Tweaking Exponential Distribution to Estimate the Chance for More Survival Time if a Cancerous Kidney is Removed

Ramalingam Shanmugam

School of Health Administration, Texas State University-San Marcos, TX 78666, USA

ABSTRACT

Patients with a malignant tumor in the kidney (recognized as hypernephroma in medical practice) are treated by what is known as nephrectomy (that is a surgical removal of the kidney) in addition to a combination of chemotherapy and immunotherapy for the sake of more survival time. This articles develops and demonstrates a statistical methodology to assess the statistical significance of an estimated chance for more survival time by tweaking the exponential distribution in a novel manner. This methodology is useful for the surgeons as much as to the patients who plan to undergo kidney removal.

Keywords: Likelihood Ratio Test, p-Value, Statistical Power, Clinical Trial

1. INTERDICTION

Clinical trials are performed periodically in U.S.A. and elsewhere to identify the best medical treatment for the patients who have malignant cancerous kidney. A complete cure is the ultimate aim. At least, more survival time of such patients since the diagnosis could be considered a medical success. With these aims, a clinical trial was performed with a random sample of patients in the University of Oklahoma Health Sciences Center (UOHSC) in Lee and Wang (2003) for the description and data of the clinical trial. These patients had malignant tumor in the kidney. This phenomenon is called hypernephroma in the medical community. The hypernephroma patients are treated with a combination of chemotherapy and immunotherapy in addition to what is called nephrectomy in the medical discipline. The nephrectomy amounts to a removal of the malignant kidney with an intention to provide more survival time for the patients. Not all patients undergo nephrectomy because there is a chance it might not work favorably. Only when the chance for nephrectomy to work is high enough, the surgeon is probably going to perform the procedure. A best assessment of such a chance requires an appropriate model for the data analysis. A literature

search reveals that the data are analyzed routinely by the Exponential Distribution (ED) for the treatment and the control group. What is ED?

Let Y≥0 be a continuous random variable representing the time of a hypernephroma patient since her/his diagnosis of cancerous kidney. The probability pattern of the random variable, Y is governed by the ED Equation 1:

$$f(y|\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}; y \ge 0; \theta > 0 \tag{1}$$

where, the parameter is the expected time of Y. That is $E(Y|\theta) = \int_0^\infty yf(y|\theta)dy = \theta$. Notice in **Table 1** below that the average time the patients without and with the nephrectomy lived are different. Does it warn us to doubt the appropriateness of using the ED in (1) for both control and treatment groups. Shouldn't the model for the patients with nephrectomy identify explicitly its impact? This impact information is washed out in the process of imposing ED in (1) for the patients with nephrectomy. We need to think "outside the box". to be realistic.

Table 1. Survival time of patients who had nephrectomy

Group	Y in months
Group 1 with	104, 9, 56, 35, 52, 68, 77, 84,
$n_1 = 16$ below 60 years old	8, 38, 72, 36, 48, 26, 108, 5
Group 2 with $n_2 = 9$ above	108, 26, 14,
60 but below 70 years old	115, 52, 5, 18, 36, 9

What is model? The model is an abstraction of a reality. If so, the model we select for the group should accommodate the effect of nephrectomy for the treatment group. It is worth remembering a proverb in this context and it is: "How can one size fit all?" There has to be a conceptual fallacy in using the ED for both the groups which we will become clear soon in a later segment of the article. In otherwords, the ED lacks the capability to address the chance for the nephrectomy to increase the survival time. A tweaking of the ED is, therefore, necessary for the sake of addressing the chance for the nephrectomy to extend the survival time of the treated patients. This tweaking is done and explained in this article with the UOHSC data. The full version of the methodology based on the tweaked exponential distribution is developed and illustrated which will be useful not only to the surgeons but also to the patients who are deciding to go through the nephrectomy procedure.

1.1. Tweaked Exponential Distribution

Let $0 \geq \phi < 1$ be an unknown probability that the nephrectomy might increase the expected time of Y. With $0 \geq \phi < 1$ denoting the chance for helpful impact of nephrectomy, the new expected time due to nephrectomy ought to be $\frac{\theta}{(1-\phi)}$. Hence, the Probability Density Function (PDF) for the patients with nephrectomy could be Equation 2:

$$f(y|\theta, \phi) = \frac{(1-\phi)}{\theta} e^{-\frac{(1-\phi)y}{\theta}}; y \geq 0; \theta > 0; 0 < \phi < 1 \quad (2)$$

The PDF in (2) is named Tweaked Exponential Distribution (TED). The ED in (1) is nested within the TED in (2) when $\phi = 0$. In other words, the event $\phi = 0$ is synonymous to refer the control group patients without nephrectomy. The TED is versatile enough to indicate an increased expected time. That is, the expected value of the TED in (2) is Equation 3:

$$E(y|\theta, \phi) = \int_0^\infty y \frac{(1-\phi)}{\theta} e^{-\frac{(1-\phi)y}{\theta}} dy = E(y|\theta, \phi = 0) + \Delta_m \quad (3)$$

where, Equation 4:

$$\Delta_m = \left(\frac{\phi}{1-\phi}\right)\theta \quad (4)$$

In (3) denotes the increment in the average survival time due to nephrectomy. When $\phi = 0$, the increment in (4) is zero. Notice that the co-factor $\frac{\theta}{(1-\phi)}$ in the increment (4) is called the odds of the impact of nephrectomy. The odds is an important concept in the medical and epidemiology practices. In the medical context of treating hypernephroma patients, when the odds is less than one, the increment is a fraction of the average survival time. When the odds is greater than one, the increment is amplified. Such knowledge becomes unavailable when the ED in (1) is blindly and incorrectly to applied to the data of patients with nephrectomy.

Furthermore, the TED in (2) offers additional advantages via its variance. The variance is a measure of volatility. The variance of the TED in (2) is Equation 5 and 6:

$$\text{var}(y|\theta, \phi) = \text{var}(y|\theta, \phi = 0) + \Delta_v \quad (5)$$

Where:

$$\Delta_v = \left(1 + \frac{1}{1-\phi}\right)\Delta_m \quad (6)$$

In (5) denotes the increment in the volatility of the survival time due to nephrectomy. It is important to notice in (6) that the volatility increment is proportional to the increment of survival time due to nephrectomy. In other words, the nephrectomy increases the survival time but with more volatility.

Of additional interest to the surgeon as much as to the hypernephroma patients is an answer to a question: How likely a patient with nephrectomy to live a specified m more months? This question is answerable once we configure the survival functions of TED in (2) and ED in (1). The survival function of the TED in (2) is Equation 7:

$$\Pr\left(Y \geq m|\theta, \phi\right) = e^{\frac{\phi m}{\theta}} \Pr(Y \geq m|\theta, \phi = 0) \quad (7)$$

where, $\Pr(Y \geq m|\theta, \phi = 0)$ denotes the survival function of the ED in (1). Note that the co-factor $e^{\frac{\phi m}{\theta}}$ is greater than

one. It means that the patients with nephrectomy have more chance to survive the m months than the patients without it in the control group. The TED is also holding the memory less property. That is:

$$\Pr(Y \geq m + q \mid \theta, \phi, Y \geq m) = \Pr(Y \geq q \mid \theta, \phi)$$

Given a specified percent, 1-p of the hypernephroma patients to survive a marked month, the mark is called the pth quantile. The pth quantile of the TED in (2) is $Q_{P|\theta,\phi} = Q_{P|\theta,\phi=0} + \Delta_Q$ where Equation 8:

$$\Delta_Q = -\phi\theta\ln(1-p) \tag{8}$$

Is positive and denotes the increment in the quantile due to nephrectomy. The quantile in (8) is influenced by both parameters.

Now, the estimation of the parameters of the TED in (2) need to be worked out. The Maximum Likelihood Estimator (MLE) is preferable over others because the MLE has invariance property (Stuart and Ord, 2009). That is the MLE of a function of the parameters is the function of MLE of the parameters. Consider a random sample $y_1, y_2,..,y_n$ from an IED in (2). Let $\bar{y}$ and s_y^2 are the sample average and variance respectively. Then by maximizing the log-likelihood Equation 9:

$$\ln L = n\left[\ln(1-\phi) - \ln\theta - \frac{(1-\phi)}{\theta}\bar{y}\right] \tag{9}$$

Setting its derivatives with respect to the parameters ϕ and θ to zero and simplifying, the MLEs Equation 10:

$$\hat{\phi}_{mle} = \frac{\left|\bar{y}^2 - s_y^2\right|}{1 + \left|\bar{y}^2 - s_y^2\right|} \tag{10}$$

And Equation 11:

$$\hat{\theta}_{mle} = \left(\frac{1}{1 + \left|\bar{y}^2 - s_y^2\right|}\right)\bar{y} \tag{11}$$

After algebraic simplifications. It is easily recognized that when $s_y^2 = \bar{y}^2$, not only $\hat{\phi}_{mle} = 0$ but also $\hat{\theta}_{mle} = \bar{y}$ which is the MLE of the parameter of ED in (1). Recall that the ED in (1) possesses a property that $\text{Var}(Y|\theta) = [E(Y|\theta)]^2$ which is validated by their sample counter parts.

Now, one would wonder whether an estimated MLE in (10) in a given data is statistically significant. This can be answered with a p-value for $\hat{\phi}_{mle}$. The MLE is asymptotically unbiased and more efficient. No significant impact of nephrectomy to have more survival time is implied by the null statement H_0: $\phi = 0$. There is a significant impact of nephrectomy so far as having more survival time by the alternative statement H_1: $\phi > 0$. An assessment of the statistical significance of the MLE in (10) amounts to performing a testing of the hypothesis H_0: $\phi = 0$ against the alternative H_1: $\phi > 0$. For this purpose, we first obtain the information matrix in (12), using (9) Equation 12:

$$I = \begin{bmatrix} -E(\partial_{\phi\phi}^2 \ln L) & -E(\partial_{\phi\theta}^2 \ln L) \\ -E(\partial_{\theta\phi}^2 \ln L) & -E(\partial_{\theta\theta}^2 \ln L) \end{bmatrix}$$
$$= n\begin{bmatrix} (1-\phi)^{-2} & (\theta[1-\phi])^{-1} \\ (\theta[1-\phi])^{-1} & \theta^{-2} \end{bmatrix} \tag{12}$$

Which is a singular matrix. The variance-covariance matrix of the MLE is the inverse of the information matrix in (12). Because of the singularity of the information matrix, only a generalized inverse matrix is possible. The generalized inverse may not be unique but the conclusion based on a generalized matrix is robust and acceptable (Bapat *et al.*, 2013). Based on a generalized inverse of (12), the variance of the MLE in (10) is $\hat{var}(\hat{\phi}_{mle}) = \frac{1}{n\hat{\theta}^2}$. Hence, the p-value for the MLE in (10) is Equation 13:

$$p - value = \Pr\left(Z > \frac{\sqrt{n}\left|\bar{y}^2 - s_y^2\right|\bar{y}}{(1 + \left|\bar{y}^2 - s_y^2\right|)^2}\right) \tag{13}$$

In an event a specific value ϕ^* is the true value for the impact chance ϕ, the probability of accepting the specific value is called statistical power. With a significance level α, the statistical power is therefore Equation 14:

$$power = \Pr\left(Z < z_\alpha - \phi^* \frac{\sqrt{n}\bar{y}}{(1 + \left|\bar{y}^2 - s_y^2\right|)^2}\right) \tag{14}$$

All these results are illustrated next with a clinical trial's data.

1.2. Illustration with UOHSC's Clinical Data

To illustrate, the data on the the survival times of $n_1 = 16$ and $n_2 = 9$ patients who underwent nephrectomy in

the age bracket less than 60 years old and 60 through 70 years old respectively in a clinical trial of the University of Oklahoma Health Sciences Center are considered. The data are displayed in **Table 1** and the results in **Table 2**.

The younger group (that is with the age below 60 years) survived longer than the older group (that is with the age above 60 but below 70 years). The volatility (that is the variance) is more in the older group. The estimated impact parameter $\hat{\phi}_{mle}$ is about the same in both groups. But, the estimated incidence parameter $\hat{\theta}_{mle}$ is more in the older group than in the younger group. The p-value of $\hat{\phi}_{mle}$ is significant in the older group but not in the younger group. This is to be interpreted that the chance impact of the nephrectomy to have more survival time is significant in the older group but not in the younger group. If the true value of the chance impact parameter is $\phi^* = 0.75$, the statistical power of accepting the true value, according to (14) with a chosen $\alpha = 0.05$, is better

in both the groups though it is slightly more in the younger group than in the older group.

Using (4), the expected increment Δ_m in survival time is found to be 51.59 and 41.99 months respectively in the younger and older groups. Using (6), the expected increment Δ_v in the volatility is found to be 83999 and 3210.15 respectively in the younger and older groups suggesting that the volatility increment is more in the younger than in the older group.

Using (7), (10) and (11), the estimated chance to survive a specified months for the patients with and without the nephrectomy are computed and sketched for the younger group in **Fig. 1** and for the older group in **Fig. 2**. According to the configurations in **Fig. 1 and 2**, the chance for the patients to survive is consistently more with the nephrectomy than without the nephrectomy in both groups. The importance of nephrectomy is supported by the collected data. The surgeons and the patients would be more pleased with these data support for the nephrectomy to survive longer.
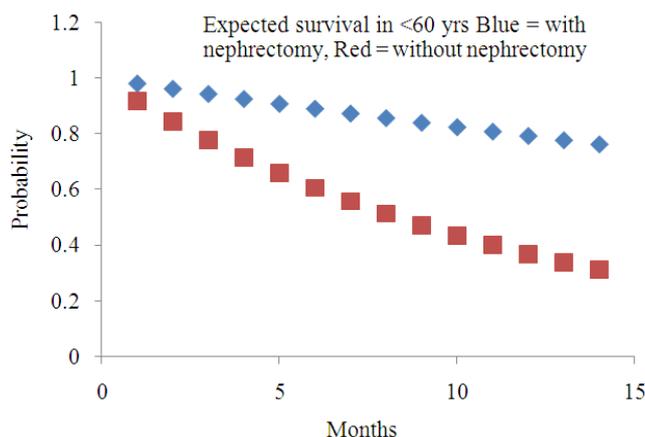
Fig. 1. Expected survival in younger group
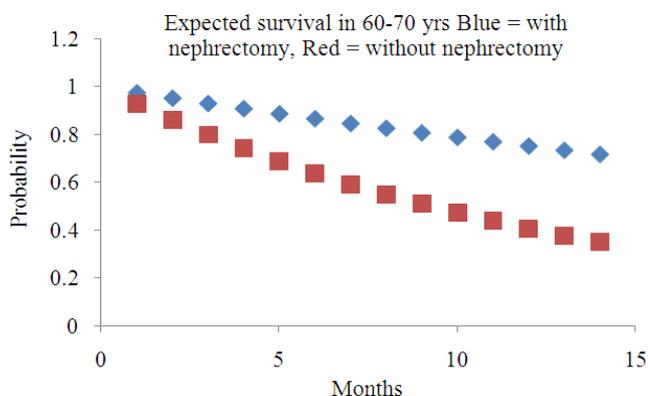
Fig. 2. Expected survival in older group

Table 2. Summary for the two groups with nephrectomy

Group	$\bar{y}$	s_y^2	$\hat{\phi}_{mle}$	$\hat{\theta}_{mle}$	p-value	Power for $\phi^* = 0.75$
1	51.6	1039.1	0.99	0.03	0.44	0.93
2	42.5	1736.5	0.98	0.56	0.04	0.65

2. CONCLUSION

In conclusion, we point out that it makes sense to use the exponential distribution for the survival time of the patients in the placebo group. Intuitively thinking outside the blindful thinking that the treatment must have an impact on the survival time and hence the impact should be captured exclusively. For this purpose, a tweking of the exponential distribution is necessary and it helps as done and demonstrated in this article. The chance oriented impact of the treatment is worth to be captured and interpreted as done in our illustration with the data of patients who underwent nephrectomy. The tweaked exponential distribution and the statistical methodology based on it of this article is versatile enough for application in other medical, health, engineering, business, economics, finance or social studies.

3. REFERENCES

Bapat, R.B., S.J. Kirkland, K.M. Prasad and S. Puntanen, 2013. Combinatorial Matrix Theory and Generalized Inverses of Matrices. 1st Edn., Springer London, Limited, New Delhi, ISBN-10: 8132210522, pp: 277.

Lee, E.T. and J.W. Wang, 2003. Statistical Methods for Survival Data Analysis. 3rd Edn., John Wiley and Sons, Hoboken, ISBN-10: 0471458554, pp: 534.

Stuart, A. and K. Ord, 2009. Kendall's Advanced Theory of Statistics. 6th Edn., Edward Arnold, London, ISBN-10: 047023380X, pp: 676.