ISSN NO. 2320-5407

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

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

CONTROLLED CLINICAL TRIALS FOR TUBERCULOSIS PATIENTS BASED ON ACCELERATED FAILURE TIME (AFT) MODELS

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 June 2015
Final Accepted: 26 July 2015
Published Online: August 2015

Key words:

Accelerated Failure Time (AFT), Clinical Trails, Cox Proportional, Kaplan Meier Analysis, Log Rank Test, Survival Analysis.

Corresponding Author

T. Leo Alexander

Abstract

In this Paper we present the analysis of the data considered for the study on 1240 tuberculosis patients admitted into the randomized control trial. Survival techniques such as Kaplan Meier estimator and accelerated failure time models have been used to study about the sputum conversion (positive to negative) times. The effects of different treatment regimen are also tested. The analysis is done in SPSS and STATA.

In Section 3.1 Kaplan Meier estimates and survival curves for gender of patients are presented and in the following Section 3.2 Kaplan Meier analysis and survival curves for treatment groups are presented and log-rank test is used to test the equality of the survival distributions in both cases. In Section 4 we present the STATA outputs of the applications of Accelerated failure time models of the results are discussed. In section 5 the overall results are discussed.

1.INTRODUCTION

Survival analysis is an area of Statistics that was developed to analyze data representing times from a time origin until the occurrence of the event of interest. In medical research, the time origin is often the time of recruitment into a clinical trial or study. Although, the event of interest can be the death of the patient, recurrence of symptoms or any other particular event, the event of interest is usually named death and the time since the time origin until the event of interest is named survival time.

Let T be the nonnegative random variable representing the failure time of an arbitrary individual. We assume that the probability distribution of T is described by a density function $f(t)$. We shall introduce the Survival function $S(t)$ and the hazard function $\lambda(t)$ which characterize the distribution of T as well.

The survival function $S(t)$ is defined by

$$S(t) = P(T \geq t)$$

and is equal to $1-F(t)$, where $F(t)$ is the cumulative distribution function of T. (Note $P(X=t)=0$ for each number t in case of a density function.)

Since the cumulative distribution function $F(t)$ specifies the distribution of T, the distribution of T is specified as well by the survival function $S(t)=1-F(t)$.

The hazard function $\lambda(t)$ specifies the instantaneous rate of failure at $T=t$ conditional upon survival to time t and is defined by the limit for $\delta \downarrow 0$ of the following ratio:

$$\frac{P(t \le T < t+\delta \mid T \ge t)}{\delta} = \frac{P(t \le T < t+\delta)}{P(T \ge t) \times \delta}$$

$$= \frac{S(t) - S(t+\delta)}{\delta} \times \frac{1}{S(t)}$$

Taking limit we get $\lambda(t) = \frac{f(t)}{S(t)}$.

Note that the derivative of the survival function $S(t)$ is equal to $-f(t)$. The distribution of T is specified by its hazard function as well because the survivor function is determined by the hazard function:

$$\frac{d}{dt}\ln(S(t)) = -\frac{f(t)}{S(t)} = -\lambda(t)$$

$$\ln(S(t)) = -\int_0^t \lambda(u)du \quad \text{(Note: } S(0) = 1\text{)}$$

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

2 Accelerated Failure Time model

For many decades the Cox proportional hazard model has been used to examine the covariate effects on the hazard function for the failure time variable. An alternative method to Cox PH is Accelerated Failure Time. It is one of the very important regression models in survival analysis where censoring is present. The Accelerated Failure Time model which regresses the logarithm of the survival time over the covariates, the Accelerated Failure Time model is applied in the field of Reliability, Industry, etc. If the Cox PH assumptions does not hold good then Accelerated Failure Time model can be used (Miller 1981, Cox and Oakes 1984 , Collet 1994 , Everitt 1998, Lee and Wang 2003).

The Accelerated Failure Time model is a linear regression model in which the response variable is the logarithm or a known monotone transformation of a failure time (Kalbfleisch and Prentice, 1980). Semi-Parametric estimation in the Accelerated Failure Time model with an unspecified error distribution has been studied extensively for the right censored data. The Accelerated Failure Time models can be used to describe the influence of unobserved heterogeneity in a non- parametric and parametric PH models. The under lying assumption of Accelerated Failure Time models is that the effect of covariates is multiplicative (proportional) with respect to the hazard. The acceleration factor is the key measure of association obtained in an Accelerated Failure Time model. These models allow evaluating the effect of predictor variable on survival time just as the hazard ratio allows the evaluation of predictor variable on the hazard.

3. Kaplan-Meier Analysis

The Kaplan–Meier estimator, also known as the product limit estimator, is an estimator for estimating the survival function from lifetime data. In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment. In economics, it can be used to measure the length of time people remain unemployed after a job loss. In engineering, it can be used to measure the time until failure of machine parts. In ecology, it can be used to estimate how long fleshy fruits remain on plants before they are removed by frugivor

3.1 Kaplan-Meier Analysis for the sex of the patients

In this section we perform Kaplan Meier analysis for the sex of the patients included in the study and obtain their means and median survival times and the survival curve for the sex of the patients is also obtained. The equality of survival distributions is tested using the log-rank test.

In the table below the significance values based on different tests for equality of the survival distributions of gender of patients is given.

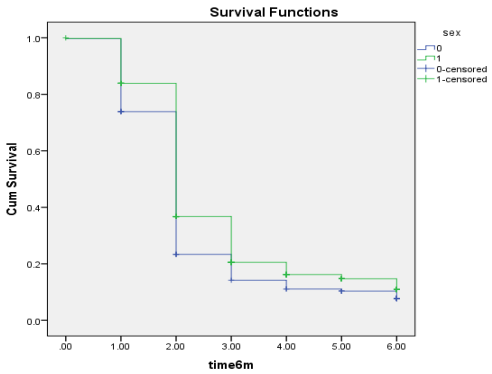
Table 3.1.1
Test for equality of survival distributions for sex of the patients**.**

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	16.892	1	0.000
Breslow (Generalized Wilcoxon)	24.524	1	0.000
Tarone-Ware	22.282	1	0.000

The above Table 3.1.1 shows that all the three tests for the equality of survival distributions are highly significant, which means that there is a significant difference in survival between the male and female. From the survival curve we can also infer that males have better survival than females.

Having tested for the survival distributions we now present the survival curve for the gender of the patients in the figure below.

Figure 3.1.1 **Survival curve for sex of patients**



The survival curve for the gender of patients shown above figure 3.1.1 tells that the survival for males is higher than the survival for females. Kaplan Meier method is a paradigm for nonparametric type; it has no assumption about the shape of hazard function. Hazard function is estimated based on empirical data, showing change over time and the Kaplan-Meier method in survival analysis is the best example. From the above Kaplan-Meier curves, we note that a substantial difference between the sex exists. From the graph we see that the survival function for each group of **sex** are not perfectly parallel but separate except at the beginning and The overlap at the middle will not cause too much concern because it is determined by only a very few number of censored subjects out of a sample with 1240 subjects.

3.2 Kaplan-Meier Analysis for treatment groups

In this Section we perform Kaplan Meier analysis for the different treatment regimen administered to the patients included in the study and obtain their means and median survival times and the survival curve for the regimen of the patients is also obtained. The equality of survival distributions is tested using the log-rank test.

In the table below the significance values based on different tests for equality of the survival distributions for treatment regimen is given.

Table 3.2.1
Test for equality of the treatment regimens.

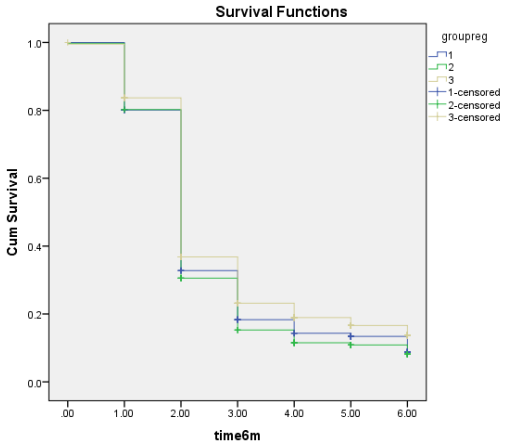
	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	8.169	2	0.017
Breslow (Generalized Wilcoxon)	5.847	2	0.054

Tarone-Ware	6.954	2	0.031

The above Table 3.2.1 shows that all the three tests for the equality of survival distributions for the different treatments are significant, which means that there is a significant difference in survival times of patients receiving different treatments.

Having tested for the survival distributions of treatment regimen we now present the survival curve for the treatment regimen administered to the patients in the figure below.

Figure 3.2.1 **Survival Curve for treatment regimen.**



The survival curve for the treatment groups of patients shown above Figure 3.2.1 tells that the survival differs for different regimen .Kaplan Meier method is a paradigm for nonparametric type; it has no assumption about the shape of hazard function. Hazard function is estimated based on empirical data, showing change over time and the Kaplan-Meier method in survival analysis is the best example. From the above Kaplan-Meier curve, we note that a substantial difference between the treatments exists.

4. Analysis of AFT 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.Exponential regression – log relative-hazard form

In the following table a brief summary of the data and the loglikelihood measures under the exponential regression 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 87.2 and which is highly significant. The loglikelihood of the model is -1427.1052

Table 4.1.1

No. of subjects	1237
Number of observations	1237

No. of failures	1062
Time at risk	3054
LR chi2(5)	87.2
Log likelihood	-1427.1052
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	0.79458	0.0601	-3.04	0.002	0.6851	0.92155
Age	0.9979	0.00279	-0.75	0.452	0.99244	1.00338
wt_0	1.01073	0.00476	2.27	0.023	1.00145	1.0201
Present	1.99569	0.17799	7.75	0	1.67562	2.3769
Groupreg	0.93993	0.03515	-1.66	0.098	0.87351	1.01141

From the above STATA output we observe that the variables such as sexcode, weight, present and groupreg 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 -1427.1052. 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(-1427.1052) = 2854.21.

Remark 1

The Hazard 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 Exponential model is 1.99569. So we now have exp (1.99569) = 7.357278. Those who are sensitive to the drug are likely to reduce the sputum conversion time by 7 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 Hazard 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 weibull model.

4.2 Weibull regression -- log relative-hazard form

In the following table a brief summary of the data and the loglikelihood measures under the Weibull regression 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 185.08 and which is highly significant. The loglikelihood of the model is -1157.1355.

Table 4.2.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	185.08
Log likelihood	-1157.1355
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 4.2.2

Table 4.2.2

_t	Haz. Ratio	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.736376	0.05509	-4.09	0	0.63594	0.85267
Age	0.99742	0.00279	-0.93	0.354	0.99197	1.00289
wt_0	1.01459	0.00452	3.25	0.001	1.00576	1.0235
Present	2.74432	0.24829	11.16	0	2.29839	3.27677
Groupreg	0.91278	0.03386	-2.46	0.014	0.84876	0.98162
/ln_p	0.63105	0.02283	27.64	0	0.5863	0.67579
P	1.87958	0.04291			1.79733	1.96558
1/p	0.53204	0.01215			0.50875	0.55638

From the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg and ln_p 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 -1157.1355. 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 weibull model we have -2LL = -2(-1157.1355) = 2314.271.

Remark 2

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 is 2.74432. So we now have exp(2.74432)= 15.55403. 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.

4.3. Lognormal regression – Accelerated Failure-Time form

In the following table a brief summary of the data and the loglikelihood measures under the exponential regression 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 116.92 and which is highly significant. The loglikelihood of the model is -1029.2903.

Table 4.3.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	116.92
Log likelihood	-1029.2903
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	0.18213	0.03989	4.57	0	0.10395	0.26031
Age	0.00333	0.00142	2.34	0.019	0.00054	0.00613

wt_0	-0.0072	0.00257	-2.82	0.005	-0.0123	-0.0022
Present	-0.375	0.04148	-9.04	0	-0.4563	-0.2937
Groupreg	0.03702	0.01938	1.91	0.056	-0.001	0.075
_cons	1.10243	0.12056	9.14	0	0.86614	1.33871
/In_sig	-0.6108	0.02232	-27.37	0	-0.6546	-0.5671
Sigma	0.54291	0.01212			0.51967	0.56718

From the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, _cons and ln_sig are found to be significant. The Coefficient for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -1029.2903. 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 Lognormal model we have -2LL = -2(-1029.2903) = 2058.581.

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 is -0.375. So we now have exp(-0.375)= 0.6872893. Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.6873 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 coefficients 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.

4.4. Loglogistic regression – Accelerated Failure-Time form

In the following table a brief summary of the data and the loglikelihood measures under the exponential regression 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 102.31 and which is highly significant. The loglikelihood of the model is -1024.5177.

Table 4.4.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	102.31
Log likelihood	-1024.5177
Prob > chi2	0.000

In the following table Coefficient, standard error, significance and confidence intervals for each of the variables are displayed in table 4.4.2

Table 4.4.2

_t	Coet.	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.18637	0.03912	4.76	0	0.10969	0.26305
Age	0.00343	0.00139	2.47	0.013	0.00071	0.00614
wt_0	-0.007	0.00257	-2.74	0.006	-0.0121	-0.002
Present	-0.3528	0.04474	-7.89	0	-0.4405	-0.2651
Groupreg	0.03336	0.01889	1.77	0.077	-0.0037	0.07039

_cons	1.058144	0.1219	8.68	0	0.81922	1.29707
/In_gam	-1.1756	0.02586	-45.46	0	-1.2263	-1.1249
Gamma	0.30863	0.00798			0.29337	0.32467

From the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, _cons and ln_gam are found to be significant. The Coefficient for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -1024.5177. 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 Loglogistic model we have -2LL = -2(-1024.5177) = 2049.035.

Remark 4

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 model is -0.3528. So we now have exp(-0.3528)= 0.7027177. Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.7027 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 coefficients 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 Gamma model

4.5. Gamma regression – Accelerated Failure-Time form

In the following table a brief summary of the data and the loglikelihood measures under the exponential regression 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 65.29 and which is highly significant. The loglikelihood of the model is -1003.2559.

Table 4.5.1

No. of subjects	1237
Number of observations	1237
No. of failures	1062
Time at risk	3054
LR chi2(5)	65.29
Log likelihood	-1003.2559
Prob > chi2	0.000

In the following table Coefficient, standard error, significance and confidence intervals for each of the variables are displayed in table 4.5.2

Table 4.5.2

_t	Coet.	Std. Err.	Z	P>|Z|	[95% Conf. Interval]	
Sexcode	0.15624	0.03835	4.07	0	0.08107	0.2314
Age	0.00409	0.00131	3.11	0.002	0.00151	0.00666
wt_0	-0.0054	0.00245	-2.19	0.028	-0.0102	-0.0006
Present	-0.2314	0.04223	-5.48	0	-0.3141	-0.1486
Groupreg	0.02985	0.01838	1.62	0.104	-0.0062	0.06587
_cons	0.73696	0.12291	6	0	0.49605	0.97786
/In_sig	-0.6592	0.02499	-26.38	0	-0.7082	-0.6102
/Kappa	-0.708	0.09764	-7.25	0	-0.8993	-0.5166

Sigma	0.51726	0.01293			0.49254	0.54322

From the above STATA output we observe that the variables such as age, sexcode, weight, present, groupreg, _cons and ln_sig are found to be significant and also Kappa statistics be significant for this model. The Coefficient for each of the variables and their corresponding confidence intervals are also displayed.

The loglikelihood (LL) of the model is found to be -1003.2559. 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 Gamma model we have -2LL = -2(-1003.2559) = 2006.512.

Remark 5

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 Gamma model is -0.2314. So we now have exp(-0.2314)=0.793442 . Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.7934 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 coefficients of variables in the model.

5.CONCLUSION

Lower values of -2LogLikelihood suggest a better model. With the exception of the Weibull and log normal distribution, 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 -2LL. For the parametric models presented in the Tables -2LL of **gamma is 2006.5118** and for log normal distribution -2LL is 2058.5806 and for log logistic it is 2049.0354. Decision based on -2LL, **Gamma distribution is the most suitable model for our data** set. In Gamma distribution all the variables like age of the patient, sex to which the patient belongs, presens etc , and weight of the patient 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 Gamma model is -0.2314. So we now have exp(-0.2314)=0.7934. Those who are sensitive to the drug are likely to increase the sputum conversion time by 0.7934 times compared to patients who are resistant. In the Gamma model, the weight of the patient also has a significant impact on their sputum conversion times. Similar interpretations can be made using the coefficients of variables in all the models.

Acknowledgement

Authors Thank **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. 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
3. Civil society perspective of TB care and control in India: Challenges and Solutions. Partnership for TB care and Control. Stop TB Partnership. Available from:http://www.stoptb.org/assets/documents/countries/partnerships/Civil%20Society%20contribution%20to%20RNTCP%20Phase%20III%20planning.pdf .
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. Jin Z, Lin DY, Wei LJ and Ying ZL (2003) Rank-based inference for the accelerated failure time model. Biometrika. 90, 341-353

10. Kalbfleisch JD and Prentice RL (1980) The statistical analysis of failure time data, John Wiley & Sons, NY
11. 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.
12. Lee ET and Wang JW. (2003) Statistical Methods for Survival Data Analysis.
13. Miller RG. (1981) Survival Analysis. *Wiley & Sons,* New York.