A competing risks regression model based on the exponential Gompertz-like subdistribution

A Kudus*, S Sulidi and M Herlina
Dept. of Statistics, Bandung Islamic University, Jl. Ranggamalela No. 1 Bandung, West Java 40116 Indonesia

*abdul.kudus@unisba.ac.id

Abstract. In some situations, research activities are conducted to measure the length of time until the event occurred. If it is possible to have not only the event of interest but also the competing events, then it is a kind of competing risks data. The main parameter of interest is the subdistribution function that shows the percentage of failing individuals due to a cause of interest. If the subdistribution needs to be predicted by predictor variables, then the regression model for competing risks data should be developed. In this paper, the regression model for the exponential Gompertz-like subdistribution is proposed. We use maximum likelihood estimation method to estimate the model parameters. Specifically, the subdistribution of the main event and all the rest competing events are estimated simultaneously. The proposed model was applied to the data of contraceptive discontinuation and gave a good result. Goodness-of-fit of the estimated regression model measured by the likelihood and Akaike Information Criterion (AIC) also gave the acceptable values.

1. Introduction

It is common that we observed the length of time until the occurrence of event. Sometimes no occurrence of event after waiting for a specific length of time, hence the observation was finished without experiencing the event. This is a kind of censored survival time data which need a special method to analyze and make inferences. The statistical issues become more complicated in study that have multiple event. In this case, several risks are competing each other to terminate the survival of a unit, and only one which finally terminate the survival of unit. This setting is called as the competing risks. The objective of competing risk data analysis is to isolate the effect of a given (a subset) of risk that acting on population.

For a single sample competing risk data, the common approach is to summarize the various endpoints by generating a series of estimated subdistribution function for each endpoint [1]. Whereas for competing risks data with covariate variables, we intend to formulate regression models for association analysis or prediction purposes [2].

One kind of regression model is direct parametric subdistribution regression by using Gompertz distribution for baseline distribution [3]. It is interesting that Gompertz distribution is suitable for capturing the improper properties of subdistribution function which is the asymptotic value might be less than one.

In spirit of modelling the improperness of subdistribution function, a non-mixture cure model has been developed for years ago [4]. In this model, the improperness property is captured by cure fraction.
which represent the individuals that are not yet experiencing the event of interest. Furthermore, it still probable that the event of interest occurred at a steady rate which resulted in the form of proper subdistribution function just like the ordinary distribution function. But, the subdistribution of competing event still has the form of improper subdistribution function with asymptotic value less than one. For the simultaneous modelling of the event of interest and the competing event subdistributions, we can use reparameterization of non-mixture cure models such that it looks like a Gompertz subdistribution. We call this as Gompertz-like subdistribution model [5].

In this paper, we extend the Gompertz-like subdistribution model to consider some covariates through regression models. Specifically, we develop a direct regression model based on non-mixture cure model by using reparametrized exponential kernel that mimic the Gompertz subdistribution. We call it an exponential Gompertz-like subdistribution regression model.

Section 2 gives materials and method of this research work. Result and discussion are presented in Section 3. Finally, Section 4 is about the conclusion of paper.

2. Materials and method
The materials which are used in this research came from survey study. We consider the Indonesian Demography and Health Survey (IDHS) data which contain of 2631 women respondent. We concerned with the data on time to contraceptive discontinuation.

Contraceptive method is one kind of mode in family planning program. This program aims to decrease the rate of reproduction by means of controlled birth scheduling. During usage the contraceptive method, women are expected not to be pregnant. Thus, it is important to identify characteristics that may relate to the discontinuation of contraceptive method, which may cause the pregnancy.

The interest of this kind of research is focus on the last episode of time of contraceptive used until it discontinues [6]. The outcome variables are measured in time scale from the usage to the discontinuation. There are three types of contraceptive discontinuation: (i) discontinuation due to failure, (ii) abandonment while in need of family planning, and (iii) switching to another contraceptive method. The occurrence of unintended pregnancy was caused solely by the contraceptive failure. Whereas the rest of two competing events suggest some decision-making and choice on the part of the woman.

We consider some covariates that might affect the rate of discontinuation as showed in Table 1. These covariates are household economic status, age of the women at the start of the episode of usage, area of residence, religion, woman’s education and the contraceptive method.

| No. | Variable name | Description |
|-----|---------------|-------------|
| 1.  | soceco        | household social and economic status (score 1-7) |
| 2.  | age           | age of start of contraceptive (years) |
| 3.  | resid         | area of residence (0=rural, 1=urban) |
| 4.  | relig         | religion (0=Moslem, 1=non-Moslem) |
| 5.  | educ          | woman’s education (0= primary or lower, 1=secondary, 2=university) with 0 (primary) as reference |
| 6.  | method        | contraceptive methods (1=pills and injectables, 2=IUDs and implants, 0= other modern methods (mainly condoms)) with 0 (other methods) as reference |

There are two important concepts which contained in the method that we used. The concept of univariate competing risks data analysis is presented in subsection 2.1. and then the concept of regression for competing risks data analysis in subsection 2.2.
2.1. Univariate competing risks data analysis

Let $T_i, i = 1, \ldots, n$ be $n$ independent random variables with positive value and common continuous distribution $F$. Independent of $T_i$'s, let $U_i, i = 1, \ldots, n$ be also independent positive random variable with possibly non-continuous with common distribution $G$, and $\delta_i, i = 1, \ldots, n$ be the event type associated with $T_i$, where $\delta_i = 1, \ldots, J$. A typical competing risks problem is to make statistical inference on $F$ based on censored observation $(Y_i, \Delta_i)$, defined by:

$$Y_i = \min(T_i, U_i), \text{ and } \Delta_i = I(T_i \leq U_i)$$

where $I (\bullet)$ is an indicator function of the specified event. The following points should be emphasized:

- The pair $(T_i, \delta_i)$ from different subjects in the sample are assumed to be independently identically distributed.
- The different event types within each subject are not assumed to be independent.
- Each subject can experience at most one event at most one event type.

The overall hazard function is defined by:

$$\lambda(t) = \lim_{\Delta \to 0} \frac{P(t \leq T_i \leq t + \Delta | T_i > t)}{\Delta}$$

and the overall survival function is given by:

$$S(t) = P(T_i > t) = \exp\left(- \int_0^t \lambda(u) du \right)$$

The hazard function for a particular event is defined by:

$$\lambda_j(t) = \lim_{\Delta \to 0} \frac{P(t \leq T_i \leq t + \Delta, \delta_i = j | T_i > t)}{\Delta}$$

This quantity indicates the rate of experiencing the $j^{th}$ competing event.

The event time density function for event type $j$ is defined by:

$$f_j(t) = \lambda_j(t) S(t)$$

The subdistribution (cumulative incidence) for the event type $j$ is defined by:

$$F_j(t) = P(T_i \leq t, \delta_i = j) = \int_0^t \lambda_j(u) S(u) du$$

$$= \int_0^t \lambda_j(u) \exp(- \Lambda(u)) du$$

where $\Lambda(u) = \int_0^u \lambda(v) dv = \sum_{j=1}^J \lambda_j(v) dv$

This is the probability of experiencing the $j^{th}$ event in the setting where competing risks are acknowledged to exist. The value of $F_j(t)$ depends on the rate of event of interest and the rest of event. $F_j(t)$ is not a true distribution function due to its properties: it is a non-decreasing function with $F_j(0) = 0$ and $F_j(\infty) = P(\delta = j) < 1$. These curves have a straightforward interpretation that is:

$$S(t) = 1 - \sum_{j=1}^J F_j(t)$$
In the analysis of competing risks data, it is common to model the hazard rate for particular event (3). But it is also worth to model the subdistribution function directly. Hence, we model the subdistribution for event type $j$ conditional on the covariate vector $Z$, $F_j(t; Z) = P(T \leq t, \delta = j|Z)$. The development of subdistribution model is carried out by considering non-mixture cure model [2]. Non-mixture cure model assumed bounded cumulative hazard $H(t)$ as $t \to \infty$,

$$H(t) \leq \theta^*, \lim_{t \to \infty} H(t) = \theta^*$$

To make sure the fulfillment of the above property, we write $H(t) = \theta^* F^*(t)$, where $F^*(t)$ is the distribution function of a nonnegative random variable, called kernel distribution. Then the survival distribution is $S(t) = e^{-\theta^* F^*(t)}$ and the distribution function is as follows

$$F(t) = 1 - \left[\exp(-\theta')\right]^{F(t)} , \theta' > 0$$

(7)

In this model, the distribution function converges to $1 - \pi$, where $\pi = \exp(-\theta^*)$, overtime; hence, this is not a proper distribution. The parameter $\pi$ is interpreted to be the cure fraction. In the competing risks framework, we utilized cure fraction to model the percentage of individuals that do not experience the event of interest. For this purpose, the subdistribution function for the cause $j$ can be formulated by

$$F_j(t) = 1 - \left[\exp(-\theta^*_j)\right]^{F(t)}$$

(8)

where $P(\delta^* = j) = \lim_{t \to \infty} F_j(t) = 1 - \exp(-\theta^*_j)$ which is the probability of event type $j$. On the other hand, $P(\delta^* \neq j) = \exp(-\theta^*_j)$ which is the proportion of individuals that do not experience the $j$th cause and $\sum_{j=1}^J F_j(t) = F(t)$.

Suppose $F^*(t)$ be an exponential distribution with parameter $\kappa$. Then the resulted subdistribution is

$$F_j(t; \theta^*_j, \kappa) = 1 - \left[\exp(-\theta^*_j)\right]^{1 - \exp(-\kappa)}$$

(9)

It is obvious that subdistribution (9) is improper when $\kappa_j > 0$ and $0 < \theta^*_j < \infty$. This model encompasses Gompertz subdistribution [3].

Now, let us revisit the Gompertz distribution,

$$F(t) = 1 - \exp\left[\tau_1 \left[1 - \exp(\rho t)\right]\right] / \rho_j$$

(10)

One of the nice properties of Gompertz distribution is the upper bound of the distribution function which may be less than 1. If $\rho \geq 0$, then asymptote of (10) for large $t$ is 1 which shows that it is a proper distribution. Whereas, the improper case of (10) occurs when $\rho < 0$ with asymptote $1 - \exp(\rho t)$.

Next, we will develop subdistribution function which may be proper or improper depend on the sign of parameter. Such kind of subdistribution will be called as Gompertz-like subdistribution. In the case of subdistribution model (9), the corresponding Gompertz-like subdistribution for the $j$th cause can be derived by using reparameterization $\theta^*_j = -\tau_1 / \rho_j$ and $\kappa_j = -\rho_j$.

The resulted subdistribution is

$$F_j(t; \tau_j, \rho_j) = 1 - \exp\left[\tau_j \left[1 - \exp(\rho_j t)\right] / \rho_j\right]$$

(11)

2.2. Regression for competing risks data analysis

Sometimes it is interesting to investigate the relationship between a particular risk and some factors $Z = (Z_1, Z_2, \ldots, Z_K)^T$. Such kind of statistical procedure is called regression. The Cox proportional hazards model is the most common regression model used to analyse survival data [7] that models the hazard rate of a given outcome. The alternative modelling is flexible parametric regression [8] and generalized Weibull regression for competing risks [9].

Extension of (11) for regression model is by incorporating the covariates $Z$ into model parameter as follows

$$F_j(t; \tau_j, \rho_j, \beta_j, Z) = 1 - \exp\left[\beta_j \cdot Z\right] \left[1 - \exp(\rho_j t)\right] / \rho_j$$

(12)
Given competing risks data as well as its covariates, we can fit the regression model (12) by means maximum likelihood estimation. The likelihood function is given by

\[
l(\psi) = \sum_{j=1}^{J} \left[ \sum_{t=1}^{T} \delta_{jt} \left[ \log(\tau_j) + z_j' \beta_j + \rho_j t + \frac{r_j}{\rho_j} \left( 1 - e^{-\rho_j t} \right) \right] + \left( 1 - \sum_{j=1}^{J} \delta_{jt} \right) \log \left[ \sum_{j=1}^{J} \exp \left( \frac{r_j}{\rho_j} e^{\rho_j t} \left( 1 - e^{-\rho_j t} \right) \right) \right] (J - 1) \right]
\]

where \( \psi = (\psi_1, \ldots, \psi_J) \) with \( \psi_j = (r_j, \rho_j, \beta_{j1}, \ldots, \beta_{jK}) \) for \( j = 1, \ldots, J \).

Whereas the Bayesian inference for this kind of model is exist in the literature [10].

3. Results and discussion

We fitted the competing risks regression model based on the Exponential Gompertz-like subdistribution as we expressed in equation (12) to contraceptive discontinuation data. The regression considers the effect of covariates to probability of three types of discontinuation.

Result of model the fitting model is shown in Table 2. This fitting gives likelihood value of 10416.58 which equals to AIC value of 20893.15. Under 5% significance level, we found that probability of discontinuation due to failure is affected by Religion. For the discontinuation due to abandonment, Age and University education level are two covariates which are significant. Whereas, for the discontinuation due to switching, in addition to education and contraceptive method, there are two other significant covariates, namely social economic status and age.

Table 2. Estimated of coefficient’s regression.

| Event type | Parameter          | Estimate  | Std. error | P-value     |
|------------|-------------------|-----------|------------|-------------|
| 1 (Failure)| \( \tau_1 \)       | 0.00168   | 0.001672   | 3.25\times10^{-1} |
|            | \( \rho_1 \)       | 0.005914  | 0.006681   | 3.76\times10^{-1} |
| Social Economic Status | -0.06044 | 0.082202 | 4.62\times10^{-1} |
| Age        | -0.04112          | 0.021564  | 5.67\times10^{-2} |
| Residence  | 0.180885          | 0.271054  | 5.05\times10^{-1} |
| Religion   | 1.475557          | 0.657311  | 2.49\times10^{-2} |
| Secondary  | 0.29854           | 0.377395  | 4.29\times10^{-1} |
| University | 0.524297          | 0.438745  | 2.32\times10^{-1} |
| Pills/Injection | 0.04961 | 0.741995 | 9.50\times10^{-1} |
| IUDs/Implants | -1.63706 | 0.913911 | 7.34\times10^{-2} |
| 2 (Abandonment)| \( \tau_2 \)       | 0.002379  | 0.000708   | 7.92\times10^{-4} |
|            | \( \rho_2 \)       | 0.023856  | 0.001732   | 9.79\times10^{-42} |
| Social Economic Status | 0.039698 | 0.021039 | 5.93\times10^{-2} |
| Age        | 0.020537          | 0.004715  | 1.38\times10^{-5} |
| Residence  | 0.014608          | 0.067415  | 8.28\times10^{-1} |
| Religion   | 0.062844          | 0.177222  | 7.23\times10^{-1} |
| Secondary  | -0.15573          | 0.087524  | 7.53\times10^{-2} |
| University | -0.26154          | 0.114962  | 2.30\times10^{-2} |
| Pills/Injection | 0.094764 | 0.238678 | 6.91\times10^{-1} |
| IUDs/Implants | -0.0178  | 0.247231 | 9.43\times10^{-1} |
Table 3. Estimated of coefficient’s regression (continued).

| Event type   | Parameter                     | Estimate | Std. error | P-value  |
|--------------|-------------------------------|----------|------------|----------|
| (Switching)  | \( \tau_3 \)                  | 0.012674 | 0.003193   | 7.40\times10^{-5} |
|              | \( \rho_3 \)                  | -0.00141 | 0.001817   | 4.37\times10^{-1} |
| Social Economic Status |                      | -0.05281 | 0.021714   | 2.73\times10^{-2} |
| Age          |                               | 0.011296 | 0.005114   | 1.51\times10^{-1} |
| Residence    |                               | 0.091762 | 0.068808   | 8.21\times10^{-1} |
| Religion     |                               | 0.055785 | 0.180476   | 6.57\times10^{-1} |
| Secondary    |                               | 0.302452 | 0.096844   | 3.28\times10^{-1} |
| University   |                               | 0.670814 | 0.118122   | 7.56\times10^{-1} |
| Pills/Injection |                           | -0.78003 | 0.173551   | 7.27\times10^{-6} |
| IUDs/Implants |                               | -0.9433  | 0.185076   | 3.70\times10^{-7} |

4. Conclusion

We have demonstrated the use of standard parametric survival models in case of competing risks by modeling subdistribution functions through non-mixture cure model. Standard exponential distributions can be employed for kernel of the non-mixture cure model, then be used to specify the improper or proper subdistribution functions.

Inferential procedure by using maximum likelihood estimation indicates that likelihood function cannot be factorized into a product of cause-specific functions, hence we fitted all subdistribution jointly. Another issue is regarding the event which occurs at a fairly steady rate over the entire time period. In this case, the subdistribution is better described by a proper distribution. To solve this problem, we may use Gompertz-like subdistribution model as proposed above. This model is useful when the form of subdistribution might be proper. However, we cannot extrapolate the subdistribution model to estimate the long-term probabilities, because the subdistribution value will reach one as time goes to infinity.

We can incorporate covariates through parameter other than cure fraction. It may increase the fitting, but we may lose the linear form of the complementary log-log transformation.

Acknowledgement

This research was supported by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia. The authors thank the referee for the insightful comments and suggestions.

References

[1] Zhang Z 2017 Survival analysis in the presence of competing risks Annals of Translational Medicine 5(3) 47.
[2] Lau B, Cole S R and Gange S J 2009 Competing risk regression models for epidemiologic data American Journal of Epidemiology 170(2) 244-56.
[3] Jeong J-H and Fine J 2006 Direct parametric inference for the cumulative incidence function Journal of the Royal Statistical Society. Series C: Applied Statistics 55(2) 187-200.
[4] Yakovlev A and Tsodikov A 1996 Stochastic Models of Tumor Latency and Their Biometrical Applications Singapore: World Scientific.
[5] Kudus A 2010 The exponential gompertz-like subdistribution model for competing risk survival time data Proceedings the Third International Conference on Mathematics and Natural Sciences 23–25.
[6] Steele F, Goldstein H and Browne W 2004 A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics Statistical Modelling 4 145-159.
[7] Cox D R 1972 Regression models and life-tables J. Royal Statist. Soc. B 34 187-220.
[8] Nelson C P, Lambert P C, Squire I B and Jones D R 2007 Flexible parametric models for relative
survival, with application in coronary heart disease Statistics in Medicine 26(30) 5486-98.

[9] Moamer S, Baghestani A R and Pourhoseingholi M A 2018 Regression Modeling of Competing Risks Survival Data in the Presence of Covariates Based on a Generalized Weibull Distribution: A Simulation Study Pakistan Journal of Statistics and Operation Research 14 (2).

[10] Gupta C, Cobre J, Polpo A and Sinha D 2016 Semiparametric Bayesian estimation of quantile function for breast cancer survival data with cured fraction Biometrical Journal 58(5) 1164-77.