This study proposes a new method for a prognostic analysis on the course of diseases. This method aims to separate, among same disease population, several prognostic groups showing different temporal patterns in the course of the disease and evaluate the influence of prognostic factors separately for each group. To do this, we develop a new temporal distribution that we call multiple component Weibull distribution, which is given by a linear combination of Weibull distributions with different shape and scale parameters.

This method is applied to the prognostic analysis of cardiac death of acute myocardial infarction (AMI) during hospitalization. The subjects are the first year study population enrolled in the 5-year prospective survey conducted by the Heart Institute, Tokyo Women’s University. We separated the prognosis of AMI into two prognostic groups, one with early death of initial failure type and the other with late death of weak abrasion failure type. It was found that a prognostic factor of early death, mostly caused by cardiogenic shock or cardiac failure, was Killip classification, whereas, in the late death, mostly caused by cardiac rupture or perforation, Q type MI and age other than Killip classification were found to be influential prognostic factors.

Key words: Weibull distribution, Prognostic analysis, Cox life regression method, Acute myocardial infarction, Killip classification.

Introduction

Many prominent statistical methods have been proposed to investigate prognostic factors influencing the course of diseases, such as multiple logistic regression analysis\(^1^-^3\) and Cox life regression model\(^4\). These methods evaluate important patient covariates, or predictors that have statistically significant impacts on the outcome of disease ("proportional hazard hypothesis")\(^5\). But even in the Cox life regression analysis that estimates the influences of the covariates on the temporal profile of a disease, it is assumed that the way in which these covariates affect the course of disease is uniform over time. But, in the actual cases of a prognostic analysis, each prognostic factor often exerts its influence in the temporally different way: it might have an impact only at the onset of disease, or later stage of the disease. Even in the population of same disease, there might be several groups having different prognoses. Current prognostic analysis methods cannot be efficient in dealing with such situations.

In this study we propose a new alternative method that we call temporal decomposition of course of diseases. This method aims to separate, among same disease population, several prognostic groups showing different temporal patterns in the course of the disease and evaluate the influences of predictor covariates separately for each prognostic group. To do this, we develop a new temporal distribution that we call multi-
component Weibull distribution. In this distribution, a linear combination of Weibull distributions with different shape and scale parameters is used to decompose course of disease. It is thought to be appropriate to use Weibull distribution as a basis function for temporal decomposition because its shape parameter can describe the various temporal modes of occurrence of events.

To evaluate its efficiency, our method is applied to the first year results of the five year prospective survey of acute myocardial infarction (AMI), which is planed and being conducted by the Heart Institute, Tokyo Women’s Medical University. The enrolled patients in the first year are 1295 with 117 cardiac deaths during hospitalization. We applied our method to this population of AMI patients to separate early and late death group and clarify the cause of the difference in the course of disease of two groups.

Materials

The acute myocardial infarction was selected as the subjects of the disease for our prognostic analysis. In January 1999, the Heart Institute of Japan in Tokyo Women’s Medical University began to conduct a five-year prospective study of AMI with other 16 related hospitals. The purpose of this study was to clarify prognostic factors related to myocardial infarction and patients were enrolled into the study, who were admitted with confirmed AMI to the 17 hospitals including the Heart Institute of Japan and Sakakibara Memory Hospital. The diagnosis of AMI requires the definite criteria of AMI so that we used definition WHO MONICA project.

A standardized data set was collected prospectively from the onset of AMI and during the hospitalization. It includes details of clinical presentation (age, sex, admission, Killip classification), medical history (time and date of onset and arrival at hospitals), coronary risk factors (previous ischemic heart disease, hypertension, diabetes mellitus, hyperlipidemia etc.), cigarette smoking status, place of the MI (prior, anterior, inferior MI), non-Q type or Q type myocardial infarction, treatment (whether or not, reperfusion therapy, percutaneous transluminal coronary angioplasty, or coronary artery bypass graft), and prognosis (discharge, cardiac death, non-cardiac death). As to the prognosis after discharge, clinical events such as recurrence of the myocardial infarction were also recorded.

This survey is now still being continued. We used the patient population who were enrolled in the first year of our survey, 1999. After five years follow up is completely finished, we will publish the whole results of the study. Among 1321 patients, 27 patients attacked with the acute myocardial infarction during hospitalization were excluded in the study object, so in our study, we used 1294 patients as study subjects. The average age was 67.08 ± 12.36 with 893 males (70.2%) and 380 females (29.8%) as shown in Table 1. The distribution of major attributes of patients were: for abnormal Q wave in ECG, Non-Q type (33.4%) and Q type (65.8%), and for Killip classification, Killip (77.0%), Killip (8.1%), Killip (7.6%) and Killip (7.3%), respectively. For other major attributes are shown in Table 1.

Methods

In our prognostic analysis of the course of the disease, we propose a new method to decompose the course of the disease. This method is comprised of three steps and it separates different modes of disease progresses contaminated in the overall course of the disease. Our method can be applied to any kinds of diseases so that we describe it in general way below,

<table>
<thead>
<tr>
<th>Table 1. Clinical characteristics of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (age years)</td>
</tr>
<tr>
<td>Female (age years)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Onset to arrival (h)</td>
</tr>
<tr>
<td>Time to reperfusion therapy(h)</td>
</tr>
<tr>
<td>Reperfusion therapy</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Prior MI</td>
</tr>
<tr>
<td>Anterior MI</td>
</tr>
<tr>
<td>Inferior MI</td>
</tr>
<tr>
<td>Killip class I</td>
</tr>
<tr>
<td>Killip class II</td>
</tr>
<tr>
<td>Killip class III</td>
</tr>
<tr>
<td>Killip class IV</td>
</tr>
<tr>
<td>NQMI</td>
</tr>
<tr>
<td>QMI</td>
</tr>
</tbody>
</table>

*Means±SD
MI: myocardial infarction
Killip class indicated the degree of congestive heart failure
NQMI: non-Q wave myocardial infarction
QMI: Q-wave myocardial infarction
WEIBULL PROGNOSTIC ANALYSIS

referring also to the application to AMI.

Description of overall profile of disease progression

First step in our prognostic analysis is to describe the overall profile of the course of disease by using the following functions:

(1) Survival function \( S(t) \): Most widely used description for disease profile is Kaplan-Meier estimate of time-to-event distribution in the presence of censored cases. In our prognostic analysis, Kaplan-Meier method is used to construct the survival function \( S(t) \) of the patients, with cardiac deaths as events and with discharges and non-cardiac deaths as censored cases.

(a) Hazard function \( h(t) \): Another method to describe profile of disease progression was hazard function that calculates the (instantaneous) mortality \( h(t) \), the rate of the death within designated time unit \((t, t + \Delta t)\), given that the subject survives until time point \( t \). We depict the hazard function for cardiac death during hospitalization with non-cardiac deaths and discharges as the censored data.

Decomposition into the different modes of temporal progression

The overall profile of the course of a disease is, in some diseases, thought to be the sum of those of different prognostic groups of the study population: there might be a poor prognostic group with a very fast disease progress, or a moderate prognostic group with a slow disease progress. To decompose those different modes of a disease progress, we made use of the statistical theory of a failure analysis.

(a) Weibull distribution and failure mode: In the failure analysis, the mode of a failure occurrence is typically classified into 3 types, which are an initial failure, abrasion (wore-away) failure and random failure according to the temporal pattern of failure rate: in the initial failure type, a failure rate is larger at the beginning and decreases monotonously, whereas in the abrasion failure type, the failure rate increases with the elapse of time and in the random failure type, the failure rate is constant regardless of time. The failure occurrence probability can be mathematically modeled by Weibull distribution, which is given as

\[
W(t \mid m, \tau) = 1 - \exp \left\{ - \left( \frac{t}{\tau} \right)^m \right\} \quad (t \geq 0)
\]

where \( W(t \mid m, \tau) \) is Weibull distribution that describes the cumulative probability of failure with time \( t \), \( m \) is shape parameter and \( \tau \) is scale parameter which gives us the time constant (time needed to 63.3% of total change). The most important feature of Weibull distribution is that it can describe three types of the failure curve according to the value of the parameter \( m \): it represents the initial failure type when the value of \( m \) is \( 0 < m < 1 \), random failure when \( m = 1 \), and the abrasion failure type when the value of \( m \) is \( m > 1 \) (Fig. 1).

(b) Multiple component Weibull distribution: We use this Weibull distribution for our temporal disease decomposition of a disease progression. In order to describe the case where the overall disease profile has mixed modes of a disease progression, we propose multiple component Weibull distribution. This distribution is composed of more than two components of Weibull distribution with some constant survival components. For example, the survival function for 2-component Weibull distribution is given by

\[
S(t) = p_1 \exp \left\{ - \left( \frac{t}{\tau_1} \right)^{m_1} \right\} + p_2 \exp \left\{ - \left( \frac{t}{\tau_2} \right)^{m_2} \right\} + p_0.
\]

where \( S(t) \) is survival function, \( p_1 \) and \( p_2 \) are ratios or prior probabilities of first and second component of different prognostic groups, and \( p_0 \) is the constant bias ratio of the study population that would not die by the studied disease at the end of the study. Obviously, \( p_0 = 1 - p_1 - p_2 \). We can generalize this composite distribution to that having more than two components. In a more general form, multiple component Weibull distribution can be described by

\[
S(t) = \sum_{i=1}^{n} p_i \exp \left\{ - \left( \frac{t}{\tau_i} \right)^{m_i} \right\} - p_0.
\]

In typical cases of temporal profile decomposition of a disease progression, the overall survival function is composed of fast (early) component and slow (late) component. In that case, the 2-component Weibull distribution of equation \( (2) \) is sufficient, with fast and slow Weibull distribution.

(c) Temporal profile decomposition of disease progression: To decompose the overall survival function, we applied this multiple component Weibull model to the overall function estimated by Kaplan-Meier method in the first step of our method. In concrete, if we use the 2-component Weibull distribution, we should estimate 6 parameters, \( p_1, p_2, m_1, m_2, \tau_1, \tau_2 \) to fix the model. To determine these parameters, we use maximum likelihood method. In this method the following likelihood function, is maximized with respect to parameters, given data set of the time of the death \((t_1, t_2, ..., t_d)\) with the times of censorship \((t^*_1, t^*_2, ..., t^*_c)\).
\[ L(p_1, p_2, m_1, m_2, \tau_1, \tau_2, | t_1, t_2, \ldots, t_d, t_1^*, t_2^*, \ldots, t_c) \]

\[ = \prod_{k=1}^{d} f(t_k) \prod_{k=1}^{c} S(t_k), \]

where \( f(t) \) is the probability density function (PDF) of the multiple component Weibull distribution, and is given by differentiating \( S(t) \) with respect to \( t \).

In this prognostic study of AMI, we used 2-component Weibull distribution. We can determine the appropriateness of this model by inspecting the shape of the overall survival function. But in more general and strict approach, the number of the components can be determined by applying Akaike Informative Criteria (AIC) method that selects the optimal number of components from the survival data.

**Prognostic factor analysis of each component**

After we separate the different components of disease progression, we extract the prognostic factors influencing each component. We apply Cox life regression method to each component separated in the previous step (c) of this analysis. In the Cox life regression model, the variation of the hazard function \( h(t) \) or survival function \( S(t) \) is explained by patient's covariates, with the following proportional hazard model,

\[ h(t) = h_0(t) \exp \{ \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_s x_s \}, \]

where \( h_0(t) \) is baseline function which gives the hazard
function of the patient having mean covariate values. The variation of the hazard function when covariates of a patient are varied from their mean values is described by a proportional factor that is given by an exponent function of linear combination of the covariates. Covariates are standardized with the mean 0 and standard deviation 1.0, so that coefficient $b_i (i = 1, \ldots, s)$ describes the magnitude of contribution to the hazard function. The proportional hazard function can be rewritten in the terms of a survival function, 

$$ S(t) = \{ S_0(t) \}^{\exp (b_1 x_1 + b_2 x_2 + \cdots + b_s x_s)}, $$

where $S_0(t)$ is the baseline survival function. The difference of the disease progression profile can be explained by the contribution pattern of the covariates.

**Result**

**Temporal profile decomposition to overall survival function of AMI**

Following our general prognostic analysis method mentioned above, we first made the Kaplan-Meier estimate of survival function with respect to the cardiac death of AMI during hospitalization (Fig. 2). The discharge and non-cardiac death were treated as censored cases. Furthermore, the instantaneous mortality rate (hazard function) of the cardiac death was depicted (Fig. 3).

It is easy to see that the overall distribution of the cardiac death of AMI has obviously two components: the early component and moderate abrasion component. We applied 2-component Weibull distribution to the survival function. The maximum likelihood method gives estimates for 6 parameter values (Table 2). From these parameter values, the following results were obtained.

1. The fast component that describes the early cardiac death group of AMI follows the initial failure type, since the value of the estimated shape parameter $m_1 = 0.9654 < 1.0$ and the time constant $\tau_1$ is 2.12 days. This means that the fast component represents early death group that was thought to die within 2 or 3 days.
2. The slow component that describes the relatively late cardiac death group of AMI follows the abrasion failure type.
type with shape parameter $m_2 = 1.3463 > 1.0$, and time constant $\tau_2$ was 102.54 days. This means that slow component represents the late death that is rather scattered in a relatively wide range.

3. The fast component of the early cardiac death group substantially disappears at the third day after the onset of AMI, so that the overall survival function shows the inflection around this third day. As the time constant of the early cardiac death $\tau_1$ was found to be 2.12 days, we could separate the early death group from the late death one by inquiring group on whether the patient was dead within 3 days from the onset of AMI or not. Hence, we divided three prognostic groups: early death group, late death group, and survived group. The early death group was defined as the patient group who died within 3 days, the late death group was defined as the patient group who died after 3 days during hospital and survived group as the patient group who discharged or died from non-cardiac death. The prognostic analysis was made in compar-

<table>
<thead>
<tr>
<th>Table D. Multiple Cox regression and estimated prognostic factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prognostic factors in early stage</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Killip</td>
</tr>
<tr>
<td>NQMI/QMI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Prognostic factors in late stage</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Killip</td>
</tr>
<tr>
<td>NQMI/QMI</td>
</tr>
</tbody>
</table>

Figure 3. Mortality rate distribution (hazard function) of cardiac death, the horizontal axis is elapsed days of hospital admission in 1999 (n=1273).
son with prognostic factors among these groups.

**Prognostic analysis of two modes of the AMI progress**

To clarify the clinical features of the separated modes of cardiac death of AMI, we applied Cox life regression model to each prognostic group. Cox life regression analysis clarifies the prognostic factors that have statistically significant influences on the variation of the temporal profile of cardiac death in AMI. The results are as follows (Table 3, Fig. 4).

1. For the early death group, Cox life regression model extracted only one significant prognostic covariate: Killip classification. In the distribution of the Killip classification in the early death group, the ratio of Killip class IV was much greater than that of the average distribution of study population (Fig. 5). When only one covariate is included in the Cox regression model, not only Killip classification but also age shows the significant contribution. But if we take these two prognostic factors together, age loses its significance because the contribution to the early death is absorbed into that of Killip classification.

2. For the late death group, we extracted three statistically significant covariates: Killip classification, Q-type or non-Q type myocardial infarction (QMI) and age. Of course the classification is most influential in the early death group, but QMI is also an independently influential prognostic factor with significant level less than 0.1% in the late death group. Age is also an independently influential prognostic factor in the late death group with significant level less than 5%. The distribution of Killip classification is not so much different than that of whole study population.

3. To characterize the early death group and late death group, we investigated their difference of temporal profile among the cause of cardiac death shown in Table

---

**Table 3. Causes of cardiac death**

<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>The number of patient</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiogenic shock</td>
<td>50</td>
<td>42.74%</td>
</tr>
<tr>
<td>Cardiac failure</td>
<td>28</td>
<td>23.93%</td>
</tr>
<tr>
<td>Ventricular free wall rupture</td>
<td>17</td>
<td>14.53%</td>
</tr>
<tr>
<td>Ventricular septum perforation</td>
<td>6</td>
<td>5.13%</td>
</tr>
<tr>
<td>Papillary muscle disconnection</td>
<td>1</td>
<td>0.85%</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>11</td>
<td>9.40%</td>
</tr>
<tr>
<td>Other cardiac deaths</td>
<td>4</td>
<td>3.42%</td>
</tr>
<tr>
<td>Total patients</td>
<td>117</td>
<td></td>
</tr>
</tbody>
</table>
IV by depicting the temporal change of mortality (hazard function) (Fig. 6). We can point out that the cardiac death from dysfunction of heart (cardiogenic shock, cardiac failure) follows the hazard function of the initial failure type, whereas the morphological breaks of the heart muscle (cardiac rupture, perforation, papillary disconnection) show the mostly constant hazard function, which is thought to be a random or weakly abrasion failure type. The causes of cardiac death are related to the two different components of temporal profile of AMI.

Discussion

Temporal profile decomposition

In the widely used prognostic analysis methods such as multiple logistic regression analysis and Cox regression analysis, it is assumed that the way in which these prognostic factors affect the prognosis is uniform in the course of diseases. But, in the actual prognostic analysis, prognostic factors often exert their influence only at the onset of the disease, or at later stage of the diseases. Or, even within the population of same disease, there are several groups having different courses of diseases and the different ways in which the prognostic factors affect the disease progress. So far there has never been such a prognostic method that can deal with this kind of situation. Recently several studies have noticed this shortcoming of the Cox regression model and tried to overcome these difficulties. Bolard et.al proposed alternative hazard models that can describe the different hazard functions by different causes of deaths. But their models are based on an empirical piecewise hazard function and cannot describe different failure modes of the course of diseases.

Our temporal profile decomposition method proposed in this study aims to separate, among same disease population, prognostic groups showing different courses of the diseases and evaluate the influences of covariates within different prognostic groups. To do so, we propose a new method that uses what we call the multiple component Weibull distribution to decompose the overall profile of the course of diseases. Weibull distribution is frequently used in a failure analysis of reliab-
bility in the field engineering because a shape parameter can describe the type of failure. Hence, the Weibull distribution would be expected to become the basis for the prognostic analysis, but there has been very few application of Weibull distribution to medical field even in the case of applying a simple Weibull function to describe the course of diseases\textsuperscript{10}. In this study multiple component Weibull distribution was used for the first time for the prognostic analysis of AMI to investigate its validity.

Prognostic analysis of AMI
As this is an essentially methodological one study and the survey project itself is still on-going, we should refrain from making definite conclusion on the prognostic analysis of AMI, but still we could make substantial discussion on results we obtained by application of our temporal profile decomposition method to AMI so far as it is concerned with the prognosis during hospitalization at the first year of the survey.

Despite remarkable advances in the treatment of acute myocardial infarction, there still exists substantially early death patient group of AMI. Although considerable information is available on the prognosis after acute myocardial infarction in western populations, little is known about the fate of Japanese subjects after AMI, though there is a prospective study such as that in Yamagata\textsuperscript{11}. Hence, it is appropriate to take AMI as a subject of our new prognostic analysis of temporal decomposition of disease progression.

Our prognostic analysis decomposed the study population into the early death group and late death group on the basis of differences of disease progression profiles. The prognostic analysis of Cox life regression method applied to each group shows that in the early death group only Killip class is significant prognostic factor. Although there have been many studies pointing out that the Killip classification is the most influential prognostic factor\textsuperscript{12-14}, they did not refer to the temporal way how the difference of the Killip classification exerts its influence on the course of AMI.

On the contrary to this in late death group Q type MI and age also proved to be significant prognostic factors. There have also been many studies clarifying that the Q-type MI is a next influential prognostic factor\textsuperscript{15-17}. No other prognostic factors such as coronary risk factors or places of MI were found to be significant as long as the prognosis (death) during hospitalization is concerned, though several previous studies pointed out that other factors might be influential\textsuperscript{18-20}.

From the hazard function analysis, the early death is
mostly caused by functional failure such as cardiogenic shock or cardiac failure so that the Killip classification proved to be significant covariates but after the risk of early death is gone, another risk of death emerges which is related to morphological failures such as cardiac rupture, perforation or papillary muscle disconnection. The risk of the cardiac death from the latter causes is found to be mostly constant during the course of a disease so that it gives a random or a weak abrasion failure type. The temporal decomposition of a disease profile gave the estimate of a shape parameter in the late death group m = 1.34, which suggests a weak abrasion failure type.

Conclusion

In this study we proposed a new prognostic analysis which we call temporal profile decomposition of a disease progression. To do this, we developed a new profile decomposition method using a multiple component Weibull distribution. In our method, after decomposition of the disease profile, the Cox life regression method is applied to each profile group to clarify the influential prognostic factors for a disease progression.

This method is applied to prognostic analysis of a cardiac death during the hospitalization, where the subjects are the first year population enrolled in the prospective survey of AMI, which is being conducted by the Heart Institute, Tokyo Women's University. We separated the prognosis of AMI into the early death of an initial failure type and the late death of a weak abrasion failure type. It was found that the prognostic factors of the early death, mostly from the cardiogenic shock or cardiac failure, was Killip classification, whereas, in the late death, mostly from a cardiac rupture or perforation, Q type MI and age other than Killip classification were found to be influential prognostic factors.

From these results, it is proved that the temporal profile decomposition is of great value to evaluate prognostic factors when there exist several different prognostic groups.

References