BMJ Open

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicentre registry in China

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023724
Article Type:	Protocol
Date Submitted by the Author:	24-Apr-2018

Complete List of Authors:	Meng, Fanqi; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Xiamen Cardiovascular Hospital, Xiamen University, Department of Cardiology Zhang, Zhihua; The First Affiliated Hospital of Nanjing Medical University,, Department of Cardiology; Jiangning Hospital Affiliated to Nanjing Medical University, Department of Cardiology Hou, Xiaofeng; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Qian, Zhiyong; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Wang, Yao; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Chen, Yanhong; Wuhan Asia Heart Hospital, Department of Cardiology Wang, Yilian; The Second People's Hospital of Lianyungang, Department of Cardiology, Zhou, Ye; The Affiliated Hospital of Jiangsu University, Department of Cardiology Chen, Zhen; Taixing People's Hospital, Department of Cardiology Zhang, Xiwen; The First People's Hospital of Huaian, Department of Cardiology Yang, Jing; The First People's Hospital of Huaian, Department of Cardiology Zhang, Jinlong; The First People's Hospital of Yancheng, Department of Cardiology Guo, Jianghong; Rugao People's Hospital, Department of Cardiology Li, Kebei; The First People's Hospital of Zhangjiagang, Department of Cardiology Chen, Lu; The Third People's Hospital of Suzhou, Department of Cardiology Zhuang, Ruijuan; The Third People's Hospital of Wuxi, Department of Cardiology Jiang, Hai; The Second Affiliated Hospital of Nanjing Medical University, Department of Cardiology Zhou, Weihua; University of Southern Mississippi, School of Computing Tang, Shaowen; Nanjing Medical University, Department of Epidemiology Wei, Yongyue; Nanjing Medical University, Department of Biostatistics Zou, Jiangang; the First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Nanjing Medical University, Key Laboratory of Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine

Keywords:	Heart failure < CARDIOLOGY, sudden cardiac death, machine learning, risk model

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicentre registry in China

Fanqi Meng,[1,2] Zhihua Zhang, [1,3] Xiaofeng Hou,[1] Zhiyong Qian,[1] Yao Wang,[1] Yanhong Chen, [4] Yilian Wang,[5] Ye Zhou,[6] Zhen Chen,[7] Xiwen Zhang,[8] Jing Yang, [8] Jinlong Zhang,[9] Jianghong Guo,[10] Kebei Li, [11] Lu Chen,[12] Ruijuan Zhuang,[13] Hai Jiang,[14] Weihua Zhou, [15] Shaowen Tang, [16] Yongyue Wei, [17] Jiangang Zou[1, 18*]

Author affiliations

[1]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

[2]Department of Cardiology, Xiamen Cardiovascular Hospital, Xiamen University, Xiamen, China

[3]Department of Cardiology, Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing, China

[4]Department of Cardiology, Wuhan Asia Heart Hospital, Wuhan, China

[5]Department of Cardiology, The Second People's Hospital of Lianyungang, Lianyungang, China

[6]Department of Cardiology, The Affiliated Hospital of Jiangsu University, Zhenjiang, China

[7]Department of Cardiology, Taixing People's Hospital, Taixing, China

[8]Department of Cardiology, The First People's Hospital of Huaian, Huaian, China

[9]Department of Cardiology, The First People's Hospital of Yancheng, Yancheng, China

[10]Department of Cardiology, Rugao People's Hospital, Rugao, China

[11]Department of Cardiology, The First People's Hospital of Zhangjiagang, Zhangjiagang, China

[12]Department of Cardiology, The Third People's Hospital of Suzhou, Suzhou, China

[13]Department of Cardiology, The Third People's Hospital of Wuxi, Wuxi, China

[14]Department of Cardiology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, China

[15]School of Computing, University of Southern Mississippi, Hattiesburg, USA.

[16]Department of Epidemiology, Nanjing Medical University, Nanjing, China

[17]Department of Biostatistics, Nanjing Medical University, Nanjing, China

[18]Key Laboratory of Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine, Nanjing Medical University, China

***Correspondence to**

Dr Jiangang Zou**:** Department of Cardiology, the First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, 210029, P.R.China; E-mail: jgzou@njmu.edu.cn; Tel. phone: +86 13605191407

Keywords: heart failure, sudden cardiac death, machine learning, risk model

Number of words:

ABSTRACT

Introduction

Left ventricular ejection fraction (LVEF) ≤35%, as current major implantable cardioverter-defibrillator (ICD) indication for primary prevention of sudden cardiac death (SCD) in heart failure (HF) patients has been widely recognized to be inefficient. More precisely selecting patients with low LVEF (≤35%) is needed to optimize deployment of ICD. Most of the existing prediction models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Compared to traditional statistical analysis, machine learning (ML) can employ computer algorithms to identify patterns in large datasets, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions. This study is aimed to develop and validate new models using ML to improve the prediction of SCD in HF patients with low LVEF.

Methods and analysis

We will conduct a retro-prospective, multicentre, observational registry of Chinese HF patients with low LVEF. The HF patients with LVEF ≤35% after optimized medication at least 3 months will be enrolled in this trial. The primary end points are all-cause death and SCD. The secondary end points are malignant arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF. The baseline demographic, clinical, biological, electrophysiological, social and psychological variables will be collected. ML techniques will be used to develop a SCD prediction model. As compared to traditional multivariable COX regression model derived from the same database, the performance of ML model will be evaluated.

Ethics and dissemination

The study protocol has been approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All results will be published in international peer-reviewed journals and presented at relevant conferences.

Clinical trial registration ChiCTR-POC-17011842.

INTRODUCTION

Heart failure (HF), with increased prevalence in Asia as well as western countries, has become a major public health problem. The prevalence of HF in Asia is 1.2%-6.7% depending on the population studied.[1] In China, there are 4.2 million HF patients, and 500,000 new cases are being diagnosed each year.[1] Although the survival rate after HF diagnosis has been increased due to obvious improvement in medical therapy, the mortality of HF remains high and around 50% of people diagnosed with HF will die within 5 years.[2] Sudden cardiac death (SCD), as a result of cardiac arrest, mainly due to lethal arrhythmias like ventricular tachycardia or ventricular fibrillation, is responsible for over 50% of all HF deaths.[3]

The most effective strategy for prevention of SCD in patients with HF, is the implantable cardioverter-defibrillator (ICD), associated with 50% relative risk reduction in arrhythmia-related death in secondary prevention,[4] and 54% relative risk reduction in primary prevention.[5] At present, left ventricular ejection fraction (LVEF) ≤35% is the major ICD indication for primary prevention of SCD.[6] However, real-world data show that only 3% to 5% of ICD patients for primary prevention with LVEF<30% to 35% receive life-saving therapies on an annual basis,[7] whereas majority of SCD victims have LVEF>35%.[8, 9] Furthermore, obvious decline in rate of SCD for HF patients with reduced LVEF, which was consistent with the cumulative benefit of optimizing medication including angiotensin converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB), beta-blocker, and mineralocorticoid receptor antagonist (MRAs), has been observed.[10] Therefore, the utility of this criterion has dwindled. New strategies of identifying high risk HF patients most likely to benefit from primary prevention ICD is urgently needed. Because of higher risk SCD in patients with LVEF≤ 35% than with LVEF>35%,[11] identifying patients with low EF at high risk of SCD will be more efficient and economically important.

Over the last decade, lots of multivariate prognostic models derived for chronic HF patients have been proposed (Table 1).[12-22] However， these models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Firstly, chronic heart failure both with reduced LVEF (HFrEF) and with

preserved LVEF (HFpEF) were represented in all models except I-PRESERVE score, which included only HFpEF. To the best of our knowledge, there is no specific trial for the prognosis of low LVEF population, which is extremely important. Additionally, although all the scores are "not parsimonious", some critical factors have not been incorporated into all prognostic models, for example, medications are contained in only 3 out of 11 scores,[15, 19, 21] and optimized medication was not required as inclusion criteria in all 11 studies. Furthermore, the major limitation of most prognostic models is inability to predict SCD risk. Electrical risk factors are not involved in all these models. Although currently some non-invasive factors, including mechanical dyssynchrony measured by echocardiography, myocardial fibrosis detected with cardiovascular magnetic resonance, and cardiac autonomic dysfunction assessed by 123-metaiodobenzylguanidine scintigraphy, have been evaluated to predict SCD in HF patients, [23] it is difficult to widely use them to predict SCD in large HF population. Resting 12-lead ECG and Holter, as the longest surviving, broadly available, easily deployed and inexpensive tests, can provide a measure of cumulative electrical risk, combination of which may significantly improve the SCD risk prediction beyond EF.[24]

Table 1. Risk model for HF in the literature

Author	Database	Year	Variables (n)	Patients (n)	End points
Agostoni	MECKI	2012	6	2716	Cardiovascular death; urgent cardiac transplant
Barlera	GISSI-HF	2013	14	6975	all-cause mortality
Collier	EMPHASIS-HF	2013	10	2737	all-cause mortality
Komajda	I-PRESERVE	2011	12	4128	all-cause mortality
Levy	SHFM	2006	14	1125	Survival
O'Connor	HF-ACTION	2012	4	2331	all-cause mortality
Pocock	CHARM	2006	21	7599	all-cause mortality
Pocock	MAGGIC	2012	13	39372	all-cause mortality
Senni	CVM-HF	2006	13	292	all-cause mortality
Senni	3C-HF	2013	11	2016	all-cause mortality; urgent heart transplant (1year)
Vazquez	MUSIC	2009	10	992	all-cause mortality; cardiac mortality; pump failure death, sudden death

Based on above reasons, the novel risk assessment tools should meet the following

requirements: (1) the risk model should be developed from special low LVEF population (≤35%) to accelerate its clinical application and promote the innovation of ICD indications for primary prevention. (2) The narrow focus on the LVEF as main predictor of SCD risk should be broadened, and more cardiac and non-cardiac factors should be included. (3) Electrical risk factors, which are broadly available, easily deployed and inexpensive, should be considered as candidate predictors to evaluate the risk of sudden arrhythmic death. (4) Although sometimes it is not easy to determine the cause of death, SCD as the primary end point should be defined whenever possible.

Data processing is the next key step to develop prognostic models. This study involves non-linear prediction models, a large number of patients and numerous predictors among which there may be complicated correlations. Traditional hypothesis-driven statistical analysis is difficult to overcome these challenges. The machine learning (ML) approaches have great potential to improve the solution. They employ computer algorithms to identify patterns in large datasets with a large number of variables, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions or decisions. Weng et al. [25] found that ML significantly improved accuracy of cardiovascular risk prediction, increased the number of patients who could benefit from preventive treatment and avoided unnecessary treatment. However, ML has not been reported to be applied to SCD risk prediction based on large HF population. Therefore, application of ML for prediction of SCD in HF patients with low LVEF will be innovative and clinically significant.

AIMS

The purpose of our study is to develop and validate new models to improve the prediction of SCD in HF patients with low LVEF. The new strategies of identifying HF patients most likely to benefit from primary prevention ICD, will impulse the revolution of ICD indications. The specific research objectives are:

1. To develop prediction models to evaluate prognosis and SCD risk respectively by ML methods and traditional COX regression in HF patients with low LVEF

(≤35%).

2. To validate these prediction models in a prospective cohort and evaluate the performance of ML models, as compared to multivariable COX regression model derived from the same database.

METHODS AND ANALYSIS

Study design

This study is a retro-prospective, multicentre, non-interventional, observational clinical registry trial. The primary sponsor is The First Affiliated Hospital of Nanjing Medical University. This trial has been registered in the Chinese Clinical Trial Registry (ChiCTR-POC-17011842). The trial will be conducted across 14 cardiovascular departments in Tertiary A hospitals throughout the People's Republic of China:

The First Affiliated Hospital of Nanjing Medical University, Nanjing

Xiamen Cardiovascular Hospital, Xiamen University, Xiamen

Wuhan Asia Heart Hospital, Wuhan

Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing

The Second People's Hospital of Lianyungang, Lianyungang

The Affiliated Hospital of Jiangsu University, Zhenjiang

Taixing People's Hospital, Taixing

The First People's Hospital of Huaian, Huaian

The First People's Hospital of Yancheng, Yancheng

Rugao People's Hospital, Rugao

The First People's Hospital of Zhangjiagang, Zhangjiagang

The Third People's Hospital of Suzhou, Suzhou

The Third People's Hospital of Wuxi, Wuxi

The Second Affiliated Hospital of Nanjing Medical University, Nanjing,

The data of the First Affiliated Hospital of Nanjing Medical University and Xiamen Cardiovascular Hospital Xiamen University will be retrospectively collected from

January 2016 to December 2017. About 500 retrospective cases meet the inclusion criteria according to preliminary estimation. The prospective recruitment has started in the above 14 hospitals since January 2018. The retrospective cases and the first 1000 prospective cases will be used to develop the prediction models. And the next 1000 prospective cases will be used for model validation. The flow diagram of the progress is illustrated in Figure 1.

Figure 1 Flow diagram of progress

Inclusion criteria

To participate in this trial, patients must comply with all of the following

1. Diagnosis of HFrEF according to the 2016 ESC HF Guideline[6]

2. LVEF $\leqslant$ 35% (measured by Simpson's methods) after optimized medication including ACEI or ARB, beta-blocker and MRA if available and not contraindicated at least 3 months.

3. Signed informed consent.

Exclusion criteria

1. Hypertrophic cardiomyopathy

2. Rheumatic heart disease

3. Congenital heart disease

4. Pulmonary heart disease

5. Pericardial diseases and myocarditis

6. Acute myocardial infarction in recent 3 months, including STEMI and NSTEMI

7. Aortic dissection

8. Severe hematologic disease including leukemia, lymphoma, aplastic anemia, etc.

9. Autoimmune disease

10. Malignant tumor

11. Hormone replacement

12. Application of other interventional clinical trial

13. Non-drug therapies for improving heart function: CRT-P/D, ICD, heart

transplantation, surgical resection of ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation

Endpoints

Primary end point

All-cause death and SCD, including cardiac death and death from other causes

Secondary end point

Lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation, rehospitalization due to HF

Recruitment and consent

Participants will be identified and recruited at each of the participating centers. The clinical status of potential participants will be assessed, and their medical records will also be reviewed to confirm the eligibility according to inclusion and exclusion criteria.

The study details will be explained to all potentially eligible and interested subjects. The patients who agree to attend this study will sign the informed consent form (ICF) indicating that they fully understand the study and their rights of confidentiality and withdrawal from the study without giving a reason.

Baseline evaluation

Prognostic models of HF in the last 10 years have been reviewed and the involved risk factors have been ranked according to their corresponding hazard ratio in respective risk model (Table 1, Figure 2). Age, sex, New York Heart Association (NYHA) class, LVEF, prior HF hospitalization, course of HF, severe valvular heart disease, atrial fibrillation, prior myocardial infarction / coronary artery bypass grafting (CABG), renal dysfunction, chronic obstructive pulmonary disease (COPD), diabetes mellitus, ischemic etiology, decreased systolic pressure, low body mass index (BMI),

anemia, hyponatremia, high N-terminal pro brain natriuretic peptide (NT-proBNP), uricemia, current smoker were included. Variables which were not listed in previous models but appear relevant to higher risk of SCD in HF patients, and would therefore merit consideration, include syncope or pre-syncope, frequent premature ventricular beat, non-sustained ventricular tachycardia, complete left bundle branch block (CLBBB), long QT interval, increased QT dispersion. In addition, self-care ability, social support and psychological state including depression and anxiety, are also predictors for subsequent poor prognosis in HF patients. The above risk factors have been assessed and confirmed by an expert panel of cardiologists and statistician, and will be collected in this study particularly.

Figure 2 Hazard ratio of variables in different risk models

The baseline data of all eligible subjects will be collected as following.

- Demographic characteristics: date of birth, gender, height and weight
- Life style behavior: smoking and drinking status
- Vital signs: blood pressure and heart rate
- NYHA class
- Etiology of heart failure: The ischemic etiology will be confirmed if any following point is met: a. prior myocardial infarction or revascularization history (coronary artery bypass grafting, CABG / percutaneous coronary intervention, PCI); b. left main or proximal segment of left anterior descending artery stenosis ≥75% showed by coronary angiogram (CAG); c. at least two main coronary artery branches stenosis ≥75% showed by CAG. Otherwise non-ischemic HF should be identified.
- Prior HF hospitalization history: first HF hospitalization or not, times of prior HF hospitalization, course of HF (since the HF symptoms appear, if unavailable, since the decreased EF was found).
- Coronary heart disease history: myocardial infarction or angina history, CAG result, revascularization history, recent angina.
- Arrhythmia history: atrial fibrillation, atrial flutter, premature atrial contraction (PAC), premature ventricular contraction (PVC), non-sustained VT (NSVT),

sustained VT (SVT), ventricular fibrillation, and some bradyarrhythmias.

- Syncope or pre-syncope history
- Cardiac arrest/ cardiopulmonary resuscitation history
- Other history: hypertension, diabetes mellitus, COPD, etc.
- Echocardiography: LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV) and LVEF measured by Simpson's method; Left atrial diameter, LV end-diastolic diameter (LVEDD) and LV end-systolic diameter (LVESD), pulmonary artery systolic pressure. The status of valve regurgitation will be evaluated (0-none; 1-mild; 2-mild to moderate; 3-moderate; 4- severe)
- ECG: Left / right bundle branch block will be recorded. QRS duration and QT interval will be tested and QT dispersion will be calculated.
- Holter: total heart beat of the whole day, minimum/ maximum/ average HR, onset of PVC, PAC, NSVT, VT, atrial fibrillation/ flutter.
- Laboratory tests results: serum creatinine, blood urea nitrogen (BUN), serum natrium, hemoglobin, Thyroid Stimulating Hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), NT-proBNP.
- medication: ACEI/ARB、beta-blocker, aldosterone antagonist, diuretic, digoxin, antiplatelet agent, anticoagulant, statin, calcium channel blocker, antiarrhythmics, Ivabradine and angiotensin receptor blocker-neprilysin inhibitor (ARNI).
- Evaluation of self-care behavior and social support: 9-item European Heart Failure Self-care Behavior Scale (9-EHFScBS） [26] will be used to determine the self-care levels in HF patients. Social Support Rating Scale (SSRS)[27] will be used to evaluate the social support condition in HF patients.
- Assessment of psychological status: Hamilton Depression Scale (HAMD) and Hamilton Anxiety Scale (HAMA).

Patient visits

After enrolled in this research, all the subjects, will be followed-up periodically in outpatient department or by telephone interview every 3 months. As primary end point, all-cause death and SCD will be focused. Cause of death will be analyzed in

detail. SCD is defined by the World Health Organization as unexpected death that occurs within 1 hour from the onset of new or worsening symptoms (witnessed arrest) or, if unwitnessed, within 24h from when the individual was last observed alive and asymptomatic[28]. The lethal arrhythmia including VT/VF, sudden cardiac arrest, cardiopulmonary resuscitation, and rehospitalization due to HF will be recorded carefully.

During follow-up, lethal arrhythmia will be recognized more precisely for patients who receive ICD or CRT/D implantation, and will be recorded as adverse event. The patients, who receive CRT-P/D, heart transplantation, surgical resection of ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation, or some other non-drug therapy to improve heart function, will be followed up as usual.

Data collection

In prospective part, clinical data of subjects will be collected and filled in the electrical data capture (EDC) system at baseline and particular follow-up visit. In retrospective part, the same baseline information, except for 9-EHFScBS, SSRS, HAMD, HAMA questionnaires, will also be captured and input into the EDC system. The following prospective visits (every three months) will conducted regularly and will be recorded in the EDC system. Investigator will record all the information of adverse event (AE), study bias, withdraw from the study or death in EDC system. In this study, the participants will be identified by study codes, and their names will not appear in EDC system. All the personal information including contact information, medical record and outcome, will not be revealed to any person who has not been authorized by principle investigator. Professional staffs are responsible for database management, data maintenance and regular data backup. Data quality will be monitored regularly. The data collection checklist is showed in Table 2.

Table 2 The checklist for data collection

Data collection	Baseline		regular visit	withdraw /death
	Retrospective cases	Prospective cases		
Informed consent	√	√		
Quantification verification	√	√		

(inclusion and exclusion)				
Baseline evaluation	√	√		
Medication	√	√		
Questionnaires 9-EHFScBS SSRS HAMD HAMA		√		
Regular follow-up visit (every 3 months)			√	
Survival state	√		√	√
Adverse event	Once happen √			
Study bias	Once happen √			
Withdraw from the study	Once happen √			
Death	Once happen √			

Statistical analysis

Data classification and pre-processing

All above collected variables, which might be predictor of all adverse prognosis of HF described in end point events, will be classified as uncontrollable variables (e.g. age, gender, history), controllable variables associated with heart (e.g. NYHA class, LVEF, increased heart rate) and controllable variables beyond heart (e.g. smoking, anemia, diabetes mellitus). Appropriate dummy variables will be used for binary variables and categorical variables, and quantitative variables will be fitted as a single continuous measurement (e.g. age, heart rate, NT-proBNP), unless there is clear evidence of nonlinearity. In order to create possible simple risk score, some continuous variables will also be categorized into several groups according to both usual clinical cutpoints and expert advice.

Machine learning

Variable selection is the process of selecting a subset of relevant variables for use in model construction, which can substantially reduce the abundant information and decrease the number of variables that are input to the prediction model. In this study, the technique named as “information gain ranking” will be used to select appropriate variables. Information gain represents the effectiveness of a variable based on entropy,

which characterizes the unpredictability of a system. The information gain of a variable is evaluated as the entropy difference of the system when including and excluding this variable. Then the variables whose information gain scores are less than a threshold are considered to be insignificant and will be excluded in the prediction.

Predictive classifiers for SCD prediction in HF patients will be developed by the following classification algorithms respectively: decision tree, logistic regression, support vector machine, artificial neural network. [29] The performance and general error estimation of these ML models will be assessed by 10-fold cross-validation. The dataset will be randomly divided into 10 equal folds. 9 folds will be used as training set with the remaining one fold as validation set. The validation results from 10 repeats will be combined to provide a measure of the overall performance. The prediction models derived from the four classification algorithms above will be evaluated based on the accuracy and interpretability. Finally, clinical experts and computer specialists will discuss and choose the best model to predict the prognosis and SCD in HF patients and then perform the further validation with the prospective dataset.

COX proportional-hazards regression

Univariable COX proportional--hazards modeling will be used to identify strong independent baseline candidate predictors for the primary outcome and secondary outcome. We will use both forward and backward stepwise procedure to derive the multivariable COX proportional--hazards model with $p<0.05$ as the inclusion criterion. Every variable in the model will be multiplied by its β-coefficient and the products will be summed to calculate the risk score. Risk function will be used to estimate the level of risk. The calculating formula is as follows.

$$P = h(t_j; X_k) = h_0(t_j) \exp(SCORE)$$

$$SCORE = X_k \beta_k = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_p x_p$$

External validation

The dynamic prospective cases will be used for external validation of the optimal ML and COX models. The validation will be performed using the models to calculate the

probability of the outcome of interest occurring for each individual included in the validation sample, when compared with the events actually observed to occur in this sample. The discrimination of each model will be estimated by receiver-operating characteristic (ROC) curve. The calibration of the models will be assessed by Hosmer-Lemeshow goodness-of-fit test. The ML prediction model will be compared with the COX regression model .

Patient and public involvement statement

During the design of this study, a survey of patient requirements, including communication needs, follow-up frequency, visit cost etc., was conducted in population of potential HF participants, which provided important evidence for drawing up this study protocol to meet most of the patients' needs, build close contact with patients, enhance the overall adherence and improve the accuracy of endpoint event. This study is not the patient-led research, patients are not involved in the recruitment to and conduct of the study. The participants will be informed of the study results by phone at the end of this study. The alive patients will be evaluated with the new prediction model, and the ICD intervention will be recommended to the high SCD risk patients.

Study time frame

The retrospective data collection in the two sub-centers started in March 2017, and prospective enrollment in all 14 sub-centers has started in January 2018. The follow-up period is scheduled to end in December 2019. The major part of data analysis will be performed from January to March 2020. The study framework and process is summarized in Figure 3.

Figure 3 Study framework and process

ETHICS AND DISSEMINATION

The study protocol has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All necessary information

about this study will be disclosed to the patients. Every subject will be asked to sign the ICF, indicating that they fully understand the study and voluntarily participate in this study. All results of this study will be published in international peer-reviewed journals and presented at relevant conferences.

STRENGTHS AND LIMITATIONS

This is the first retro-prospective, multicentre registry study in China, aimed to investigate the feasibility and accuracy of applying ML to predict SCD in HF patients with low LVEF. Except for SCD, a broad range of outcomes, including all-cause death, lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF, will be evaluated in this study, and the corresponding prediction models will also be developed, if available. The traditional multivariable COX regression model will be derived from the same database and will be compared to machine learning. This project has great promise to improve ICD patient selection.

The limitations of this study are as follows: 1. The SCD prediction of HF patients with LVEF>35% will not be evaluated based on the design of this study. 2. Sometimes, It might be difficult to determine the end point of this study for some patients, when dealing with SCD, lethal arrhythmia and sudden cardiac arrest, especially outside the hospital.

CONCLUSION

This study is aimed to investigate the feasibility and accuracy of machine learning to predict SCD in HF patients with low LVEF. By the completion of this study, it is expected to derive and validate the new prediction models. Our study has promise to improve selection of ICD candidates for primary prevention in HF patients.

FOOTNOTES

Acknowledgements The authors thank Weihua Zhou, PhD of the University of Southern Mississippi for providing machine learning techniques; Rongbin Yu,

Shaowen Tang and Yongyue Wei, PhD of Nanjing Medical University for providing advice regarding study design and statistical analysis; Xiamen Cardiovascular Hospital, Xiamen University (Xiamen, China), Wuhan Asia Heart Hospital (Wuhan, China), Jiangning Hospital Affiliated to Nanjing Medical University, (Nanjing, China), The Second People's Hospital of Lianyungang (Lianyungang, China), The Affiliated Hospital of Jiangsu University (Zhenjiang, China), Taixing People's Hospital (Taixing, China), The First People's Hospital of Huaian (Huaian, China), The First People's Hospital of Yancheng (Yancheng, China), Rugao People's Hospital (Rugao, China), The First People's Hospital of Zhangjiagang (Zhangjiagang, China), The Third People's Hospital of Suzhou (Suzhou, China), The Third People's Hospital of Wuxi (Wuxi, China), The Second Affiliated Hospital of Nanjing Medical University (Nanjing, China) for collaboration including recruitment and follow-up of HF patients. The authors also thank the HF patients who participated in the survey of patient requirements during the design of this study.

Contributors JGZ and FQM conceived and designed the study. ZHZ, XFH, ZYQ, YW, YHC, YLW, YZ, ZC, XWZ, JY, JLZ, JHG, KBL, LC, RJZ, HJ participated in different phases of the protocol design. WHZ provide expertise on data processing and machine learning. SWT and YYW provided their expertise for traditional statistical analysis. JGZ obtained funding. FQM drafted the final manuscript. All authors have read the manuscript and provided feedback. JGZ approved the final version of the manuscript prior to submission. FQM took responsibility for the submission process.

Funding This study was supported mainly by Jiangsu Province's Key Medical Center (grant number: YXZXA2016010; PI: Jiangang Zou, MD, PhD); in part by National Natural Science Foundation of China (grant number: 81470457; PI: Jiangang Zou, MD, PhD) and Jiangsu Science and Technology Department (grant number: BE2016764; PI: Jiangang Zou, MD, PhD). This research was also supported in part by an AHA grant (grant number: 17AIREA33700016; PI: Weihua Zhou, PhD) and a New Faculty startup grant from the University of Southern Mississippi (grant number: DE01791; PI: Weihua Zhou, PhD).

Competing interests We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Ethics approval This study has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCE

[1] Sato N. Epidemiology of Heart Failure in Asia[J]. Heart Fail Clin, 2015,11(4):573-579.

[2] Mozaffarian D, Benjamin E J, Go A S, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association[J]. Circulation, 2015,131(4):e29-e322.

[3] Tomaselli G F, Zipes D P. What causes sudden death in heart failure?[J]. Circ Res, 2004,95(8):754-763.

[4] Connolly S J, Hallstrom A P, Cappato R, et al. Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials. AVID, CASH and CIDS studies. Antiarrhythmics vs Implantable Defibrillator study. Cardiac Arrest Study Hamburg. Canadian Implantable Defibrillator Study[J]. Eur Heart J, 2000,21(24):2071-2078.

[5] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[6] Ponikowski P, Voors A A, Anker S D, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC[J]. Eur Heart J, 2016,37(27):2129-2200.

[7] Kober L, Thune J J, Nielsen J C, et al. Defibrillator Implantation in Patients with Nonischemic Systolic Heart Failure[J]. N Engl J Med, 2016,375(13):1221-1230.

[8] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[9] Stecker E C, Vickers C, Waltz J, et al. Population-based analysis of sudden cardiac death with and without left ventricular systolic dysfunction: two-year findings from the Oregon Sudden Unexpected Death Study[J]. J Am Coll Cardiol, 2006,47(6):1161-1166.

[10] Shen L, Jhund P S, Petrie M C, et al. Declining Risk of Sudden Death in Heart Failure[J]. N Engl J Med, 2017,377(1):41-51.

[11] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[12] Agostoni P, Corra U, Cattadori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis[J]. Int J

Cardiol, 2013,167(6):2710-2718.

[13] Barlera S, Tavazzi L, Franzosi M G, et al. Predictors of mortality in 6975 patients with chronic heart failure in the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure trial: proposal for a nomogram[J]. Circ Heart Fail, 2013,6(1):31-39.

[14] Collier T J, Pocock S J, McMurray J J, et al. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms: insight from a novel risk score for prognosis derived from the EMPHASIS-HF trial[J]. Eur Heart J, 2013,34(36):2823-2829.

[15] Komajda M, Carson P E, Hetzel S, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE)[J]. Circ Heart Fail, 2011,4(1):27-35.

[16] Levy W C, Mozaffarian D, Linker D T, et al. The Seattle Heart Failure Model: prediction of survival in heart failure[J]. Circulation, 2006,113(11):1424-1433.

[17] O'Connor C M, Whellan D J, Wojdyla D, et al. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction: the HF-ACTION predictive risk score model[J]. Circ Heart Fail, 2012,5(1):63-71.

[18] Pocock S J, Wang D, Pfeffer M A, et al. Predictors of mortality and morbidity in patients with chronic heart failure[J]. Eur Heart J, 2006,27(1):65-75.

[19] Pocock S J, Ariti C A, McMurray J J, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies[J]. Eur Heart J, 2013,34(19):1404-1413.

[20] Senni M, Santilli G, Parrella P, et al. A novel prognostic index to determine the impact of cardiac conditions and co-morbidities on one-year outcome in patients with heart failure[J]. Am J Cardiol, 2006,98(8):1076-1082.

[21] Senni M, Parrella P, De Maria R, et al. Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score[J]. Int J Cardiol, 2013,163(2):206-211.

[22] Vazquez R, Bayes-Genis A, Cygankiewicz I, et al. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure[J]. Eur Heart J, 2009,30(9):1088-1096.

[23] Halliday B P, Cleland J, Goldberger J J, et al. Personalizing Risk Stratification for Sudden Death in Dilated Cardiomyopathy: The Past, Present, and Future[J]. Circulation, 2017,136(2):215-231.

[24] Aro A L, Reinier K, Rusinaru C, et al. Electrical risk score beyond the left ventricular ejection fraction: prediction of sudden cardiac death in the Oregon Sudden Unexpected Death Study and the Atherosclerosis Risk in Communities Study[J]. Eur Heart J, 2017,38(40):3017-3025.

[25] Weng S F, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data?[J]. PLoS One, 2017,12(4):e174944.

[26] Jaarsma T, Arestedt K F, Martensson J, et al. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument[J]. Eur J Heart Fail, 2009,11(1):99-105.

[27] Hu X, Hu X, Su Y, et al. The changes and factors associated with post-discharge self-care behaviors among Chinese patients with heart failure[J]. Patient Prefer Adherence, 2015,9:1593-1601.

[28] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[29] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. Springer, New York; 2013.

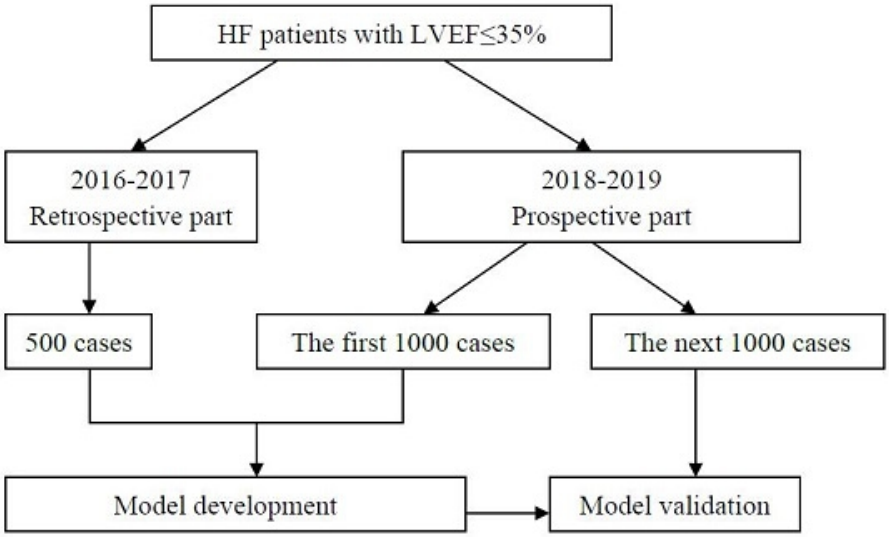

Figure 1 Flow diagram of progress

158x97mm (96 x 96 DPI)

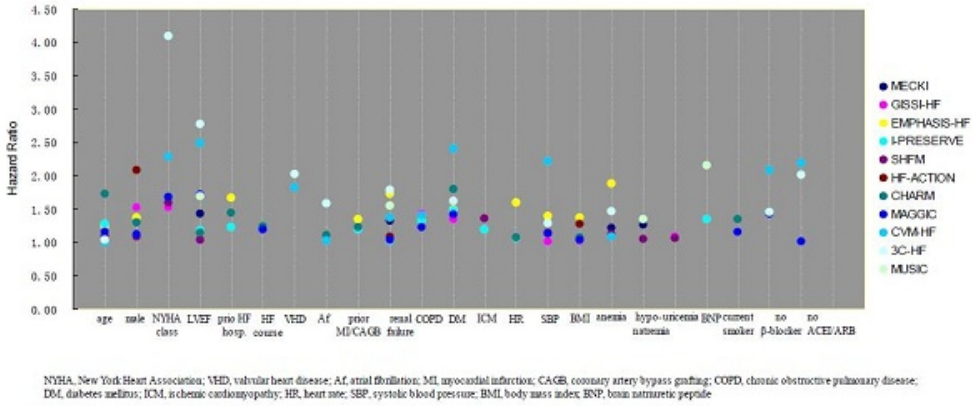

Figure 2 Hazard ratio of variables in different risk models

158x78mm (96 x 96 DPI)

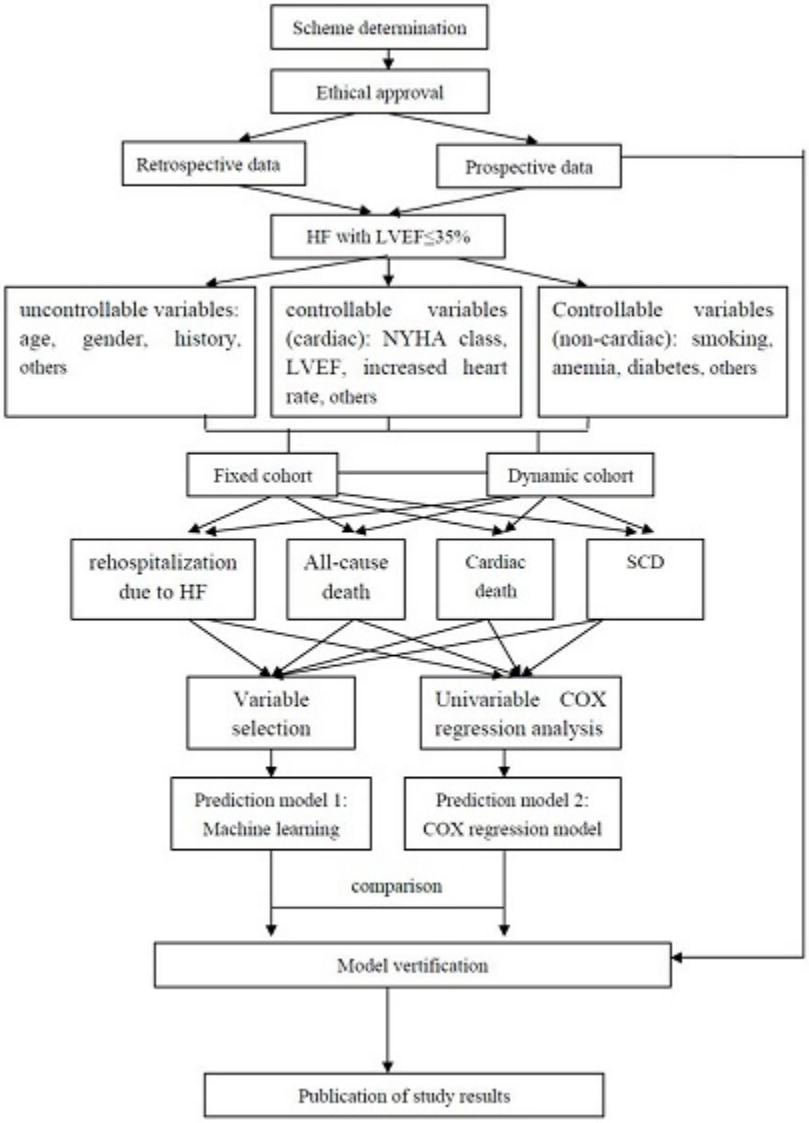

Figure 3 Study framework and process

76x101mm (150 x 150 DPI)

SPIRIT

STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym **(P1)**
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry **(P6)**
	2b	All items from the World Health Organization Trial Registration Data Set **(P6)**
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support **(P16)**
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors **(P1, P15-16)**
	5b	Name and contact information for the trial sponsor **(P1, P6)**
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities **(P6, P15)**
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) **(P6-7, P11)**
Introduction		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention **(P3-P5)**
	6b	Explanation for choice of comparators **(P5)**
Objectives	7	Specific objectives or hypotheses **(P5-6)**

Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) **(P6-7)**
Methods: Participants, interventions, and outcomes		
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained **(P6)**
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) **(P7-P8)**
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered **(Not applicable)**
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) **(Not applicable)**
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) **(Not applicable)**
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial **(Not applicable)**
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended **(P8)**
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended **(see Figure1)**
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations **(P6-7)**
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size **(P8)**
Methods: Assignment of interventions (for controlled trials)		
Allocation:		

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions **(Not applicable)**
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned **(Not applicable)**
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions **(Not applicable)**
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how **(Not applicable)**
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial **(Not applicable)**
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol **(P8-11)**
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols **(P11-12)**
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol **(P11)**
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol **(P12-P14)**
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses) **(Not applicable)**
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) **(Not applicable)**

Methods: Monitoring

Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct **(P11)**
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval **(P14)**
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) **(Not applicable)**
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) **(P8)**
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable **(Not applicable)**
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial **(P11)**
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site **(P16)**
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators **(P11)**
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation **(Not applicable)**

Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions **(P14)**
	31b	Authorship eligibility guidelines and any intended use of professional writers **(Not applicable)**
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code **(Not applicable)**
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates **(see ICF)**
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable **(Not applicable)**

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicenter registry in China

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023724.R1
Article Type:	Protocol
Date Submitted by the Author:	25-Oct-2018
Complete List of Authors:	Meng, Fanqi; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Xiamen Cardiovascular Hospital, Xiamen University, Department of Cardiology Zhang, Zhihua; The First Affiliated Hospital of Nanjing Medical University,, Department of Cardiology; Jiangning Hospital Affiliated to Nanjing Medical University, Department of Cardiology Hou, Xiaofeng; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Qian, Zhiyong; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Wang, Yao; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Chen, Yanhong; Wuhan Asia Heart Hospital, Department of Cardiology Wang, Yilian; The Second People's Hospital of Lianyungang, Department of Cardiology, Zhou, Ye; The Affiliated Hospital of Jiangsu University, Department of Cardiology Chen, Zhen; Taixing People's Hospital, Department of Cardiology Zhang, Xiwen; The First People's Hospital of Huaian, Department of Cardiology Yang, Jing; The First People's Hospital of Huaian, Department of Cardiology Zhang, Jinlong; The First People's Hospital of Yancheng, Department of Cardiology Guo, Jianghong; Rugao People's Hospital, Department of Cardiology Li, Kebei; The First People's Hospital of Zhangjiagang, Department of Cardiology Chen, Lu; The Third People's Hospital of Suzhou, Department of Cardiology Zhuang, Ruijuan; The Third People's Hospital of Wuxi, Department of Cardiology Jiang, Hai; The Second Affiliated Hospital of Nanjing Medical University, Department of Cardiology Zhou, Weihua; University of Southern Mississippi, School of Computing Tang, Shaowen; Nanjing Medical University, Department of Epidemiology Wei, Yongyue; Nanjing Medical University, Department of Biostatistics Zou, Jiangang; the First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Nanjing Medical University, Key Laboratory of

	Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine
Primary Subject Heading:	Medical management
Secondary Subject Heading:	Cardiovascular medicine
Keywords:	Heart failure < CARDIOLOGY, sudden cardiac death, machine learning, risk model

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicenter registry in China

Fanqi Meng,[1,2] Zhihua Zhang, [1,3] Xiaofeng Hou,[1] Zhiyong Qian,[1] Yao Wang,[1] Yanhong Chen, [4] Yilian Wang,[5] Ye Zhou,[6] Zhen Chen,[7] Xiwen Zhang,[8] Jing Yang, [8] Jinlong Zhang,[9] Jianghong Guo,[10] Kebei Li, [11] Lu Chen,[12] Ruijuan Zhuang,[13] Hai Jiang,[14] Weihua Zhou, [15] Shaowen Tang, [16] Yongyue Wei, [17] Jiangang Zou[1, 18*]

Author affiliations

[1]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

[2]Department of Cardiology, Xiamen Cardiovascular Hospital, Xiamen University, Xiamen, China

[3]Department of Cardiology, Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing, China

[4]Department of Cardiology, Wuhan Asia Heart Hospital, Wuhan, China

[5]Department of Cardiology, The Second People's Hospital of Lianyungang, Lianyungang, China

[6]Department of Cardiology, The Affiliated Hospital of Jiangsu University, Zhenjiang, China

[7]Department of Cardiology, Taixing People's Hospital, Taixing, China

[8]Department of Cardiology, The First People's Hospital of Huaian, Huaian, China

[9]Department of Cardiology, The First People's Hospital of Yancheng, Yancheng, China

[10]Department of Cardiology, Rugao People's Hospital, Rugao, China

[11]Department of Cardiology, The First People's Hospital of Zhangjiagang, Zhangjiagang, China

[12]Department of Cardiology, The Third People's Hospital of Suzhou, Suzhou, China

[13]Department of Cardiology, The Third People's Hospital of Wuxi, Wuxi, China

[14]Department of Cardiology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, China

[15]School of Computing, University of Southern Mississippi, Hattiesburg, USA.

[16]Department of Epidemiology, Nanjing Medical University, Nanjing, China

[17]Department of Biostatistics, Nanjing Medical University, Nanjing, China

[18]Key Laboratory of Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine, Nanjing Medical University, China

***Correspondence to**

Dr Jiangang Zou**:** Department of Cardiology, the First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, 210029, P.R.China; E-mail: jgzou@njmu.edu.cn; Tel. phone: +86 13605191407

Keywords: heart failure, sudden cardiac death, machine learning, risk model

ABSTRACT

Introduction

Left ventricular ejection fraction (LVEF) ≤35%, as current significant implantable cardioverter-defibrillator (ICD) indication for primary prevention of sudden cardiac death (SCD) in heart failure (HF) patients, has been widely recognized to be inefficient. Improvement of patient selection for low LVEF (≤35%) is needed to optimize deployment of ICD. Most of the existing prediction models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Compared to traditional statistical analysis, machine learning (ML) can employ computer algorithms to identify patterns in large datasets, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions. This study is aimed to develop and validate new models using ML to improve the prediction of SCD in HF patients with low LVEF.

Methods and analysis

We will conduct a retro-prospective, multi-center, observational registry of Chinese HF patients with low LVEF. The HF patients with LVEF ≤35% after optimized medication at least 3 months will be enrolled in this study. The primary endpoints are all-cause death and SCD. The secondary endpoints are malignant arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF. The baseline demographic, clinical, biological, electrophysiological, social and psychological variables will be collected. Both ML and traditional multivariable COX proportional hazards regression models will be developed and compared in the prediction of SCD. Moreover, the ML model will be validated in a prospective study.

Ethics and dissemination

The study protocol has been approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All results of this study will be

published in international peer-reviewed journals and presented at relevant conferences.

Clinical trial registration ChiCTR-POC-17011842.

Strengths and limitations of this study

- This study is the first multicenter registry study in China, aimed to investigate the feasibility and accuracy of applying ML to predict SCD in HF patients with low LVEF.
- A broad range of outcomes, including SCD, all-cause death, lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF, will be evaluated in this study, and the corresponding prognostic models will be developed.
- Machine learning and the traditional multivariable COX proportional hazards regression model will be derived from the same database and be compared.
- HF patients with LVEF>35% will not be included based on the design of this study, which will restrict the application of the results of this study to the HF with low LVEF,
- It might be difficult to determine the endpoint of this study sometimes for some patients, when dealing with SCD, lethal arrhythmia and sudden cardiac arrest, especially when outside the hospital.

INTRODUCTION

Heart failure (HF), has become a major public health problem with increased prevalence in both Asia and Western countries. The prevalence of HF in Asia is 1.2%-6.7% depending on the population studied.[1] In China, there are 4.2 million HF patients, and 500,000 new cases are being diagnosed each year.[1] Although the survival rate after HF diagnosis has been increased due to improvement in medical therapy, the mortality of HF remains high. Around 50% of people diagnosed with HF will die within 5 years.[2] The two most common causes of death in patients with HF are sudden cardiac death (SCD) and progressive pump failure. SCD in HF patients is usually caused by lethal arrhythmias such as ventricular tachycardia or ventricular fibrillation, and is reported to be responsible for nearly 50% of all cardiovascular death in HF patients.[3, 4]

The most effective strategy for prevention of SCD in patients with HF is the implantable cardioverter-defibrillator (ICD), associated with 54% relative risk reduction in primary prevention[5], and 50% relative risk reduction in arrhythmia-related death in secondary prevention.[6] There is a higher risk of SCD in patients with LVEF≤ 35% than with LVEF>35%.[7] At present, left ventricular ejection fraction (LVEF) ≤35% is the major ICD indication for primary prevention of SCD.[8] However, real-world data show that only 3% to 5% of ICD patients for primary prevention with LVEF≤35% receive shock therapies on an annual basis,[9] whereas some SCD victims have LVEF>35%.[10, 11] Identifying the patients who will be most likely to benefit from primary prevention ICD, is urgently needed. Based on the latest literature, LVEF≤35% is still an independent predictor of all-cause and cardiovascular mortality in chronic systolic HF, and displays a

better combination of sensitivity and specificity than 40% cut-off.[12] Finding ways to evaluate the SCD risk in patients with lower EF will be more efficient and economically significant. Furthermore, a noticeable decline in the rate of SCD for HF patients with reduced LVEF has been observed, which was consistent with the cumulative benefit of optimizing medication including angiotensin converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB), beta-blocker, and mineralocorticoid receptor antagonist (MRAs).[13] Therefore, it is imperative to update the criterion for ICD implantation

Over the last decade, lots of multivariate prognostic models derived for chronic HF patients have been proposed (Table 1).[14-25] However, these models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Most above prognostic scores were developed form trial databases, and the subjects included various types of heart failure. There is no specific study for the prognosis of low LVEF population. Additionally, although all the scores are "not parsimonious", some critical factors are not incorporated into the prognostic models, for example, medications are contained in I-PRESERVE[17], MAGGIC[21] and 3C-HF[23]. Optimized medication was not required as inclusion criteria in all 12 studies. Furthermore, the most above prognostic models are not able to predict SCD risk. In recent years, the advances in strain echocardiography[26, 27], cardiac magnetic resonance[26, 27] and cardiac radionuclide imaging[28, 29] have provided essential insights into the mechanisms of ventricular arrhythmias, and have been recommended to predict the SCD in patients with HF. Although these new methods are effective and noninvasive, the widespread use in large HF population to predict SCD is difficult, due to high equipment and technical requirements. Resting 12-lead ECG and Holter, as the longest surviving, broadly available, quickly deployed and inexpensive tests, can provide a measure of cumulative electrical risk, which may be combined with other factors to improve the SCD risk prediction.[30]

Based on above reasons, the novel risk assessment tools should meet the following requirements: (1) the risk model should be developed from the population with low LVEF (≤35%) to accelerate its clinical application and promote the accuracy of ICD indications for primary prevention. (2) More cardiac and non-cardiac factors beyond LVEF should be included. (3) Electrical risk factors should be included as candidate predictors to evaluate the risk of sudden arrhythmic death. (4) Although sometimes it is not easy to determine the cause of death, SCD as the primary endpoint should be defined whenever possible.

Data processing is the crucial step to develop the prognostic models. This study involves non-linear prediction models, a large number of patients and numerous predictors with complicated correlations. Traditional hypothesis-driven statistical analysis is difficult to overcome these challenges. The machine learning (ML) approaches have great potential to improve the solution. They employ computer algorithms to identify patterns in large datasets with a large number of variables, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions or decisions.[31] Weng et al. [32] found that ML significantly improved the accuracy of cardiovascular risk prediction, increased the number of patients who could benefit from preventive treatment and avoided unnecessary treatment. Recent studies have shown that the application of machine learning techniques may have the potential to improve heart failure outcomes and management, including cost savings by improving existing diagnostic and treatment support systems.[33] However, ML has not been reported to be applied to SCD risk prediction based on large HF population. Therefore, the application of ML for the prediction of SCD in HF patients with low LVEF is technically innovative and clinically significant.

AIMS

The purpose of our study is to develop and validate new models to improve the prediction

of SCD in HF patients with low LVEF. The new strategies of identifying HF patients most likely to benefit from primary prevention ICD will improve the revolution of ICD indications. The specific research objective is to develop prediction models to evaluate prognosis and SCD risk respectively by ML methods and traditional COX proportional hazard regression in HF patients with low LVEF (≤35%).

METHODS AND ANALYSIS

Study design

This study is a retro-prospective, multi-center, non-interventional, observational clinical registry. The primary sponsor is The First Affiliated Hospital of Nanjing Medical University. This study has been registered in the Chinese Clinical Trial Registry (ChiCTR-POC-17011842). The study will be conducted across 14 cardiovascular departments in Tertiary A hospitals throughout the People's Republic of China. (see Supplement)

The cases from January 2016 to December 2017 in the First Affiliated Hospital of Nanjing Medical University and Xiamen Cardiovascular Hospital Xiamen University will be collected retrospectively and followed-up prospectively. About 500 retrospective cases meet the inclusion criteria according to preliminary estimation. The prospective recruitment has started in the above 14 hospitals since January 2018. The retrospective cases and the first 1000 prospective cases will be used to develop the prediction models. And the next 1000 prospective cases will be used for model validation. The flow diagram of the progress is illustrated in Figure 1.

Inclusion criteria

To participate in this study, patients must comply with all of the following

1. Diagnosis of HFrEF according to the 2016 ESC HF Guideline[8]

2. LVEF≤35% (measured by Simpson's methods) after optimized medication including

ACEI or ARB, beta-blocker and MRA if available and not contraindicated at least 3 months.

3. Signed informed consent.

Exclusion criteria

The patient with any of the following will be excluded:

1. Hypertrophic cardiomyopathy
2. Rheumatic heart disease
3. Congenital heart disease
4. Pulmonary heart disease
5. Pericardial diseases and myocarditis
6. Acute myocardial infarction in recent 3 months, including STEMI and NSTEMI
7. Aortic dissection
8. Severe hematologic disease including leukemia, lymphoma, aplastic anemia, etc.
9. Autoimmune disease
10. Malignant tumor
11. Hormone replacement
12. Application of other interventional clinical trials
13. Non-drug therapies for improving heart function: CRT-P/D, ICD, heart transplantation, surgical resection of ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation

Endpoints

Primary endpoint

All-cause death and SCD, including cardiac death and death from other causes

Secondary endpoint

Lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation, rehospitalization due to HF

Recruitment and consent

Participants will be identified and recruited at each of the participating centers. The clinical status of potential participants will be assessed, and their medical records will also be reviewed to confirm the eligibility according to the inclusion and exclusion criteria.

The study details will be explained to all potentially eligible and interesting subjects. The patients who agree to attend this study will sign the informed consent form (ICF) indicating that they fully understand the study and their rights of confidentiality and withdrawal from the study without giving a reason.

Baseline evaluation

Prognostic models of HF in the last 10 years have been reviewed, and the associated risk factors have been ranked according to their corresponding hazard ratio in respective risk models (Table 1, Figure 2). Age, sex, New York Heart Association (NYHA) class, LVEF, prior HF hospitalization, course of HF, severe valvular heart disease, atrial fibrillation, prior myocardial infarction / coronary artery bypass grafting (CABG), renal dysfunction, chronic obstructive pulmonary disease (COPD), diabetes mellitus, ischemic etiology, decreased systolic pressure, low body mass index (BMI), anemia, hyponatremia, high N-terminal pro-brain natriuretic peptide (NT-proBNP), uricemia, current smoker were included. Variables which were not listed in previous models but appear relevant to higher risk of SCD in HF patients, and would therefore, merit consideration, including syncope or pre-syncope, frequent premature ventricular beat, non-sustained ventricular tachycardia, complete left bundle branch block (CLBBB), long QT interval, increased QT dispersion. In addition, self-care ability, social support and psychological state including

depression and anxiety, are also predictors for subsequent poor prognosis in HF patients. The above risk factors have been assessed and confirmed by an expert panel of cardiologists and statisticians and will be collected in this study particularly.

The baseline data that will be collected in all eligible subjects are as follows.

- Demographic characteristics: date of birth, gender, height and weight
- Lifestyle behavior: smoking and drinking status
- Vital signs: blood pressure and heart rate
- NYHA class
- Etiology of heart failure: The ischemic etiology will be confirmed if any following point is met: a. prior myocardial infarction or revascularization history (coronary artery bypass grafting, CABG / percutaneous coronary intervention, PCI); b. left main or proximal segment of the left anterior descending artery stenosis ≥75% showed by coronary angiogram (CAG); c. at least two main coronary artery branches stenosis ≥75% showed by CAG. Otherwise, non-ischemic HF should be identified.
- Prior HF hospitalization history: first HF hospitalization or not, times of prior HF hospitalization, the course of HF (since the HF symptoms appear; if unavailable, since the decreased EF was found).
- Coronary heart disease history: myocardial infarction or angina history, CAG result, revascularization history, recent angina.
- Arrhythmia history: atrial fibrillation, atrial flutter, premature atrial contraction (PAC), premature ventricular contraction (PVC), non-sustained VT (NSVT), sustained VT (SVT), ventricular fibrillation, and some bradyarrhythmias.
- Syncope or pre-syncope history
- Cardiac arrest/ cardiopulmonary resuscitation history
- Other histories: hypertension, diabetes mellitus, COPD, etc.
- Echocardiography: LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV) and LVEF measured by Simpson's method; Left atrial diameter, LV

end-diastolic diameter (LVEDD) and LV end-systolic diameter (LVESD), pulmonary artery systolic pressure. The status of valve regurgitation will be evaluated (0-none; 1-mild; 2-mild to moderate; 3-moderate; 4- severe)

- ECG: Left / right bundle branch block will be recorded. QRS duration and QT interval will be tested, and QT dispersion will be calculated.
- Holter: total heartbeat of the whole day, minimum/ maximum/ average HR, onset of PVC, PAC, NSVT, VT, atrial fibrillation/ flutter.
- Laboratory tests results: serum creatinine, blood urea nitrogen (BUN), serum natrium, hemoglobin, Thyroid Stimulating Hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), NT-proBNP.
- Medication: ACEI/ARB、beta-blocker, aldosterone antagonist, diuretic, digoxin, antiplatelet agent, anticoagulant, statin, calcium channel blocker, antiarrhythmics, Ivabradine and angiotensin receptor blocker-neprilysin inhibitor (ARNI).
- Evaluation of self-care behavior and social support: 9-item European Heart Failure Self-care Behavior Scale (9-EHFScBS）[34] will be used to determine the self-care levels in HF patients. Social Support Rating Scale (SSRS)[35] will be used to evaluate the social support condition in HF patients.
- Assessment of psychological status: Hamilton Depression Scale (HAMD) and Hamilton Anxiety Scale (HAMA).
- Socioeconomic and educational status: marital status, educational status, monthly income, sources of medical expenses, medical insurance

Patient visits

After being enrolled in this research, all the subjects will be followed-up periodically in the outpatient department or by telephone interview every 3 months. The compliance with medications will be evaluated. As the primary endpoint, all-cause death and SCD will be focused. Cause of death will be analyzed in detail. SCD is defined by the World

Health Organization as unexpected death that occurs within 1 hour from the onset of new or worsening symptoms (witnessed arrest) or, if unwitnessed, within 24h from when the individual was last observed alive and asymptomatic[36]. The lethal arrhythmia including VT/VF, sudden cardiac arrest, cardiopulmonary resuscitation, and rehospitalization due to HF will be recorded carefully.

During follow-up, lethal arrhythmia will be recognized more precisely for patients who receive ICD or CRT/D implantation, and will be recorded as an adverse event. The patients, who receive CRT-P/D, heart transplantation, surgical resection of a ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation, or some other non-drug therapy to improve heart function, will be followed up as usual.

Data collection

In the prospective part, clinical data of subjects will be collected and filled in the electrical data capture (EDC) system at baseline and particular follow-up visit. In the retrospective part, the same baseline information, except for 9-EHFScBS, SSRS, HAMD, HAMA questionnaires, will also be captured and input into the EDC system. The following prospective visits (every three months) will be conducted regularly and will be recorded in the EDC system. Investigators will record all the information of adverse events (AE), study bias, withdrawal from the study or death in EDC system. In this study, the participants will be identified by study codes, and their names will not appear in the EDC system. All the personal information including contact information, medical record and outcome, will not be revealed to any person who has not been authorized by a principal investigator. Professional staffs are responsible for database management, data maintenance and regular data backup. Data quality will be monitored regularly. The data collection checklist is showed in Table 2.

Data pre-processing

All above-collected variables, which might be predictors of all adverse prognosis of HF described in endpoint events, will be classified as uncontrollable variables (e.g. age, gender, history), controllable variables associated with heart (e.g. NYHA class, LVEF, increased heart rate) and controllable variables beyond heart (e.g. smoking, anemia, diabetes mellitus). Appropriate dummy variables will be used for binary variables and categorical variables, and quantitative variables will be fitted as a single continuous measurement (e.g. age, heart rate, NT-proBNP), unless there is clear evidence of nonlinearity. In order to create a practice simple risk score, some continuous variables will also be categorized into several groups according to both common clinical cut points and expert advice.

Machine learning

Variable selection is the process of selecting a subset of relevant variables for use in model construction, which can substantially reduce the abundant information and decrease the number of variables that are input to the prediction model. In this study, the technique named as "information gain ranking" will be used to select appropriate variables. Information gain represents the effectiveness of a variable based on entropy, which characterizes the unpredictability of a system. The information gain of a variable is evaluated as the entropy difference of the system when including and excluding this variable. Then the variables whose information gain scores are less than a threshold are considered to be insignificant and will be excluded from the prediction.

Prediction models for SCD in HF patients will be developed by the following classification algorithms respectively: decision trees, logistic regression, support vector machine, random forest, and artificial neural network. [29] The performance and general error estimation of these ML models will be assessed by 10-fold cross-validation. The dataset will be randomly divided into 10 equal folds. 9 folds will be used as the training

set with the remaining one fold as the validation set. The validation results from 10 repeats will be combined to provide a measure of the overall performance. The prediction models derived from the above classification algorithms above will be evaluated based on the accuracy and interpretability. Finally, clinical experts and computer specialists will discuss and choose the best model to predict the prognosis of SCD in HF patients and then perform further validation with the prospective dataset.

COX proportional hazards regression

Univariable COX proportional hazards modeling will be used to identify strong independent baseline candidate predictors for the primary and secondary outcomes. We will use both forward and backward stepwise procedure to derive the multivariable COX proportional hazards model with $p<0.05$ as the inclusion criterion. Every variable in the model will be multiplied by its β -coefficient, and the products will be summed to calculate the risk score. Risk function will be used to estimate the level of risk. The calculating formula is as follows.[37]

$$P = h(t_j; X_k) = h_0(t_j) \exp(SCORE)$$

$$SCORE = X_k \beta_k = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_p x_p$$

Model validation

The dynamic prospective cases will be used for external validation of the optimal ML and COX proportional hazards models. The validation will be performed using the models to calculate the probability of the outcome of interest occurring for each individual included in the validation sample when compared with the events actually observed to occur in this sample. The discrimination of each model will be estimated by receiver-operating characteristic (ROC) curve. The calibration of the models will be assessed by the Hosmer-Lemeshow goodness-of-fit test. The ML prediction model will be compared with the COX proportional hazards regression model.

Patient and public involvement statement

During the design of this study, a survey of patient requirements, including communication needs, follow-up frequency, visit cost etc., was conducted in population of potential HF participants, which provided important evidence for drawing up this study protocol to meet most of the patients' needs, build close contact with patients, enhance the overall adherence and improve the accuracy of endpoint event. This study is not a patient-led research, and patients are not involved in the recruitment of the study. The study results will be informed to the participants by phone at the end of this study. The alive patients will be evaluated with the new prediction model, and the ICD intervention will be recommended to the high SCD risk patients.

Study timeframe

The retrospective data collection in the two sub-centers started in March 2017, and prospective enrollment in all 14 sub-centers has started in January 2018. The follow-up period is scheduled to end in December 2019. The major part of data analysis will be performed from January to March 2020. The study framework and process is summarized in Figure 3.

ETHICS AND DISSEMINATION

The study protocol has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All necessary information about this study will be disclosed to the patients. Every subject will be asked to sign the ICF, indicating that they fully understand the study and voluntarily participate in this study. All results of this study will be published in international peer-reviewed journals and presented at relevant conferences.

DISCUSSION

The evaluation of SCD risk in HF patients is a problem that urgently needed to be solved. The existing prediction strategies for the SCD risk in HF patients lack clinical practice value for various reasons. ICD indication for primary prevention of SCD could be optimized by identifying the high SCD risk patients in HF with low LVEF (≤35%). It is of great practical value and economic significance.

We reviewed some predictive studies of HF in the past years and ranked the risk factors according to their corresponding hazard ratio, which have been included in our study as candidate risk factors. Otherwise, some other variables which appear relevant to risk of SCD in HF patients are also collected. Therefore, the efficiency and practicality of predictive model development has been highly improved.

This study is the first multicenter registry study in China, aimed to investigate the feasibility and accuracy of applying ML to predict SCD in HF patients with low LVEF. A broad range of outcomes, including SCD, all-cause death, lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF, will be evaluated in this study, and the corresponding prognostic models will be developed. Machine learning and the traditional multivariable COX proportional hazards regression model will be derived from the same database and will be compared.

The limitations of this study are as follows: 1. HF patients with LVEF>35% will not be included based on the design of this study, which will restrict the application of the results of this study to the HF with low LVEF. 2. It might be difficult to determine the endpoint of this study sometimes for some patients, when dealing with SCD, lethal arrhythmia and sudden cardiac arrest, especially when outside the hospital.

Table 1. The risk model for HF in the literature

Author	Database	Year	Variables (n)	Patients (n)	Endpoints
Agostoni[14]	MECKI	2012	6	2716	Cardiovascular death; urgent cardiac transplant
Barlera[15]	GISSI-HF	2013	14	6975	all-cause mortality

Collier[16]	EMPHASIS-HF	2013	10	2737	all-cause mortality
Komajda[17]	I-PRESERVE	2011	12	4128	all-cause mortality
Levy[18]	SHFM	2006	14	1125	Survival
O'Connor[19]	HF-ACTION	2012	4	2331	all-cause mortality
Pocock[20]	CHARM	2006	21	7599	all-cause mortality
Pocock[21]	MAGGIC	2012	13	39372	all-cause mortality
Senni[22]	CVM-HF	2006	13	292	all-cause mortality
Senni[23]	3C-HF	2013	11	2016	all-cause mortality; urgent heart transplant (1year)
Vazquez[24]	MUSIC	2009	10	992	all-cause mortality; cardiac mortality; pump failure death, sudden death
Nicole[25]	BARDICHE-index	2017	8	1811	all-cause mortality; all-cause hospitalization; CHF-related hospitalization

Table 2 The checklist for data collection

Data collection	**Baseline**		**regular visit**	**withdraw /death**
	Retrospective cases	**Prospective cases**		
Informed consent	√	√		
Quantification verification (inclusion and exclusion)	√	√		
Baseline evaluation	√	√		
Medication	√	√		
Questionnaires 9-EHFScBS SSRS HAMD HAMA socioeconomic and educational status		√		
Regular follow-up visit (every 3 months)			√	
Survival state	√		√	√
Adverse event	Once happen √			
Study bias	Once happen √			
Withdraw from the study	Once happen √			
Death	Once happen √			

Figure 1 Flow diagram of progress

Figure 2 Hazard ratio of variables in different risk models

NYHA, New York Heart Association; VHD, valvular heart disease; Af, atrial fibrillation; MI, myocardial infarction; CAGB, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease;

DM, diabetes mellitus; ICM, ischemic cardiomyopathy; HR, heart rate; SBP, systolic blood pressure; BMI, body mass index; BNP, brain natriuretic peptide

Figure 3 Study framework and process

FOOTNOTES

Acknowledgements The authors thank Xiamen Cardiovascular Hospital, Xiamen University (Xiamen, China), Wuhan Asia Heart Hospital (Wuhan, China), Jiangning Hospital Affiliated to Nanjing Medical University, (Nanjing, China), The Second People's Hospital of Lianyungang (Lianyungang, China), The Affiliated Hospital of Jiangsu University (Zhenjiang, China), Taixing People's Hospital (Taixing, China), The First People's Hospital of Huaian (Huaian, China), The First People's Hospital of Yancheng (Yancheng, China), Rugao People's Hospital (Rugao, China), The First People's Hospital of Zhangjiagang (Zhangjiagang, China), The Third People's Hospital of Suzhou (Suzhou, China), The Third People's Hospital of Wuxi (Wuxi, China), The Second Affiliated Hospital of Nanjing Medical University (Nanjing, China) for collaboration including recruitment and follow-up of HF patients. The authors also thank the HF patients who participated in the survey of patient requirements during the design of this study.

Contributors JGZ and FQM conceived and designed the study. ZHZ, XFH, ZYQ, YW, YHC, YLW, YZ, ZC, XWZ, JY, JLZ, JHG, KBL, LC, RJZ, HJ participated in different phases of the protocol design. WHZ provide expertise in data processing and machine learning. SWT and YYW provided their expertise for traditional statistical analysis. JGZ obtained funding. FQM drafted the final manuscript. All authors have read the manuscript and provided feedback. JGZ approved the final version of the manuscript before submission. FQM took responsibility for the submission process.

Funding This study was supported mainly by Jiangsu Province's Key Medical Center (grant number: YXZXA2016010; PI: Jiangang Zou, MD, Ph.D.); in part by National Natural Science Foundation of China (grant number: 81470457; PI: Jiangang Zou, MD, Ph.D.) and Jiangsu Science and Technology Department (grant number: BE2016764; PI: Jiangang Zou, MD, Ph.D.). This research was also supported in part by an AHA grant (grant number: 17AIREA33700016; PI: Weihua Zhou, Ph.D.) and a New Faculty startup grant from the University of Southern Mississippi (grant number: DE01791; PI: Weihua Zhou, Ph.D.).

Competing interests We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Ethics approval This study has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Reference

[1] Sato N. Epidemiology of Heart Failure in Asia[J]. Heart Fail Clin, 2015,11(4):573-579.

[2] Mozaffarian D, Benjamin E J, Go A S, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association[J]. Circulation, 2015,131(4):e29-e322.

[3] Tomaselli G F, Zipes D P. What causes sudden death in heart failure?[J]. Circ Res, 2004,95(8):754-763.

[4] Solomon S D, Wang D, Finn P, et al. Effect of candesartan on cause-specific mortality in heart failure patients: the Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity (CHARM) program[J]. Circulation, 2004,110(15):2180-2183.

[5] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[6] Connolly S J, Hallstrom A P, Cappato R, et al. Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials. AVID, CASH and CIDS studies. Antiarrhythmics vs Implantable Defibrillator study. Cardiac Arrest Study Hamburg . Canadian Implantable Defibrillator Study[J]. Eur Heart J, 2000,21(24):2071-2078.

[7] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[8] Ponikowski P, Voors A A, Anker S D, et al. 2016 ESC Guidelines for the diagnosis and treatment of

acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC[J]. Eur Heart J, 2016,37(27):2129-2200.

[9] Kober L, Thune J J, Nielsen J C, et al. Defibrillator Implantation in Patients with Nonischemic Systolic Heart Failure[J]. N Engl J Med, 2016,375(13):1221-1230.

[10] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[11] Stecker E C, Vickers C, Waltz J, et al. Population-based analysis of sudden cardiac death with and without left ventricular systolic dysfunction: two-year findings from the Oregon Sudden Unexpected Death Study[J]. J Am Coll Cardiol, 2006,47(6):1161-1166.

[12] Aimo A, Januzzi J J, Vergaro G, et al. Left ventricular ejection fraction for risk stratification in chronic systolic heart failure[J]. Int J Cardiol, 2018.

[13] Shen L, Jhund P S, Petrie M C, et al. Declining Risk of Sudden Death in Heart Failure[J]. N Engl J Med, 2017,377(1):41-51.

[14] Agostoni P, Corra U, Cattadori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis[J]. Int J Cardiol, 2013,167(6):2710-2718.

[15] Barlera S, Tavazzi L, Franzosi M G, et al. Predictors of mortality in 6975 patients with chronic heart failure in the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure trial: proposal for a nomogram[J]. Circ Heart Fail, 2013,6(1):31-39.

[16] Collier T J, Pocock S J, McMurray J J, et al. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms: insight from a novel risk score for prognosis derived from the EMPHASIS-HF trial[J]. Eur Heart J, 2013,34(36):2823-2829.

[17] Komajda M, Carson P E, Hetzel S, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE)[J]. Circ Heart Fail, 2011,4(1):27-35.

[18] Levy W C, Mozaffarian D, Linker D T, et al. The Seattle Heart Failure Model: prediction of survival in heart failure[J]. Circulation, 2006,113(11):1424-1433.

[19] O'Connor C M, Whellan D J, Wojdyla D, et al. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction: the HF-ACTION predictive risk score model[J]. Circ Heart Fail, 2012,5(1):63-71.

[20] Pocock S J, Wang D, Pfeffer M A, et al. Predictors of mortality and morbidity in patients with chronic heart failure[J]. Eur Heart J, 2006,27(1):65-75.

[21] Pocock S J, Ariti C A, McMurray J J, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies[J]. Eur Heart J, 2013,34(19):1404-1413.

[22] Senni M, Santilli G, Parrella P, et al. A novel prognostic index to determine the impact of cardiac conditions and co-morbidities on one-year outcome in patients with heart failure[J]. Am J Cardiol, 2006,98(8):1076-1082.

[23] Senni M, Parrella P, De Maria R, et al. Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score[J]. Int J Cardiol, 2013,163(2):206-211.

[24] Vazquez R, Bayes-Genis A, Cygankiewicz I, et al. The MUSIC Risk score: a simple method for

predicting mortality in ambulatory patients with chronic heart failure[J]. Eur Heart J, 2009,30(9):1088-1096.

[25] Uszko-Lencer N, Frankenstein L, Spruit M A, et al. Predicting hospitalization and mortality in patients with heart failure: The BARDICHE-index[J]. Int J Cardiol, 2017,227:901-907.

[26] Delgado V, Bucciarelli-Ducci C, Bax J J. Diagnostic and prognostic roles of echocardiography and cardiac magnetic resonance[J]. J Nucl Cardiol, 2016,23(6):1399-1410.

[27] Halliday B P, Cleland J, Goldberger J J, et al. Personalizing Risk Stratification for Sudden Death in Dilated Cardiomyopathy: The Past, Present, and Future[J]. Circulation, 2017,136(2):215-231.

[28] Kelesidis I, Travin M I. Use of cardiac radionuclide imaging to identify patients at risk for arrhythmic sudden cardiac death[J]. J Nucl Cardiol, 2012,19(1):142-152, 153-157.

[29] Martins D S M, Vidigal F M, Morao M A. Iodine-123-metaiodobenzylguanidine scintigraphy in risk stratification of sudden death in heart failure[J]. Rev Port Cardiol, 2013,32(6):509-516.

[30] Aro A L, Reinier K, Rusinaru C, et al. Electrical risk score beyond the left ventricular ejection fraction: prediction of sudden cardiac death in the Oregon Sudden Unexpected Death Study and the Atherosclerosis Risk in Communities Study[J]. Eur Heart J, 2017,38(40):3017-3025.

[31] Quinlan J R. Induction of decision trees. Machine learning 1.1[M]. 1986.

[32] Weng S F, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data?[J]. PLoS One, 2017,12(4):e174944.

[33] Awan S E, Sohel F, Sanfilippo F M, et al. Machine learning in heart failure: ready for prime time[J]. Curr Opin Cardiol, 2018,33(2):190-195.

[34] Jaarsma T, Arestedt K F, Martensson J, et al. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument[J]. Eur J Heart Fail, 2009,11(1):99-105.

[35] Hu X, Hu X, Su Y, et al. The changes and factors associated with post-discharge self-care behaviors among Chinese patients with heart failure[J]. Patient Prefer Adherence, 2015,9:1593-1601.

[36] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[37] Harrell F J, Lee K L, Califf R M, et al. Regression modelling strategies for improved prognostic prediction[J]. Stat Med, 1984,3(2):143-152.

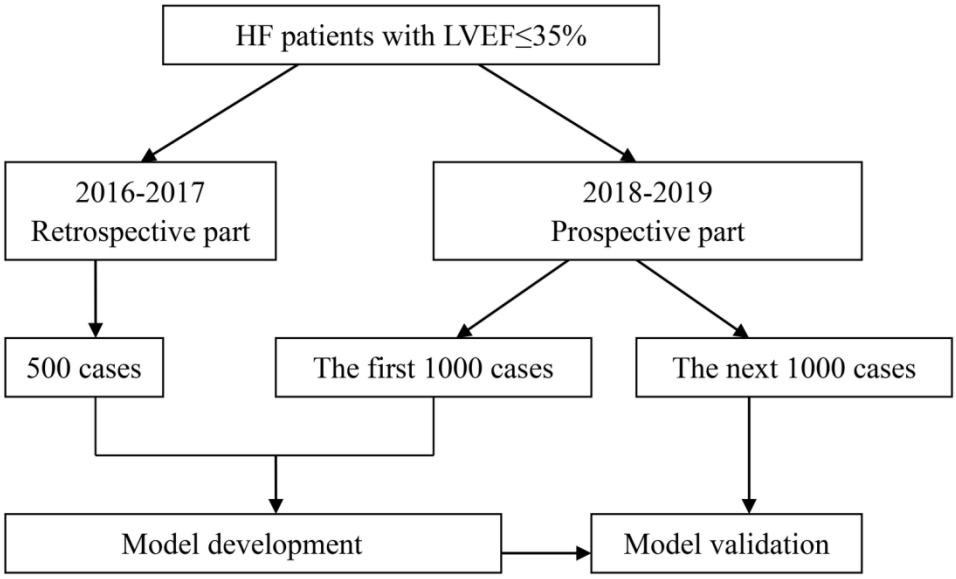

Figure 1 Flow diagram of progress

140x104mm (300 x 300 DPI)

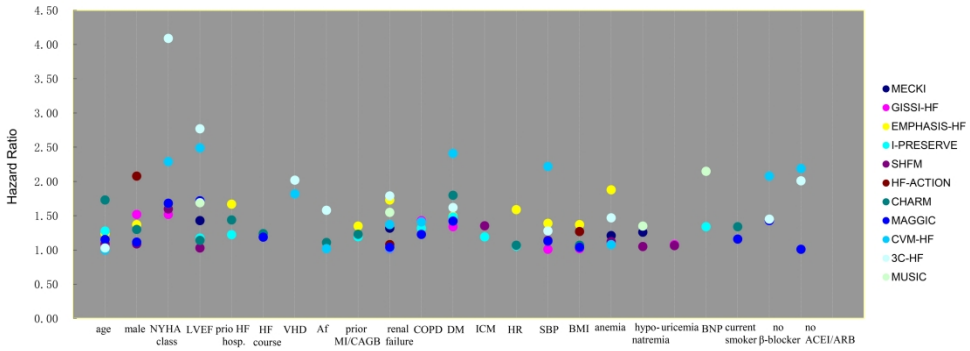

Figure 2 Hazard ratio of variables in different risk models
NYHA, New York Heart Association; VHD, valvular heart disease; Af, atrial fibrillation; MI, myocardial infarction; CAGB, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; ICM, ischemic cardiomyopathy; HR, heart rate; SBP, systolic blood pressure; BMI, body mass index; BNP, brain natriuretic peptide

325x126mm (300 x 300 DPI)

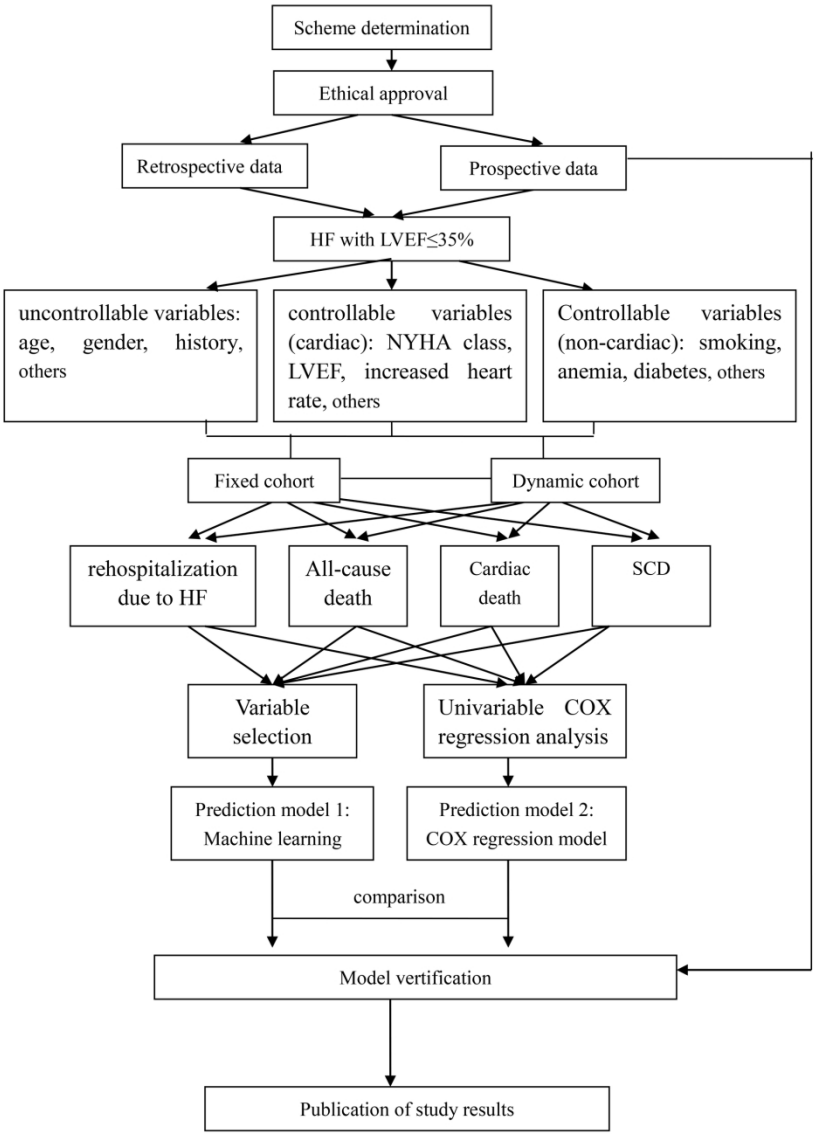

Figure 3 Study framework and process

158x226mm (300 x 300 DPI)

List of Hospitals

The First Affiliated Hospital of Nanjing Medical University, Nanjing

Xiamen Cardiovascular Hospital, Xiamen University, Xiamen

Wuhan Asia Heart Hospital, Wuhan

Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing

The Second People's Hospital of Lianyungang, Lianyungang

The Affiliated Hospital of Jiangsu University, Zhenjiang

Taixing People's Hospital, Taixing

The First People's Hospital of Huaian, Huaian

The First People's Hospital of Yancheng, Yancheng

Rugao People's Hospital, Rugao

The First People's Hospital of Zhangjiagang, Zhangjiagang

The Third People's Hospital of Suzhou, Suzhou

The Third People's Hospital of Wuxi, Wuxi

The Second Affiliated Hospital of Nanjing Medical University, Nanjing

SPIRIT
STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym **(P1)**
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry **(P7)**
	2b	All items from the World Health Organization Trial Registration Data Set **(P7)**
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support **(P17-18)**
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors **(P1, P17)**
	5b	Name and contact information for the trial sponsor **(P1, P7)**
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities **(P7, P17)**
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) **(P7, P11-12)**
Introduction		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention **(P4-6)**
	6b	Explanation for choice of comparators **(P5-6)**
Objectives	7	Specific objectives or hypotheses **(P6)**

Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) **(P6-7)**

Methods: Participants, interventions, and outcomes

Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained **(P6-7)**
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) **(P7-8)**
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered **(Not applicable)**
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) **(Not applicable)**
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) **(Not applicable)**
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial **(Not applicable)**
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended **(P8)**
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended **(see Figure1)**
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations **(P7)**
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size **(P8)**

Methods: Assignment of interventions (for controlled trials)

Allocation:

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions **(Not applicable)**
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned **(Not applicable)**
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions **(Not applicable)**
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how **(Not applicable)**
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial **(Not applicable)**
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol **(P8-12)**
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols **(P11)**
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol **(P11-12)**
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol **(P12-P14)**
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses) **(Not applicable)**
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) **(Not applicable)**

Methods: Monitoring

Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct **(P11-12)**
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval **(P14)**
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) **(Not applicable)**
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) **(P8)**
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable **(Not applicable)**
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial **(P11-12)**
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site **(P18)**
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators **(P11-12)**
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation **(Not applicable)**

Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions **(P14)**
	31b	Authorship eligibility guidelines and any intended use of professional writers **(Not applicable)**
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code **(Not applicable)**
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates **(see ICF)**
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable **(Not applicable)**

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicenter registry in China

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023724.R2
Article Type:	Protocol
Date Submitted by the Author:	04-Feb-2019
Complete List of Authors:	Meng, Fanqi; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Xiamen Cardiovascular Hospital, Xiamen University, Department of Cardiology Zhang, Zhihua; The First Affiliated Hospital of Nanjing Medical University,, Department of Cardiology; Jiangning Hospital Affiliated to Nanjing Medical University, Department of Cardiology Hou, Xiaofeng; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Qian, Zhiyong; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Wang, Yao; The First Affiliated Hospital of Nanjing Medical University, Department of Cardiology Chen, Yanhong; Wuhan Asia Heart Hospital, Department of Cardiology Wang, Yilian; The Second People's Hospital of Lianyungang, Department of Cardiology, Zhou, Ye; The Affiliated Hospital of Jiangsu University, Department of Cardiology Chen, Zhen; Taixing People's Hospital, Department of Cardiology Zhang, Xiwen; The First People's Hospital of Huaian, Department of Cardiology Yang, Jing; The First People's Hospital of Huaian, Department of Cardiology Zhang, Jinlong; The First People's Hospital of Yancheng, Department of Cardiology Guo, Jianghong; Rugao People's Hospital, Department of Cardiology Li, Kebei; The First People's Hospital of Zhangjiagang, Department of Cardiology Chen, Lu; The Third People's Hospital of Suzhou, Department of Cardiology Zhuang, Ruijuan; The Third People's Hospital of Wuxi, Department of Cardiology Jiang, Hai; The Second Affiliated Hospital of Nanjing Medical University, Department of Cardiology Zhou, Weihua; University of Southern Mississippi, School of Computing Tang, Shaowen; Nanjing Medical University, Department of Epidemiology Wei, Yongyue; Nanjing Medical University, Department of Biostatistics Zou, Jiangang; the First Affiliated Hospital of Nanjing Medical University, Department of Cardiology; Nanjing Medical University, Key Laboratory of

	Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine
Primary Subject Heading:	Medical management
Secondary Subject Heading:	Cardiovascular medicine
Keywords:	Heart failure < CARDIOLOGY, sudden cardiac death, machine learning, risk model

Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retro-prospective multicenter registry in China

Fanqi Meng,[1,2] Zhihua Zhang, [1,3] Xiaofeng Hou,[1] Zhiyong Qian,[1] Yao Wang,[1] Yanhong Chen, [4] Yilian Wang,[5] Ye Zhou,[6] Zhen Chen,[7] Xiwen Zhang,[8] Jing Yang, [8] Jinlong Zhang,[9] Jianghong Guo,[10] Kebei Li, [11] Lu Chen,[12] Ruijuan Zhuang,[13] Hai Jiang,[14] Weihua Zhou, [15] Shaowen Tang, [16] Yongyue Wei, [17] Jiangang Zou[1, 18*]

Author affiliations

[1]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

[2]Department of Cardiology, Xiamen Cardiovascular Hospital, Xiamen University, Xiamen, China

[3]Department of Cardiology, Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing, China

[4]Department of Cardiology, Wuhan Asia Heart Hospital, Wuhan, China

[5]Department of Cardiology, The Second People's Hospital of Lianyungang, Lianyungang, China

[6]Department of Cardiology, The Affiliated Hospital of Jiangsu University, Zhenjiang, China

[7]Department of Cardiology, Taixing People's Hospital, Taixing, China

[8]Department of Cardiology, The First People's Hospital of Huaian, Huaian, China

[9]Department of Cardiology, The First People's Hospital of Yancheng, Yancheng, China

[10]Department of Cardiology, Rugao People's Hospital, Rugao, China

[11]Department of Cardiology, The First People's Hospital of Zhangjiagang, Zhangjiagang, China

[12]Department of Cardiology, The Third People's Hospital of Suzhou, Suzhou, China

[13]Department of Cardiology, The Third People's Hospital of Wuxi, Wuxi, China

[14]Department of Cardiology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, China

[15]School of Computing, University of Southern Mississippi, Hattiesburg, USA.

[16]Department of Epidemiology, Nanjing Medical University, Nanjing, China

[17]Department of Biostatistics, Nanjing Medical University, Nanjing, China

[18]Key Laboratory of Targeted Intervention of Cardiovascular Disease, Collaborative Innovation Center for Cardiovascular Disease Translational Medicine, Nanjing Medical University, China

***Correspondence to**

Dr Jiangang Zou**:** Department of Cardiology, the First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, 210029, P.R.China; E-mail: jgzou@njmu.edu.cn; Tel. phone: +86 13605191407

Keywords: heart failure, sudden cardiac death, machine learning, risk model

ABSTRACT

Introduction

Left ventricular ejection fraction (LVEF) ≤35%, as current significant implantable cardioverter-defibrillator (ICD) indication for primary prevention of sudden cardiac death (SCD) in heart failure (HF) patients, has been widely recognized to be inefficient. Improvement of patient selection for low LVEF (≤35%) is needed to optimize deployment of ICD. Most of the existing prediction models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Compared to traditional statistical analysis, machine learning (ML) can employ computer algorithms to identify patterns in large datasets, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions. This study is aimed to develop and validate new models using ML to improve the prediction of SCD in HF patients with low LVEF.

Methods and analysis

We will conduct a retro-prospective, multi-center, observational registry of Chinese HF patients with low LVEF. The HF patients with LVEF ≤35% after optimized medication at least 3 months will be enrolled in this study. The primary endpoints are all-cause death and SCD. The secondary endpoints are malignant arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF. The baseline demographic, clinical, biological, electrophysiological, social and psychological variables will be collected. Both ML and traditional multivariable COX proportional hazards regression models will be developed and compared in the prediction of SCD. Moreover, the ML model will be validated in a prospective study.

Ethics and dissemination

The study protocol has been approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All results of this study will be published in international peer-reviewed journals and presented at relevant conferences.

Clinical trial registration ChiCTR-POC-17011842.

Strengths and limitations of this study

- This study is the first multicenter registry study in China, aimed to investigate the feasibility and accuracy of applying ML to predict SCD in HF patients with low LVEF.
- A broad range of outcomes, including SCD, all-cause death, lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF, will be evaluated in this study, and the corresponding prognostic models will be developed.
- Machine learning and the traditional multivariable COX proportional hazards regression model will be derived from the same database and be compared.
- HF patients with LVEF>35% will not be included based on the design of this study, which will restrict the application of the results of this study to the HF with low LVEF,
- It might be difficult to determine the endpoint of this study sometimes for some patients, when dealing with SCD, lethal arrhythmia and sudden cardiac arrest, especially when outside the hospital.

INTRODUCTION

Heart failure (HF), has become a major public health problem with increased prevalence in both Asia and Western countries. The prevalence of HF in Asia is 1.2%-6.7% depending on the population studied.[1] In China, there are 4.2 million HF patients, and 500,000 new cases are being diagnosed each year.[1] Although the survival rate after HF diagnosis has been increased due to improvement in medical therapy, the mortality of HF remains high. Around 50% of people diagnosed with HF will die within 5 years.[2] The two most common causes of death in patients with HF are sudden cardiac death (SCD) and progressive pump failure. SCD in HF patients is usually caused by lethal arrhythmias such as ventricular tachycardia or ventricular fibrillation, and is reported to be responsible for nearly 50% of all cardiovascular death in HF patients.[3, 4]

The most effective strategy for prevention of SCD in patients with HF is the implantable cardioverter-defibrillator (ICD), associated with 54% relative risk reduction in primary prevention[5], and 50% relative risk reduction in arrhythmia-related death in secondary prevention.[6] There is a higher risk of SCD in patients with LVEF≤ 35% than with LVEF>35%.[7] At present, left ventricular ejection fraction (LVEF) ≤35% is the major ICD indication for primary prevention of SCD.[8] However, real-world data show that only 3% to 5% of ICD patients for primary prevention with LVEF≤35% receive shock therapies on an annual basis,[9] whereas some SCD victims have LVEF>35%.[10, 11] Identifying the patients who will be most likely to benefit from primary prevention ICD, is urgently needed. Based on the latest literature, LVEF≤35% is still an independent predictor of all-cause and cardiovascular mortality in chronic systolic HF, and displays a better combination of sensitivity and specificity than 40% cut-off.[12] Finding ways to evaluate the SCD risk in patients with lower EF will be more efficient and economically significant. Furthermore, a noticeable decline in the rate of SCD for HF patients with reduced LVEF has been observed, which was consistent with the cumulative benefit of optimizing medication including angiotensin converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB), beta-blocker, and mineralocorticoid receptor

antagonist (MRAs).[13] Therefore, it is imperative to update the criterion for ICD implantation

Over the last decade, lots of multivariate prognostic models derived for chronic HF patients have been proposed (Table 1).[14-25] However, these models are not appropriate to identify ICD candidates at high risk of SCD in HF patients with low LVEF. Most above prognostic scores were developed form trial databases, and the subjects included various types of heart failure. There is no specific study for the prognosis of low LVEF population. Additionally, although all the scores are "not parsimonious", some critical factors are not incorporated into the prognostic models, for example, medications are contained in I-PRESERVE[17], MAGGIC[21] and 3C-HF[23]. Optimized medication was not required as inclusion criteria in all 12 studies. Furthermore, the most above prognostic models are not able to predict SCD risk. In recent years, the advances in strain echocardiography[26, 27], cardiac magnetic resonance[26, 27] and cardiac radionuclide imaging[28, 29] have provided essential insights into the mechanisms of ventricular arrhythmias, and have been recommended to predict the SCD in patients with HF. Although these new methods are effective and noninvasive, the widespread use in large HF population to predict SCD is difficult, due to high equipment and technical requirements. Resting 12-lead ECG and Holter, as the longest surviving, broadly available, quickly deployed and inexpensive tests, can provide a measure of cumulative electrical risk, which may be combined with other factors to improve the SCD risk prediction.[30]

Based on above reasons, the novel risk assessment tools should meet the following requirements: (1) the risk model should be developed from the population with low LVEF (≤35%) to accelerate its clinical application and promote the accuracy of ICD indications for primary prevention. (2) More cardiac and non-cardiac factors beyond LVEF should be included. (3) Electrical risk factors should be included as candidate predictors to evaluate the risk of sudden arrhythmic death. (4) Although sometimes it is not easy to determine the cause of death, SCD as the primary endpoint should be defined whenever possible.

Data processing is the crucial step to develop the prognostic models. This study

involves non-linear prediction models, a large number of patients and numerous predictors with complicated correlations. Traditional hypothesis-driven statistical analysis is difficult to overcome these challenges. The machine learning (ML) approaches have great potential to improve the solution. They employ computer algorithms to identify patterns in large datasets with a large number of variables, analyze rules automatically and build both linear and non-linear models in order to make data-driven predictions or decisions.[31] Weng et al. [32] found that ML significantly improved the accuracy of cardiovascular risk prediction, increased the number of patients who could benefit from preventive treatment and avoided unnecessary treatment. Recent studies have shown that the application of ML techniques may have the potential to improve heart failure outcomes and management, including cost savings by improving existing diagnostic and treatment support systems.[33] ML algorithms also have been applied to predict SCD in some recent studies and results indicate their significant advantages for predicting SCD.[34, 35] However, more studies based on large-scale cohort are needed to evaluate ML for prediction of SCD in HF patients. Therefore, the application of ML for the prediction of SCD in HF patients with low LVEF is technically innovative and clinically significant.

AIMS

The purpose of our study is to develop and validate new models to improve the prediction of SCD in HF patients with low LVEF. The new strategies of identifying HF patients most likely to benefit from primary prevention ICD will improve the revolution of ICD indications. The specific research objective is to develop prediction models to evaluate prognosis and SCD risk respectively by ML methods and traditional COX proportional hazard regression in HF patients with low LVEF (≤35%).

METHODS AND ANALYSIS

Study design

This study is a retro-prospective, multi-center, non-interventional, observational clinical registry. The primary sponsor is The First Affiliated Hospital of Nanjing Medical University. This study has been registered in the Chinese Clinical Trial Registry (ChiCTR-POC-17011842). The study will be conducted across 14 cardiovascular departments in Tertiary A hospitals throughout the People's Republic of China. (see Supplement)

The cases from January 2016 to December 2017 in the First Affiliated Hospital of Nanjing Medical University and Xiamen Cardiovascular Hospital Xiamen University will be collected retrospectively and followed-up prospectively. About 500 retrospective cases meet the inclusion criteria according to preliminary estimation. The prospective recruitment has started in the above 14 hospitals since January 2018. The retrospective cases and the first 1000 prospective cases will be used to develop the prediction models. And the next 1000 prospective cases will be used for model validation. The flow diagram of the progress is illustrated in Figure 1.

Inclusion criteria

To participate in this study, patients must comply with all of the following

1. Diagnosis of HFrEF according to the 2016 ESC HF Guideline[8]

2. LVEF≤35% (measured by Simpson's methods) after optimized medication including ACEI or ARB, beta-blocker and MRA if available and not contraindicated at least 3 months.

3. Signed informed consent.

Exclusion criteria

The patient with any of the following will be excluded:

1. Hypertrophic cardiomyopathy

2. Rheumatic heart disease

3. Congenital heart disease

4. Pulmonary heart disease

5. Pericardial diseases and myocarditis

6. Acute myocardial infarction in recent 3 months, including STEMI and NSTEMI

7. Aortic dissection

8. Severe hematologic disease including leukemia, lymphoma, aplastic anemia, etc.

9. Autoimmune disease

10. Malignant tumor

11. Hormone replacement

12. Application of other interventional clinical trials

13. Non-drug therapies for improving heart function: CRT-P/D, ICD, heart transplantation, surgical resection of ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation

Endpoints

Primary endpoint

All-cause death and SCD, including cardiac death and death from other causes

Secondary endpoint

Lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation, rehospitalization due to HF

Recruitment and consent

Participants will be identified and recruited at each of the participating centers. The clinical status of potential participants will be assessed, and their medical records will also be reviewed to confirm the eligibility according to the inclusion and exclusion criteria.

The study details will be explained to all potentially eligible and interesting subjects. The patients who agree to attend this study will sign the informed consent form (ICF) indicating that they fully understand the study and their rights of confidentiality and withdrawal from the study without giving a reason.

Baseline evaluation

Prognostic models of HF in the last 10 years have been reviewed, and the associated risk factors have been ranked according to their corresponding hazard ratio in respective risk models (Table 1, Figure 2). Age, sex, New York Heart Association (NYHA) class, LVEF, prior HF hospitalization, course of HF, severe valvular heart disease, atrial fibrillation, prior myocardial infarction / coronary artery bypass grafting (CABG), renal dysfunction, chronic obstructive pulmonary disease (COPD), diabetes mellitus, ischemic etiology, decreased systolic pressure, low body mass index (BMI), anemia, hyponatremia, high N-terminal pro-brain natriuretic peptide (NT-proBNP), uricemia, current smoker were included. Variables which were not listed in previous models but appear relevant to higher risk of SCD in HF patients, and would therefore, merit consideration, including syncope or pre-syncope, frequent premature ventricular beat, non-sustained ventricular tachycardia, complete left bundle branch block (CLBBB), long QT interval, increased QT dispersion. In addition, self-care ability, social support and psychological state including depression and anxiety, are also predictors for subsequent poor prognosis in HF patients. The above risk factors have been assessed and confirmed by an expert panel of cardiologists and statisticians and will be collected in this study particularly.

The baseline data that will be collected in all eligible subjects are as follows.

- Demographic characteristics: date of birth, gender, height and weight
- Lifestyle behavior: smoking and drinking status
- Vital signs: blood pressure and heart rate
- NYHA class
- Etiology of heart failure: The ischemic etiology will be confirmed if any following point is met: a. prior myocardial infarction or revascularization history (coronary artery bypass grafting, CABG / percutaneous coronary intervention, PCI); b. left main or proximal segment of the left anterior descending artery stenosis ≥75% showed by coronary angiogram (CAG); c. at least two main coronary artery branches stenosis ≥75% showed by CAG. Otherwise, non-ischemic HF should be identified.

- Prior HF hospitalization history: first HF hospitalization or not, times of prior HF hospitalization, the course of HF (since the HF symptoms appear; if unavailable, since the decreased EF was found).
- Coronary heart disease history: myocardial infarction or angina history, CAG result, revascularization history, recent angina.
- Arrhythmia history: atrial fibrillation, atrial flutter, premature atrial contraction (PAC), premature ventricular contraction (PVC), non-sustained VT (NSVT), sustained VT (SVT), ventricular fibrillation, and some bradyarrhythmias.
- Syncope or pre-syncope history
- Cardiac arrest/ cardiopulmonary resuscitation history
- Other histories: hypertension, diabetes mellitus, COPD, etc.
- Echocardiography: LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV) and LVEF measured by Simpson's method; Left atrial diameter, LV end-diastolic diameter (LVEDD) and LV end-systolic diameter (LVESD), pulmonary artery systolic pressure. The status of valve regurgitation will be evaluated (0-none; 1-mild; 2-mild to moderate; 3-moderate; 4- severe)
- ECG: Left / right bundle branch block will be recorded. QRS duration and QT interval will be tested, and QT dispersion will be calculated.
- Holter: total heartbeat of the whole day, minimum/ maximum/ average HR, onset of PVC, PAC, NSVT, VT, atrial fibrillation/ flutter.
- Laboratory tests results: serum creatinine, blood urea nitrogen (BUN), serum natrium, hemoglobin, Thyroid Stimulating Hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), NT-proBNP.
- Medication: ACEI/ARB、beta-blocker, aldosterone antagonist, diuretic, digoxin, antiplatelet agent, anticoagulant, statin, calcium channel blocker, antiarrhythmics, Ivabradine and angiotensin receptor blocker-neprilysin inhibitor (ARNI).
- Evaluation of self-care behavior and social support: 9-item European Heart Failure Self-care Behavior Scale (9-EHFScBS）[36] will be used to determine the self-care levels in HF patients. Social Support Rating Scale (SSRS)[37] will be used to evaluate the social support condition in HF patients.

- Assessment of psychological status: Hamilton Depression Scale (HAMD) and Hamilton Anxiety Scale (HAMA).
- Socioeconomic and educational status: marital status, educational status, monthly income, sources of medical expenses, medical insurance

Patient visits

After being enrolled in this research, all the subjects will be followed-up periodically in the outpatient department or by telephone interview every 3 months. The compliance with medications will be evaluated. As the primary endpoint, all-cause death and SCD will be focused. Cause of death will be analyzed in detail. SCD is defined by the World Health Organization as unexpected death that occurs within 1 hour from the onset of new or worsening symptoms (witnessed arrest) or, if unwitnessed, within 24h from when the individual was last observed alive and asymptomatic[38]. The lethal arrhythmia including VT/VF, sudden cardiac arrest, cardiopulmonary resuscitation, and rehospitalization due to HF will be recorded carefully.

During follow-up, lethal arrhythmia will be recognized more precisely for patients who receive ICD or CRT/D implantation, and will be recorded as an adverse event. The patients, who receive CRT-P/D, heart transplantation, surgical resection of a ventricular aneurysm, interventional left ventricular restoration with Revivent(TM) / Parachute(TM) system), MitraClip therapy for recurrent mitral regurgitation, or some other non-drug therapy to improve heart function, will be followed up as usual.

Data collection

In the prospective part, clinical data of subjects will be collected and filled in the electrical data capture (EDC) system at baseline and particular follow-up visit. In the retrospective part, the same baseline information, except for 9-EHFScBS, SSRS, HAMD, HAMA questionnaires, will also be captured and input into the EDC system. The following prospective visits (every three months) will be conducted regularly and will be recorded in the EDC system. Investigators will record all the information of

adverse events (AE), study bias, withdrawal from the study or death in EDC system. In this study, the participants will be identified by study codes, and their names will not appear in the EDC system. All the personal information including contact information, medical record and outcome, will not be revealed to any person who has not been authorized by a principal investigator. Professional staffs are responsible for database management, data maintenance and regular data backup. Data quality will be monitored regularly. The data collection checklist is showed in Table 2.

Data pre-processing

All above-collected variables, which might be predictors of all adverse prognosis of HF described in endpoint events, will be classified as uncontrollable variables (e.g. age, gender, history), controllable variables associated with heart (e.g. NYHA class, LVEF, increased heart rate) and controllable variables beyond heart (e.g. smoking, anemia, diabetes mellitus). Appropriate dummy variables will be used for binary variables and categorical variables, and quantitative variables will be fitted as a single continuous measurement (e.g. age, heart rate, NT-proBNP), unless there is clear evidence of nonlinearity. In order to create a practice simple risk score, some continuous variables will also be categorized into several groups according to both common clinical cut points and expert advice.

Machine learning

Variable selection is the process of selecting a subset of relevant variables for use in model construction, which can substantially reduce the abundant information and decrease the number of variables that are input to the prediction model. In this study, the technique named as “information gain ranking” will be used to select appropriate variables. Information gain represents the effectiveness of a variable based on entropy, which characterizes the unpredictability of a system. The information gain of a variable is evaluated as the entropy difference of the system when including and excluding this variable. Then the variables whose information gain scores are less than a threshold are considered to be insignificant and will be excluded from the

prediction.

Prediction models for SCD in HF patients will be developed by the following classification algorithms respectively: decision trees, logistic regression, support vector machine, random forest, and artificial neural network. [29] The performance and general error estimation of these ML models will be assessed by 10-fold cross-validation. The dataset will be randomly divided into 10 equal folds. 9 folds will be used as the training set with the remaining one fold as the validation set. The validation results from 10 repeats will be combined to provide a measure of the overall performance. The prediction models derived from the above classification algorithms above will be evaluated based on the accuracy, sensitivities, specificities and the area under the Receiver operating characteristic curve. Finally, clinical experts and computer specialists will discuss and choose the best model to predict the prognosis of SCD in HF patients and then perform further validation with the prospective dataset.

COX proportional hazards regression

Univariable COX proportional hazards modeling will be used to identify strong independent baseline candidate predictors for the primary and secondary outcomes. We will use both forward and backward stepwise procedure to derive the multivariable COX proportional hazards model with $p<0.05$ as the inclusion criterion. Every variable in the model will be multiplied by its β -coefficient, and the products will be summed to calculate the risk score. Risk function will be used to estimate the level of risk. The calculating formula is as follows.[39]

$$P = h(t_j; X_k) = h_0(t_j) \exp(SCORE)$$

$$SCORE = X_k \beta_k = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots \beta_p x_p$$

Model validation

The dynamic prospective cases will be used for external validation of the optimal ML and COX proportional hazards models. The validation will be performed using the models to calculate the probability of the outcome of interest occurring for each

individual included in the validation sample when compared with the events actually observed to occur in this sample. The discrimination of each model will be estimated by receiver-operating characteristic (ROC) curve. The calibration of the models will be assessed by the Hosmer-Lemeshow goodness-of-fit test. The ML prediction model will be compared with the COX proportional hazards regression model.

Patient and public involvement statement

During the design of this study, a survey of patient requirements, including communication needs, follow-up frequency, visit cost etc., was conducted in population of potential HF participants, which provided important evidence for drawing up this study protocol to meet most of the patients' needs, build close contact with patients, enhance the overall adherence and improve the accuracy of endpoint event. This study is not a patient-led research, and patients are not involved in the recruitment of the study. The study results will be informed to the participants by phone at the end of this study. The alive patients will be evaluated with the new prediction model, and the ICD intervention will be recommended to the high SCD risk patients.

Study timeframe

The retrospective data collection in the two sub-centers started in March 2017, and prospective enrollment in all 14 sub-centers has started in January 2018. The follow-up period is scheduled to end in December 2019. The major part of data analysis will be performed from January to June 2020. The study framework and process is summarized in Figure 3.

ETHICS AND DISSEMINATION

The study protocol has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06). All necessary information about this study will be disclosed to the patients. Every subject will be asked to sign the ICF, indicating that they fully understand the study and voluntarily participate in

this study. All results of this study will be published in international peer-reviewed journals and presented at relevant conferences.

DISCUSSION

The evaluation of SCD risk in HF patients is a problem that urgently needed to be solved. The existing prediction strategies for the SCD risk in HF patients lack clinical practice value for various reasons. ICD indication for primary prevention of SCD could be optimized by identifying the high SCD risk patients in HF with low LVEF (≤35%). It is of great practical value and economic significance.

We reviewed some predictive studies of HF in the past years and ranked the risk factors according to their corresponding hazard ratio, which have been included in our study as candidate risk factors. Otherwise, some other variables which appear relevant to risk of SCD in HF patients are also collected. Therefore, the efficiency and practicality of predictive model development has been highly improved.

This study is the first multicenter registry study in China, aimed to investigate the feasibility and accuracy of applying ML to predict SCD in HF patients with low LVEF. A broad range of outcomes, including SCD, all-cause death, lethal arrhythmia, sudden cardiac arrest, cardiopulmonary resuscitation and rehospitalization due to HF, will be evaluated in this study, and the corresponding prognostic models will be developed. Machine learning and the traditional multivariable COX proportional hazards regression model will be derived from the same database and will be compared.

The limitations of this study are as follows: 1. HF patients with LVEF>35% will not be included based on the design of this study, which will restrict the application of the results of this study to the HF with low LVEF. 2. It might be difficult to determine the endpoint of this study sometimes for some patients, when dealing with SCD, lethal arrhythmia and sudden cardiac arrest, especially when outside the hospital.

Table 1. The risk model for HF in the literature

Author	Database	Year	Variables (n)	Patients (n)	Endpoints
Agostoni[14]	MECKI	2012	6	2716	Cardiovascular death; urgent cardiac transplant
Barlera[15]	GISSI-HF	2013	14	6975	all-cause mortality
Collier[16]	EMPHASIS-HF	2013	10	2737	all-cause mortality
Komajda[17]	I-PRESERVE	2011	12	4128	all-cause mortality
Levy[18]	SHFM	2006	14	1125	Survival
O'Connor[19]	HF-ACTION	2012	4	2331	all-cause mortality
Pocock[20]	CHARM	2006	21	7599	all-cause mortality
Pocock[21]	MAGGIC	2012	13	39372	all-cause mortality
Senni[22]	CVM-HF	2006	13	292	all-cause mortality
Senni[23]	3C-HF	2013	11	2016	all-cause mortality; urgent heart transplant (1year)
Vazquez[24]	MUSIC	2009	10	992	all-cause mortality; cardiac mortality; pump failure death, sudden death
Nicole[25]	BARDICHE-index	2017	8	1811	all-cause mortality; all-cause hospitalization; CHF-related hospitalization

Table 2 The checklist for data collection

Data collection	Baseline		regular visit	withdraw /death
	Retrospective cases	Prospective cases		
Informed consent	√	√		
Quantification verification (inclusion and exclusion)	√	√		
Baseline evaluation	√	√		
Medication	√	√		
Questionnaires 9-EHFScBS SSRS HAMD HAMA socioeconomic and educational status		√		
Regular follow-up visit (every 3 months)			√	
Survival state	√		√	√
Adverse event	Once happen √			
Study bias	Once happen √			
Withdraw from the study	Once happen √			
Death	Once happen √			

Figure 1 Flow diagram of progress

Figure 2 Hazard ratio of variables in different risk models

NYHA, New York Heart Association; VHD, valvular heart disease; Af, atrial fibrillation; MI, myocardial infarction; CAGB, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease;

DM, diabetes mellitus; ICM, ischemic cardiomyopathy; HR, heart rate; SBP, systolic blood pressure; BMI, body mass index; BNP, brain natriuretic peptide

Figure 3 Study framework and process

FOOTNOTES

Acknowledgements The authors thank Xiamen Cardiovascular Hospital, Xiamen University (Xiamen, China), Wuhan Asia Heart Hospital (Wuhan, China), Jiangning Hospital Affiliated to Nanjing Medical University, (Nanjing, China), The Second People's Hospital of Lianyungang (Lianyungang, China), The Affiliated Hospital of Jiangsu University (Zhenjiang, China), Taixing People's Hospital (Taixing, China), The First People's Hospital of Huaian (Huaian, China), The First People's Hospital of Yancheng (Yancheng, China), Rugao People's Hospital (Rugao, China), The First People's Hospital of Zhangjiagang (Zhangjiagang, China), The Third People's Hospital of Suzhou (Suzhou, China), The Third People's Hospital of Wuxi (Wuxi, China), The Second Affiliated Hospital of Nanjing Medical University (Nanjing, China) for collaboration including recruitment and follow-up of HF patients. The authors also thank the HF patients who participated in the survey of patient requirements during the design of this study.

Contributors JGZ and FQM conceived and designed the study. ZHZ, XFH, ZYQ, YW, YHC, YLW, YZ, ZC, XWZ, JY, JLZ, JHG, KBL, LC, RJZ, HJ participated in different phases of the protocol design. WHZ provide expertise in data processing and machine learning. SWT and YYW provided their expertise for traditional statistical analysis. JGZ obtained funding. FQM drafted the final manuscript. All authors have read the manuscript and provided feedback. JGZ approved the final version of the manuscript before submission. FQM took responsibility for the submission process.

Funding This study was supported mainly by Jiangsu Province's Key Medical Center (grant number: YXZXA2016010; PI: Jiangang Zou, MD, Ph.D.); in part by National

Natural Science Foundation of China (grant number: 81470457; PI: Jiangang Zou, MD, Ph.D.) and Jiangsu Science and Technology Department (grant number: BE2016764; PI: Jiangang Zou, MD, Ph.D.). This research was also supported in part by an AHA grant (grant number: 17AIREA33700016; PI: Weihua Zhou, Ph.D.) and a New Faculty startup grant from the University of Southern Mississippi (grant number: DE01791; PI: Weihua Zhou, Ph.D.).

Competing interests We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Ethics approval This study has been approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (2017-SR-06).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Reference

[1] Sato N. Epidemiology of Heart Failure in Asia[J]. Heart Fail Clin, 2015,11(4):573-579.

[2] Mozaffarian D, Benjamin E J, Go A S, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association[J]. Circulation, 2015,131(4):e29-e322.

[3] Tomaselli G F, Zipes D P. What causes sudden death in heart failure?[J]. Circ Res, 2004,95(8):754-763.

[4] Solomon S D, Wang D, Finn P, et al. Effect of candesartan on cause-specific mortality in heart failure patients: the Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity (CHARM) program[J]. Circulation, 2004,110(15):2180-2183.

[5] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[6] Connolly S J, Hallstrom A P, Cappato R, et al. Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials. AVID, CASH and CIDS studies. Antiarrhythmics vs Implantable Defibrillator study. Cardiac Arrest Study Hamburg . Canadian Implantable Defibrillator Study[J]. Eur Heart J, 2000,21(24):2071-2078.

[7] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[8] Ponikowski P, Voors A A, Anker S D, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC[J]. Eur Heart J, 2016,37(27):2129-2200.

[9] Kober L, Thune J J, Nielsen J C, et al. Defibrillator Implantation in Patients with Nonischemic Systolic Heart Failure[J]. N Engl J Med, 2016,375(13):1221-1230.

[10] Myerburg R J, Spooner P M. Opportunities for sudden death prevention: directions for new clinical and basic research[J]. Cardiovasc Res, 2001,50(2):177-185.

[11] Stecker E C, Vickers C, Waltz J, et al. Population-based analysis of sudden cardiac death with and without left ventricular systolic dysfunction: two-year findings from the Oregon Sudden Unexpected Death Study[J]. J Am Coll Cardiol, 2006,47(6):1161-1166.

[12] Aimo A, Januzzi J J, Vergaro G, et al. Left ventricular ejection fraction for risk stratification in chronic systolic heart failure[J]. Int J Cardiol, 2018.

[13] Shen L, Jhund P S, Petrie M C, et al. Declining Risk of Sudden Death in Heart Failure[J]. N Engl J Med, 2017,377(1):41-51.

[14] Agostoni P, Corra U, Cattadori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis[J]. Int J Cardiol, 2013,167(6):2710-2718.

[15] Barlera S, Tavazzi L, Franzosi M G, et al. Predictors of mortality in 6975 patients with chronic heart failure in the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure trial: proposal for a nomogram[J]. Circ Heart Fail, 2013,6(1):31-39.

[16] Collier T J, Pocock S J, McMurray J J, et al. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms: insight from a novel risk score for prognosis derived from the EMPHASIS-HF trial[J]. Eur Heart J, 2013,34(36):2823-2829.

[17] Komajda M, Carson P E, Hetzel S, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE)[J]. Circ Heart Fail, 2011,4(1):27-35.

[18] Levy W C, Mozaffarian D, Linker D T, et al. The Seattle Heart Failure Model: prediction of survival in heart failure[J]. Circulation, 2006,113(11):1424-1433.

[19] O'Connor C M, Whellan D J, Wojdyla D, et al. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction: the HF-ACTION predictive risk score model[J]. Circ Heart Fail, 2012,5(1):63-71.

[20] Pocock S J, Wang D, Pfeffer M A, et al. Predictors of mortality and morbidity in patients with chronic heart failure[J]. Eur Heart J, 2006,27(1):65-75.

[21] Pocock S J, Ariti C A, McMurray J J, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies[J]. Eur Heart J, 2013,34(19):1404-1413.

[22] Senni M, Santilli G, Parrella P, et al. A novel prognostic index to determine the impact of cardiac conditions and co-morbidities on one-year outcome in patients with heart failure[J]. Am J Cardiol, 2006,98(8):1076-1082.

[23] Senni M, Parrella P, De Maria R, et al. Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score[J]. Int J Cardiol, 2013,163(2):206-211.

[24] Vazquez R, Bayes-Genis A, Cygankiewicz I, et al. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure[J]. Eur Heart J, 2009,30(9):1088-1096.

[25] Uszko-Lencer N, Frankenstein L, Spruit M A, et al. Predicting hospitalization and mortality in patients with heart failure: The BARDICHE-index[J]. Int J Cardiol, 2017,227:901-907.

[26] Delgado V, Bucciarelli-Ducci C, Bax J J. Diagnostic and prognostic roles of echocardiography and cardiac magnetic resonance[J]. J Nucl Cardiol, 2016,23(6):1399-1410.

[27] Halliday B P, Cleland J, Goldberger J J, et al. Personalizing Risk Stratification for Sudden Death in Dilated Cardiomyopathy: The Past, Present, and Future[J]. Circulation, 2017,136(2):215-231.

[28] Kelesidis I, Travin M I. Use of cardiac radionuclide imaging to identify patients at risk for arrhythmic sudden cardiac death[J]. J Nucl Cardiol, 2012,19(1):142-152, 153-157.

[29] Martins D S M, Vidigal F M, Morao M A. Iodine-123-metaiodobenzylguanidine scintigraphy in risk stratification of sudden death in heart failure[J]. Rev Port Cardiol, 2013,32(6):509-516.

[30] Aro A L, Reinier K, Rusinaru C, et al. Electrical risk score beyond the left ventricular ejection fraction: prediction of sudden cardiac death in the Oregon Sudden Unexpected Death Study and the Atherosclerosis Risk in Communities Study[J]. Eur Heart J, 2017,38(40):3017-3025.

[31] Quinlan J R. Induction of decision trees. Machine learning 1.1[M]. 1986.

[32] Weng S F, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data?[J]. PLoS One, 2017,12(4):e174944.

[33] Awan S E, Sohel F, Sanfilippo F M, et al. Machine learning in heart failure: ready for prime time[J]. Curr Opin Cardiol, 2018,33(2):190-195.

[34] Ebrahimzadeh E, Foroutan A, Shams M, et al. An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal[J]. Comput Methods Programs Biomed, 2019,169:19-36.

[35] Au-Yeung W M, Reinhall P G, Bardy G H, et al. Development and validation of warning system of ventricular tachyarrhythmia in patients with heart failure with heart rate variability data[J]. PLoS One, 2018,13(11):e207215.

[36] Jaarsma T, Arestedt K F, Martensson J, et al. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument[J]. Eur J Heart Fail, 2009,11(1):99-105.

[37] Hu X, Hu X, Su Y, et al. The changes and factors associated with post-discharge self-care behaviors among Chinese patients with heart failure[J]. Patient Prefer Adherence, 2015,9:1593-1601.

[38] Yousuf O, Chrispin J, Tomaselli G F, et al. Clinical management and prevention of sudden cardiac death[J]. Circ Res, 2015,116(12):2020-2040.

[39] Harrell F J, Lee K L, Califf R M, et al. Regression modelling strategies for improved prognostic prediction[J]. Stat Med, 1984,3(2):143-152.

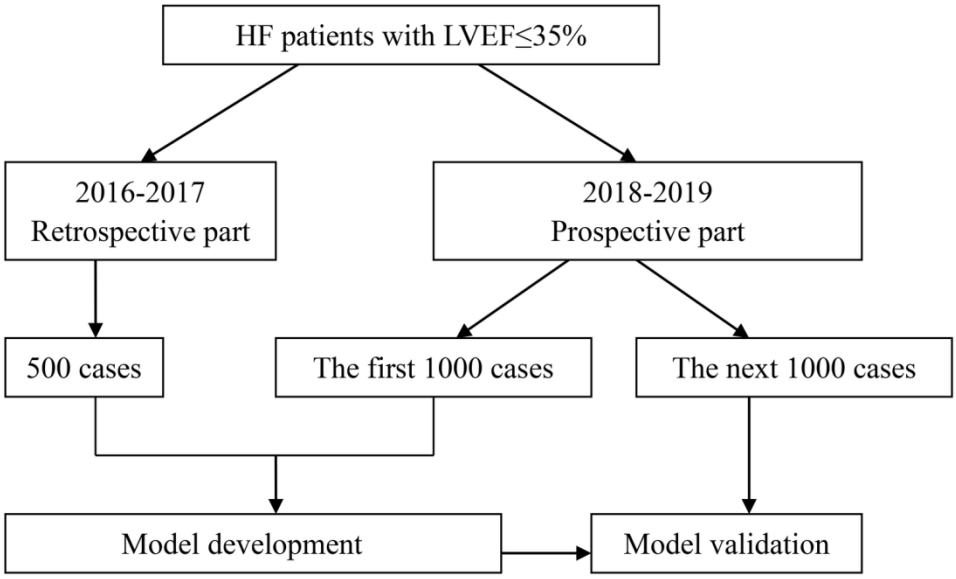

Figure 1 Flow diagram of progress

140x104mm (300 x 300 DPI)

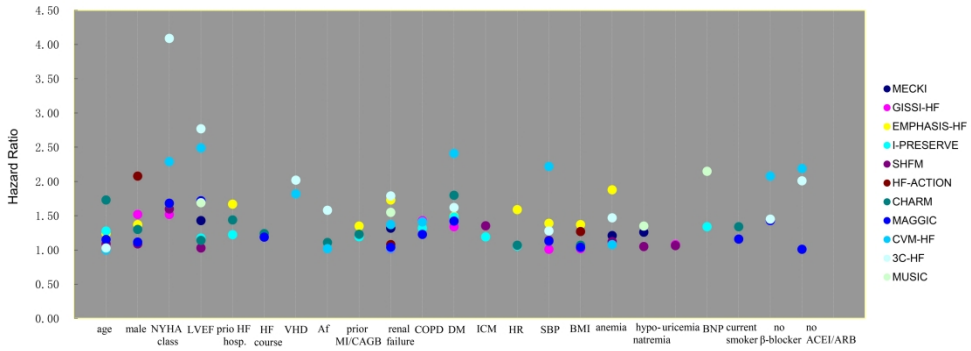

Figure 2 Hazard ratio of variables in different risk models
NYHA, New York Heart Association; VHD, valvular heart disease; Af, atrial fibrillation; MI, myocardial infarction; CAGB, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; ICM, ischemic cardiomyopathy; HR, heart rate; SBP, systolic blood pressure; BMI, body mass index; BNP, brain natriuretic peptide

325x126mm (300 x 300 DPI)

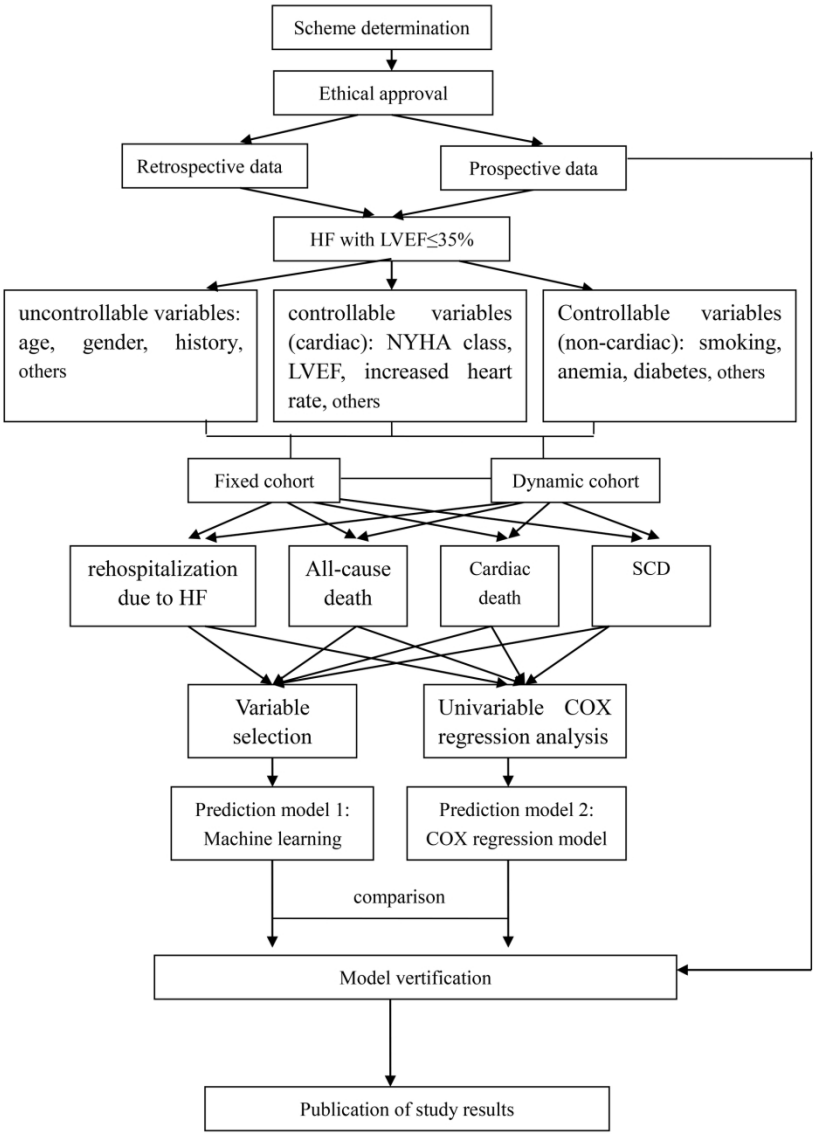

Figure 3 Study framework and process

158x226mm (300 x 300 DPI)

List of Hospitals

The First Affiliated Hospital of Nanjing Medical University, Nanjing

Xiamen Cardiovascular Hospital, Xiamen University, Xiamen

Wuhan Asia Heart Hospital, Wuhan

Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing

The Second People's Hospital of Lianyungang, Lianyungang

The Affiliated Hospital of Jiangsu University, Zhenjiang

Taixing People's Hospital, Taixing

The First People's Hospital of Huaian, Huaian

The First People's Hospital of Yancheng, Yancheng

Rugao People's Hospital, Rugao

The First People's Hospital of Zhangjiagang, Zhangjiagang

The Third People's Hospital of Suzhou, Suzhou

The Third People's Hospital of Wuxi, Wuxi

The Second Affiliated Hospital of Nanjing Medical University, Nanjing

SPIRIT

STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym **(P1)**
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry **(P7)**
	2b	All items from the World Health Organization Trial Registration Data Set **(P7)**
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support **(P17-18)**
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors **(P1, P17)**
	5b	Name and contact information for the trial sponsor **(P1, P7)**
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities **(P7, P17)**
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) **(P7, P11-12)**
Introduction		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention **(P4-6)**
	6b	Explanation for choice of comparators **(P5-6)**
Objectives	7	Specific objectives or hypotheses **(P6)**

Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) **(P6-7)**
Methods: Participants, interventions, and outcomes		
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained **(P6-7)**
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) **(P7-8)**
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered **(Not applicable)**
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) **(Not applicable)**
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) **(Not applicable)**
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial **(Not applicable)**
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended **(P8)**
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended **(see Figure1)**
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations **(P7)**
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size **(P8)**
Methods: Assignment of interventions (for controlled trials)		
Allocation:		

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions **(Not applicable)**
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned **(Not applicable)**
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions **(Not applicable)**
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how **(Not applicable)**
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial **(Not applicable)**
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol **(P8-12)**
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols **(P11)**
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol **(P11-12)**
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol **(P12-P14)**
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses) **(Not applicable)**
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) **(Not applicable)**

Methods: Monitoring

Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct **(P11-12)**
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval **(P14)**
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) **(Not applicable)**
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) **(P8)**
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable **(Not applicable)**
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial **(P11-12)**
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site **(P18)**
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators **(P11-12)**
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation **(Not applicable)**

Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions **(P14)**
	31b	Authorship eligibility guidelines and any intended use of professional writers **(Not applicable)**
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code **(Not applicable)**
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates **(see ICF)**
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable **(Not applicable)**

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.