Thailand Statistician
July 2018; 16(2): 221-229
http://statassoc.or.th
Short communication

Scientific note: Similarities between Survival Analysis using Kaplan-Meier and ANOVA

Hossam F. Abou-Shaara
Department of Plant Protection, Faculty of Agriculture, Damanhour University, Damanhour, 22516, Egypt.
*Corresponding author; e-mail: hossam.farag@agr.dmu.edu.eg

Received: 13 August 2017
Accepted: 4 March 2018

Abstract

Survival experiments are essential for researches in various biological fields. Survival analysis can be done using Kaplan-Meier. This test is not well known than ANOVA. It is hypothesized that ANOVA can give similar results to Kaplan-Meier. In this study, three survival experiments using different number of treatments or cases were analyzed to compare ANOVA and Kaplan-Meier. Similarities between Duncan after ANOVA and Kaplan-Meier with Mantel-Cox, Breslow, or Tarone-Ware were found. Also, percentages of survived individuals from Kaplan-Meier were similar to calculated cumulative mortality.

Keywords: ANOVA, Kaplan-Meier, Mantel-Cox, Post Hoc, survival.

1. Introduction

In many biological studies, recording the event (i.e. death) over specific period is important. Such survival studies are essential in many research fields including: medical studies, zoology, and entomology. Kaplan-Meier as a non-parametric estimator is used for survival analysis. In case that the weights are related to the survival function of times, the Kaplan-Meier can be used as a weighted average of identically distributed terms (Satten and Datta 2001). However, Kaplan-Meier analysis alone is not sufficient especially when estimates of the degree of separation of the different subgroups are required (Ture et al. 2009). Cumulative incidence competing risks (CICR) has been found to be most appropriate than Kaplan-Meier methods which overestimate event rates when an analysis focuses on first events with competing risks (Southern et al. 2006). The logistic regression estimates are closely related to Kaplan-Meier curves (Efron 1988).

It is also possible to record the percentages of event (i.e. mortality) over the study period and then compare test groups according to their mortality. The analysis of percentages can be accomplished using Analysis of Variance (ANOVA) and Post-Hoc tests. ANOVA is well known by researchers than Kaplan-Meier. Therefore, it is hypothesized that results of Kaplan-Meier are similar to ANOVA for survival experiments. To test this hypothesis, three survival experiments were analyzed using the two methods. The results of the two methods were then compared.

2. Methods

Presumptive data for three survival experiments were used to compare Kaplan-Meier with ANOVA.

2.1. Three groups with equal number of cases

In this experiment the number of dead individuals (bees) was recorded daily and up to 6 days for three groups. Three replicates (10 individuals per replicate) per group (the total number of cases = 30 per group). For ANOVA, the daily death percentages were calculated for each replicate. For Kaplan-Meier, the event (death = 1) was recorded per each individual.

2.2. Four groups with unequal number of cases

Four groups were used in this experiment. Three replicates per each group but for the first three groups 10 individuals per replicate (30 per group) was used while for the fourth group 15 individuals were used per replicate with a total of 45 per group. The daily death percentage was recorded over 6 days and then groups were compared using ANOVA. But for Kaplan-Meier analysis, the event (death = 1) was recorded per each individual.

2.3. Six groups with equal number of cases

In this experiment, six groups were compared over 6 days using 3 replicates (cases) per group. In each replicate 10 individuals with a total of 30 per group. The daily death percentage was calculated per each group and compared using ANOVA. The event (death = 1) was recorded per each individual to perform the Kaplan-Meier analysis.

2.4. Statistical analysis

SPSS v.16 (Released 2007. SPSS for Windows, Version 16.0. Chicago, USA) was used to perform the analysis. The groups based on the events (death = 1) were compared using Kaplan-Meier followed by Mantel-Cox, Breslow, and Tarone-Ware. The difference between groups was considered significant when $p \leq 0.05$. ANOVA at significant level of 0.05 followed by Duncan at significant level of 0.1 was used to compare groups (as independent factor) based on their death percentages (as dependent factor).

3. Results and Discussion

3.1. Three groups with equal number of cases

Cumulative mortality for the groups over the 6 days was placed in Table 1 and Figure 1. Kaplan-Meier showed that censored percentages of survived individuals (Table 2 and Figure 2) were similar to percentages of survived individuals at day 6 (= 100- death percentages at day 6) presented in Table 1. The overall comparisons between survival curves of the groups using Mantel-Cox, Breslow, and Tarone-Ware (Table 3) showed the presence of significant differences ($p < 0.05$). Pairwise comparison (Table 4) showed no significant differences ($p > 0.05$) between group 2 and 3 while the differences between group 1 and each of group 2 and 3 were significance ($p < 0.05$).

Table 1 Cumulative mortality over 6 days for the three groups

Days	Groups		
	1	2	3
1	3.3	10.0	10.0
2	16.7	26.7	36.7
3	26.7	40.0	43.3
4	30.0	56.7	63.3
5	43.3	73.3	86.7
6	46.7	80.0	90.0
Percentage of survived individuals at day 6	53.3	20	10

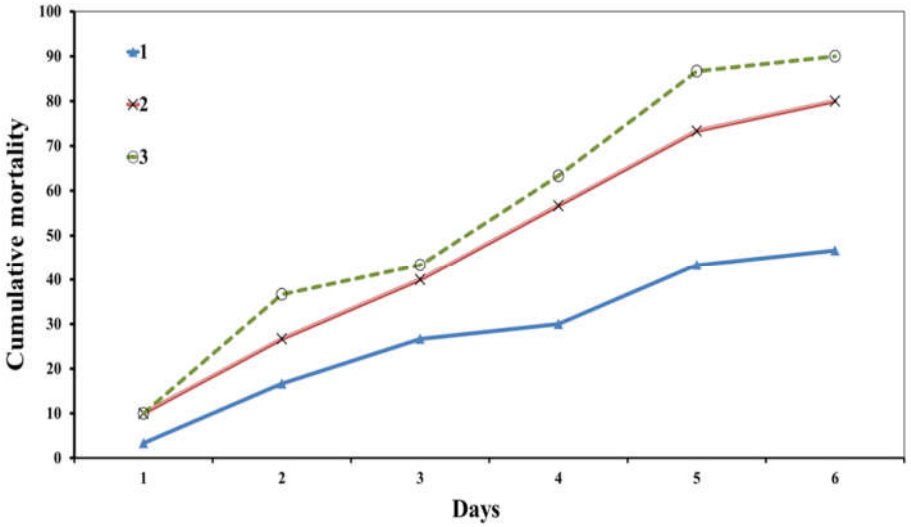

Figure 1 Cumulative mortality over 6 days for the three groups

Table 2 Case processing summary of Kaplan-Meier analysis

			Censored	
Treatments	Total N	N of Events	N	Percentage
1	30	14	16	53.3
2	30	24	6	20.0
3	30	27	3	10.0
Overall	90	65	25	27.8

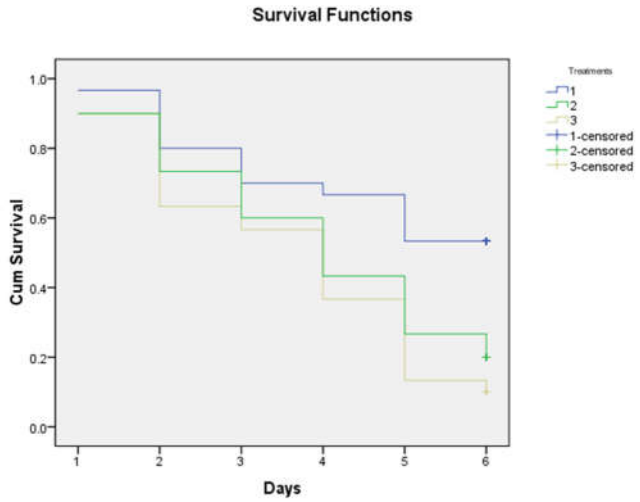

Figure 2 Cumulative survival over 6 days for the three groups

Table 3 Groups comparison over strata

	Chi-Square	df	p-value
Log Rank (Mantel-Cox)	11.782	2	.003
Breslow (Generalized Wilcoxon)	8.171	2	.017
Tarone-Ware	9.897	2	.007

Test of equality of survival distributions for the different levels of Treatments.

Table 4 Pairwise comparison over strata

		1		2		3	
	Treatments	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value
Log Rank (Mantel-Cox)	1			5.951	.015	10.932	.001
	2	5.951	.015			.896	.344
	3	10.932	.001	.896	.344		
Breslow	1			4.160	.041	7.912	.005
	2	4.160	.041			.655	.418
	3	7.912	.005	.655	.418		
Tarone-Ware	1			5.012	.025	9.360	.002
	2	5.012	.025			.777	.378
	3	9.360	.002	.777	.378		

According to ANOVA, there were significant differences between groups ($p < 0.05$) (Table 5), similar to findings in Table 3. Post Hoc tests using Duncan showed that group 1 was significantly

different than group 2 and 3 while no significant differences was existed between group 2 and 3 (Table 6). This finding is very similar to Table 4 which presented pairwise comparisons between groups using survival analysis. This reflected the high similarities between ANOVA and survival analysis using Kaplan-Meier.

Table 5 ANOVA for the 3 groups

Source	Sum of Squares	df	Mean Square	F	p-value
Between Groups	5959.259	2	2979.630	4.089	.023
Within Groups	37161.111	51	728.649		
Total	43120.370	53			

Table 6 Post Hoc using Duncan 0.1 after ANOVA for the 3 groups

		Subset for alpha = 0.1	
Treatments	N	1	2
1	18	30.00	
2	18		47.78
3	18		55.00
p-value		1.000	.426

Means for groups in homogeneous subsets are displayed.

3.2. Four groups with unequal number of cases

As presented in Table 7, the cumulative mortality increased over the 6 days for the 4 groups. As in the above mentioned experiment the censored percentages (Table 8) are the same as percentage of survived individuals at day 6 (Table 7). Unlike the previous experiment the number of cases of group 4 was higher than the other 3 groups (Table 8). The comparisons between groups showed the presence of significant differences (Table 9). The pairwise comparison over strata (Table 10) showed that Mantel-Cox and Tarone-Ware detected the same significant differences between group 1 and groups 2, 3, and 4, and between group 4 and 3 while Breslow detected only the significant differences between group 1 and groups 2 and 3 and between groups 3 and 4.

Table 7 Cumulative mortality over 6 days for the 4 groups

Days	Groups			
	1	2	3	4
1	3.3	10.0	10.0	2.3
2	20.0	26.7	36.7	6.7
3	30.0	40.0	43.3	26.7
4	33.3	56.7	63.3	31.1
5	46.7	73.3	86.7	48.9
6	46.7	80.0	90.0	93.3
Percentage of survived individuals at day 6	53.3	20.0	10.0	6.7

Table 8 Case processing summary of Kaplan-Meier analysis

Treatments	Total N	N of Events	Censored	
			N	Percentage
1	30	14	16	53.3
2	30	24	6	20.0
3	30	27	3	10.0
4	45	42	3	6.7
Overall	135	107	28	20.7

Table 9 Groups comparison over strata

	Chi-Square	df	p-value
Log Rank (Mantel-Cox)	16.873	3	.001
Breslow (Generalized Wilcoxon)	13.875	3	.003
Tarone-Ware	15.134	3	.002

Test of equality of survival distributions for the different levels of Treatments.

Table 10 Pairwise comparison over strata

		1		2		3		4	
	Treatments	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value
Log Rank (Mantel-Cox)	1			5.951	.015	10.932	.001	6.666	.010
	2	5.951	.015			.896	.344	1.142	.285
	3	10.932	.001	.896	.344			7.861	.005
	4	6.666	.010	1.142	.285	7.861	.005		
Breslow (Generalized Wilcoxon)	1			4.160	.041	7.912	.005	2.386	.122
	2	4.160	.041			.655	.418	3.025	.082
	3	7.912	.005	.655	.418			9.987	.002
	4	2.386	.122	3.025	.082	9.987	.002		
Tarone-Ware	1			5.012	.025	9.360	.002	4.208	.040
	2	5.012	.025			.777	.378	2.119	.145
	3	9.360	.002	.777	.378			9.223	.002
	4	4.208	.040	2.119	.145	9.223	.002		

The ANOVA (Table 11) showed the presence of significant differences between the groups in a similar way to survival analysis (Table 9). Duncan (Table 12) showed the presence of significant differences between group 1 and groups 2 and 3 and between group 3 and 4. This significant separation is similar to Breslow. Duncan was different than Mantel-Cox and Tarone-Ware when 4 groups with unequal number of cases were compared.

Table 11 ANOVA for the 4 groups

	Sum of Squares	df	Mean Square	F	p-value
Between Groups	7163.379	3	2387.793	2.939	.039
Within Groups	55254.617	68	812.568		
Total	62417.996	71			

Table 12 Post Hoc using Duncan after ANOVA for the 4 groups

Treatments	N	Subset for alpha = 0.1		
		1	2	3
1	18	30.00		
4	18	34.82	34.82	
2	18		47.78	47.78
3	18			55.00
p-value		.614	.177	.450

Means for groups in homogeneous subsets are displayed.

3.3. Six groups with equal number of cases

The percentages of survived individuals at day 6 in Table 13 were the same as censored percentages in Table 14. Groups comparison over strata (Table 15) using Breslow and Tarone-Ware showed the absence of significant differences between groups ($p > 0.05$) except Mantel-Cox which detected the presence of significant differences ($p < 0.05$). Pairwise comparison over strata (Table 16) showed the significant differences between group 1 and groups 2, 3, 5 and 6 using Mantel-Cox, Breslow, and Tarone-Ware.

Table 13 Cumulative mortality over 6 days for the three groups

Days	Groups					
	1	2	3	4	5	6
1	3.3	10.0	10.0	3.3	3.3	3.3
2	16.7	26.7	36.7	10.0	16.7	6.7
3	26.7	40.0	43.3	40.0	40.0	30.0
4	30.0	56.7	63.3	46.7	46.7	60.0
5	43.3	73.3	86.7	73.3	63.3	86.7
6	46.7	80.0	90.0	73.3	90.0	96.7
Percentage of survived individuals at day 6	53.3	20.0	10.0	26.7	10.0	3.3

Table 14 Case processing summary of Kaplan-Meier

Treatments	Total N	N of Events	Censored	
			N	Percentage
1	30	14	16	53.3
2	30	24	6	20.0
3	30	27	3	10.0
4	30	22	8	26.7
5	30	27	3	10.0
6	30	29	1	3.3
Overall	180	143	37	20.6

Table 15 Groups comparison over strata

	Chi-Square	df	p-value
Log Rank (Mantel-Cox)	17.448	5	.004
Breslow (Generalized Wilcoxon)	10.449	5	.063
Tarone-Ware	13.404	5	.020

Test of equality of survival distributions for the different levels of Treatments.

Table 16 Pairwise comparison over strata

	Treatments	1		2		3		4		5		6	
		Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value	Chi-Square	p-value
Log Rank (Mantel-Cox)	1			5.951	.015	10.932	.001	2.752	.097	7.440	.006	11.411	.001
	2	5.951	.015			.896	.344	.698	.403	.004	.947	.641	.423
	3	10.932	.001	.896	.344			3.275	.070	1.770	.183	.110	.740
	4	2.752	.097	.698	.403	3.275	.070			.603	.437	2.621	.105
	5	7.440	.006	.004	.947	1.770	.183	.603	.437			1.574	.210
	6	11.411	.001	.641	.423	.110	.740	2.621	.105	1.574	.210		
Breslow (Generalized Wilcoxon)	1			4.160	.041	7.912	.005	1.679	.195	3.868	.049	5.309	.021
	2	4.160	.041			.655	.418	.649	.420	.208	.649	.006	.940
	3	7.912	.005	.655	.418			2.936	.087	2.276	.131	1.103	.294
	4	1.679	.195	.649	.420	2.936	.087			.136	.713	.515	.473
	5	3.868	.049	.208	.649	2.276	.131	.136	.713			.399	.528
	6	5.309	.021	.006	.940	1.103	.294	.515	.473	.399	.528		
Tarone-Ware	1			5.012	.025	9.360	.002	2.196	.138	5.462	.019	7.946	.005
	2	5.012	.025			.777	.378	.648	.421	.090	.764	.097	.756
	3	9.360	.002	.777	.378			3.051	.081	2.158	.142	.575	.448
	4	2.196	.138	.648	.421	3.051	.081			.269	.604	1.220	.269
	5	5.462	.019	.090	.764	2.158	.142	.269	.604			.897	.344
	6	7.946	.005	.097	.756	.575	.448	1.220	.269	.897	.344		

The comparison between groups using ANOVA (Table 17) showed the absence of significant differences in a way similar to group comparisons using Breslow and Tarone-Ware. Duncan (Table 18) showed the presence of significant differences only between group 1 and 3. This was different than pairwise comparison between groups using Mantel-Cox, Breslow, and Tarone-Ware. It was clear that when 6 groups were compared ANOVA showed few similarities to survival analysis. Also, it was clear that results of Mantel-Cox were not always the same as Breslow, and Tarone-Ware. It could be said that comparisons between the 6 groups can be done using ANOVA or survival analysis. But pairwise comparisons better to be done using survival analysis.

Table 17 ANOVA for the 6 groups

	Sum of Squares	df	Mean Square	F	p-value
Between Groups	6307.407	5	1261.481	1.339	.254
Within Groups	96100.000	102	942.157		
Total	102407.407	107			

Table 18 Post Hoc using Duncan after ANOVA for the 6 groups

	Sum of Squares	df	Mean Square	F	p-value
Between Groups	6307.407	5	1261.481	1.339	.254
Within Groups	96100.000	102	942.157		
Total	102407.407	107			

4. Conclusion

Group comparisons using Kaplan-Meier showed similarity to ANOVA apart from number of groups. Pairwise comparisons using Mantel-Cox were not always the same as Breslow, and Tarone-Ware. Similarities were found between Kaplan-Meier (Mantel-Cox, Breslow, and Tarone-Ware) and ANOVA (Duncan) when 3, 4 and 6 groups were compared. ANOVA can be used to compare groups using cumulative mortality as alternative to Kaplan-Meier.

References

Efron B. Logistic Regression, survival analysis, and the Kaplan-Meier curve. J Am Stat Assoc. 1988; 83 (402): 414-425.

Satten GA, Datta S. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. Am Stat. 2001; 55(3): 207-210.

Southern DA, Faris PD, Brant R, Galbraith PD, Norris CM, Knudtson ML, Ghali WA. Kaplan-Meier methods yielded misleading results in competing risk scenarios. J Clin Epi. 2006; 59(10): 1110-1114.

Ture M, Tokatli F, Kurt I. Using Kaplan-Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. Expert Syst Appl. 2009; 36(2):2017-2026.