An exponential tilt mixture model for time-to-event data to evaluate treatment effect heterogeneity in randomized clinical trials

Abstract

Evaluating the effect of a treatment on a time to event outcome is the focus of many randomized clinical trials. It is often observed that the treatment effect is heterogeneous, where only a sub group of the patients may respond to the treatment due to some unknown mechanism such as genetic polymorphism. In this paper, we propose a semi parametric exponential tilt mixture model to estimate the proportion of patients who respond to the treatment and to assess the treatment effect. Our model is a natural extension of parametric mixture models to a semi parametric setting with a time to event outcome. We propose a nonparametric maximum likelihood estimation approach for inference and establish related asymptotic properties. Our method is illustrated by a randomized clinical trial on biodegradable polymer delivered chemotherapy form malignant gliomas patients.

Keywords: exponential tilt model, mixture model, time to event data, treatment heterogeneity, randomized clinical trial

Introduction

Survival time is a primary end point for assessing the treatment effect in many randomized clinical trials. It is often observed that the treatment may only be effective to a sub group of the population due to some unknown mechanism such as genetic polymorphism. For patients in that sub group, which we call responders; their survival time is associated with the treatment assignment. For patients not in that subgroup, which we call non responders, their survival time follows the same distribution regardless of the treatment assignment.

One such example is from a randomized trial conducted by Brem et al., Two hundred and twenty two recurrent brain malignant gliomas patients were randomized to receive either biodegradable polymer delivered bis-chloroethyl nitro-sourea (BCNU), a chemotherapy, or place bowafers at the time of primary surgical resection. The histograms of survival time comparing the BCNU treated group and the place bogroup are shown in (Figure1). The histogram for the BCNU group appears to be a mixture of two unimode distributions, where one of the two modes is at roughly the same location as the mode for the place bogroup. Therefore, the BCNU treatment effect on patients’ survival appears to be heterogeneous, the distribution of survival time in the treatment group is a mixture of two unimode distributions, one for non-responders and the other for responders.

Clinically, three questions are of great interest. First, what is the treatment effect for responders? Third, how to test the existence of the proportion of responders in the population? Second, how to estimate the treatment effect for responders at the time of primary surgical resection. The histograms for the place bogroup are shown in (Figure1). The histogram for the place bogroup is a log-linear function. Built up on the ETM, the ETMM can be viewed as a natural generalization of parametric mixture models. The ETM accommodates a broad class of parametric models except for the Weibull mixture model.

For completely observed data, another interesting semi parametric model, referred to as the Exponential Tilt Mixture Model (ETMM), has been proposed by Qin and Zou et al., This model is based on the Exponential Tilt Model (ETM), which assumes a log-linear function for the density ratio of two mixture components, leaving the base line density unspecified. The ETM accommodates a broad class of parametric models because the density ratio functions from many parametric models, such as normal model, gamma model, and log normal model, have very simple forms that can be easily modelled by a log-linear function. Built up on the ETM, the ETMM can be viewed as a natural generalization of parametric mixture models. The ETM without a mixture structure has been extended to handle survival data with censored observations. See McLachlan & McGiffin for a review of the development in this field. More recently, semi parametric models have become an interesting alternative because of the more flexible model frame work and weaker model assumptions.

The proportional hazards model, which is the most frequently used semi parametric model for survival data, has been extended to handle mixture data. However, one concern of using the proportional hazards mixture model is that it does not have clear connections to parametric mixture models except for the Weibull mixture model.

For completely observed data, another interesting semi parametric model, referred to as the Exponential Tilt Mixture Model (ETM), has been proposed by Qin and Zou et al., This model is based on the Exponential Tilt Model (ETM), which assumes a log-linear function for the density ratio of two mixture components, leaving the base line density unspecified. The ETM accommodates a broad class of parametric models because the density ratio functions from many parametric models, such as normal model, gamma model, and log normal model, have very simple forms that can be easily modelled by a log-linear function. Built up on the ETM, the ETMM can be viewed as a natural generalization of parametric mixture models. The ETM without a mixture structure has been extended to handle censored, time-to-event data Shen et al., Wang et al., However, to our knowledge, the extension of ETMM to time-to-event outcome has not been studied yet.

One important issue for mixture models is testing for the existence of treatment effect. This hypothesis testing problem is not regular because the parameter space is degenerated under the null hypothesis. The asymptotic distribution of the regular likelihood ratio test statistic is very complicated Hartigan, Chen & Chen. Alternatively, Chen proposed a penalized likelihood ratio test, which solved the degenerate problem by adding a penalizing term to the likelihood function. The asymptotic distribution of the test statistic turned out to be much
simpler. However, there is no clear standard how to choose an appropriate penalizing function. Lemdani & Pons\textsuperscript{15} and Liang & Rathouez\textsuperscript{16} proposed another testing method based on the following observation although jointly testing for two parameters, the mixture proportion and the distributional parameter of the mixture component, is not regular, the test for one parameter given the other is regular. By utilizing this nice property, they proposed testing statistics with very simple asymptotic distributions for binomial mixture models.

In this paper, we extend the ETMM to censored, time-to-event data. We propose a non parametric maximum likelihood estimation (NPMLE) approach to estimate model parameters and study the associated asymptotic properties. We also develop a statistical testing procedure to assess the existence of treatment effect based on the ETMM. The application of our methods is illustrated by using the BCNU data.

Materials and methods

Exponential tilt mixture model (ETMM)

Let $F_0(t)$ be the distribution of survival time in the control group. As for the treatment group, to characterize the hetero generous treatment effect for responders and non responders, we assume the survival time follows a mixture distribution, $\lambda F_0(t) + (1-\lambda) F_t(t)$ where $1-\lambda$ is the proportion of responders and $F_t(t)$ is the distribution of survival time for responders after receiving the treatment. We consider an ETM for the treatment effect on responders, $d F_t(t) \propto d F_0(t) \exp \{h(t, \beta)\}$, where $h(t, \beta)$ is a pre specified parametric function with a vector of parameters $\beta$. Our model is semi parametric because $F_0(t)$ is completely unspecified. The ETMM can be regarded as a semi parametric generalization of parametric mixture models. For example, it reduces to the normal mixture model with two mixture components if $F_0$ is a normal distribution and $h(t, \beta) = \beta t + \beta^2 t^2$.

Parameter estimation

We assume the survival time and the censoring time are independent given treatment assignment and further assume the censoring time to be discrete with a finite number of values $0, 1, \ldots, C$, for the control group and $0, 1, \ldots, C$, for the treatment group. Let $C = (C_0, C_1, \ldots, C) = 1, 0, 1, 0, 1, 0, \ldots, 1, 0, 1$ be consistent estimator of $\theta$. Suppose the observed data consist of $n = n_0 + n_1$ uncensored, independent observations $x_1, \ldots, x_n$ where the first $n_0$ observations come from the place bogroup with density $d F_0(x)$ and the next $n_1$ come from the treatment group with density $d F_t(x) + (1-\lambda) d F_0(x)$ . The data also contain $m_0$ censored, independent observations from the place bogroup and $m_1$ independent observations from the treatment group, with $m_0$, $m_1$ observations censored at time $d F_0(x)$ and $d F_t(x)$. Let $N_0 = n_0 + m_0$ and $N_1 = n_1 + m_1$ be the numbers of observations in the two groups and $N = n_0 + n_1$ be the total number of observations. Furthermore, $P_0 = N_0 / N$, $P_1 = N_1 / N$. In this sub section, we assume there is a treatment effect so that $\lambda > 1$. We will discuss the test for the existence of treatment effect in Section 2.3. Our parameters of interest include the mixture proportion $\lambda$ and the distributional parameter $\beta$ characterizing the difference between the two mixture components.

Now consider discrete distributions with point masses only at the observed failure times and let $P_i = F_0\{x_i\}$. Then on parametric log likelihood function\textsuperscript{15,16} is,

$$l(\theta, F_0) = \log \sum_{j=1}^{p_1} \log \left( \lambda + (1-\lambda) \frac{\exp\{h(x_j, \beta)\}}{\sum_{i=1}^{P_0} \exp\{h(x_i, \beta)\}} \right) + \sum_{j=1}^{p_0} \log \sum_{i=1}^{n_0} \frac{\exp\{h(x_i, \beta)\}}{\sum_{i=1}^{P_0} \exp\{h(x_i, \beta)\}}$$

with

$$d \sum_{j=1}^{p_1} \sum_{i=1}^{n_0} \log \frac{\exp\{h(x_i, \beta)\}}{\sum_{i=1}^{P_0} \exp\{h(x_i, \beta)\}} + \sum_{j=1}^{p_0} \sum_{i=1}^{n_0} \log \frac{\exp\{h(x_i, \beta)\}}{\sum_{i=1}^{P_0} \exp\{h(x_i, \beta)\}}$$

(1)

With a constraint $\sum_{i=1}^{n_0} p_i = 1$ let $\theta = (\lambda, \beta)^T$ and define the profile likelihood of $\theta$ as $P(\theta) = \max F_{0\theta}(\theta, F_0)$. The NPMLE of $\theta$ is obtained by maximizing $P(\theta)$, that is, $\hat{\theta} = \arg \max P(\theta)$.

One appealing feature of the NPMLE method route is that it does not have the problem of “unbounded likelihood”. Day\textsuperscript{19} pointed out that, for normal mixture models, the global MLE may not exist because the likelihood goes to infinity if one mixture component concentrates on a single sample point. This problem does not arise in our non parametric likelihood approach. By considering discrete distributions, the point mass at any sample point is almost one, which yields a bounded likelihood. In practice, $p(\theta)$ can be obtained via the EM algorithm\textsuperscript{20} see Supplementary Materials for the details. Then, $\hat{\theta}$ can be obtained via Newton-Raphson method by maximizing $p(\theta)$.

The asymptotic properties of $\theta$ are as follows,

Proposition 1. Assume that all the elements in $\partial h(x, \beta) / \partial \beta$, $\partial^2 h(x, \beta) / \partial \beta^2$ and $(\partial h(x, \beta) / \partial \beta) (\partial \lambda (x, \beta) / \partial \beta)^T$ are continuous and bounded by some function $k(x)$ for $x \in (0, \infty)$ in a neighbourhood of the true value of $\beta$ satisfying $\int_0^x k(\lambda) dF_0 < \infty$ and $\int_0^x k(\lambda) dF_1 < \infty$. We have $\hat{\theta}$ is a consistent estimator of $\theta$.

$$\sqrt{N}(\hat{\theta} - \theta) \rightarrow N(0, U^{-1})$$

Where $\theta$ the true parameter value and $U = -\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{n_0} \frac{\partial^2 \Phi(\theta)}{\partial \theta^2}$ is the Fisher information. Proof. See Supplementary Materials.

To estimate survival probabilities for responders and non responders at a give time point $t_0$, it is natural to consider the following estimators:

$$\hat{h}(t_0) = \frac{\sum_{i=1}^{n_0} p_i I\{x_i > t_0\}}{\sum_{i=1}^{n_0} p_i \exp\{h(x_i, \beta)\} I\{x_i > t_0\}}$$

An exponential tilt mixture model for time-to-event data to evaluate treatment effect heterogeneity in randomized clinical trials

Proposition 2 shows that the survival probability estimators $\left\{ \hat{S}_i(t) \right\}$ for non responders and $\left\{ \hat{S}_i(t) \right\}$ for responders in the treatment group are a symptotically linear, which can be used to calculate standard errors of the estimators and to construct point wise confidence intervals for survival functions.

**Proposition 2.** Under the same conditions as in Proposition 1,

$$\sqrt{N} \left( \frac{\hat{S}(t)}{S(t)} - 1 \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi(t_0, x_i, \delta_i, z_i, \theta, n) + o_p(1)$$

Where the expression of $\phi(t_0, x_i, \delta_i, z_i, \theta, n)$ is defined in equation (S_i) in Supplementary Materials.

Proof. See Supplementary Materials.

**Testing the existence of treatment effect**

There are two equivalent ways to express the null hypothesis for testing the existence of treatment effect. One way, expressed as $\lambda=1$, is to test whether the data come from a mixture of two distributions or from a single distribution. The other way, expressed as $\beta=0$, is to test whether the two mixture components are the same. The diversity in expressing the null hypothesis indicates their regularity of this testing problem. Under the null hypothesis, the joint parameter space for ($\lambda$, $\beta$) is degenerated. It includes two lines $\lambda=1$ and $\beta=0$ and $\lambda$ is arbitrary. The asymptotic distribution for the regular likelihood ratio statistics very complicated.

In this sub section, we propose a simple solution to this problem based on the observation that under the null hypothesis, if we fix one parameter, testing for the other parameter is regular. The idea is an along us to Lemdani & Pons; Liang & Rathouz used for parametric mixture models. Two tests can be constructed

a. one is the likelihood ratio test for $\lambda$ with affixed $\beta$.

b. The other is the likelihood ratio test for $\beta$ with a fixed $\lambda$.

These two tests correspond to the two different yet equivalent expressions of the null hypothesis, $\beta=0$ and $\lambda=1$. Since $\lambda$ is simply a scalar and easier to be set to affixed value, for convenient, we will focus on testing $\tilde{\beta}=0$ for a fixed $\lambda$.

Proposition 3. Under the same conditions as in Proposition 1, for any given $0 \leq \lambda < 1$,

$$LRT_{\lambda}(\lambda) = 2 \left[ p_{F_1} \left( \lambda, \hat{\beta} \right) - p_{F_1}(0) \right] \rightarrow \chi^2_1,$$

Where $\hat{\beta}_\lambda$ is the NPMLE of $\beta$ for a given $\lambda$ and $p_{F_1}(0)$ is the profile likelihood under the null hypothesis. Proof. See Supplementary Materials.

In practice, it is also of interest to estimate the proportions of non responders and responders by constructing a 95% confidence interval for $\lambda$. However, the regular Wald confidence interval constructed from Proposition 2 may underestimate $\lambda$ because it always excludes one based on the assumption of $\lambda<1$ we made in the proposition. Alternatively, we consider the following procedure,

i. Perform the test in Proposition 3. If the p-value is smaller than 0.05, go to step 2; otherwise, go to step ii. Construct the regular Wald confidence interval, which is

$$\left( \exp \left\{ G_{0.025} / (1 + \exp \{ G_{0.025} \} \right\}, \exp \left\{ G_{0.975} / (1 + \exp \{ G_{0.975} \} \right\} \right).$$

Construct the confidence interval as

$$\left( \exp \left\{ G_{0.05} / (1 + \exp \{ G_{0.05} \} \right\}, 1 \right).$$

Where $G_q$ is the q quantile of the asymptotic distribution of

$$\log \left( \hat{\lambda} / (1 - \hat{\lambda}) \right).$$

When $\lambda < 1$, the confidence intervals in both Step 2 and Step 3 can cover $\lambda$ with probability 0.95. When $\lambda=1$, there is 95% of the chance to choose the interval in Step3, which always contains one. Thus, the fore mentioned procedure can always covert the true value of $\lambda$ with probability 0.95.

**Results**

**Simulations**

The first set of simulations evaluates the finite sample performance of the estimate or for the mixture proportion $\lambda$. We considered the first three simulation scenarios listed in (Table 1). Specifically, $\lambda$ was chosen as 0.25, 0.50 or 0.75, representing different proportions of responders in the treatment group. Observations in the placebo group were simulated from log-normal distribution $F_2(.) \sim LN \left( 3.2, 0.9^2 \right)$, and those in the treatment group were simulated from $\lambda F_2(.) + (1 - \lambda) F_1(\cdot)$, where $\lambda F_2(.) \sim LN \left( 3.7, 0.2^2 \right)$. The simulation setting is similar to the BCU data in the sense that the distribution of survival time for responders in the treatment group has larger mean and smaller variance than that for nonresponders. We assumed censoring to occur at 6 fixed time points, the 30%, 40%, 50%, 60%, 70% and 80% quantile of $F_2(.)$. The censoring probabilities were 14% for the placebo group and 16% to 20% for the treatment group, depending on the mixture proportion. Separate simulations were run for $N_t = N_c = 15$ or 500. Under each setting, 1,000 independent data sets were simulated. We considered the following two specifications of the $h(.)$ function in the ETM,

$$h(t; \beta) = \beta_1 \log(t) + \beta_2 \left( \log(t) \right)^2$$

The first $h(.)$ function is proportional to the density ratio for lognormal distributions. The corresponding ETM, which we call the lognormal ETM, reduces to alog-normal model when $F_1$ is alog normal distribution. The second $h(.)$ function is a more general expression that in corporate density ratios for log-normal, exponential, and gamma distributions. The corresponding ETM is called the general ETM. Simulation results are summarized in (Table 1&2).

For $\lambda=0.25$ or 0.50, estimates from both the log-normal ETM and the general ETM are virtually unbiased. Coverage probabilities are also close to the desired value 0.95. For $\lambda=0.75$ and when sample size is small ($N_t = N_c = 150$), biases are large and coverage probabilities are lower than 0.95. This is due to the limited number of responders in the data. The estimation becomes better when the sample size increases to 500.

The second set of simulations evaluates the performance of the following statistic $LRT_{\lambda}(\lambda)$ proposed in Section 2.3. We considered the

same three scenarios as in the first set of simulations. In addition, we considered a null scenario where both $F_0(\lambda)$ and $F_1(\lambda) = LN(3.2, 0.9^2)$.

The simulation scenarios are listed in (Table 1). We used the log-normal ETM as our working model and considered five choices of $\lambda$ in $LRT_\lambda(\lambda)$, 0.00, 0.25, 0.50, 0.75 and 0.90. Results are summarized in (Table 3).

### Table 1 Simulation scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$F_0$</th>
<th>$F_1$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>LN(3.2, 0.9)</td>
<td>LN(3.7, 0.2)</td>
<td>0.25</td>
</tr>
<tr>
<td>(ii)</td>
<td>LN(3.2, 0.9)</td>
<td>LN(3.7, 0.2)</td>
<td>0.50</td>
</tr>
<tr>
<td>(iii)</td>
<td>LN(3.2, 0.9)</td>
<td>LN(3.7, 0.2)</td>
<td>0.75</td>
</tr>
<tr>
<td>(iv)</td>
<td>LN(3.2, 0.9)</td>
<td>LN(3.7, 0.2)</td>
<td>0.90</td>
</tr>
</tbody>
</table>

### Table 2 Simulation results for estimating $\lambda$

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Log-normal ETM</th>
<th>General ETM scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_0=N_1$</td>
<td>$N_0=N_1$</td>
<td>$N_0=N_1$</td>
</tr>
<tr>
<td>Bias</td>
<td>SSE</td>
<td>SEE</td>
</tr>
<tr>
<td>(i)</td>
<td>150</td>
<td>-0.002</td>
</tr>
<tr>
<td>(ii)</td>
<td>500</td>
<td>-0.002</td>
</tr>
<tr>
<td>(iii)</td>
<td>150</td>
<td>-0.014</td>
</tr>
<tr>
<td>(iv)</td>
<td>500</td>
<td>-0.002</td>
</tr>
</tbody>
</table>

Bias is the mean difference between $\lambda$ and the true value, SSE is the standard error of $\lambda$, SEE is the square root of the mean of the variance estimate of $\lambda$, and CP is the coverage probability of the 95% confidence interval of $\lambda$.

### Table 3 Empirical Type I error/power for testing the existence of treatment effect.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$N_0=N_1$</th>
<th>$LRT_\lambda(0)$</th>
<th>$LRT_\lambda(0.25)$</th>
<th>$LRT_\lambda(0.5)$</th>
<th>$LRT_\lambda(0.75)$</th>
<th>$LRT_\lambda(0.9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>150</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(ii)</td>
<td>500</td>
<td>0.86</td>
<td>0.94</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>(iii)</td>
<td>150</td>
<td>0.28</td>
<td>0.32</td>
<td>0.52</td>
<td>0.77</td>
<td>0.70</td>
</tr>
<tr>
<td>(iv)</td>
<td>500</td>
<td>0.75</td>
<td>0.78</td>
<td>0.87</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The power appears to be higher for larger $\lambda$. For example, the empirical power is 1.00 for $\lambda = 0.90$ in simulation scenario with $N_0 = N_1 = 500$. But the power for $\lambda = 0.00$ is only 0.75. However, the test statistic with larger $\lambda$ is also less stable. The Type 1 error is larger than the expected value when $\lambda = 0.75$ or 0.90 and $N_0 = N_1 = 150$.

### Data analysis

Malignant gliomas are a common type of malignant primary brain tumors. Despite of the disproportionately high morbidity and mortality, drug treatment is hampered by the difficulty in crossing the blood-brain barrier and the severe complications from systemic exposure to drugs targeted for the brain. To overcome these problems, a method was developed to incorporate BCNU, a very effective chemotherapeutic drug, into biodegradable polymers. Implantation of the BCNU-incorporated polymers at the tumor site allows local sustained release of BCNU with minimal systemic exposure. In a clinical trial conducted by Brem et al., two hundred and twenty two patients with recurrent malignant brain tumors were randomized in to a BCNU-treatment group and control group. Patients’ survival time was recorded. In this section, we evaluate the effectiveness of BCNU treatment and estimate the proportion of patients that benefited from the treatment (Table 4).

We consider two ETMs, the log-normal ETM and the general ETM. The results are summarized in (Table 4). Since the likelihood
ratio test comparing the two ETMs is not significant (p-value=0.168), we choose the log-normal ETM as our working model. Based on this model, we test the hypothesis of no treatment effect (β=0) using the test statistic LRT λ(λ) described in Proposition 3. Fixing λ at 0.00, 0.25, 0.50, 0.75, or 0.90, we obtain the p-value as 0.38, 0.36, 0.29, 0.10, or 0.12. The test with smaller λ seems to be less powerful, which is consistent with our simulation findings in Section 3. Since none of the tests is significant at 5% level, we follow Step 3 of the procedure described in Section 2.3 to calculate a 95% confidence interval for λ as (0.595,1). The survival function estimates for the first year of follow-up comparing treatment responders vs. non responders are shown in (Figure 2). The survival fraction for treatment responders is significantly higher than that for non responders in the first thirty weeks. But after that period, the survival fractions for the two sub groups become close.

| Analysis results for the BCNU data. SE is the estimated error of $\lambda$. |
|---------------------------------|--------|-----------------|-----------------|
|                                | $\hat{\lambda}$ | SE   | 95% Confidence interval | Log-likelihood |
| Log-normal ETM                 | 0.804   | 0.099          | (0.595,1)       | -1151.355      |
| General ETM                    | 0.738   | 0.111          | (0.524,1)       | -1150.405      |

Figure 1 Histograms of the BCNU data.
An exponential tilt mixture model for time-to-event data to evaluate treatment effect heterogeneity in randomized clinical trials

An exponential tilt mixture model for time-to-event data to evaluate treatment effect heterogeneity in randomized clinical trials. Based on the choice of $\lambda$, $0 < e < 1$ has for a given $\lambda$. The test

Acknowledgements

None

Conflicts of interest

Author declares that there are no conflicts of interest.

Funding

None.

References

