The impact of creating mathematical formula to predict cardiovascular events in patients with heart failure

Mari Sakamoto1,2, Hiroki Fukuda1,2, Jiyoong Kim1, Tomomi Ide3, Shintaro Kinugawa4, Arata Fukushima4, Hiroyuki Tsutsui3,4, Akira Ishii5, Shin Ito1, Hiroshi Asanuma6, Masanori Asakura1,7, Takashi Washio5 & Masafumi Kitakaze1

Since our retrospective study has formed a mathematical formula, \( \alpha = f(x_1, \ldots, x_{252}) \), where \( \alpha \) is the probability of cardiovascular events in patients with heart failure (HF) and \( x_i \) is each clinical parameter, we prospectively tested the predictive capability and feasibility of the mathematical formula of cardiovascular events in HF patients. First of all, to create such a mathematical formula using limited number of the parameters to predict the cardiovascular events in HF patients, we retrospectively determined \( f(x) \) that formulates the relationship between the most influential 50 clinical parameters \( (x) \) among 252 parameters using 167 patients hospitalized due to acute HF; the nonlinear optimization could provide the formula of \( \alpha = f(x_1, \ldots, x_{50}) \) which fitted the probability of the actual cardiovascular events per day. Secondly, we prospectively examined the predictability of \( f(x) \) in other 213 patients using 50 clinical parameters in 3 hospitals, and we found that the Kaplan–Meier curves using actual and estimated occurrence probabilities of cardiovascular events were closely correlated. We conclude that we created a mathematical formula \( f(x) \) that precisely predicted the occurrence probability of future cardiovascular outcomes of HF patients per day. Mathematical modelling may predict the occurrence probability of cardiovascular events in HF patients.

Heart failure (HF), one of the leading causes of mortality and morbidity worldwide1, is the end stage of many cardiovascular diseases. Although the cause of HF is usually unique for each patient, numerous clinical and social factors, including disease severity, treatment protocols, comorbidity, lifestyle and social environment, independently link to the patients’ prognoses2–6, implicated in ‘precision medicine’7. However, many studies have not considered the relative contribution of such factors and have not examined the contribution of the unexpected and unknown factors to cardiovascular events. Even if the multiple factors are identified as necessary for the occurrence of cardiovascular events, the results are qualitative and not quantitative. Finally, these factors are usually proved retroactively because researchers usually do not test the reproducibility of the results in the prospective study, which may not lead the definite conclusion for the identified factors. To overcome these limitations, we devised a mathematical formula using all the parameters and factors in the medical records to provide the occurrence probability of cardiovascular events and revealed that more than 250 factors are linked to the occurrence probability of cardiovascular events in patients with HF8. However, one might argue that this formula is merely the result of the fitting of the clinical data with the function of occurrence probability of cardiovascular outcomes...

1Department of Clinical Research and Development, National Cerebral and Cardiovascular Center, 5-7-1 Fujishirodai, Suita, Osaka, Japan. 2Department of Cell Biology and National Cerebral and Cardiovascular Center, 5-7-1 Fujishirodai, Suita, Osaka, Japan. 3Department of Cardiovascular Medicine, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka, Japan. 4Department of Cardiovascular Medicine, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Kita 14-jo Nishi 5-chome, Kita-ku, Sapporo, Hokkaido, Japan. 5The Institute of Scientific and Industrial Research, Osaka University, 1-1 Yamadaoka, Suita, Osaka, Japan. 6Department of Internal Medicine, Meiji University of Integrative Medicine, Hiyoshi-ho Nantan, Kyoto, Japan. 7Present address: Hyogo College of Medicine, 1-1 Mukogawacho, Nishinomiya, Hyogo, Japan. Mari Sakamoto and Hiroki Fukuda contributed equally to this work. Correspondence and requests for materials should be addressed to M.K. (email: kitakaze@zf6.so-net.ne.jp)
and that this mathematical formula may not predict the future clinical outcomes such as a law of gravity—the law of gravity guarantees the time for an object to reach the ground.

To clarify that our mathematical model prospectively provides the probability of cardiovascular events, we devised a mathematical formula using the clinical retrospective data of patients with HF and tested whether this formula can predict the probability of future cardiovascular events per day in patients with HF. If this is proved, we can obtain the formula to predict the occurrence probability of cardiovascular events using many clinical or social parameters beforehand, leading to the precision medicine of HF.

**Methods**

**Ethics statement.** This study was approved by National Cerebral and Cardiovascular Center Research Ethics Committee (M22-49, M24-51). The Committee decided that the acquisition of informed consent from 167 patients was not required according to the Japanese Clinical Research Guideline because this was a retrospective observational study. Instead, we made a public announcement using both internet homepage of our institution and bulletin boards of our out-patient and in-patient clinics in accordance with the request of the Ethics Committee and the Guideline. For the prospective observational study of 213 patients, we obtained written informed consent after the approval of Research Ethics Committees in three institutes of National Cerebral and Cardiovascular Center and Hokkaido and Kyushu Universities. Registration number of the clinical trial is UMIN000018691 at https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?func=brows&action=brows&recptno=R000021637&type=summary&language=j.

**Protocols.**

Protocol I: The creation of the mathematical formula using the retrospective data. Since we retrospectively obtained 252 clinical parameters among 402 parameters in 152 patients with acute decompensated HF (ADHF), calculated the formula to provide the probability of cardiovascular events (the hospitalization or death due to HF) and added 16 patients in the patients’ cohort after sorting the data, we enrolled 167 patients with ADHF admitted between November 2007 and October 2009. We followed up these patients until the time of cardiovascular events or December 2014. The diagnosis of HF was confirmed by an expert team of cardiologists using the Framingham criteria.

Here, we showed how to create the mathematical formula to predict the cardiovascular events in the previous study. First of all, our hypothesis in the previous study is that we can derive a mathematical formula for the estimation of prognosis, i.e., the equation \( \tau = f(x_1, \ldots, x_n) \), where \( x_1, \ldots, x_n \) are clinical features and \( \tau \) represents the day for the cardiovascular event in the patients with HF, and we showed the positive evidence to support such a hypothesis in the previous study. In the present study, we prospectively tested the predictive capability and feasibility of the mathematical formula of cardiovascular events in HF patients to strengthen the feasibility of the creation of the mathematical formula to predict the probability of the cardiovascular events.

Then we explained how we performed to create the mathematical formula of \( \tau = f(x_1, \ldots, x_n) \) in the previous study. Since we obtained 402 parameters at the discharge following the hospitalization due to ADHF from the data of careful history-taking, physical examinations, laboratory tests, chest X-rays, electrocardiograms, complete Doppler echocardiographic studies, coronary angiography, right heart catheterization, cardiac scintigraphy, cardiovascular magnetic resonance, cardiopulmonary exercise testing and polysomnography in patients with HF, we hypothesized that all or some of the parameters influence the time of cardiovascular events to some extents, and we quantitatively assessed the occurrence probability of the cardiovascular events using the probability model based on the Poisson process. Thus, the probability density \( p(t) \) for the cardiovascular events of patient \( i \) at an elapsed time \( t \) after discharge is represented by the following exponential formula:

\[
p_i(t) = \frac{1}{\tau_i} \exp\left(-\frac{t}{\tau_i}\right)
\]

A mean elapsed time \( \tau_i \) from discharge to the rehospitalization of patient \( i \) depends on some of the given clinical factors \( X' = \{x_1', \ldots, x_n'\} \) of the patient, i.e., a common subset \( X'_i \subseteq X' \) over all patients. The dependency is primarily approximated by the following inverse linear relation:

\[
\tau_i \cong \frac{1}{\sum_{j \in X'_i} \beta_j x_j' + \gamma}
\]

(2)

where the denominator represents the expected frequency of cardiovascular rehospitalization per day. \( X'_i \) is a set of values of the factors in \( X \) for patient \( i \); \( \beta_j \) is the contributing weight of the \( j \)th factor to the frequency, and \( \gamma \) is the intrinsic frequency for any patient. We considered that \( \tau_i \) of the patients are sampled from a common population distribution \( p_\tau(\tau) \). Therefore, the total probability distribution of the rehospitalization time \( P(t) \) is expected to be a superposition of Eq. (1) for various \( \tau \) sampled from \( p_\tau(\tau) \), as follows, where \( p(t) \) is \( p_i(t) \) in Eq. (1) for a general \( \tau \):

\[
P(t) = \int_0^\infty p_\tau(\tau)p(t)d\tau = \int_0^\infty p_\tau(\tau)\frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right)d\tau
\]

(3)

Then we used the following natural conjugate prior distribution for the unknown \( p_\tau(\tau) \):
where \( \tau_i \) is given by the dataset \( D \).

After several steps of the manipulation, we finally described the modeling algorithm. First, the value of every factor \( x_i^j \) for all patients \( i = 1, \ldots, n \) in \( D \) was normalized to fit into the interval \([0, 1]\) using the maximum and minimum values. This normalization to eliminate differences in the factor scales was necessary to allow for the measurement of the essential contribution of each factor’s variation to \( \tau_i \). Subsequently, we applied the equations (1) and (2) to the normalized dataset \( D_N \) to model the probabilistic rehospitalization process and we determined the model parameters \( \beta_j \) and \( \gamma \) in the equation (1) to maximize the following objective function:

\[
L(\beta_1, \ldots, \beta_p, \gamma) = \ln \left[ \prod_{i=1}^{n} \left( \sum_{j=1}^{p} \beta_j x_i^j + \gamma \right) \exp \left( -\left( \sum_{j=1}^{p} \beta_j x_i^j + \gamma \right) \gamma \right) \right] - \lambda \left( \sum_{j=1}^{p} |\beta_j| + |\gamma| \right)
\]

The first term is the log-likelihood of the model consisting of the previous equations over \( D_N \). The second term is called an \( L_1 \)-regularization term, which penalizes the coefficients of negligible factors by setting them equal to zero when the larger hyper-parameter \( \lambda \) eliminates more factors. This term avoids the over-fitting of the model to the dataset by selecting a set of effective factors \( X_i^j \) from a given \( X_i \). In our study, \( \lambda \) is tuned to be 0.02 to maintain the largest value of the equation (5) similarly to the other parameters \( \beta_j \) and \( \gamma \).

To seek the optimum parameter values of \( \beta_1, \ldots, \beta_p, \gamma \) that maximize the objective function \( L(\beta_1, \ldots, \beta_p, \gamma) \), we applied a simple greedy hill-climbing algorithm, in which the parameter values are iteratively modified toward their gradient direction \( \partial L / \partial \beta_j, \ldots, \partial L / \partial \beta_p, \partial L / \partial \gamma \). When the improvement of \( L \) becomes nearly negligible, the resulting parameter values are taken as the optima. Because this process depends on the initial values of the parameters, we repeated this optimization 100 times starting with random initial values and selected the result providing the maximum \( L \). This was how we selected 252 influential parameters among the 402 clinical parameters in the previous study.

Then we selected the most influential 50 parameters among 252 parameters and revised the mathematical formula. The 50 most influential parameters in the present study are defined as the clinical parameters with the 50 highest coefficients values shown in the previous manuscript. The number of the 50 is arbitrary and the realistic values to be collected for the prospective study.

**Protocol II: The prospective study to validate the mathematical formula.** We prospectively enrolled 213 patients with ADHF admitted between May 2013 and March 2015 in three different hospitals of National Cerebral and Cardiovascular Center (\( n = 114 \)) and Hokkaido (\( n = 80 \)) and Kyushu Universities (\( n = 19 \)) and followed...
up these patients until the time of cardiovascular events or the end of April 2016. The timing of patients’ discharge was determined by an expert team of cardiologists in charge of the HF department; discharge was recommended when patients presented no signs of decompensation such as the New York Heart Association (NYHA) Functional Classification < 3, no sign of rales, no galloping rhythm, stable blood pressure and an improvement in renal function due to an optimal treatment that followed international guidelines13. Rehospitalization was defined as hospitalization for decompensated HF and cardiovascular death was defined as the death due to the worsening of HF. The primary endpoint was the first cardiovascular event of either rehospitalization or death due to the worsening of HF.

Then we created the mathematical model for the occurrence probability of cardiovascular events. First of all, we assumed that the probability of cardiovascular events per day of patients does not change significantly from its discharge to its cardiovascular events. We defined the mathematical formula to predict the constant occurrence probability of cardiovascular events per day as follows:

$$\alpha = f(x_1, ..., x_p; \beta, c) = \beta^T X + c = \sum_{j=1}^{p} \beta_j x_j + c$$

where \( \alpha \) is the estimated occurrence probability of cardiovascular events per day for a patient, \( X = (x_1, ..., x_p)^T \) is a clinical feature vector of the patient, \( \beta = (\beta_1, ..., \beta_p)^T \) is a weight vector of the features, and \( c \) is an intercept of \( \alpha \). In this study, 50 clinical features, that is, \( p = 50 \), was used. As any event occurring with a constant probability in a given time period is generated by a Poisson process14, cardiovascular events of a patient also occur through this process with its individual \( \alpha \). Thus, the probability density for cardiovascular events of a patient at an elapsed time \( t \) after discharge is represented by the following exponential formula:

$$P(t|X; \beta, c) = \exp(-\alpha t) = \exp(- (\beta^T X + c)t) = \exp \left\{ - \left( \sum_{j=1}^{p} \beta_j x_j + c \right) t \right\}$$

(7)

Given a retrospective dataset \( D = \{(X_i, t_i)| i = 1, ..., N_R\} \) where \( X_i \) and \( t_i \) are the clinical feature vector and the elapsed days at the cardiovascular event from the discharge of a patient \( i \), respectively, the expected survival curve of patients in \( D_R \) is represented as:

$$P_{R_R}(t|\beta, c) = \int_{D_R} P(t|X; \beta, c) P_R(X) dX = \int_{D_R} \exp(- (\beta^T X + c)t) P_R(X) dX$$

$$= \frac{1}{N_{R1}} \sum_{i \in D_R} \exp[- (\beta^T X_i + c) \cdot t] = \frac{1}{N_{R1}} \sum_{j=1}^{p} \exp \left\{ - \left( \sum_{j=1}^{p} \beta_j x_j + c \right) t \right\}$$

(8)
| Clinical factors | Population (n = 167) |
|------------------|---------------------|
| Age, (years)     | 72 (60–79)          |
| Gender, Male/Female | 98/69              |

**Signs at admission**

|            |          |
|------------|----------|
| NYHA class (II/III/IV) | 52/54/61   |
| Elevated jugular venous pressure | 97 (58)    |
| S3 gallop   | 107 (64) |
| Leg edema   | 91 (54)  |

**Cause of ADHF**

|                   |          |
|-------------------|----------|
| Cardiomyopathy    | 56 (34)  |
| Hypertensive heart disease | 25 (15) |
| Ischemic heart disease | 16 (10)   |
| Valvular heart disease | 47 (28)   |
| Cardiac sarcoidosis | 7 (4)    |
| Myocarditis       | 1 (1)    |
| Others            | 15 (9)   |

**Comorbidity**

|                  |          |
|------------------|----------|
| Hypertension     | 81 (49)  |
| Diabetes mellitus| 39 (23)  |
| Hyperlipidemia   | 47 (28)  |
| Chronic Af       | 67 (40)  |
| Cerebrovascular disease | 31 (19) |
| Obstructive pulmonary disease | 10 (6) |

**Baseline characteristics at admission/at discharge**

|                      |          |
|----------------------|----------|
| Heart rate (beats/min) | 81 (69–103)/68 (60–76) |
| Systolic BP (mmHg)    | 126 (108–152)/110 (100–120) |
| Diastolic BP (mmHg)   | 70 (60–88)/69 (56–70) |
| Body weight (kg)      | 58 (49–68)/53 (44–61) |

**Laboratory factors at discharge**

|                     |          |
|---------------------|----------|
| Hemoglobin, (g/dl)  | 12.0 (10.7–13.4) |
| WBC, (/µl)          | 5600 (4600–6600) |
| Albumin, (g/dl)     | 3.9 (3.6–4.1) |
| AST, (U/l)          | 25.0 (20.5–31.5) |
| ALT, (U/l)          | 19.0 (13.0–18.0) |
| BUN, (mg/dl)        | 21.0 (16.0–30.8) |
| Creatinine, (mg/dl) | 0.9 (0.8–1.3) |
| Sodium, (mEq/l)     | 138 (135–140) |
| Uric acid, (mg/dl)  | 7.0 (5.7–8.4) |
| CRP, (mg/dl)        | 0.18 (0.04–0.53) |
| HbA1c, (%)          | 5.5 (5.2–6.2) |
| BNP, (pg/ml)        | 191 (102–413) |

**Echocardiographic factors at admission/at discharge**

|                   |          |
|-------------------|----------|
| LVDd (mm)         | 59 (49–66)/58 (49–66) |
| LVDs (mm)         | 48 (36–57)/45 (34–58) |
| %FS (%)           | 19 (11–29)/20 (13–31) |
| IVST (mm)         | 9 (8–11)/9 (8–11) |
| PWT (mm)          | 10 (8–11)/10 (8–11) |
| LAD (mm)          | 49 (43–57)/45 (39–54) |
| Pressure across tricuspid valve (mmHg) | 34 (25–42)/22 (14–27) |

**Medication at admission**

|                |          |
|----------------|----------|
| Use of dopamine | 6 (4)    |
| Use of dobutamine | 35 (21)  |
| Use of phosphodiesterase inhibitor | 28 (17) |
| Use of hANP     | 38 (23)  |
| Use of nitroglycerin | 27 (16)  |
| Use of intravenous diuretic | 76 (46)  |

**Oral medications at discharge**

|                |          |
|----------------|----------|
|                | Continued |
Clinical factors | Population (n = 167)
---|---
ACE inhibitor | 80 (48)
ARB | 55 (21)
Beta-blockers | 
Digitalis | 48 (29)
Diuretics | 151 (90)

| Clinical factors | Population (n = 167) |
|---|---|
| ACE inhibitor | 80 (48) |
| ARB | 55 (21) |

Beta-blockers
- Digitalis: 48 (29)
- Diuretics: 151 (90)

Table 1. Patient characteristics in the retrospective study. Data are given as the median (interquartile range) or n (%). ACE inhibitor, angiotensin-converting enzyme inhibitor; ADHF, acute decompensated heart failure; AF, atrial fibrillation; ARB, angiotensin II receptor blocker; BNP, B-type natriuretic peptide; BUN, Blood urea nitrogen; BP, blood pressure; CRP, C-reactive protein; FS, fractional shortening; hANP, human atrial natriuretic peptide; LVST, Interventricular septum thickness; LAD, left atrial dimension; LVDd, Left ventricular end-diastolic dimension; LVDs, Left ventricular end-systolic dimension; NYHA, New York Heart Association; PWT, posterior wall thickness.

where 

\[ P_R(X) \] is the population distribution of the retrospective dataset \( D_R \). Conversely, we directly derived the Kaplan–Meier survival curve \( P_R(t) \) using \( D_R \) by following a standard procedure. Then, we estimated the best parameter values of \( \beta \) and \( c \), which minimize the following Kullback–Leibler divergence (KL-divergence)\[16\]. The KL-divergence is a well-known statistical measure to reveal the discrepancy between two probability distributions.

\[ KL(P_R, P_R(\beta, c)) = \int P_R(t) \log \left( \frac{P_R(t)}{P_R(\beta, c)} \right) dt \]

\[ = \sum_{N_R \in \mathbb{R}_{\geq0}} \left[ \log \left( \frac{P_R(t)}{P_R(\beta, c)} \right) \right] \]

\[ = \min \left( \sum_{N_R \in \mathbb{R}_{\geq0}} \left[ \log \left( \frac{P_R(t)}{P_R(\beta, c)} \right) \right] \right) \]

where \( D_R \) is a dataset excluding the patients whose observations are censored and, thus, do not have \( t \), in \( D_R \). The parameters \( \beta \) and \( c \) minimizing this measure are determined by using the Nelder–Mead method\[17\], which is a renowned non-linear optimization algorithm.

We used these estimated parameter values of \( \beta \) and \( c \) to predict the survival curve of a given prospective dataset \( D_P = \{ X_i, t_i \}_{i=1}^{N_P} \) where \( N_P = 213 \) in our case. The predicted survival curve was obtained by substituting the above-mentioned best values of \( \beta \) and \( c \) and the clinical feature vectors \( X_i \) of patients in \( D_p \) to the following 

\[ P_R(\beta, c) = \int_{D_P} P(t; \beta, c) dX = \int_{D_P} \exp\left( -((\beta^T X + c) t) \right) P(X) dX \]

\[ = \frac{1}{N_p} \sum_{X_i \in D_P} \exp\left( -((\beta^T X_i + c) t) \right) = \frac{1}{N_p} \sum_{X_i \in D_P} \exp\left( -\left( \sum_{j=1}^{p} \beta_j x_{ij} + c \cdot t \right) \right) \]

We compared this predicted curve for the prospective dataset \( D_P \) and the Kaplan–Meier survival curve\[15\] \( P_R(t) \) directly derived from \( D_R \).

**Statistical Analysis.** Normally distributed data were expressed as mean ± standard deviation; other values were reported as a median and interquartile range (IQR). We conducted the goodness-of-fit test and used the coefficient of determination as a measure to assess the significant relationships between the predictive curves and actual Kaplan–Meier curves of the cardiovascular event-free rate. The differences in the predictive curves were tested using the Wilcoxon signed-rank test. We estimated the error bounds of the parameters, \( \alpha \) and \( \beta \), by applying the standard bootstrap sampling\[16\]. All tests were two-tailed, and \( P < 0.05 \) was considered significant. All analyses were performed using the JMP software for Windows (version 8.0.2, SAS Inc., Cary, NC).

**Results**

**Patients characteristics.** In the retrospective study (Protocol I), the clinical characteristics of the patients are summarized in Table 1. In 78 patients, cardiovascular events (\( n = 71 \) for HF rehospitalization, and \( n = 14 \) for HF-related death) occurred at a median time of 260 days after discharge and the remaining 89 patients had no cardiovascular events by a median time of 859 days after discharge (range, 515–1194 days). Among clinical parameters, we selected the highest coefficient values of 50 parameters without the data of cardiac catheterization; the 50 clinical parameters with coefficient values for constructing the mathematical formula are depicted in Table 2. In the prospective study (Protocol II), the clinical characteristics of 213 patients are summarized in Table 3. Of these, 84 patients were readmitted to each hospital at a median time of 161 days after discharge, and 21 patients died due to worsening of HF at a median time of 275 days; the remaining 114 patients had no cardiovascular events by a median time of 636 days after discharge (range, 183–898 days).
| Clinical factors                           | Maximum Values | coefficient values |
|------------------------------------------|----------------|--------------------|
| Age, (years)                             | 72 (60–79)     | 91 1.789           |
| Gender, Male/Female                      | 98/69          | 1 0.362            |
| NYHA class (II/III/IV) at admission      | 52/54/61       | 4 0.022            |
| Heart rate at admission (beats/min)      | 81 (69–104)    | 142 0.539          |
| Leg edema                                | 91 (54)        | 1 –0.692           |
| Cause of ADHF                            |                |                    |
| Cardiomyopathy                           | 56 (34)        | 1 –0.439           |
| Hypertensive heart disease               | 25 (15)        | 1 –0.363           |
| Ischemic heart disease                   | 16 (10)        | 1 10.842           |
| Valvular heart disease                   | 47 (28)        | 1 –0.282           |
| Comorbidity                              |                |                    |
| Hypertension                             | 81 (49)        | 1 0.134            |
| Hyperlipidema                            | 47 (28)        | 1 0.288            |
| Chronic Af                               | 67 (40)        | 1 –0.185           |
| Cerebrovascular disease                  | 31 (19)        | 1 –0.108           |
| Obstructive pulmonary disease            | 10 (6)         | 1 –0.604           |
| CRT                                      | 35 (20)        | 1 –0.719           |
| ICD                                      | 35 (20)        | 1 –0.554           |
| Pacemaker                                | 14 (8)         | 1 0.752            |
| Number of family members                 | 1 (1–2)        | 6 –0.525           |
| Albumin at admission, (g/dl)             | 3.7 (3.4–4.0)  | 4.9 –1.773         |
| CRP at admission, (mg/dl)                | 0.3 (0.1–0.9)  | 24.5 –1.000        |
| LVDs at admission, (mm)                  | 48 (36–57)     | 79 –0.085          |
| %FS at admission, (%)                    | 19 (11–29)     | 65 0.567           |
| WBC at admission, (×10⁶)                 | 6500 (5000–8850) | 23500 0.661    |
| AST at discharge, (U/l)                  | 25.0 (20.5–21.5) | 575 –3.374       |
| BUN at discharge, (mg/dl)                | 21.0 (16–30.8) | 133 0.772          |
| Uric acid at discharge, (mg/dl)          | 7.0 (5.7–8.4)  | 13.1 –0.051        |
| CRP at discharge, (mg/dl)                | 0.18 (0.04–0.53) | 17.22 –1.265    |
| BNP at discharge, (pg/ml)                | 191 (102–413)  | 3257 –0.826        |
| %FS at discharge, (%)                    | 20 (13–31)     | 53 –1.013          |
| IVST at discharge, (mm)                  | 9 (8–11)       | 17 –0.889          |
| AR grade (≥II) at discharge              | 21 (13)        | 3 –0.422           |
| MR grade (≥II) at discharge              | 48 (29)        | 4 0.090            |
| TR grade (≥II) at discharge              | 43 (26)        | 4 0.379            |
| Oral medications at discharge            |                |                    |
| ACE inhibitor                             | 80 (48)        | 1 0.017            |
| Anti-allergic                             | 12 (7)         | 1 –0.335           |
| Anti-inflammatory drug                    | 5 (3)          | 1 –0.563           |
| Antiplatelet                              | 45 (27)        | 1 0.252            |
| Antithyroid drug                          | 2 (1)          | 1 –1.751           |
| Beta-blockers                             | 109 (65)       | 1 0.720            |
| Bronchodilator                           | 7 (4)          | 1 –0.489           |
| Cardiotoxic drug                          | 22 (13)        | 1 –0.466           |
| Choleretic drug                           | 10 (6)         | 1 –0.847           |
| Digitalis                                 | 48 (29)        | 1 0.194            |
| Diuretics                                 | 151 (90)       | 1 0.971            |
| Intestinal disease drug                   | 4 (2)          | 1 –0.163           |
| Lipid-lowering drug                       | 37 (22)        | 1 0.307            |
| Proton pump inhibitor                     | 60 (36)        | 1 –0.023           |
| Purgative                                 | 49 (29)        | 1 0.279            |
| Sedative-hypnotic (benzodiazepin)         | 36 (22)        | 1 –0.392           |
| Vitamins                                  | 14 (8)         | 1 0.211            |
| (Constant)                                |               | 0.9014             |

Table 2. The 50 clinical parameters with coefficient values for constructing the mathematical formula. Data are given as the Median (interquartile range) or n (%). ACE inhibitor, angiotensin-converting enzyme inhibitor; ADHF, acute decompensated heart failure; Af, atrial fibrillation; AR, aortic regurgitation; BNP, B-type natriuretic peptide; BUN, Blood urea nitrogen; CRT, cardiac resynchronization therapy; CRP, C-reactive protein; FS, fractional shortening; ICD, Implantable Cardioverter Defibrillator; VST, interventricular septum thickness; LVDs, Left ventricular end-systolic dimension MR, mitral regurgitation; NYHA, New York Heart Association; TR, tricuspid regurgitation.
| Clinical factors                      |     |     |
|---------------------------------------|-----|-----|
| Age, (years)                          | 71  | (60–79) |
| Gender, Male/Female                   | 135/78 |
| NYHA class (II/III/IV) at admission   | 47/108/58 |
| Heart rate at admission (beats/min)   | 80  | (68–97) |
| Leg edema                             | 136 | (64) |

**Cause of ADHF**

|                      |     |     |
|----------------------|-----|-----|
| Cardiomyopathy       | 76  | (36) |
| Hypertensive heart disease | 29 | (14) |
| Ischemic heart disease | 26 | (12) |
| Valvular heart disease | 47 | (22) |

**Comorbidity**

|                      |     |     |
|----------------------|-----|-----|
| Hypertension         | 108 | (51) |
| Hyperlipidemia       | 73  | (34) |
| Chronic AF           | 104 | (49) |
| Cerebrovascular disease | 14 | (7) |
| Obstructive pulmonary disease | 2 | (3) |
| CRT                  | 24  | (11) |
| ICD                  | 31  | (15) |
| Pacemaker            | 25  | (12) |
| Number of family members in the same household | 1 | (1–2) |

**Albumin at admission, (g/dl)**

|                      |     |     |
|----------------------|-----|-----|
| 3.8 (3.5–4.1)        |     |     |

**CRP at admission, (mg/dl)**

|                      |     |     |
|----------------------|-----|-----|
| 0.4 (0.1–1.1)        |     |     |

**LVDs at admission, (mm)**

|                      |     |     |
|----------------------|-----|-----|
| 47 (34–58)           |     |     |

**%FS at admission, (%)**

|                      |     |     |
|----------------------|-----|-----|
| 18 (10.3–30.6)       |     |     |

**WBC at admission, (µl)**

|                      |     |     |
|----------------------|-----|-----|
| 5240 (4200–6400)     |     |     |

**AST at discharge, (U/l)**

|                      |     |     |
|----------------------|-----|-----|
| 23 (18–30)           |     |     |

**BUN at discharge, (mg/dl)**

|                      |     |     |
|----------------------|-----|-----|
| 23 (18–35)           |     |     |

**Uric acid at discharge, (mg/dl)**

|                      |     |     |
|----------------------|-----|-----|
| 6.5 (5.3–7.9)        |     |     |

**CRP at discharge, (mg/dl)**

|                      |     |     |
|----------------------|-----|-----|
| 0.2 (0.1–0.6)        |     |     |

**BNP at discharge, (pg/ml)**

|                      |     |     |
|----------------------|-----|-----|
| 244.8 (117–457)      |     |     |

**%FS at discharge, (%)**

|                      |     |     |
|----------------------|-----|-----|
| 20.9 (11–31.7)       |     |     |

**IVST at discharge, (mm)**

|                      |     |     |
|----------------------|-----|-----|
| 10 (9–11)            |     |     |

**AR grade (≥II) at discharge**

|                      |     |     |
|----------------------|-----|-----|
| 26 (12)              |     |     |

**MR grade (≥II) at discharge**

|                      |     |     |
|----------------------|-----|-----|
| 93 (44)              |     |     |

**TR grade (≥II) at discharge**

|                      |     |     |
|----------------------|-----|-----|
| 59 (28)              |     |     |

**Oral medications at discharge**

|                      |     |     |
|----------------------|-----|-----|
| ACE inhibitor         | 111 | (52) |
| Anti-allergic         | 8   | (4)  |
| Anti-inflammatory drug | 48 | (23) |
| Antiplatelet          | 27  | (13) |
| Antithyroid drug      | 3   | (1)  |
| Beta-blockers         | 161 | (76) |
| Bronchodilator        | 2   | (1)  |
| Cardiotonic drug      | 36  | (17) |
| Choleretic drug       | 11  | (5)  |
| Digitalis             | 42  | (20) |
| Diuretics             | 181 | (85) |
| Intestinal disease drug | 19 | (9)  |
| Lipid-lowering drug   | 79  | (37) |
| Proton pump inhibitor  | 119 | (56) |
| Purgative             | 63  | (30) |
| Sedative-hypnotic (benzodiazepin) | 12 | (6)  |
| Vitamins              | 7   | (3)  |

Table 3. The selected 50 clinical parameters in the prospective study. Data are given as the median (interquartile range) or n (%). ACE inhibitor, angiotensin-converting enzyme inhibitor; ADHF, acute decompensated heart failure; AF, atrial fibrillation; AR, aortic regurgitation; BNP, B-type natriuretic peptide; BUN, Blood urea nitrogen; CRT, cardiac resynchronization therapy; CRP, C-reactive protein; FS, fractional shortening; ICD, Implantable Cardioverter Defibrillator; VST, interventricular septum thickness; LVDs, Left ventricular end-systolic dimension MR, mitral regurgitation; NYHA, New York Heart Association; TR, tricuspid regurgitation.
Predictive capability of the mathematical formula for the prospective outcomes. We confirmed that the Kaplan–Meier curves using this formula and actual data in the retrospective study revealed the proper fitting of the probability of cardiovascular outcomes (Fig. 1). Then, in the prospective study, we just analyzed the prospective data using only our institute. Figure 2 shows that the mathematical formula obtained from the retrospective study can predict the clinical outcomes observed in the prospective study. Thus, we tested whether our formula can predict the probability of cardiovascular events in all the institutes, and we found that our formula can predict the clinical outcomes for three institutes (Fig. 3).

The factors that provoke or prevent cardiovascular events in 50 clinical factors. Since we found that the mathematical formula applies to predict the occurrence of cardiovascular events in the prospective study, we assumed that each attribute coefficient for this mathematical formula is also essential for the clinical practice for HF (Table 2). When we investigated the contribution of each parameter for the objective measure, we found that ischemic heart disease results in a worse prognosis. In the physical examination, high heart rate or implantation of pacemaker classification was the worse factor, and the implantation of cardiac resynchronization therapy or implantable cardioverter defibrillator demonstrated better outcomes. Furthermore, the data of blood analysis, echocardiography and oral medications related to the cardiovascular events in the complex and confounding manners. Intriguingly, the number of family members resulted in a better prognosis.

Discussion
This study provided the evidence that the mathematical formula using the retrospective clinical data provides the occurrence probability of cardiovascular events in the prospective study in patients with HF. We were able to derive the formula of \( \alpha = f(x_1, \ldots, x_{50}) \), where \( \alpha \) is the probability of the cardiovascular events and \( x_1, \ldots, x_{50} \) are clinical factors observed before cardiovascular events, which could prospectively predict the occurrence probability of cardiovascular events. This study proposes the novel idea that the occurrence probability of future cardiovascular events can be mathematically formulated and deduced from the retrospective clinical and personal parameters before the time of cardiovascular events.

Importantly, we found that the occurrence probability depends not only on the cardiac dysfunction but also the dysfunction parameters of other organs, such as the kidneys and liver, and social factors, such as the number of family members living with a patient. Therefore, we can regard the occurrence probability as the overall severity of HF. This concept is well matched to the idea that we need to investigate the effect of certain treatment of HF by judging the mortality or morbidity, but not by cardiac function in large-scale clinical trials. The mortality or morbidity during a certain observation period is depicted by the Kaplan–Meier curves, which represent the occurrence probability of cardiovascular events.

What is the differences between the present and previous studies to assess clinical outcomes? The earlier studies, including ours, have merely identified the important factors for cardiovascular outcomes using the cohort data of patients with HF. In such studies, clinical data are retrospectively or prospectively collected and identified the most influential factors using the multivariate analysis. However, no researcher has tested whether such
multiple factors can quantitatively predict the occurrence probability of future cardiovascular events. Most of all, arbitrary factors, which are unintentionally collected by investigators and usually ignored, may be essential factors to explain the occurrence probability, and the investigator-intended analysis of the data cannot cover such arbitrary factors beyond expectation. This is the concept of analysis of big data or data mining analysis.22 Wang et al.23 revealed that although multiple biomarkers are associated with a high relative risk of adverse events, even the combination of these factors only moderately improved the prediction of risk in an individual. This suggests that the occurrence of cardiovascular events may not be well predictable even after the multiple factors are convoluted. In contrast, we collected almost all the numerical data in the medical records documented before the onset of cardiovascular events and solved the mathematical formula using these parameters to provide the exact probability for future cardiovascular events. Of more than 250 clinical factors that constitute the original mathematical formula,8 we selected the 50 most influential factors and re-solved the mathematical formula. The mathematical formula using these 50 factors potentially validates its plausibility for the calculation of the occurrence probability of cardiovascular events in patients with HF, suggesting that we need more clinical data to predict the future outcomes or obtain the mathematical formula for the prediction than we expected. WBC values at admission may approximately indicate the unique value of each patient. On the other hand, the most abnormal values at the admission may determine the severity of the pathophysiology of CHF.

Moreover, the pathophysiology and treatment strategy of HF are common worldwide, suggesting that such formulas should be valid to provide the future occurrence of cardiovascular events in other countries. Of course, the concept to create a mathematical formula should be translated and transmitted worldwide to know the real risk of cardiovascular events and to treat the clinical factors using their data in patients with HF.

There are several applications and limitations for the present study. First of all, since these 50 clinical parameters can be easily provided in outpatient or inpatient clinics, we can evaluate the severity of HF from the viewpoint of the probability of the onset of cardiovascular events in each patient. Secondly, we can identify what clinical factors increase the probability of cardiovascular events, suggesting that we can identify the target of the treatment of HF in each patient. Thirdly, this formula may provide the educational tool for the HF patients. Fourthly, the concept of the creation of formula to predict the clinical outcomes may be applicable to the other fields such as cerebral infarction or cancers.4 On the other hand, we have some limitations of the present formula because we created the formula using the data of the HF patients with mild to moderate severity of HF symptom. Therefore, we are not able to apply the present formula to the severe HF patients to predict the occurrence probability of the cardiovascular events because we did not derive the present equation from the cohort of severe HF patients. To respond this requirement, we need to create the mathematical formula using the data of the severe HF patients.

**Conclusions**

We created a mathematical formula that precisely provides the probability of the clinical outcomes of patients who are hospitalized with ADHF and discharged after appropriate treatment. Mathematics using the present cardiovascular big data may predict the occurrence probability of future cardiovascular events. Since we found the importance of the clinical parameters independent of cardiac function, it merits the better treatment of HF.
References

1. Go, A. S. et al. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. Circulation 129, e28–e292, https://doi.org/10.1161/CIR.0000000000000148 (2014).

2. Braunwald, E. Biomarkers in heart failure. The New England journal of medicine 358, 2148–2159, https://doi.org/10.1056/NEJMr0800239 (2008).

3. Fonarow, G. C., Peacock, W. E., Phillips, C. O., Givertz, M. M. & Lopatin, M. Admission B-type natriuretic peptide levels and inhospital mortality in acute decompensated heart failure. Journal of the American College of Cardiology 49, 1943–1950, https://doi.org/10.1016/j.jacc.2007.02.037 (2007).

4. Abraham, W. T. et al. Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). Journal of the American College of Cardiology 52, 347–356, https://doi.org/10.1016/j.jacc.2008.04.028 (2008).

5. Mancini, D. M. et al. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation 83, 778–786 (1991).

6. Ioh, H., Taniguchi, K., Koike, A. & Doi, M. Evaluation of severity of heart failure using ventilatory gas analysis. Circulation 81, ii31–37 (1990).

7. Collins, F. S. & Varmus, H. A new initiative on precision medicine. The New England journal of medicine 372, 793–795, https://doi.org/10.1056/NEJMp1500523 (2015).

8. Yoshida, A. et al. Derivation of a mathematical expression for predicting the time to cardiac events in patients with heart failure: a retrospective clinical study. Hypertension research: official journal of the Japanese Society of Hypertension 36, 450–456, https://doi.org/10.1038/hr.2012.200 (2013).

9. Newton, I. The Principia: mathematical principles of natural philosophy. (Univ of California Press, 1999).

10. Bakken, A. H., Hesse, W. B., Bates, D. H. & Green, L. J. Understanding and using electronic health records. Jama 311, 1380–1386, https://doi.org/10.1001/jama.2013.277667 (2014).

11. Shah, S. H. et al. Opportunities for the Cardiovascular Community in the Precision Medicine Initiative. Circulation 133, 226–231, https://doi.org/10.1161/circulationaha.115.019475 (2016).

12. McKee, P. A., Castelli, W. P., McNamara, P. M. & Kannel, W. B. The natural history of congestive heart failure: the Framingham study. The New England journal of medicine 285, 1441–1446, https://doi.org/10.1056/nejm197112232851471 (1971).

13. Hunt, S. A. et al. Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. Journal of the American College of Cardiology 53, e1–e90, https://doi.org/10.1016/j.jacc.2008.11.013 (2009).

14. Guillemb, I. Mathematics: from the birth of numbers. (WW Norton & Company, 1997).

15. Elron, B. Logistic regression, survival analysis, and the Kaplan-Meier curve. Journal of the American statistical Association 83, 414–425 (1988).

16. Bishop, C. M. Pattern recognition. Machine Learning 128, 1–58 (2006).

17. McKinnon, K. I. M. Convergence of the Nelder–Mead Simplex Method to a Nonstationary Point. SIAM Journal on Optimization 9, 148–158, https://doi.org/10.1137/s1052623496303482 (1998).

18. Ferreira-Gonzalez, I. et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ (Clinical research ed.) 334, 786, https://doi.org/10.1136/bmj.39136.682083.AE (2007).

19. Ohara, T. et al. Plasma adiponectin is associated with plasma brain natriuretic peptide and cardiac function in healthy subjects. Hypertension research: official journal of the Japanese Society of Hypertension 31, 825–831, https://doi.org/10.1259/hypres.31.825 (2008).

20. Chen, C. Y. et al. Serum blood urea nitrogen and plasma brain natriuretic Peptide and low diastolic blood pressure predict cardiovascular morbidity and mortality following discharge in acute decompensated heart failure patients. Circulation journal: official journal of the Japanese Society of Hypertension 76, 2372–2379 (2012).

21. Chanson-Rolle, A., Aubin, F., Brasco, V., Hamasaki, T. & Kitakaze, M. Influence of the Lactotripeptides Isoleucine-Proline-Proline and Valine-Proline-Proline on Systolic Blood Pressure in Japanese Subjects: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. PloS one 10, e0142235, https://doi.org/10.1371/journal.pone.0142235 (2015).

22. Zuo, K. et al. Predicting 30-day Hospital Readmission with Publicly Available Administrative Database. A Conditional Logistic Regression Modeling Approach. Methods of information in medicine 54, 560–567, https://doi.org/10.3414/me14-02-0017 (2015).

23. Li, G. et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet 371, 1783–1789, https://doi.org/10.1016/s0140-6736(08)60766-7 (2008).

24. Tanaka, G., Hirata, Y., Goldberg, S. L., Bruchovsky, N. & Aihara, K. Mathematical modelling of prostate cancer growth and its application to hormone therapy. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 368, 5029–5044, https://doi.org/10.1098/rsta.2010.0221 (2010).

Acknowledgements

This study is supported by Grants-in-aid from the Ministry of Health, Labor, and Welfare-Japan, Grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology-Japan, Grants-in-aid from Japan Agency for Medical Research and Development (JP17ek0210080).

Author Contributions

We would like to identify the role of each author in this manuscript as follows. Study concept and design: T.W., M.K.; Data collection, M.S., J.K., T.I., S.K., A.F., H.T.; Data analysis: H.F., A.I., S.I., H.A., M.A.; Figures and Tables: H.A., M.A.; Writing: M.K.

Additional Information

Competing Interests: Nothing to disclose for M.S., H.F., J.K., T.I., S.K., A.F., I.A., S.I., H.A. and T.W. H.T. reports personal fees from Astellas, Otsuka, Takeda, Daiichi–Sankyo, Tanabe–Mitsubishi, Boehringer Ingelheim, Novartis, Bayer, and Bristol Myers Squibb. M.A. reports personal fees from Bayer, Ono, Otsuka, Pfizer, Sanofi, and Takeda. M.K. reports grants from Japanese government, during the conduct of the study; grants from Japanese government, Japan Heart Foundation, Japan Cardiovascular Research Foundation, Novartis, and Nihon Kohden; and personal fees from Astellas, Pfizer, Tanabe-Mitsubishi, Ono, AstraZeneca and Kureha; personal fees from Daiichi-sankyo, Bayer, Kowa, MSD, Shionogi, Taisho-Toyama and Toaeyo

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018