failure (HF), their use in routine clinical practice is often constrained by their cost and limited availability.

Objective

We examined the utility of an artificial intelligence (AI) algorithm that analyses printed electrocardiograms (ECGs) for outcome prediction in patients with acute HF.

Methods

We retrospectively analysed prospectively collected data of patients with acute HF in two tertiary centres. Baseline ECGs were analysed using a deep learning system called Quantitative ECG (QCG™) trained to detect several urgent clinical conditions, including shock, cardiac arrest, and reduced left ventricular ejection fraction (LVEF).

Results

Among the 1,254 patients enrolled, in-hospital cardiac death (IHCD) occurred in 53 (4.2%) patients, and the QCG score for critical events (QCG-Critical) was significantly higher in these patients than in survivors (0.57±0.23 vs. 0.29±0.20, *P*<.001). QCG-Critical score was an independent predictor of IHCD after adjustment for age, sex, comorbidities, HF aetiology/type, atrial fibrillation, and QRS widening (adjusted odds ratio [OR], 1.68; 95% confidence interval [CI], 1.47–1.92; *P*<.001, per 0.1 increase), and even after additional adjustments for echocardiographic LVEF and N-terminal pro-B-type natriuretic peptide (adjusted OR, 1.59; 95% CI, 1.36–1.87; *P*<.001, per 0.1 increase). During long-term follow-up, patients with higher QCG-Critical scores (>0.5) had higher mortality rates than those with low QCG-Critical scores (<0.25) (adjusted hazard ratio, 2.69; 95% CI, 2.14–3.38; *P*<.001).

Conclusions

Predicting outcomes in patients with acute HF using the QCG-Critical score is feasible, indicating

that the AI-based ECG score may be a novel biomarker for these patients.

Trial Registration

The study design has been registered in ClinicalTrial.gov NCT01389843.

Keywords: acute heart failure; electrocardiography; artificial intelligence; deep learning

Introduction

Heart failure (HF) is a major global health problem affecting millions of people worldwide, leading to significant morbidity, mortality, and healthcare expenditure [1-3]. Although several valuable biomarkers, such as N-terminal pro-B-type natriuretic peptide (NT-proBNP) [4,5] and cardiac troponins [6], have been introduced for patients with HF, their use in routine clinical practice is often constrained by their cost and limited availability.

Electrocardiogram (ECG) is an essential and cost-effective tool for evaluating cardiovascular diseases. It is widely available, non-invasive, and provides real-time information about cardiac electrical activity, which is crucial for detecting arrhythmias, ischemia, and other cardiac abnormalities. With the advances in artificial intelligence (AI) and deep learning, there has been growing interest in employing AI algorithms to analyse ECG data and predict outcomes in patients with various cardiovascular conditions [7,8].

In this study, we investigated the utility of an AI algorithm that analyses printed ECG images for outcome prediction in patients with acute HF. This will demonstrate the potential of AI-assisted ECG analysis for predicting outcomes in these patients, potentially overcoming the cost and availability constraints of current biomarkers.

Methods

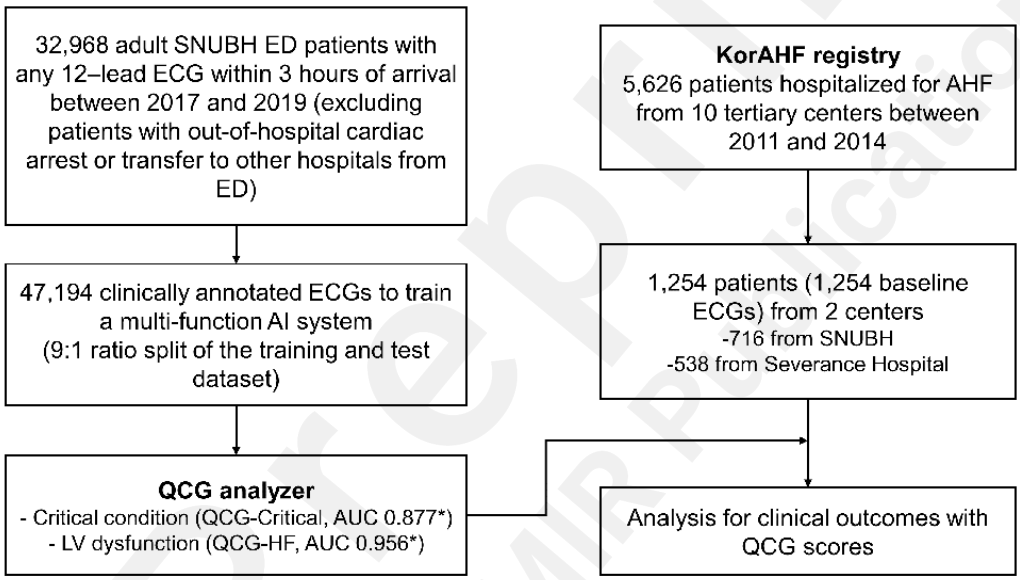
Study population

This was a sub-study of the prospective multicentre Korean Acute Heart Failure (KorAHF) registry, which enrolled consecutive 5,625 patients upon initial hospital admission for acute HF at 10 tertiary university hospitals in Korea. Details on the KorAHF registry objectives, design, and population are available on the clinical trial registration site (ClinicalTrial.gov, NCT01389843), and have been published previously [9,10]. Briefly, patients who had signs or symptoms of HF and met one of the following criteria were eligible for KorAHF registry: 1) lung congestion or 2) objective left

ventricular systolic dysfunction or structural heart disease findings. There were no exclusion criteria.

In the present study, we retrospectively analysed the prospectively collected data from 1,254 patients who were hospitalised for acute HF from March 2011 to February 2014 at two out of 10 participant tertiary centres (Seoul National University Bundang Hospital and Severance Hospital) using the KorAHF registry (Figure 1). Additional ECG image data were collected for this study.

Figure 1. Flowchart of the training and validation study population.



The AI-ECG analyzer called QCG™ was developed utilizing 47,194 annotated ECG images of over 32,968 patients who visited the emergency department of SNUBH between 2017 and 2019. The QCG analyzer was applied to ECGs from a subpopulation of KorAHF registry which enrolled AHF patients between 2011 and 2014.

*Internal validation results for two QCG scores.

AHF, acute heart failure; AI, artificial intelligence; AUC, area under the curve; ECG, electrocardiogram; ED, emergency department; LV, left ventricular; SNUBH, Seoul National

University Bundang Hospital, QCG, quantitative ECG.

This study conforms with the principles outlined in the Declaration of Helsinki. The study protocol was approved by the ethics committee or institutional review board at each hospital (approval numbers B-1104-125-014 and 2022-2166-001). The need for written informed consent was waived by the institutional review board. Our research strictly adheres to the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [11].

Clinical follow-up and endpoints

Data collection methods have been previously described [9]. Briefly, data on patients' clinical manifestations, biochemical parameters, medication, and outcome were collected using a web-based case report form up to 60 months by research nurses. Outcome data on patients lost to follow-up were additionally collected from the National Death Records.

The primary endpoint of this study was all-cause mortality. Secondary outcomes included in-hospital outcomes, in particular, in-hospital mortality. All deaths were considered cardiac unless a definite non-cardiac cause could be established. All outcome data reported from the participating centres were reviewed by an independent clinical event adjudicating committee.

AI algorithm

The AI analyzer called Quantitative ECG (QCG^{TM}) is composed of an encoder-part and multiple task-specific networks. The encoder part is a modified convolutional neural network with residual connections, squeeze excitation modules and a non-local block. The task-specific networks are multilayer percetron models. The encoder part accepts two-dimensional ECG images as input to produce a common numerical feature vector for downstream tasks. The encoder part was pretrained

on 49,731 open ECGs using self-supervised learning schemes and then fine-tuned on 47,194 annotated ECG images of over 32,968 patients who visited the Emergency Department of Seoul National University Bundang Hospital between 2017 and 2019 using multitask learning schemes. The tasks include classification of 12 rhythms (with 35 subtypes) and production of 10 digital biomarkers correlated with the risk of 1) being critically-ill (shock, respiratory failure or cardiac arrest), 2) cardiac ischemia (acute coronary syndrome, STEMI, myocardial injury as defined elevated troponin level), 3) cardiac dysfunction (pulmonary edema, left and right heart dysfunction, pulmonary hypertension and clinically significant pericardial effusion), and 4) hyperkalemia. Several validation studies of the system have been published previously [12-14]. The collection of these AI algorithms has been developed into a mobile application (ECG Buddy, ARPI), which has been approved by the Korean Ministry of Food and Drug Safety. In this study, two QCG features were evaluated: QCG-Critical for critical conditions such as shock or mortality, and QCG-HF for reduced echocardiographic left ventricular ejection fraction (LVEF) of <40%. The QCG scores, representing probability, ranged from 0 to 1.0, with 0 indicating low and 1.0 indicating high probability. In a 9:1 ratio split of the training and test datasets, the internal validation results for these two QCG features showed an area under the curve (AUC) of 0.877 for QCG-Critical and 0.956 for QCG-HF. The composition of the training and validation dataset is presented as a flowchart in Figure 1.

Statistical analysis

Categorical variables are reported as frequencies (percentages) and continuous variables are expressed as means ± standard deviations or medians with interquartile ranges. The two key AI-driven scores (QCG-Critical and QCG-HF) were analyzed as continuous variables. Student’s t-test and chi-square (or Fisher’s exact) test were used to compare the baseline clinical characteristics between the two groups. The discrimination performance of QCG scores for in-hospital outcomes was evaluated using receiver operating characteristic (ROC) curve analysis. The AUC values were

compared using the DeLong test. The logistic regression model was employed to estimate the odds ratios (ORs) and 95% confidence intervals (CIs). Survival analysis was performed using the Kaplan–Meier method, and the Cox proportional hazard model was used to estimate the hazard ratios (HRs) and 95% CIs for the clinical outcomes. Multivariable analysis was performed with the inclusion of clinically relevant variables.

All tests were two-tailed, and a *P*-value <.05 was considered statistically significant. Statistical analyses were performed using R programming version 4.3.0 (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Baseline characteristics

Data of 1,254 patients (716 from Seoul National University Bundang Hospital and 538 from Severance Hospital) were analysed. Among these, 53 patients (4.2%) experienced in-hospital cardiac death (IHCD). The baseline characteristics of the study population according to the in-hospital outcomes are shown in Table 1. Compared with survivors, these patients were older, had a higher prevalence of ischemic heart disease, lower LVEF, and higher NT-proBNP levels. Contrarily, atrial fibrillation (AF) was more frequent in survivors. The QCG-Critical and QCG-HF scores were significantly higher in patients who experienced IHCD than in survivors (0.57±0.23 vs. 0.29±0.20, *P*<.001; 0.78±0.18 vs. 0.64±0.31, *P*<.001) (Figure S1 in the multimedia appendix 1).

Table 1. Baseline characteristics

	Total (n=1,254)	**In-hospital cardiac death**	**Survivor (n=1,201)**	***P* value**

		(n=53)		
Age (years)	69.8 ± 14.7	74.0 ± 14.5	69.6 ± 14.1	.03
Male	673 (53.7%)	29 (54.7%)	644 (53.6%)	.99
Hypertension	843 (67.2%)	31 (58.5%)	812 (67.6%)	.22
Diabetes mellitus	499 (39.8%)	23 (43.4%)	476 (39.6%)	.69
Cerebrovascular disease	224 (17.9%)	7 (13.2%)	217 (18.1%)	.47
Chronic kidney disease	212 (28.3%)	11 (20.8%)	212 (28.3%)	.69
Ischemic heart disease	365 (29.1%)	25 (47.2%)	340 (28.3%)	.005
Valvular heart disease	217 (17.3%)	9 (17.0%)	208 (17.3%)	.999
De novo HF	612 (48.8%)	29 (54.7%)	583 (48.5%)	.46
Atrial fibrillation	417 (34.7%)	10 (10.9%)	417 (34.7%)	.03
QRS duration ≥120ms	318 (25.4%)	17 (32.1%)	301 (25.1%)	.32
LVEF (%)	35.3 ± 14.7	28.5 ± 11.9	35.6 ± 14.7	.002
NT-proBNP (pg/mL)	10,373 ± 11,915	17,035 ± 1,900	10,092 ± 11,879	<.001
QCG parameters				
QCG-Critical	0.30 ± 0.21	0.57 ± 0.23	0.29 ± 0.20	<.001
QCG-HF	0.65 ± 0.31	0.78 ± 0.18	0.64 ± 0.31	<.001

Values are expressed as number (%) or mean ± standard deviation.

HF, heart failure; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro B-type natriuretic peptide, QCG, quantitative electrocardiogram.

Predictors of IHCD

In the univariable logistic regression analysis, the QCG-Critical and QCG-HF scores were significant predictors of IHCD (OR, 1.66; 95% CI, 1.47–1.87; *P*<.001 and OR, 1.21; 95% CI, 1.08–1.37; *P*=.001, respectively, per 0.1 increase) (Table 2). Other than QCG scores, echocardiographic LVEF, NT-proBNP level, age, ischemic heart disease and AF were significantly correlated with IHCD.

Table 2. Predictors for In-hospital cardiac death

	OR (95% CI)	*P* value	Adjusted OR (95% CI)[a]	*P* value	Adjusted OR (95% CI)[b]	*P* value
QCG parameters (per 0.1)						
QCG-Critical	1.66 (1.47 – 1.87)	<.001	1.68 (1.47 – 1.92)	<.001	1.59 (1.36 – 1.87)	<.001
QCG-HF	1.21 (1.08 – 1.37)	.001	1.22 (1.08 – 1.39)	.002	1.02 (0.84 – 1.24)	.82
LVEF (per 5% decrease)	1.21 (1.07 –1.37)	.002	1.26 (1.09 - 1.45)	.001	*1.29 (1.10 – 1.51)*	*.02*
NT-proBNP (per 1,000)	1.03 (1.01 – 1.05)	<.001	1.04 (1.01 – 1.06)	<.001	*1.03 (1.00 – 1.05)*	*.002*
Age	1.03 (1.00 – 1.05)	.03	*1.03 (1.00 – 1.06)*	*.03*	*1.04 (1.01 – 1.08)*	*.009*
Male	1.05 (0.60 – 1.82)	.88	*1.21 (0.66 – 2.23)*	*.53*	*1.00 (0.48 – 2.07)*	*.99*
Hypertension	0.68 (0.39 – 1.18)	.17	*0.51 (0.27 – 0.99)*	*.046*	*0.42 (0.20 – 0.90)*	*.03*
Diabetes mellitus	1.17 (0.67 – 2.03)	.58	*0.72 (0.38 – 1.36)*	*.31*	*0.56 (0.26 – 1.19)*	*.13*
Chronic kidney disease	1.22 (0.62 – 2.41)	.56	*1.08 (0.49 – 2.34)*	*.85*	*1.06 (0.41 – 2.74)*	*.90*
Cerebrovascular disease	0.69 (0.31 – 1.55)	.37	*0.64 (0.26 – 1.53)*	*.31*	*0.54 (0.18 – 1.68)*	*.29*
Ischemic heart disease	2.26 (1.30 – 3.93)	.004	*3.00 (1.51 – 5.97)*	*.002*	*2.54 (1.12 – 5.76)*	*.03*
Valvular heart disease	0.98 (0.47 – 2.03)	.95	*2.00 (0.84 – 4.75)*	*.12*	*1.94 (0.68 – 5.50)*	*.21*
ADHF (vs. De novo)	0.78 (0.45 – 1.36)	.38	*0.54 (0.26 – 1.11)*	*.10*	*0.51 (0.22 – 1.19)*	*.12*
Atrial fibrillation	0.44 (0.22 – 0.88)	.02	*0.73 (0.34 – 1.58)*	*.42*	*0.80 (0.32 – 2.00)*	*.63*
QRS duration >120ms	1.41 (0.78 - 2.55)	.25	*0.94 (0.48 – 1.83)*	*.85*	*1.23 (0.57 – 2.64)*	*.60*

[a]Adjusted for age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, atrial fibrillation, and QRS duration.

[b]Further adjusted for LVEF and NT-proBNP, in addition to all the covariates above.

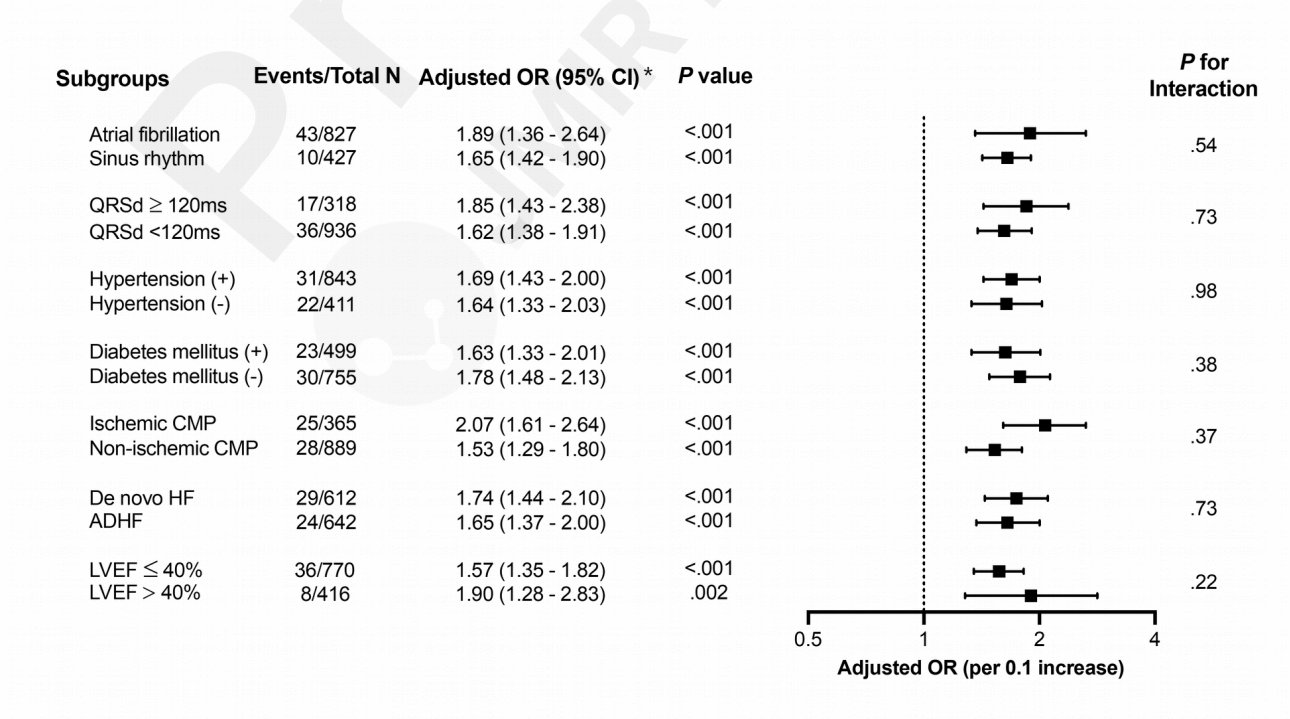
When a variable is included as a covariate for adjustment, it is not adjusted for itself and QCG-Critical is added to the adjustment model (presented in *italics*).

ADHF, acute decompensated heart failure; CI, confidence interval; OR, odds ratio. Other abbreviations are as Table 1.

After adjustment for age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, AF, and QRS duration, the two QCG scores remained significant predictors of IHCD. Moreover, the QCG-Critical score was an independent predictor of IHCD after further adjustment for echocardiographic LVEF and NT-proBNP level (OR, 1.59; 95% CI, 1.36–1.87; *P*<.001).

In a subgroup analysis, the QCG-Critical score was a significant predictor of IHCD, regardless of initial rhythm (AF or sinus rhythm), QRS width (wide or narrow), hypertension, diabetes, HF aetiology (ischemic or non-ischemic), HF type (de novo or acute decompensated HF), and LVEF (HF with reduced EF vs HF with preserved or mildly reduced EF), after adjustment for other clinical parameters (Figure 2).

Figure 2. Subgroup analysis results for predicting IHCD.



Adjusted ORs are presented for a 0.1 increase in the QCG-Critical score

*Adjusted for age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, atrial fibrillation, and QRS duration.

ADHF, acute decompensated heart failure; CI, confidence interval; CMP, cardiomyopathy; HF, heart failure; IHCD, in-hospital cardiac death; LVEF, left ventricular ejection fraction; OR, odds ratio.

QCG-Critical score and IHCD

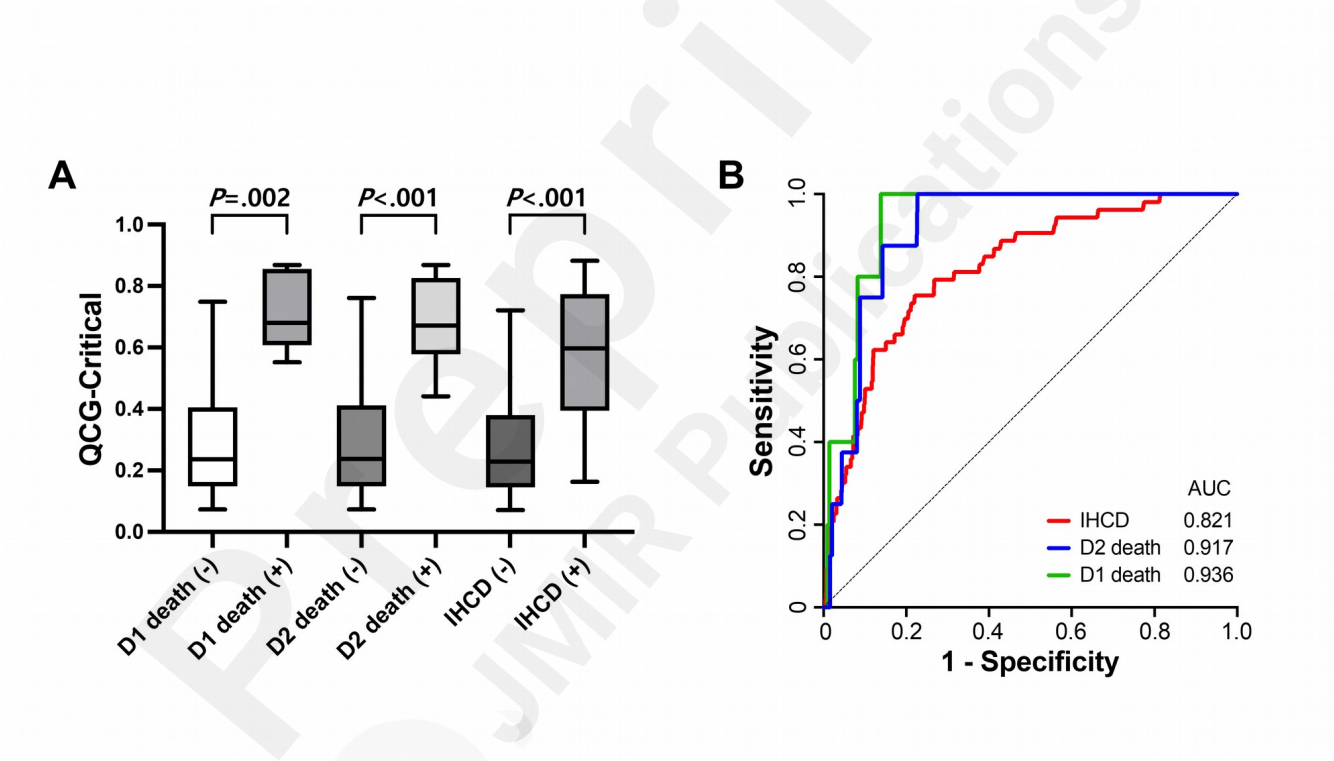
The QCG-Critical score was significantly higher in patients who experienced cardiac death within 1 day, 2 days, or during hospitalisation, than in survivors (Figure 3A). When the performance of the QCG-Critical score for predicting these events was analysed using ROC curves, the AUC values for 1- and 2-day mortality and IHCD were 0.936, 0.917, and 0.821, respectively (Figure 3B). Comparatively, the AUC values of echocardiographic LVEF and NT-proBNP level for IHCD were 0.642 (vs. QCG-Critical, $P<.001$) and 0.720 (vs. QCG-Critical, $P=.07$) (Figure 4A).

When the QCG-Critical score was compared with the prediction model utilizing traditional clinical variables (model 1) including age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, AF, and QRS duration, the AUC value of QCG-Critical score was significantly higher than the AUC of model 1 (0.821 vs. 0.705, $P=.02$). In addition, when the QCG-Critical score was added to a model 1, it significantly enhanced the prediction for IHCD (AUC of model 1 vs. model 1 with

QCG-Critical = 0.705 vs 0.843, *P*<.001) (Figure 4B). When NT-proBNP and LVEF were further included in model 1 (model 2), the QCG-Critical score again demonstrated additional predictive value for IHCD compared to model 2 alone (AUC of model 2 vs. model 2 with QCG-Critical = 0.787 vs. 0.863, *P*=.01) (Figure 4C).

--

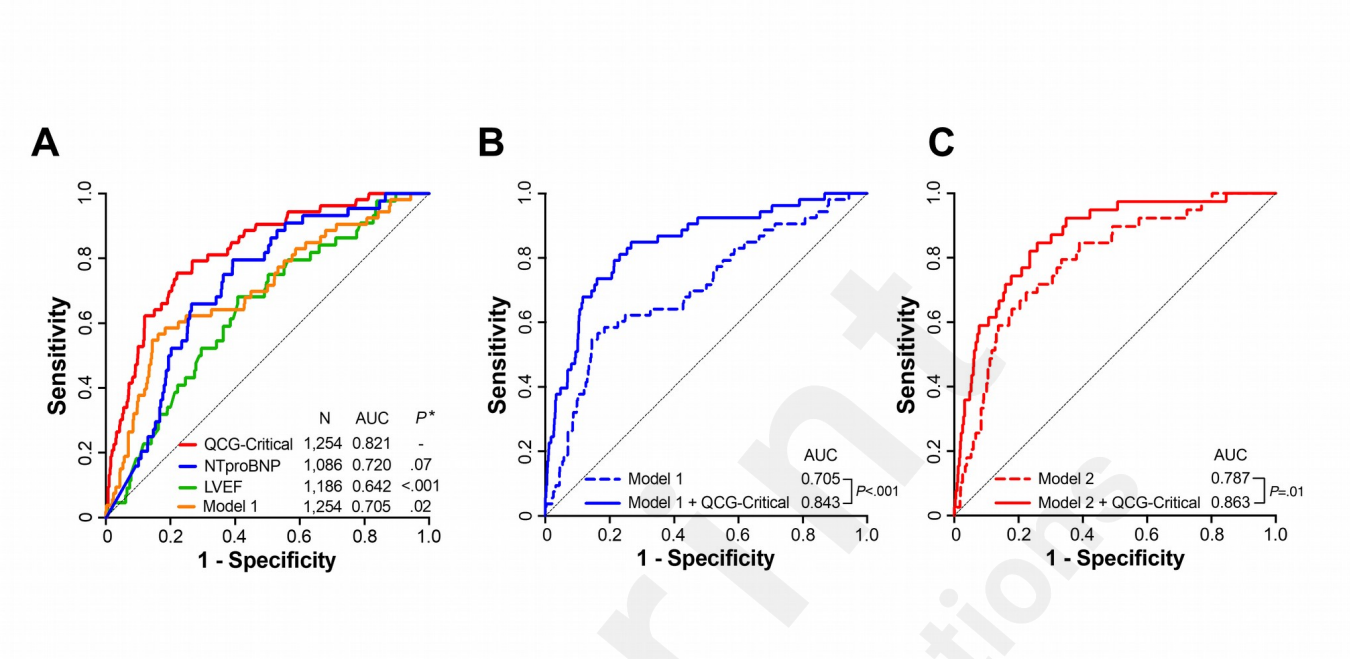
Figure 3. Performance of the QCG-Critical score for predicting IHCD



(A) The QCG-Critical score was significantly higher in patients who experienced cardiac death within 1 day, 2 days, and during hospitalisation than in survivors. The Box-and-Whisker plot is presented as 5–95 percentiles. (B) Performance of the QCG-Critical score presented as ROC curves.

AUC, area under the curve; IHCD, in-hospital cardiac death; ROC, receiver operating characteristic.

Figure 4. The ROC curves for predicting IHCD



(A) The AUC value of the QCG-Critical score was 0.821 and tended to be higher than that of echocardiographic LVEF, NT-proBNP or a model utilizing traditional clinical variables (model1). (B) The performance of the prediction model 1 and with the addition of the QCG-Critical score. (C) Upon further incorporation of NT-proBNP and LVEF into the model 1 (model 2), the QCG-Critical score demonstrated additional predictive value for IHCD than model 2 alone.

Model 1 includes age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, atrial fibrillation and QRS duration. NT-proBNP and LVEF were further incorporated into the model 2.

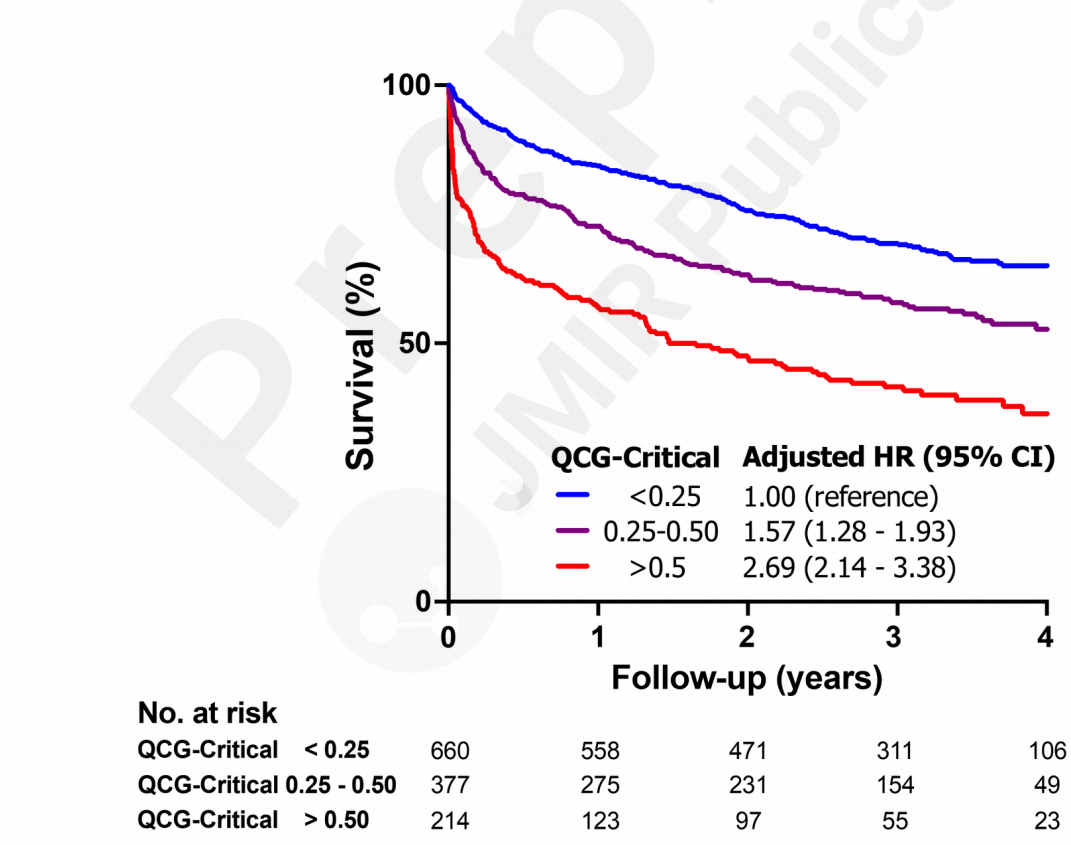
**P* value for comparison with the AUC of QCG-Critical score.

HF, heart failure; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type natriuretic peptide. Other abbreviations as Figure 3.

QCG-Critical score and long-term outcomes

During a median follow-up of 2.7 years, 508 deaths occurred in the study population. To further analyse the performance of the QCG-Critical score for outcome prediction, we divided patients in three QCG-Critical score groups based on arbitrary cut-off values of 0.25 and 0.50, and conducted survival analysis (Figure 5).

Figure 5. Kaplan–Meier curves for long-term mortality according to the QCG-Critical scores



CI, confidence interval; HR, hazard ratio.

After adjustment for age, sex, comorbidities, HF aetiology and type, AF, and QRS widening, patients with higher QCG-Critical scores had significantly higher all-cause mortality rates during follow-up than those with the lowest QCG-Critical scores (<0.25). The adjusted HRs (95% CIs) were 1.57 (1.28–1.93) for patients with QCG-Critical scores between 0.25 and 0.50, and 2.69 (2.14–3.38) for patients with QCG-Critical scores higher than 0.50 (all *P*<.001). With additional adjustment for LVEF and NT-proBNP to the previous model, adjusted HR was 1.61 and 2.27, respectively, consistent with the main analysis (Figure S2 in the multimedia appendix 1.).

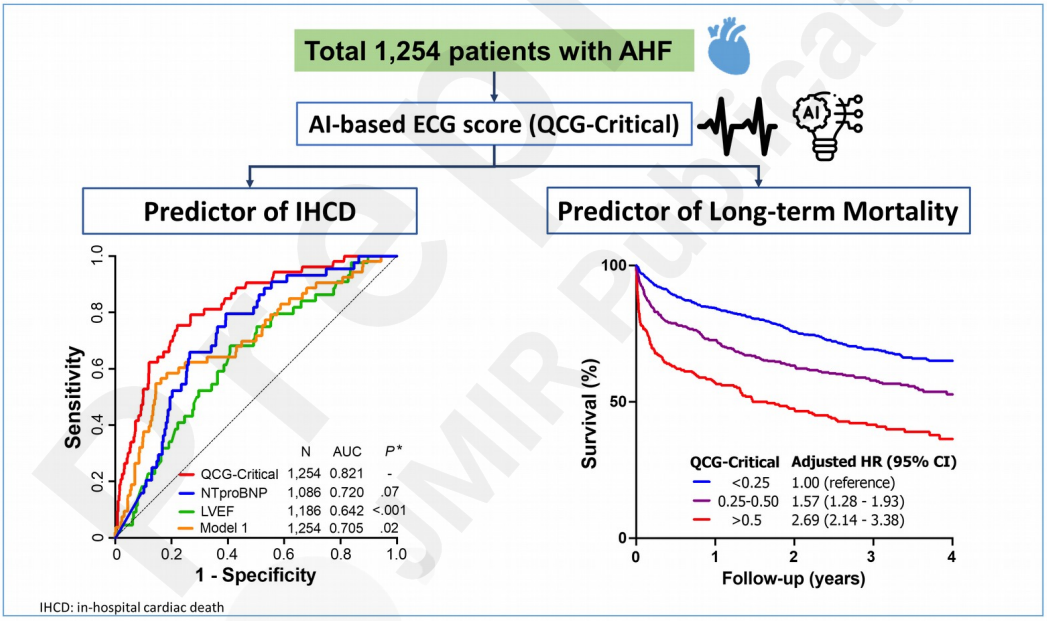
In a subgroup analysis, a higher QCG-Critical score (>0.50 vs. ≤0.50) was significantly correlated with all-cause mortality during follow-up, regardless of initial rhythm (AF or sinus rhythm), QRS width (wide or narrow), hypertension, diabetes, HF aetiology, HF type (de novo or acute decompensated HF), and LVEF (HF with reduced EF vs HF with preserved or mildly reduced EF), after adjustment for other clinical parameters (Figure S3 in the multimedia appendix 1).

Discussion

Predicting outcomes in patients with HF is important for guiding management and improving prognosis [15], but is often hindered by the complexity of HF pathophysiology and the presence of other comorbidities. Recently, AI algorithms based on big medical records have been found to be helpful in predicting the outcomes of patients with HF [16,17], but these are difficult to apply in daily practice and their performance requires further improvements. In the current study, the

QCG-Critical, a newly developed AI-ECG score (QCG™), was well correlated with early mortality and IHCD during the index after adjusting for traditional clinical risk factors. Moreover, the QCG-Critical score was an independent predictor of long-term all-cause mortality in this population, suggesting that this AI-based ECG score may serve as a novel biomarker for these patients (Graphical Abstract).

Graphical Abstract. AI-based ECG score for predicting IHCD and long-term mortality



**P* value for comparison with the AUC of QCG-Critical score.

ECG is cost-effective, widely available, and easy to perform, and is therefore often used as a first-line evaluation for patients with cardiovascular diseases. ST-elevation myocardial

infarction is a quintessential disease where ECG evaluation is critical for timely diagnosis. Although ECG is not deterministic for HF diagnosis, several studies have demonstrated that some ECG features are correlated with its characteristics [18]. In addition, the presence of atrial fibrillation or QRS widening may be ECG features reflecting unfavourable underlying haemodynamics, thus correlating with poor prognosis [19,20]. More subtle ECG changes have also been suggested as predictors of poor prognosis in patients with HF, but these require high levels of experience and skill for interpretation, which may limit their applicability [21].

Theoretically, the ECG signal may contain information regarding the electric and mechanical activities of the diseased heart beyond physicians' perception. With the assistance of AI, ECG may provide valuable information beyond its current usage. For example, Attia et al. [22] reported that LVEF reduction may be detected by ECG using AI. This new application of AI-ECG was reproduced by other researchers [23,24]. In the current study, the QCG-HF score also showed good performance in predicting reduced echocardiographic LVEF of less than 40%, with an AUC value of 0.884 (Figure S4 in the multimedia appendix 1). Notably, in the above-mentioned studies, the AI-ECG-predicted LVEF was correlated with the prognosis of patients with chronic HF, while in our study, the AI-based ECG score had a predictive value in patients with acute HF. Thus, to the best of our knowledge, the current study represents an initial effort in terms of predicting the outcomes of acute HF using AI-based ECG interpretation.

The QCG-Critical score was originally trained to detect critical medical conditions that may result in shock or mortality within a day [12]. In the present study, it predicted early cardiac mortality in patients with acute HF with high accuracy. For the prediction of IHCD, the AUC value of the QCG-Critical score was higher than that of echocardiographic LVEF. It was also higher than the AUC value of the serum NT-proBNP level, but without statistical significance.

Notably, the QCG-Critical score was available for all 1,254 patients enrolled in the KorAHF study, while LVEF and NT-proBNP results were not available in 68 (5.4%) and 168 (13.4%) patients, respectively. Considering that the KorAHF study enrolled patients from tertiary centres in Korea, a high proportion of patients with acute HF might not be able to benefit from these echocardiographic or serum biomarker tests in real-world practice. Because ECG is a widely available evaluation tool and QCG scores are derived from ECG images, the QCG-Critical score may serve as an adequate alternative biomarker for risk stratification of patients with acute HF in real-world settings with limited resources. It may also be useful even in well-equipped centres because it would be available immediately after the ECG exam, without requiring additional waiting for echocardiography or laboratory tests. This may be beneficial for timely risk stratification in the emergency department. The QCG-Critical score was not only correlated with IHCD but also showed a strong association with long-term mortality. In addition, the subgroup analysis demonstrated a consistent correlation between the QCG-Critical score and clinical outcomes. These results emphasise the potential of AI-based ECG interpretation as a novel biomarker in this field.

This study has several limitations. First, the study population predominantly consisted of Asians; hence, further studies are needed to validate our results across different ethnicities. Second, the AI algorithm tested in this study was derived from one of the participating centres (Seoul National University Bundang Hospital). However, there was a temporal difference between patient enrolment for algorithm training (2017 to 2019) and test population (KorAHF enrolment, 2011 to 2014), and another external centre (Severance Hospital) was involved in this study. Nevertheless, this may limit the generalisability of our findings. Third, the ECG format may affect the algorithm’s performance. Although the manufacturers of the ECG devices in the

two participating hospitals differed, there was no significant difference in the AI algorithm performance between the hospitals (PageWriter TC 30, and TC 70; Philips in Seoul National University Bundang Hospital, and MAC 5500, MAC VU360; GE Healthcare in Severance Hospital). However, because the system uses printed ECG images as input, there may be problematic scenarios where the qualities of the images influence the predictive power of the biomarkers. Although some recent AI algorithm-based studies suggest further interpretation analysis, the QCG system does not support gradient-weighted class activation mapping or similar visualization for model explainability due to the custom network architecture we use. Therefore, we could not evaluate which part of the ECG images the system uses for each prediction.

In conclusion, predicting outcomes in patients with acute HF using the newly developed AI-based ECG score appeared feasible. Thus, this score may serve as a novel biomarker for patients with HF, potentially overcoming the cost and availability constraints of current biomarkers.

Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00265933). This work was Supported by grant no 13-2024-0009 from the Seoul National University Bundang Hospital Research Fund.

Conflicts of Interest

The artificial intelligence algorithm used in this study is currently under development for commercial service by ARPI Inc. Professors Joonghee Kim and Youngjin Cho are employed by this company and hold dual roles as CEO and Head of Research Collaboration Center, respectively.

Multimedia appendix 1: Figure S1, Figure S2, Figure S3, Figure S4

Abbreviations

AF: atrial fibrillation

AI: artificial intelligence

AUC: area under the curve

CI: confidence intervals

ECG: electrocardiogram

HF: heart failure

HR: hazard ratios

IHCD: in-hospital cardiac death

NT-proBNP: N-terminal pro-B-type natriuretic peptide

OR: odds ratios

ROC: receiver operating characteristic

QCG: quantitative electrocardiogram

References

1. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Journal of the American College of Cardiology 2022;79(17):e263-e421. [doi:10.1016/j.jacc.2021.12.012 PMID - 35379503]
2. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. European heart journal 2021;42(36):3599-3726
3. Park JJ, Lee CJ, Park S-J, et al. Heart failure statistics in Korea, 2020: a report from the Korean Society of Heart Failure. International Journal of Heart Failure 2021;3(4):224
4. Januzzi JL, Rehman SU, Mohammed AA, et al. Use of Amino-Terminal Pro–B-Type Natriuretic Peptide to Guide Outpatient Therapy of Patients With Chronic Left Ventricular Systolic Dysfunction. Journal of the American College of Cardiology 2011;58(18):1881-1889. [doi:10.1016/j.jacc.2011.03.072 PMID - 22018299]
5. Masson S, Latini R, Anand IS, et al. Prognostic Value of Changes in N-Terminal Pro-Brain Natriuretic Peptide in Val-HeFT (Valsartan Heart Failure Trial). Journal of the American College of Cardiology 2008;52(12):997-1003. [doi:10.1016/j.jacc.2008.04.069 PMID - 18786480]
6. Peacock WFt, De Marco T, Fonarow GC, et al. Cardiac troponin and outcome in acute heart failure. N Engl J Med 2008;358(20):2117-2126. [doi:10.1056/NEJMoa0706824]
7. Verbrugge FH, Reddy YNV, Attia ZI, et al. Detection of Left Atrial Myopathy Using Artificial Intelligence-Enabled Electrocardiography. Circ Heart Fail 2022;15(1):e008176. [doi:10.1161/CIRCHEARTFAILURE.120.008176]
8. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet 2019;394(10201):861-867. [doi:10.1016/s0140-6736(19)31721-0 PMID - 31378392]

9. Lee SE, Cho H-J, Lee H-Y, et al. A multicentre cohort study of acute heart failure syndromes in Korea: rationale, design, and interim observations of the Korean Acute Heart Failure (KorAHF) registry: Interim analysis of the KorAHF registry. European Journal of Heart Failure 2014;16(6):700-708. [doi:10.1002/ejhf.91 PMID - 24797348]
10. Cho Y, Cho S-Y, Oh I-Y, et al. Implantable Cardioverter-defibrillator Utilization and Its Outcomes in Korea: Data from Korean Acute Heart Failure Registry. Journal of Korean Medical Science 2020;35(46):e397. [doi:10.3346/jkms.2020.35.e397 PMID - 33258331]
11. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. Journal of medical Internet research 2016;18(12):e323
12. Kim D, Hwang JE, Cho Y, et al. A Retrospective Clinical Evaluation of an Artificial Intelligence Screening Method for Early Detection of STEMI in the Emergency Department. Journal of Korean Medical Science 2022;37(10):e81. [doi:10.3346/jkms.2022.37.e81 PMID - 35289140]
13. Choi YJ, Park MJ, Ko Y, et al. Artificial intelligence versus physicians on interpretation of printed ECG images: Diagnostic performance of ST-elevation myocardial infarction on electrocardiography. International Journal of Cardiology 2022;363:6-10. [doi:10.1016/j.ijcard.2022.06.012 PMID - 35691440]
14. Park J, Yoon Y, Cho Y, Kim J. Feasibility of Artificial Intelligence–Based Electrocardiography Analysis for the Prediction of Obstructive Coronary Artery Disease in Patients With Stable Angina: Validation Study. JMIR cardio 2023;7(1):e44791
15. Fonarow GC, Albert NM, Curtis AB, et al. Improving evidence-based care for heart failure in outpatient cardiology practices: primary results of the Registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF). Circulation 2010;122(6):585-596. [doi:10.1161/CIRCULATIONAHA.109.934471]
16. Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 2018;18(1):44. [doi:10.1186/s12911-018-0620-z]
17. Ahmad T, Lund LH, Rao P, et al. Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to

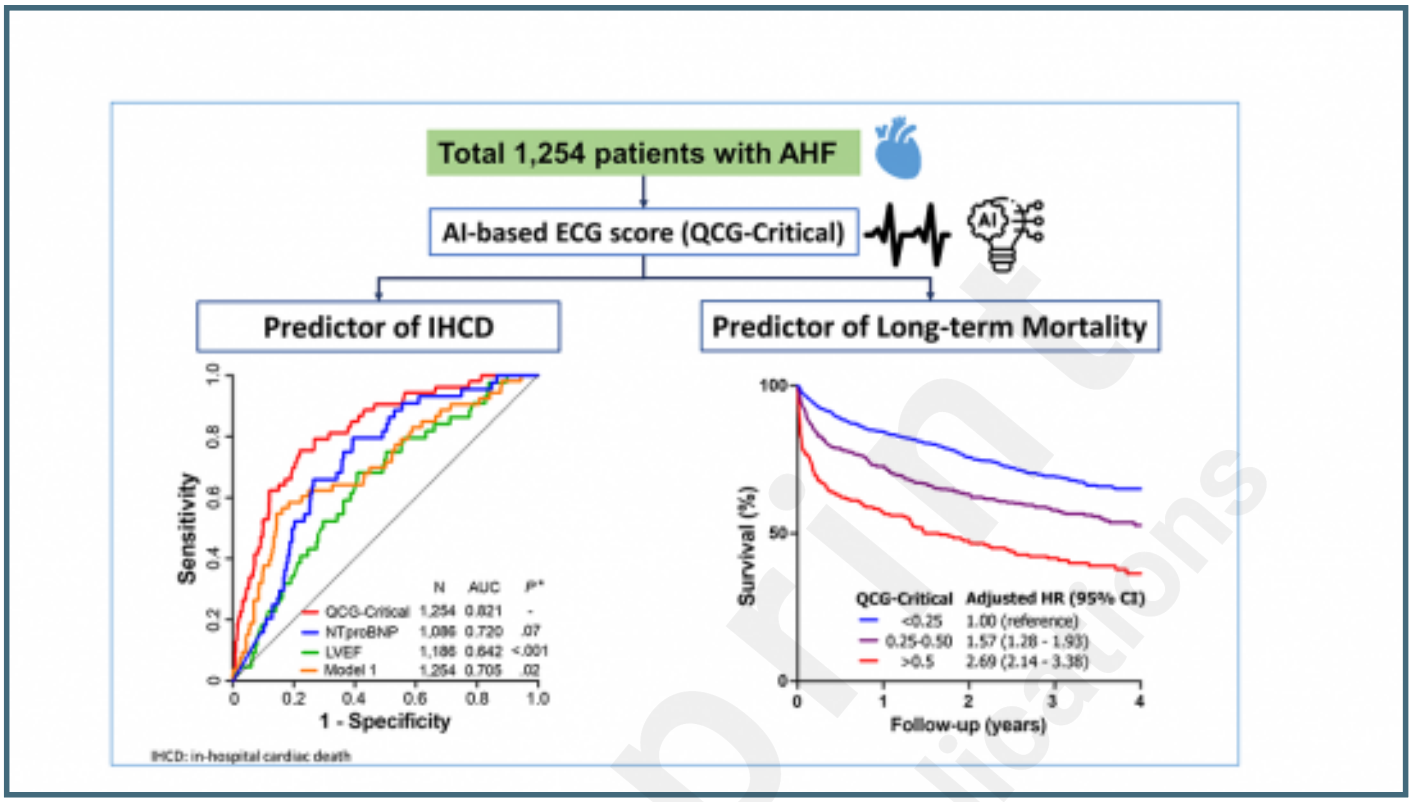
Therapy in a Large Cohort of Heart Failure Patients. J Am Heart Assoc 2018;7(8). [doi:10.1161/JAHA.117.008081]

18. Hendry PB, Krisdinarti L, Erika M. Scoring System Based on Electrocardiogram Features to Predict the Type of Heart Failure in Patients With Chronic Heart Failure. Cardiol Res 2016;7(3):110-116. [doi:10.14740/cr473w]
19. Lee JH, Park JJ, Cho Y, et al. Prognostic Implication of Ventricular Conduction Disturbance Pattern in Hospitalized Patients with Acute Heart Failure Syndrome. Korean Circ J 2019;49(7):602-611. [doi:10.4070/kcj.2018.0290]
20. Wang NC, Maggioni AP, Konstam MA, et al. Clinical implications of QRS duration in patients hospitalized with worsening heart failure and reduced left ventricular ejection fraction. JAMA 2008;299(22):2656-2666. [doi:10.1001/jama.299.22.2656]
21. Das MK, Maskoun W, Shen C, et al. Fragmented QRS on twelve-lead electrocardiogram predicts arrhythmic events in patients with ischemic and nonischemic cardiomyopathy. Heart Rhythm 2010;7(1):74-80. [doi:10.1016/j.hrthm.2009.09.065]
22. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019;25(1):70-74. [doi:10.1038/s41591-018-0240-2]
23. Choi J, Lee S, Chang M, Lee Y, Oh GC, Lee HY. Deep learning of ECG waveforms for diagnosis of heart failure with a reduced left ventricular ejection fraction. Sci Rep 2022;12(1):14235. [doi:10.1038/s41598-022-18640-8]
24. Kwon J-m, Kim K-H, Jeon K-H, et al. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. Korean circulation journal 2019;49(7):629-639

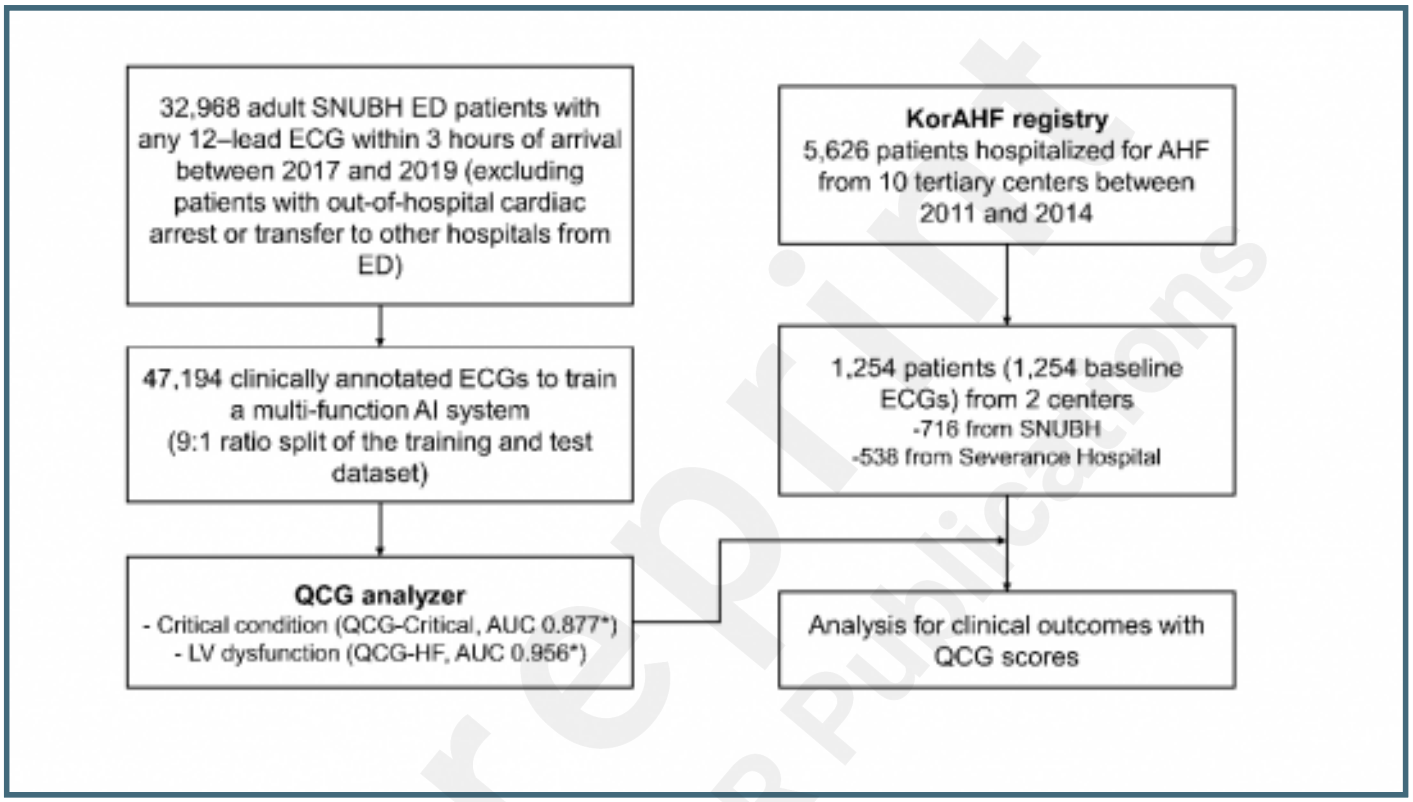
Supplementary Files

Figures

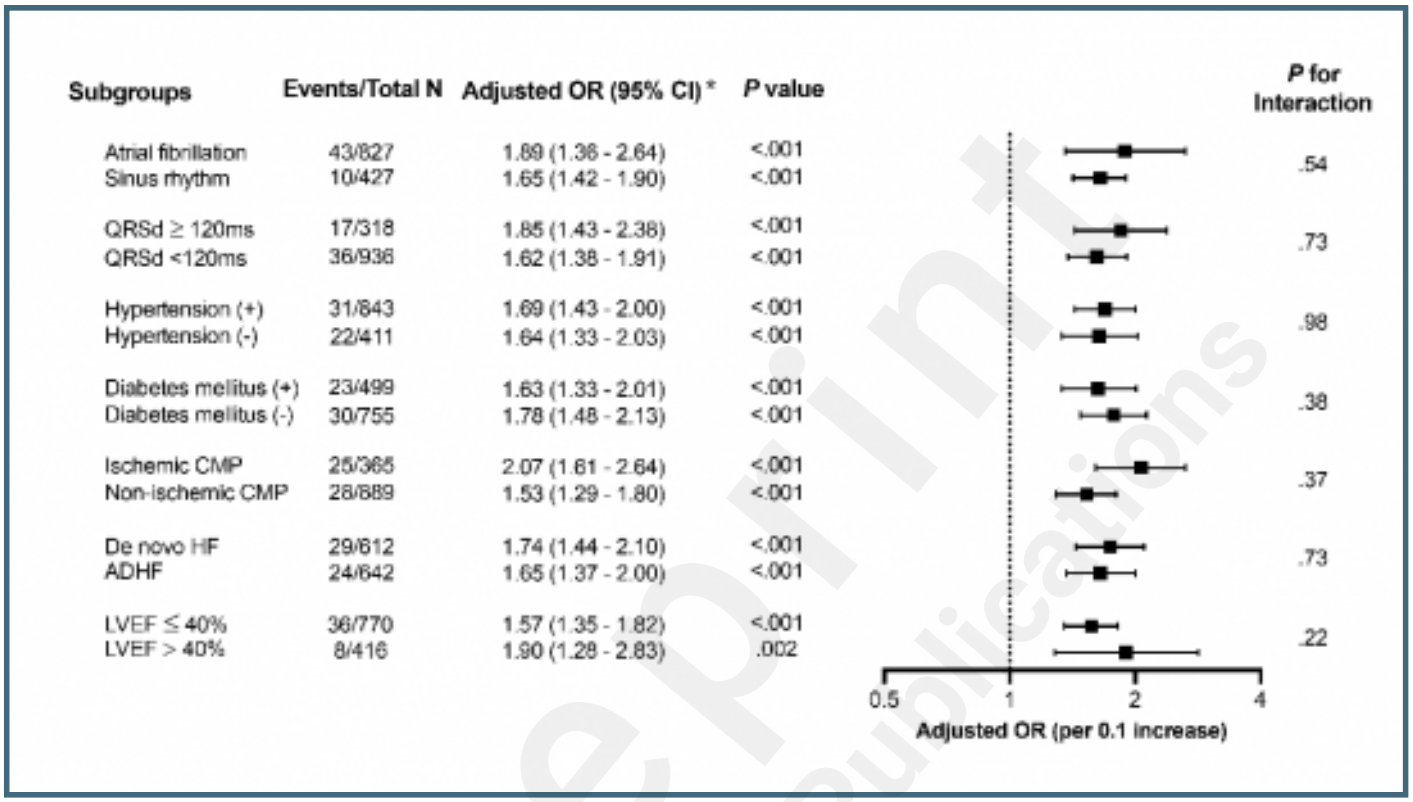
Graphical abstract. AI-based ECG score for predicting IHCD and long-term mortality *P value for comparison with the AUC of QCG-Critical score.



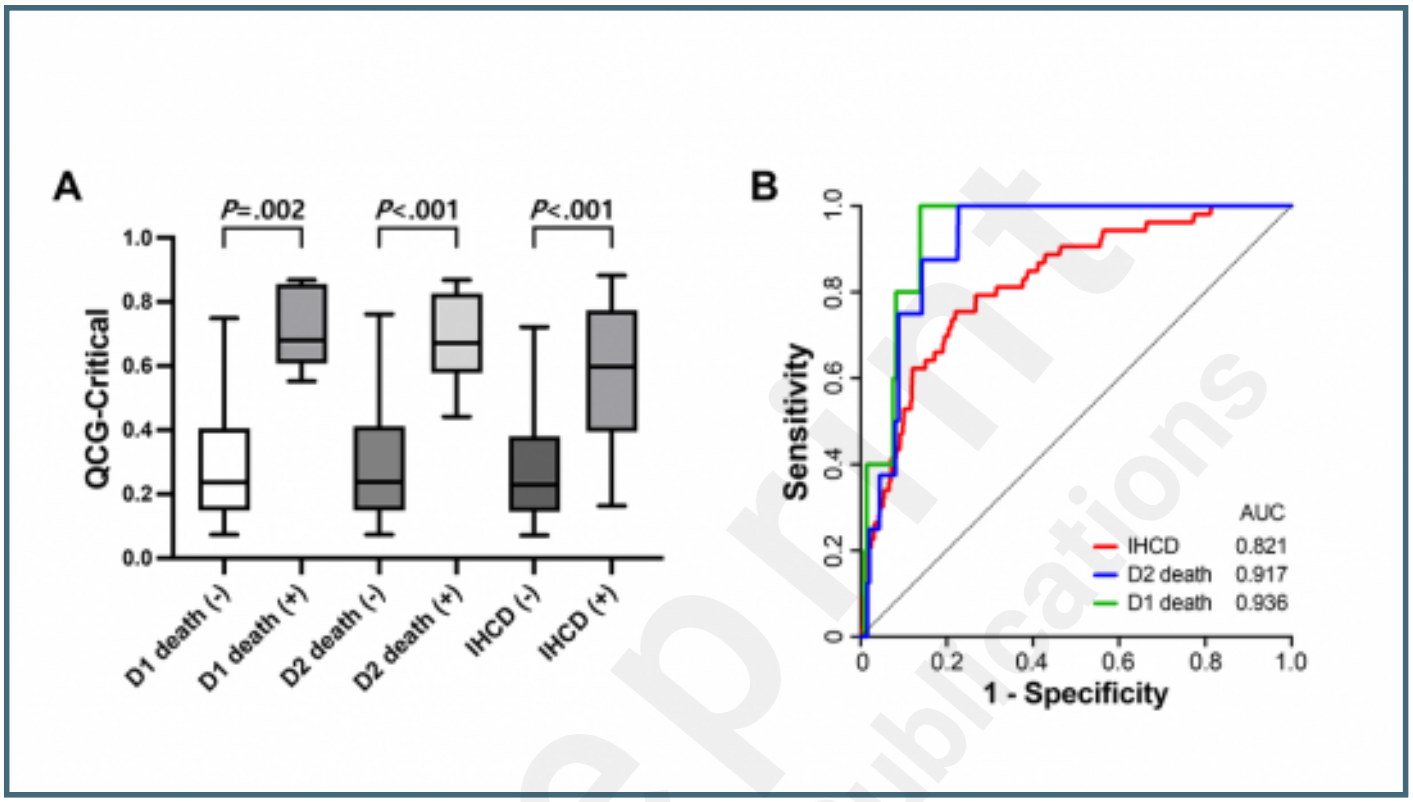
Flowchart of the training and validation study population. The AI-ECG analyzer called QCGTM was developed utilizing 47,194 annotated ECG images of over 32,968 patients who visited the emergency department of SNUBH between 2017 and 2019. The QCG analyzer was applied to ECGs from a subpopulation of KorAHF registry which enrolled AHF patients between 2011 and 2014. *Internal validation results for two QCG scores. AHF, acute heart failure; AI, artificial intelligence; AUC, area under the curve; ECG, electrocardiogram; ED, emergency department; LV, left ventricular; SNUBH, Seoul National University Bundang Hospital, QCG, quantitative ECG.



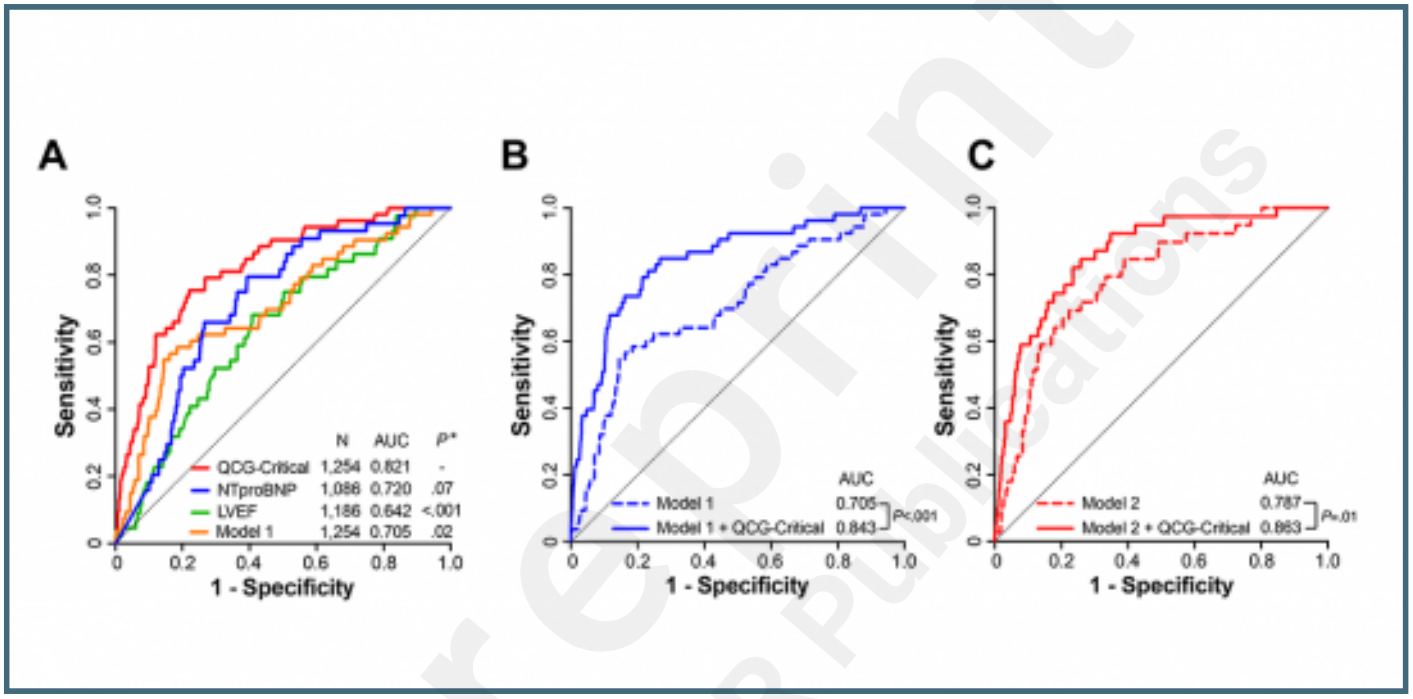
Subgroup analysis results for predicting IHCD. Adjusted ORs are presented for a 0.1 increase in the QCG-Critical score
*Adjusted for age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, atrial fibrillation, and QRS duration. ADHF, acute decompensated heart failure; CI, confidence interval; CMP, cardiomyopathy; HF, heart failure; IHCD, in-hospital cardiac death; LVEF, left ventricular ejection fraction; OR, odds ratio.



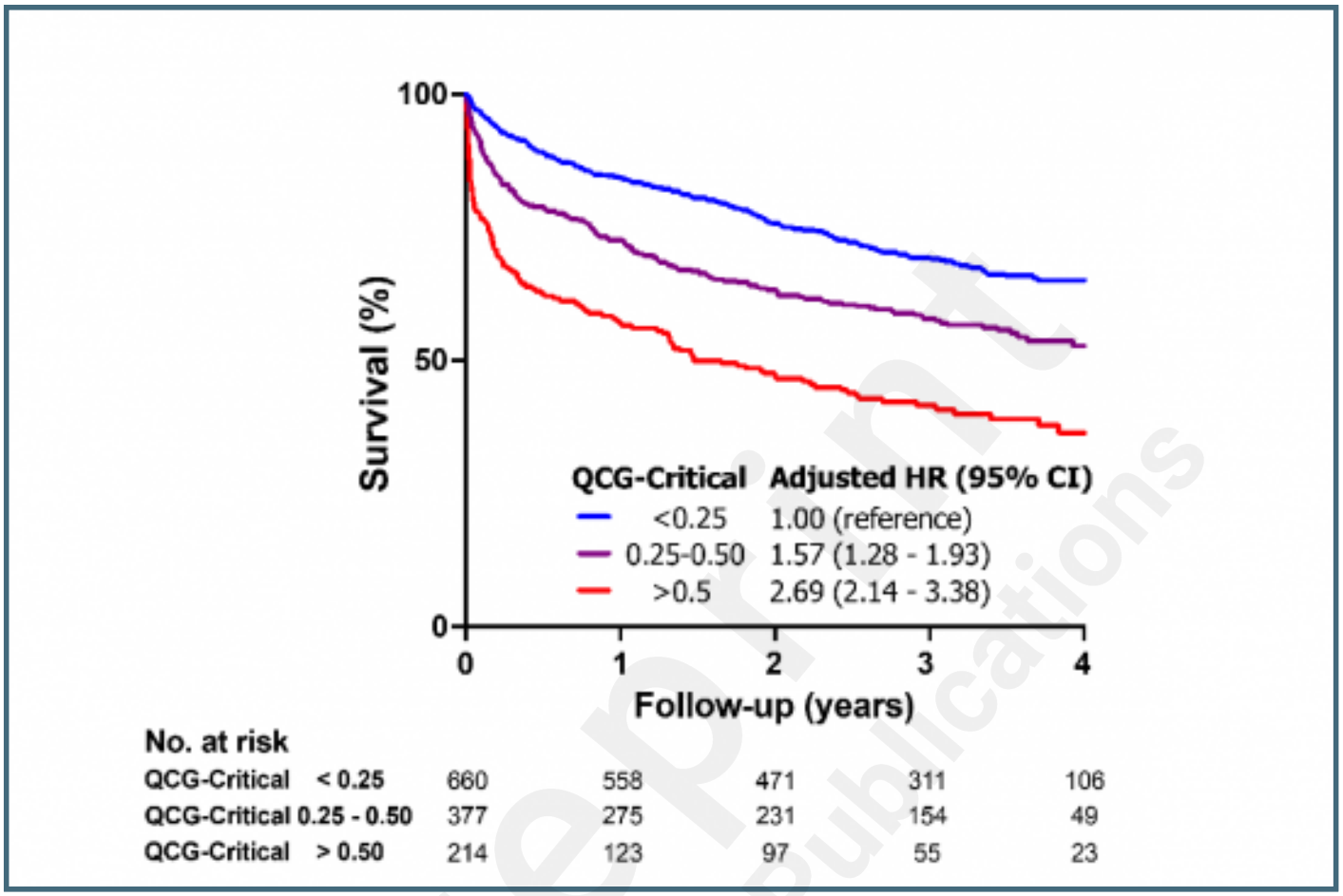
Performance of the QCG-Critical score for predicting IHCD (A) The QCG-Critical score was significantly higher in patients who experienced cardiac death within 1 day, 2 days, and during hospitalisation than in survivors. The Box-and-Whisker plot is presented as 5–95 percentiles. (B) Performance of the QCG-Critical score presented as ROC curves. AUC, area under the curve; IHCD, in-hospital cardiac death; ROC, receiver operating characteristic.



The ROC curves for predicting IHCD (A) The AUC value of the QCG-Critical score was 0.821 and tended to be higher than that of echocardiographic LVEF, NT-proBNP or a model utilizing traditional clinical variables (model1). (B) The performance of the prediction model 1 and with the addition of the QCG-Critical score. (C) Upon further incorporation of NT-proBNP and LVEF into the model 1 (model 2), the QCG-Critical score demonstrated additional predictive value for IHCD than model 2 alone. Model 1 includes age, sex, hypertension, diabetes, chronic kidney disease, cerebrovascular disease, ischemic heart disease, valvular heart disease, HF type, atrial fibrillation and QRS duration. NT-proBNP and LVEF were further incorporated into the model 2. *P value for comparison with the AUC of QCG-Critical score. HF, heart failure; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type natriuretic peptide. Other abbreviations as Figure 3.



Kaplan–Meier curves for long-term mortality according to the QCG-Critical scores CI, confidence interval; HR, hazard ratio.



Multimedia Appendixes

Supplementary material.
URL: http://asset.jmir.pub/assets/b6ab532469d5a312a21147835f9ce11c.docx