Cihan University-Erbil Scientific Journal (CUESJ)

RESEARCH ARTICLE

The Estimation of (Covid-19) Cases in Kurdistan Region Using Nelson Aalen Estimator

Sami A. Obed[1*], Parzhin A. Mohammed[1], Dler H. Kadir[1,2]

[1]*Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil, Kurdistan Region - F.R. Iraq,*
[2]*Department of Business Administration, Cihan University-Erbil, Kurdistan Region, Iraq*

ABSTRACT

It is described how the Nelson–Aalen estimator may be used to control the rate of a nonparametric estimate of the cumulative hazard rate function based on right censored as well as left condensed survival data, furthermore how the Nelson–Aalen estimator can be utilized to estimate various amounts. This technique is mostly applied to survival data and product quality data similar to the incorporated relative mortality in a multiplicative model with outer rates and the cumulative infection rate in a straightforward epidemic model. It is shown that tallying measures produce a structure that permits to a brought together treatment of all these different conditions, and the main little and massive sample properties of the assessor are summarized. This estimator is a weighted average of the Nelson-Aalen reliability estimates over 2 time periods. The suggested estimator's suitability and utility in model selection are reviewed. Morover, a real-world dataset is evaluated to demonstrate the proposed estimator's suitability and utility. This work proposes a simple and nearly unbiased estimator to fill this gap. The information was gathered from the Ministry of Health's website between October 1, 2020, and February 28, 2021. The results of the Nelson Allen Estimator demonstrated that the odds of surviving were higher during a short period of time after being exposed to the virus. As time passes, the possibilities become slimmer. The closer the estimate comes to value 1 from 0.5 upward, the greater the chances of surviving the infection.

Keywords: Covied-19, death, estimator, nelson Allen estimator, survival analysis

INTRODUCTION

Nelson-Aalen estimation is useful to separate how a assumed population develops through time. This strategy is frequently applied to survival data and product feature information. This makes the virus lethal since it can be encountered by someone who is infected without showing signs. Many papers/articles had been written on the novel coronavirus (Covid-19) within a relatively short time after its existence.[1]

This research will answer questions such as: What are the affected people's odds of survival? What is the survival probability's time frame? How can a person survive a viral outbreak, particularly in the impacted areas?

The Nelson-Aalen estimator, or more generally visualizing the hazard function over time, is not a very popular approach to survival analysis. That is because in evaluation to the survival function clarification of the curves is not so simple and intuitive. However, the risk function is of great significance for more advanced methods to survival analysis, for example, the Cox regression. That is why it is vital to appreciate the concept. Moreover, it will try to provide some insights about it. We can say that the cumulative hazard functions (CHFs).[2]

The Nelson-Aalen estimation is a method for estimating a public's hazard function without considering that the data originates from a specific distribution. The Nelson-Aalen approach produces the CHF from the hazard function, which is then used to obtain the survival function. It is a non-parametric approach of generating results due to the lack of parameters necessary in this model.

This research will serve as a foundation for health policy formulation, resulting in fewer untimely deaths and an overall improvement in quality of life.[2]

The cumulative hazard is one of the most reliable parameters in the analysis of time-to-event data with

Corresponding Author:
Sami A. Obed, Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil, Kurdistan Region - F.R. Iraq E-mail: muhsin.hmeadi@cihanuniversity.edu.iq

Received: July 7, 2021
Accepted: September 6, 2021
Published: September 30, 2021

DOI: 10.24086/cuesj.v5n2y2021. pp 24-31

independently right-censored and left-truncated survival times. Cumulative transition intensity is another name for it. The well-known Nelson–Aalen estimator is frequently used to estimate it nonparametric ally. In this context time-simultaneous confidence bands are perhaps the best interpretive tool for accounting for related estimating uncertainty in this situation.[3]

METHODOLOGY

This section studied some basic concepts of survival analysis; survival function, hazard function, and some tests and methods used to analyze survival data.

Coronavirus is a virus whose genome consists of a single strand of ribonucleic acid. This is to distinguish this disease from other coronavirus outbreak in the past or future. This was a result of genetic comparison of the animal and the infected person. Numerous review methods and strategies were designed during the course of this study to collect data/materials relevant to the study in order to gain a thorough understanding of the subject matter.[4]

A survival analysis can be used to evaluate not just the likelihood of manufacturing equipment failure based on the number of hours it has been in operation, but also to distinguish between different operating situations. If the chance varies depending on whether the equipment is operated outdoors or indoors, for example.[4]

As a result, it is important to remember that the overall quality of the sample frames and methods utilized determines the trustworthiness of the survey estimates. In general, the broader the sample frame, the greater the result's quality and validity.[5]

NELSON–AALEN ESTIMATOR

The Nelson–Aalen estimator is a nonparametric estimator which might be used to estimate the cumulative hazard rate function from censored survival information the Nelson–Aalen estimator might be applied and exemplify its use in one specific case. Moreover, it is shown how checking measures give a system, which permits to a bound together treatment of every one of these different circumstances, and summarizes the main properties of the Nelson–Aalen estimator.[2]

The Nelson-Aalen analysis strategy has a place with the enlightening strategies for survival analysis, for example, life table analysis and Kaplan-Meier investigation. The Nelson-Aalen approach can rapidly provide you with a bend of combined peril and gauge the hazard capacities dependent on unpredictable time intervals.[6]

An "occasion" can be the failure of a non-repairable component, the death of a human being, or any occurrence for which the experimental unit remains in the "failed" state (e.g., death) from the point at which it changed on.[4]

SURVIVAL FUNCTION ESTIMATOR[5]

Let $t_1 < t_2 <\ldots < t_k$ represent the observed death times in a sample of n subjects from a homogeneous population with survival function S(t). Consider S(t) as a discrete function with probability mass at each t_i,; i =1;…; k. Therefore it can be written that

$$S_{(ti)} = (1-q_1)(1-q_2)\ldots(1-q_i) = \prod\nolimits_{j=1}^{i}(1-q_j) \quad (1)$$

where q_j is the probability of subject death in the interval $[t_j\text{-}1,\ t_i]$conditional of being alive at t_j-1 that is, q_j can be written as

$$q_j = p(T \in [\,t_j\text{-}1, t_j) \,/\, T \geq t_j\text{-}1) \quad (2)$$

Suppose that d_i deaths occurs at t_i and there are n_i subjects at risk at t_i, KME is obtained from (1) and (2) as

$$\dot{S}_{KM(t)} = \prod_{i/t_i<t} \frac{(n_i - d_i)}{n_i} = \prod_{i/t_i<t} (1 - \frac{d_i}{n_i}) \quad (3)$$

It is a staircase function of t with a jump at each failure time.

The NAE is given by

$$H_{(t_j)} = \sum\nolimits_{k=1}^{j} d_k/n_k,\ R(t_j) = \exp\left(-H_{(t_j)}\right) = \coprod_{k=1}^{j} - d_k/n_k \quad (4)$$

Where H(t) is the CHF and is also a staircase function of t.

It is noted that the above definitions extend the original definitions of the NAE to the censored observations.[1]

CENSORING CATEGORIES FOR THE NELSON-AALEN ANALYSIS[3]

There are some types of censoring of survival data:

1. Left censoring: We know that an event occurred at t * t when it is reported at time t=t(i) (i)
2. Right censoring: When an occasion is reported at time t=t(i), we see that it happened at time t * t(i) if it happened at all
3. Interval censoring: We know that an event occurred during [t(i-1); t(i)] when it is recorded at time t=t(i)
4. Exact censoring: We know that an incident happened exactly at t=t(i) because it is reported at time t=t(i) (i).

INDEPENDENT CENSORING FOR THE NELSON-AALEN METHOD

The Nelson-Aalen technique requires that the explanations be unbiased. Second, the censoring should be consistent – if two random persons are enrolled in the research at time t-1 and the other is censored at time t while the other lives, each must have the same chance of surviving at time t.[7]

Independent censorship can be divided into four categories:

- Simple Type I: All individuals are censored at the same time, or individuals are followed for a set period of time
- Progressive Type I: At the same time, everybody is censored (for illustration, when the study terminates).
- Type II: The research will attempt until all of the n measures have been validated
- Random: The period when censoring occurs is unrelated to the time it takes to survive.[8]

THE NELSON-AALEN METHOD AND THE CHF

The Nelson-Aalen methodology allows for population comparisons based on their hazard curves. When studying CHFs, the Nelson-Aalen estimator should be used instead of the Kaplan-Meier estimator. The Kaplan-Meier estimator should be used when evaluating cumulative survival functions.

The CHF is a function that calculates the probability of:

$$\mathrm{H(T)} = \sum_{T_i \le T} \frac{d_i}{r_i} \qquad (5)$$

With d_i being the number of observation dropping at time Ti and r_i the number of observation at hazard (still in the study) at time Ti.[9]

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 gives the results of Nelson Allen Estimator from the first of October (2020) till 28th of February (2021) applied to a data set of size 61716 cases, and shows that the most of the percentage of disease in this study are on October (2020) (46.1%) and deaths (28.9%). After that on November (2020) the percent estimate to life (33.6%) and (27.9%) was death. On the other hand, the highest estimated time until the death on December (2020) and equal (37.5%). The lowest estimated time for death which is equal to (2.1%) on February (2021).

Application Nelson Aalen Survival Estimate Formula/Notations

$e^{-\sum_{t=0}^{m-1}\mu_t}$ Nelson Aalen survival estimate (SNA(t)

$\sum_{t=0}^{m-1}\mu_t$ Estimate of the CHF.

m day of observation

t_m date/time death observed

d_m number of deaths at tm

n_m number of people available to die at tm

$\mu_m = \frac{d_m}{n_m}$ hazard rate or probability that a life fails at tm.

Table 1 : Descriptive statistics for qualitative variables

Months	Status	%
October (2020)	Disease	46.1
	Deaths	28.9
November (2020)	Disease	33.6
	Deaths	27.9
December (2020)	Disease	11.6
	Deaths	37.5
January (2021)	Disease	4.4
	Deaths	3.4
February (2021)	Disease	4.6
	Deaths	2.1

ANALYSIS AND INTERPRETATION OF DATA

A value of 0.5 in the Nelson Aalen survival model suggests an equal chance of life and death. This survival model was chosen to analyze the data since it is a non-parametric technique with a non-informative censoring estimator assumption. From 0.5 to 1, the closer the value is to 1, the better the chance of survival, and vice versa. Nelson Aalen survival analysis of the data is shown in Table 2 Calculation of Nelson Aalen Survival Analysis.

In the Kurdistan Region, there were 61171 confirmed cases and 2476 deaths, according to Table 3. The total reported case number is the total number of people who have been infected, died, or who have been recovered/discharged. Out of the 61171 cases reported, 28234 were found to be infected in October. The remaining 28% of cases were those that had a resolution this month.

In Figure 1, observation/investigation of cases was approved out from October 1 to February 28; 2021. The descriptive statistics is concerned largely with summary calculations and graphical displays of results/data to derive reasonable decisions.

The horizontal axis shows the time to event, and the vertical axis shows the number of Cases and Deaths In this plot, drops in the survival curve occur whenever the covid-19 takes effect in a patient. Thus, any point on the survival curve shows the probability that a patient on a given treatment will not have experienced relief by that time.

Explanation:

In Figure 2, the survival estimate and the range of time it applies have been clearly stated for better understanding. In the same vein, the value of m which ∑ μt=1 applies has also been shown. That is if there are no lives remaining to be censored among the observed lives when the investigation ended, and the estimation of the last value of the estimate will be zero which means sure death at that time. From the analysis carried out in Table 3, at the initial day of the investigation, the estimate shows the value of one (1).

From the analysis, it is evident that as time passes by, the estimate figure activates to reduce, meaning that more

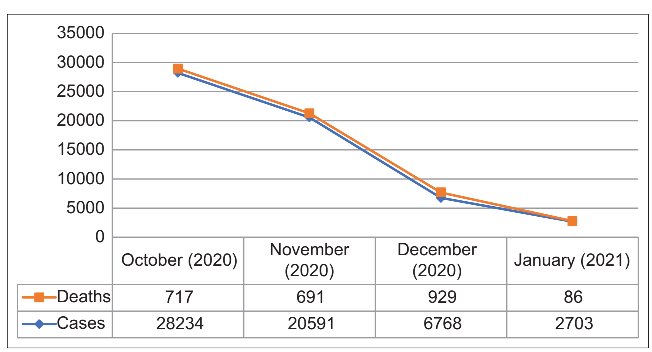

Figure 1: Monthly Deaths and Cases Growth Rate

Table 2: Results of Estimation of Covid-19 cases

Day of observation	Cases	Death	Hazard rates	Cumulative hazard rates	Estimation
1	684	21	0.0307	0.0307	0.97
2	640	21	0.0328	0.06351	0.938
3	494	16	0.0324	0.0959	0.909
4	630	19	0.0302	0.12606	0.882
5	839	31	0.0369	0.16301	0.85
6	793	30	0.0378	0.20084	0.818
7	886	29	0.0327	0.23357	0.792
8	793	31	0.0391	0.27266	0.761
9	838	12	0.0143	0.28698	0.751
10	769	19	0.0247	0.31169	0.732
11	589	24	0.0407	0.35244	0.703
12	720	26	0.0361	0.38855	0.678
13	900	23	0.0256	0.41411	0.661
14	912	20	0.0219	0.43604	0.647
15	974	28	0.0287	0.46478	0.628
16	932	18	0.0193	0.4841	0.616
17	722	23	0.0319	0.51595	0.597
18	797	32	0.0402	0.5561	0.573
19	1032	26	0.0252	0.5813	0.559
20	1010	24	0.0238	0.60506	0.546
21	1078	15	0.0139	0.61897	0.538
22	1308	17	0.013	0.63197	0.532
23	1053	20	0.019	0.65096	0.522
24	860	26	0.0302	0.6812	0.506
25	828	25	0.0302	0.71139	0.491
26	1283	23	0.0179	0.72932	0.482
27	1597	24	0.015	0.74434	0.475
28	1297	27	0.0208	0.76516	0.465
29	1285	18	0.014	0.77917	0.459
30	1002	25	0.025	0.80412	0.447
31	689	24	0.0348	0.83895	0.432
32	1054	28	0.0266	0.86552	0.421
33	1137	26	0.0229	0.88839	0.411
34	1235	23	0.0186	0.90701	0.404
35	962	28	0.0291	0.93612	0.392
36	1195	23	0.0192	0.95536	0.385
37	997	33	0.0331	0.98846	0.372
38	771	21	0.0272	1.0157	0.362
39	787	21	0.0267	1.04238	0.353
40	900	24	0.0267	1.06905	0.343
41	765	27	0.0353	1.10434	0.331
42	772	22	0.0285	1.13284	0.322
43	1170	29	0.0248	1.15763	0.314
44	686	28	0.0408	1.19844	0.302

(*Contd...*)

Table 2: (*Continued*)

Day of observation	Cases	Death	Hazard rates	Cumulative hazard rates	Estimation
45	573	27	0.0471	1.24556	0.288
46	571	28	0.049	1.2946	0.274
47	659	23	0.0349	1.3295	0.265
48	645	25	0.0388	1.36826	0.255
49	576	22	0.0382	1.40646	0.245
50	530	21	0.0396	1.44608	0.235
51	567	23	0.0406	1.48664	0.226
52	361	20	0.0554	1.54204	0.214
53	495	17	0.0343	1.57639	0.207
54	488	21	0.043	1.61942	0.198
55	500	21	0.042	1.66142	0.19
56	435	25	0.0575	1.71889	0.179
57	379	17	0.0449	1.76375	0.171
58	403	26	0.0645	1.82826	0.161
59	259	14	0.0541	1.88232	0.152
60	304	14	0.0461	1.92837	0.145
61	415	14	0.0337	1.9621	0.141
62	389	23	0.0591	2.02123	0.132
63	331	17	0.0514	2.07259	0.126
64	248	9	0.0363	2.10888	0.121
65	236	8	0.0339	2.14278	0.117
66	298	9	0.0302	2.17298	0.114
67	377	10	0.0265	2.19951	0.111
68	372	12	0.0323	2.23176	0.107
69	356	9	0.0253	2.25704	0.105
70	287	13	0.0453	2.30234	0.1
71	248	11	0.0444	2.3467	0.096
72	197	6	0.0305	2.37715	0.093
73	183	9	0.0492	2.42633	0.088
74	167	9	0.0539	2.48022	0.084
75	201	9	0.0448	2.525	0.08
76	222	3	0.0135	2.53851	0.079
77	238	11	0.0462	2.58473	0.075
78	215	6	0.0279	2.61264	0.073
79	215	8	0.0372	2.64985	0.071
80	156	4	0.0256	2.67549	0.069
81	118	8	0.0678	2.74329	0.064
82	165	3	0.0182	2.76147	0.063
83	278	5	0.018	2.77945	0.062
84	163	8	0.0491	2.82853	0.059
85	150	3	0.02	2.84853	0.058
86	164	3	0.0183	2.86683	0.057
87	128	4	0.0313	2.89808	0.055
88	118	9	0.0763	2.97435	0.051
89	132	2	0.0152	2.9895	0.05

(*Contd...*)

Table 2: (*Continued*)

Day of observation	Cases	Death	Hazard rates	Cumulative hazard rates	Estimation
90	140	3	0.0214	3.01093	0.049
91	123	3	0.0244	3.03532	0.048
92	153	1	0.0065	3.04185	0.048
93	122	4	0.0328	3.07464	0.046
94	68	2	0.0294	3.10405	0.045
95	97	1	0.0103	3.11436	0.044
96	102	6	0.0588	3.17319	0.042
97	107	6	0.0561	3.22926	0.04
98	104	3	0.0288	3.25811	0.038
99	80	4	0.05	3.30811	0.037
100	140	5	0.0357	3.34382	0.035
101	139	3	0.0216	3.3654	0.035
102	68	6	0.0882	3.45364	0.032
103	126	2	0.0159	3.46951	0.031
104	88	0	0	3.46951	0.031
105	158	1	0.0063	3.47584	0.031
106	69	4	0.058	3.53381	0.029
107	78	3	0.0385	3.57227	0.028
108	43	2	0.0465	3.61878	0.027
109	110	3	0.0273	3.64606	0.026
110	97	2	0.0206	3.66668	0.026
111	100	4	0.04	3.70668	0.025
112	68	3	0.0441	3.75079	0.023
113	77	4	0.0519	3.80274	0.022
114	85	2	0.0235	3.82627	0.022
115	49	1	0.0204	3.84668	0.021
116	58	1	0.0172	3.86392	0.021
117	94	1	0.0106	3.87456	0.021
118	67	2	0.0299	3.90441	0.02
119	70	5	0.0714	3.97584	0.019
120	73	2	0.0274	4.00324	0.018
121	68	1	0.0147	4.01794	0.018
122	48	1	0.0208	4.03877	0.018
123	50	2	0.04	4.07877	0.017
124	107	2	0.0187	4.09747	0.017
125	54	2	0.037	4.1345	0.016
126	48	3	0.0625	4.197	0.015
127	56	3	0.0536	4.25058	0.014
128	51	1	0.0196	4.27018	0.014
129	24	4	0.1667	4.43685	0.012
130	38	2	0.0526	4.48948	0.011
131	70	2	0.0286	4.51805	0.011
132	64	3	0.0469	4.56493	0.01
133	69	0	0	4.56493	0.01
134	59	0	0	4.56493	0.01

(*Contd...*)

Table 2: (*Continued*)

Day of observation	Cases	Death	Hazard rates	Cumulative hazard rates	Estimation
135	98	3	0.0306	4.59554	0.01
136	53	1	0.0189	4.61441	0.01
137	106	3	0.0283	4.64271	0.01
138	72	4	0.0556	4.69827	0.009
139	138	2	0.0145	4.71276	0.009
140	115	1	0.0087	4.72145	0.009
141	117	2	0.0171	4.73855	0.009
142	131	0	0	4.73855	0.009
143	77	1	0.013	4.75153	0.009
144	150	2	0.0133	4.76487	0.009
145	157	1	0.0064	4.77124	0.008
146	129	2	0.0155	4.78674	0.008
147	176	3	0.017	4.80379	0.008
148	199	6	0.0302	4.83394	0.008
149	202	0	0	4.83394	0.008
150	78	0	0	4.83394	0.008
151	237	0	0	4.83394	0.007

Table 3: Reported cases Covid-19 monthly in Kurdistan Region

Months	Cases	Deaths
October (2020)	28234	717
November (2020)	20591	691
December (2020)	6768	929
January (2021)	2703	86
February (2021)	2875	53

deaths are being noted. This indicates that there are balances of surviving the disease at the first period of the infection, provided the infected persons take adequate medical aid. Invariably, as shown in Table 2, the more the virus stays in the body, the lower the probability of survival or the chance of recovery from the disease.

It's worth noting that the survival rate follows a specific pattern. As the value of m increases, the estimate values drop. This is due to the fact that the number of cases and deaths included in the study follow the same trend. Both the number of cases and the number of deaths recorded increased during the study period.

In practice, if the results of this study are used to predict a future epidemic of this virus, an average of 2476 infected people out of a 61171infected people are predicted to die, because the survival chance for that duration is 0.97.

This projection will be accurate or true if interventions targeted at reducing risk factors and ensuring adequate access to health care are not implemented properly and efficiently. When life's quality is poor, it will be brief.

$$S_{NA(t)}\begin{cases} 1 & 0 \\ 0.97 & 0<m\le 1 \\ 0.93 & 1<m\le 2 \\ 0.90 & 2<m\le 3 \\ 0.882 & 3<m\le 4 \\ 0.850 & 4<m\le 5 \\ 0.818 & 5<m\le 6 \\ 0.792 & 6<m\le 7 \\ 0.761 & 7<m\le 8 \\ . & . \\ . & . \\ . & . \\ 0.009 & 142<m\le 143 \\ 0.009 & 143<m\le 144 \\ 0.009 & 144<m\le 145 \\ 0.008 & 145<m\le 146 \\ 0.008 & 146<m\le 147 \\ 0.008 & 147<m\le 148 \\ 0.008 & 148<m\le 149 \\ 0.008 & 149<m\le 150 \\ 0.007 & 150<m\le 151 \end{cases}$$

Figure 2: The Nelson Aalen estimate of S(t)

CONCLUSION AND RECOMMENDATION

Conclusion

During conducting the Nelson Allen Estimator from Covied-19 cases and according to the results from the practical part the following conclusions have been drawn:

1. The number of Novel Coronavirus cases reported was unusually high. It was in the thousands. This demonstrates how easily the virus can be spread from one person to another or from an animal to a person
2. If prevention efforts were not made in a timely manner, the number could rise to a million. In this study, it was discovered that some people may be infected with this virus without realizing it because the symptoms of infection are not visible
3. These individuals infect others without even realizing it. According to the findings, those who have recently been exposed to the disease have a better chance of surviving if medical assistance is provided without severe or time-consuming requirements.

Recommendation

This study suggests that those who are unsure if they have contracted the disease due to a lack of symptoms should consult a doctor right once for tests. Every country's government should establish a free testing center for such validation. Making it free and easily available will encourage everyone who is sick to do a confirmatory test.

This will help to slow the disease's spread. Since the virus first appeared in Wuhan, this study suggests using the WUHAN preventive approach to keep the infection from spreading or infecting humans.

WUHAN Prevention Concept

- Wash your hands/body regularly
- Use nose cover/mask
- Have your hotness/coldness of your body checked
- Avoid unnecessary crowd
- Never touch sensitive parts of your body with unclean hands/materials/equipment.

REFERENCES

1. D. Abbas. Analysis of breast cancer data using Kaplan-Meier survival analysis. *Journal of Kufa for Mathematics and Computer*, vol. 1, no. 6, pp. 7-14, 2012.
2. T. D.Bluhmki. The wild bootstrap for multivariate Nelson-Aalen estimators. *Lifetime Data Analysis*, vol. 25, no. 1, pp. 97-127, 2019.
3. R. L. Cao. Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics*, vol. 17, no. 1, pp. 31-56, 2005.
4. C. A. El-Nouty. The presmoothed Nelson-Aalen estimator in the competing risk model. *Communications in Statistics-Theory and Methods*, vol. 33, no. 1, pp. 135-151, 2005.
5. D. A. Luo. Bias and mean-square error for the Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics*, vol. 3, no. 1, pp. 37-51, 1993.
6. R. A. Jiang. A bias-corrected Nelson-Aalen estimator. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 1043. Changsha: IOP Publishing, p. 022013, 2021.
7. A. A. Winnett. Adjusted Nelson-Aalen estimates with retrospective matching. *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 245-256, 2002.
8. A. S. Rossa. The Nelson-Aalen and Kaplan-Meier estimators under a sequential sampling scheme. *Communications in Statistics-Theory and Methods*, vol. 38, no. 16-17, pp. 3077-3098, 2009.
9. H. A. Zhang. On Nelson-Aalen type estimation in the partial Koziol-Green model. *Statistics*, vol. 44, no. 5, pp. 455-465, 2010.