ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Frailty Models to Analyze Sputum Conversion of Tuberculosis Patients

T. Leo Alexander[1], S.Thobias[2] and K. Siva Naga Raju[3]
1, 2.Associate Professor, Department of Statistics, Loyola College, Chennai-34
3. Assistant Professor, Department of Statistics, Loyola College, Chennai-34

Manuscript Info

Manuscript History:

Received: 15 May 2015
Final Accepted: 14 June 2015
Published Online: July 2015

Key words:

Clinical Trails, Frailty Models, Log relative Hazard form, Randomized, Survival Analysis

Corresponding Author

T. Leo Alexander

Abstract

A frailty model is a random effect model for time to event data. This Paper deals with Gamma Frailty model to evaluate the Sputum Conversion (positive to negative) times of 1240 tuberculosis patients that are admitted into the randomized control trials. The analysis is done in STATA.

Section 4.1 is the analysis of Weibull regression from Gamma Frailty, Table 4.1.1 summarizes the Survival data and Table 4.1.2 gives hazard ratio, standard error, significance and confidence intervals for each of the variables that are displayed. In section 4.2 the analysis of Lognormal regression from Gamma Frailty is discussed, Tables 4.2.1 summarizes the Survival data and Table 4.1.2 gives hazard ratio, standard error, significance and confidence intervals for each of the variables that are displayed. Section 4.3 shows the analysis of Loglogistic regression from Gamma Frailty, Table 4.3.1 summarizes the Survival data and Table 4.3.2 discusses the model. The final result is obtained in Section 5.

1. INTRODUCTION

The statistical analysis of lifetime data (or more exactly, time-to-event, event-history or duration data) plays an important role in medicine, epidemiology, biology, demography, economics, engineering and other fields. It has expanded rapidly in the last three decades, with works having been published in various disciplines in addition to statistics.

2. A Study on Tuberculosis

Human Tuberculosis (TB) is an airborne infectious disease that may affect the lungs (pulmonary TB) or other parts of the body (extra pulmonary TB). The most common form of TB is pulmonary but both forms can also co-exist. Tuberculosis can have a wide range of symptoms such as cough, chest pain, shortness of breath, fatigue, fever or weight loss. It is transmitted when people infected with pulmonary TB cough or sneeze.

TB bacteria

The causative agents of TB are grouped in the Mycobacterium tuberculosis complex (MTBC): Mycobacterium tuberculosis (Mtb), Mycobacterium bovis, Mycobacterium africanum, Mycobacterium canneti and Mycobacterium microti. Mtb is responsible for most cases of TB and although it can affect animals, humans are the main hosts.

3. Frailty Models

In most clinical trials, survival analysis implicitly assumes a homogenous population to be studied, but in many applications, the population under study will be heterogeneous (i.e.) a mixture of individuals with different hazards.

The frailty models for survival data considers the variability in life times, more generally, time to some specific event, these come from two separate sources. One is simple randomness and it is described by a hazard function. The other source is described by a random effect (frailty), which is a random variable. It may be either an individual variable or a variable common to several individuals. The frailty approach is a statistical modelling concept which aims to account for heterogeneity, caused by unmeasured covariates. In studies, which particularly involve human subjects, there may be factors other than the measured covariates that significantly affect the distribution of survival time.

This can be referred as heterogeneity of the subjects. Some early papers on this factor was worked by Vaupel, Manton , and Stallard (1979) , who used the concept of frailty to describe the differences in survival time among apparently similar individuals. Aalen (1994) also provided a relatively non- technical summary. Hougaard (1995) presented an excellent overview of frailty models. Klein and Moeschberger (1997) presented methods based on incorporating frailty in proportional hazards models.

The basic idea of a frailty model is to incorporate an unmeasured "random" effect in the hazard function to account for heterogeneity in the subjects. When the observed data consists of triplets (t_i, X_i, C_i), i = 1,2, ….n. the observed follow-up times , the vector of p covariates and a right censoring indicator variable, the hazard function at time t for the i^{th} subject is ,under the proportional hazards model, $h(t, x, \beta) = h_o(t)\exp(X_i'\beta)$

A frailty model includes, in the hazard function the value of an additional unmeasured covariate, "frailty", denoted by z_i , yielding the hazard function as $h_f(t, X_i, \beta_i) = z_i\, h(t, X_i, \beta_i)$

the above equation represents the hazard function which has been modified by the inclusion of a frailty.

It is advantages to use fully parametric model, such as the Weibull regression model with frailty because the estimation is easier and it is possible to describe explicitly the effect that frailties have on the hazard ratios over time Aalen(1994). In particular, the fact is that, the most frail individuals tend to fail early in the follow-up, the average hazard ratio tends to decrease over time. Frailty models have often been used when groups of subjects have responses that are likely to be dependent in some general way.

Types of frailty models

1. Model with an univariate survival time as end point.
2. Model which describe multivariate survival end points.

3.1 Univariate frailty models

To account an unobserved heterogeneity in the population under study, a univariate frailty model was introduced into survival analysis. The key idea is that individuals possess different frailties, and that more patients who are most frail will die earlier than the others. When mortality rates are estimated, one may be interested in how, their rates change over time or age. It is observed that the hazard function raises at the beginning, reaches a maximum and then declines (or) levels off at a constant value. Sometimes there are too many covariates to be considered in the model, (or) in other cases, the researcher do not know (or) is not able to measure all the relevant covariates.

In both cases, there are two sources of variability in survival data.

- Variability accounted for by measurable risk factor, which is theoretically predictable.
- Heterogeneity caused by unknown covariates which is theoretically unpredictable.

Frailty distribution:

An important problem in the area of frailty models is the choice of the frailty distribution. The frailty distributions most often applied are

1. Gamma distribution
2. Positive stable distribution
3. Three-parameter distribution
4. Compound Poisson distribution
5. Log- normal distribution

3.2 Multivariate Frailty Models

These models are used in the field of multivariate survival data.

(a) Lifetimes of relatives,(twins, parent- child).

(b) Recurrent events like infections in the same individual are considered.

These models accounts for the presence of dependence between these event times. The commonly used approach is to specify independence among observed data items conditionally on a set of unobserved (or) latent variables.

Let $S(\frac{t_1}{z}, x_1)$ and $S(\frac{t_2}{z}, x_2)$ be the conditional survival functions of two related individuals with different vectors of observed covariate x_1 and $x_2$respectively, the two dimensional survival function is of the form, $S(t_1, t_2) \int S(\frac{t_1}{z}, x_1) S(\frac{t_2}{z}, x_2) g(z) dz$, where g denotes the density of the frailty 'z'.

4. Analysis of Frailty Models

1240	Total observations
3	Observation end on or before enter
1237	Observation remaining, representing
1062	Failures in single record/ single failure data
3054	Total analysis time at risk

The covariates under the study are

1. Age in years
2. Treatment: regiment
3. Gender: Male(1) Female(0)
4. Drug susceptibility pattern: Res(1) Sen(0)
5. Weight in Kg at the time of admission

4.1 Weibull regression -- log relative-hazard form Gamma frailty

In the following table a brief summary of the data and the loglikelihood measures under the Weibull regression – log relative Hazard form Gamma Frailty are displayed. The total observations used for the study are 1237 and the event of interest has occurred in 1062 patients. The Likelihood ratio Chi-Square value is 71.42 and which is highly significant. The loglikelihood of the model is -1003.76.

Table 4.1.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	71.42
Log likelihood	-1003.76
Prob > chi2	0.000

In the following table hazard ratio, standard error, significance and confidence intervals for each of the variables are displayed in table 4.1.2

Table 4.1.2

_t	Haz. Ratio	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.47918	0.08041	-4.38	0	0.34488	0.66579
Age	0.98307	0.00571	-2.94	0.003	0.97195	0.99432
wt_0	1.02303	0.01084	2.15	0.032	1.00201	1.0445
Present	2.74592	0.47933	5.79	0	1.9503	3.86609
Groupreg	0.87136	0.06869	0.081	0.081	0.74662	1.01694
/ln_p	1.44051	0.04655	30.95	0	1.34928	1.53174
/ln_the	0.66344	0.09054	7.33	0	0.48599	0.84089
P	4.22284	0.19656			3.85465	4.62621
1/p	0.23681	0.01102			0.21616	0.25943
Theta	1.94145	0.17577			1.62578	2.31842

Likelihood-ratio test of theta=0: chibar2(01) = 306.76 Prob>=chibar2 = 0.000

Using the Weibull Model form Gamma frailty (from Table 4.1.2) we could access that the following covariates are said to be the influencing factor for the time until the sputum conversion are given from the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, ln_p and In_the are found to be significant. The hazard ratio for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -1003.76. Further we need to calculate -2 logikelihood (-2LL) to decide up on the model that suits the data well. Hence we calculate -2 loglikelihood (-2LL), for the exponential model we have -2LL = -2(-1003.76) = 2007.52.

Remark 1

The Harzed Ratio of the variable Present (it is a drug susceptibility test which tells whether a particular drug works well for a particular patient or not) in Weibull Model from Gamma Frailty model is 2.74592. So we now have exp(2.74592)= 15.57894. Those who are sensitive to the drug are likely to reduce the sputum conversion time by 15 times compared to the patients who are resistant. The weight of the patient also has a significant impact on their sputum conversion times. Similar interpretations can be made using the Harzed Ratio of variables in the model.

Similarly we calculate the loglikelihood values under different distributions and make comparisons. In the next section we shall discuss the results under Lognormal model.

4.2 Lognormal regression – Accelerated Failure-Time form Gamma frailty

In the following table a brief summary of the data and the loglikelihood measures under the Lognormal Model from Gamma Frailty Model are displayed. The total observations used for the study are 1237 and the event of interest has occurred in 1062 patients. The Likelihood ratio Chi-Square value is 64.49 and which is highly significant. The loglikelihood of the model is -999.111.

Table 4.2.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	64.49
Log likelihood	-999.111
Prob > chi2	0.000

In the following table Coefficient, standard error ,significance and confidence intervals for each of the variables are displayed in table 4.2.2.

Table 4.2.2

_t	Coet.	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.16318	0.03834	4.26	0	0.08805	0.23832
Age	0.00426	0.00133	3.2	0.001	0.00165	0.00686
wt_0	-0.0052	0.00249	-2.08	0.037	-0.0101	-0.0003
Present	-0.221	0.04384	-5.04	0	-0.3069	-0.1351
groupreg	0.03059	0.01847	1.66	0.098	-0.0056	0.06679
_cons	0.77169	0.12229	6.31	0	0.53201	1.01137
/ln_sig	-0.8663	0.03686	-23.5	0	-0.9385	-0.794
/ln_the	-0.7335	0.14654	-5.01	0	-1.0207	-0.4463
Sigma	0.42052	0.0155			0.39121	0.45203
Theta	0.48022	0.07037			0.36034	0.64

Likelihood-ratio test of theta=0: chibar2(01) = 60.36 Prob>=chibar2 = 0.000

Using the Lognormal Model form Gamma frailty (from Table 4.2.2) we could access that the following covariates are said to be the influencing factor for the time until the sputum conversion are given from the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, _cons, ln_sig and In_the are found to be significant. The hazard ratio for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -999.111. Further we need to calculate -2 logikelihood (-2LL) to decide up on the model that suits the data well. Hence we calculate -2 loglikelihood (-2LL), for the exponential model we have -2LL = -2(-999.111) = -1998.222

Remark 2

The Coefficient of the variable Present (it is a drug susceptibility test which tells whether a particular drug works well for a particular patient or not) in Lognormal Model from Gamma Frailty model is -0.221. So we now have exp(-0.221)= 0.8017167. Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.8017 times compared to patients who are resistant. The weight of the patient also has a significant impact on their sputum conversion times. Similar interpretations can be made using the Coefficient of variables in the model.

Similarly we calculate the loglikelihood values under different distributions and make comparisons. In the next section we shall discuss the results under Loglogistic model.

4.3 Loglogistic regression – Accelerated Failure-Time form Gamma frailty

In the following table a brief summary of the data and the loglikelihood measures under the Loglogistic Model from Gamma Frailty Model are displayed. The total observations used for the study are 1237 and the event of interest has occurred in 1062 patients. The Likelihood ratio Chi-Square value is 52.55 and which is highly significant. The loglikelihood of the model is -992.969.

Table 4.3.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	52.55
Log likelihood	-992.969
Prob > chi2	0.000

In the following table Coefficient, standard error, significance and confidence intervals for each of the variables are displayed in table 4.3.2

Table 4.3.2

_t	Coet.	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.15236	0.0366	4.16	0	0.08063	0.22409
Age	0.0039	0.00125	3.12	0.002	0.00145	0.00636
wt_0	-0.0044	0.00237	-1.84	0.065	-0.009	0.00028
Present	-0.1651	0.04311	-3.83	0	-0.2496	-0.0806
Groupreg	0.0263	0.01735	1.52	0.13	-0.0077	0.06031
_cons	0.73375	0.11584	6.33	0	0.50671	0.96078
/ln_gam	-1.4411	0.03848	-37.45	0	-1.5165	-1.3657
/ln_the	-0.8263	0.13884	-5.95	0	-1.0985	-0.5542
Gamma	0.23666	0.00911			0.21947	0.2552
Theta	0.43765	0.06076			0.33338	0.57452

Likelihood-ratio test of theta=0: chibar2(01) = 63.10 Prob>=chibar2 = 0.000

Using the Loglogistic Model form Gamma frailty (from Table 4.3.2) we could access that the following covariates are said to be the influencing factor for the time until the sputum conversion are given from the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, _cons, ln_gam and In_the are found to be significant. The hazard ratio for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -992.969. Further we need to calculate -2 logikelihood (-2LL) to decide up on the model that suits the data well. Hence we calculate -2 loglikelihood (-2LL), for the exponential model we have -2LL = -2(-992.969) = 1985.938

Remark 3

The Coefficient of the variable Present (it is a drug susceptibility test which tells whether a particular drug works well for a particular patient or not) in Lognormal Model from Gamma Frailty model is -0.1651. So we now have exp(-0.1651)= 0.8478. Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.8478 times compared to patients who are resistant. The weight of the patient also has a significant impact on their sputum conversion times. Similar interpretations can be made using the Coefficient of variables in the model.

5. CONCLUSION

Exponential regression for Gamma frailty model is not convergent for this survival data. Lower values of -2LogLikelihood suggest a better model. it is difficult to use a formal statistical test to discriminate between parametric models. One way of selecting an appropriate parametric model is to base the decision on minimum (AIC) and also based on the -2 LL. For the parametric models presented in the Tables -2LL of Weibull regression for gamma frailty is 2007.5154, log normal regression for gamma frailty distribution -2LL is 1998.222 and log logistic regression for gamma frailty is 1985.93762. Decision based on -2LL, compare to all other models **log logistic regression for gamma frailty distribution is the most suitable model for our data set**. In log logistic distribution the variables like age of the patient, sex to which the patient belongs, present ln_gam , ln_the are all significant at 5% level. Thus we can conclude that each covariate included in the study have significant impact on the occurrence of event i.e, sputum conversion.

The coefficient of the variable Present (it is a drug susceptibility test which tells whether a particular drug works well for a particular patient or not) in Loglogistic regression for Gamma frailty model is -0.1651. So we now have exp(-0.1651)=0.8478089. Those who are sensitive to the drug are likely to increase the sputum conversion time

by 0.8478 times compared to patients who are resistant. In the Loglogistic model, the ln_gam a significant impact on their sputum conversion times. Similar interpretations can be made using the coefficients of variables in all the models.

In clinical trial applications the AFT models and Frailty Models is often a more realistic model than the PH model in the analysis of time to event data. The PH model is appropriate when there is a difference between the groups in the longer term in the context of the follow-up period. The Frailty models is more appropriate when the group differences are seen over a shorter time frame while in the longer term the probability of remaining event free is similar in the two groups. PH model is not always appropriate and the AFT model and Frailty Models in many applications provides a more appropriate modelling framework and has the added advantage of being straightforward and easier to interpret.

It is found that the Frailty Model should be considered as an alternative to the AFT Models and PH model in the analysis of time to event data, especially in applications where the effects of treatment are to accelerate (or delay) the event of interest with no permanent effect in the context of the follow-up period.

Acknowledgement

Authors Thanks **Dr C. PONNURJA,** Scientist-D, ICMR, Chennai and **Dr. Pari Dayal** Head, Department of Statistics DRBCC Hindu College, Chennai for their invaluable help and support.

References

1. Aalen, O.O. (1987) Two examples of modelling heterogeneity in survival analysis. Scandinavian Journal of Statistics 14, 19 – 25
2. AalenO.O.(1994) Effects of frailty in Survival Analysis. *Statistical Methods in Medical Research,* 3, 227-43.
3. Anderson, J.E., Louis, T.A. (1995) Survival analysis using a scale change random effects model. Journal of the American Statistical Association 90, 669 – 679
4. Collett D. (1994): Modeling of Survival Data in Medical Research. Chapman & Hall, London, U.K.
5. Cox, D.R and Oakes, D. (1984), Analysis of Survival Data. London Chapman and Hall.
6. Everitt BS (1998). Statistical analysis of medical data. *Arnold*, London
7. Global tuberculosis report 2014. 1.Tuberculosis – epidemiology. 2.Tuberculosis, Pulmonary – prevention and control. 3.Tuberculosis – economics. 4.Tuberculosis, Multidrug-Resistant. 5.Annual Reports. I.World Health Organization. ISBN 978 92 4 156480 9 (NLM classification: WF 300) World Health Organization (WHO)
8. Harrell, Jr., Frank E. (2004). SAS macro *daspline*. Mayo Clinic College of Medicine
9. Hougaard P (1995). Fraility models models for survival data. Lifetime data Analysis 1: 255-273.
10. Jin Z, Lin DY, Wei LJ and Ying ZL (2003) Rank-based inference for the accelerated failure time model. Biometrika. 90, 341-353
11. Kalbfleisch JD and Prentice RL (1980) The statistical analysis of failure time data, John Wiley & Sons, NY
12. Kay R and Kinnersley N (2002) On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. Drug Inform. J. 36, 571–579.
13. Klein, J.P. and Moeschberger, M.L. (1997): Survival Analysis Techniques for Censored and Truncated data, New York: Springer-Verlag
14. Lee ET and Wang JW. (2003) Statistical Methods for Survival Data Analysis.
15. Miller RG. (1981) Survival Analysis. *Wiley & Sons*, New York.