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. [1]. Two hundred and twenty two recurrent brain malignant gliomas patients were randomized to receive either bio degradable 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 (Figure 1). The histogram for the BCNU group appears to be a mixture of two uni-mode 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 uni-mode distributions, one for non-responders and the other for responders. Clinically, three questions are of great interest. First, what is the proportion of responders in the population? Second, how to estimate the treatment effect for responders? Third, how to test the existence of treatment effect? (Figure 1).

One approach to analyze data with a mixture structure is by using finite (parametric) mixture models [2], where each mixture component was assumed to follow a parametric distribution and the EM algorithm was used to find maximum likelihood estimates of the distribution al parameter and the mixture

Abbreviations
ETM: Exponential Tilt Model; ETMM: Exponential Tilt Mixture Model; NPMLE: Nonparametric Maximum Likelihood Estimation; BCNU: Bis-chloroethylNitrosourea

Chi Wang1,2,*, Zhiqiang Tan3 and Thomas A. Louis4
1Department of Biostatistics, University of Kentucky, USA
2Biostatistics and Bioinformatics Shared Resource Facility, University of Kentucky, USA
3Department of Statistics, Rutgers, The State University of New Jersey, USA
4Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, USA

*Corresponding author: Chi Wang, Department of Biostatistics, College of Public Health and Biostatistics and Bioinformatics Shared Resource Facility, Markey Cancer Centre, University of Kentucky, Lexington, KY 40536, USA, Tel: 001-8593232045; Email: chi.wang@uky.edu

Received: September 17, 2014  |  Published: October 07, 2014

Keywords
Exponential tilt model; Mixture model; Time to event data; Treatment heterogeneity; Randomized clinical trial

Figure 1: Histograms of the BCNU data.
Materials and Methods

Exponential tilt mixture model (ETMM)

Let $F_0(.)$ be the distribution of survival time in the control group. As for the treatment group, to characterize the heterogeneous treatment effect for responders and non-responders, we assume the survival time follows a mixture distribution:

$$
\lambda F_0(t) + (1-\lambda) F_1(t)
$$

where $\lambda\in(0,1)$ is the proportion of responders and $F_0(.)$ is the distribution of surviving time for responders after receiving the treatment. We consider an ETM for the treatment effect on responders: $d F_1(t) = d F_0(t) \exp[h(t, \beta)]$, where $h(t, \beta)$ is a pre-specified parametric function with a vector of parameters $\beta$. Our model is semiparametric because $F_0(.)$ is completely unspecified. The ETM can be regarded as a semiparametric 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 $h(t, \beta) = \beta_1 t + \beta_2 t^2$.

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 $C_0, C_1, \ldots, C_{d_0}, C_{d_0+1}, \ldots, C_{d_1}$, for the control group and $C_1, \ldots, C_{d_1}, C_{d_1+1}, \ldots, C_{d_2}$, for the treatment group. Let $CZ = (CZ_1, \ldots, CZ_{d_2})^T$, $Z = 0.1$. Suppose the observed data consist of $n = (n_0 + m_0)$ uncensored, independent observations $x_{1r}, \ldots, x_{nr}$, where the first $n_0$ observations come from the placebo group with density $d F_0(.)$ and the next $n_1$ come from the treatment group with density $\lambda d F_0(.) + (1-\lambda) d F_1(.)$. The data also contain $m_0$ censored, independent observations from the placebo group and $m_1$ independent observations from the treatment group, with $m_0$ observations censored at time $C_1(j = 1, \ldots, d_1, z = 0, 1, m_1+1, \ldots, m_2 d_2 = m_2)$. Let $N(=n_0 + m_0)$ and $N(=n_1 + m_1)$ be the numbers of observations in the two groups and $N(=n_0 + n_1)$ be the total number of observations. Furthermore, $P_0 = N_0 + N_1$, $P_1 = N_1 + N$. 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 $P_1 = F_0(\{x_{1r}\})$.
Then on parametric log likelihood function \([17,18]\) is:

\[
\ell(\theta, F_\theta) = \sum_{i=1}^{n} \log P + \sum_{i=1}^{n} \log \left( \frac{\exp[h(x_i, \beta)]}{\sum_{j} \exp[h(x_j, \beta)]} \right)
\]

\[
d_n \theta = \hat{\theta} \exp \left( \frac{\beta h(x_i, \beta)}{\sum_{j} \exp[h(x_j, \beta)]} \right)
\]

\[
\log \sum_{i=1}^{n} \left( \frac{\exp[h(x_i, \beta)]}{\sum_{j} \exp[h(x_j, \beta)]} \right) l_i(x_i > c_s) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\exp[h(x_i, \beta)]}{\sum_{j} \exp[h(x_j, \beta)]} \right) l_i(x_i > c_s) \right)
\]

Proposition 2 shows that the survival probability estimators \(S_0(t)\) for non responders and \(S_1(t)\) for responders in the treatment group are a symptomatically linear, which can be used to calculate standard errors of the estimators and to construct joint wise confidence intervals for survival functions.

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

\[
\sqrt{N} \left( \sum_{i=1}^{N} \phi(t_i, x_i, \delta_i, z_i, \theta, n) - \frac{S_0(t)}{S(t)} \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi(t_i, x_i, \delta_i, z_i, \theta, n) + o_p(1)
\]

Where the expression of \(\phi(t_i, x_i, \delta_i, z_i, \theta, n)\) is defined in equation \((S_{5})\) 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 \(\beta\) is arbitrary, and \(\beta = 0\) and \(\lambda\) is arbitrary. The asymptotic distribution for the regular likelihood hypothesis, the joint parameter space for \((\lambda, \beta)\) is degenerated.

In this subsection, 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 analog to Lemdani and Pons [15]; Liang and Rathouz [16] used for parametric mixture models. Two tests can be constructed

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

b. the other is the likelihood ratio test for \(\lambda\) for with a fixed \(\beta\)

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 scaler and easy to be set to a fixed value, for conveniences, we will focus on testing $\beta=0$ for a fixed $\lambda$.

**Proposition 3:** Under the same conditions as in Proposition 1, for any given $0 \leq \lambda < 1$, $LRT_\lambda(\lambda) = 2 \left[ \frac{p_{\lambda}(\lambda, \hat{\beta}_j) - p_{\lambda}(0)}{p_{\lambda}(0)} \right] \to X^2_2$,

where $\hat{\beta}_j$ is the NPMLE of $\beta$ for given $\lambda$ and $p_{\lambda}(0)$ 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 3.

ii. Construct the regular Wald confidence interval, which is $\exp\{G_{0.025}\} / \{1 + \exp\{G_{0.025}\}\}$. The confidence interval as $\exp\{G_{0.025}\} / \{1 + \exp\{G_{0.025}\}\}$. Where $G_q$ is the $q$-quantile of the asymptotic distribution of $\left[ \frac{\hat{\beta}}{\left(1 - \hat{\lambda}\right)} \right]$. 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 Step 3, 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 $F_0(\cdot) \sim LN\left(3.2, 0.9^2\right)$, and those in the treatment group were simulated from $\lambda F_0(\cdot) + (1-\lambda) F_1(\cdot)$, where $\lambda F_0(\cdot) \sim LN\left(3.7, 0.2^2\right)$. The simulation setting is similar to the BCNU 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 non-responders. We assumed censoring to occur at 6 fixed time points, the 30%, 40%, 50%, 60%, 70% and 80% quantile of $F_0$. The censoring probabilities were 14% for the place group and 16% to 20% for the treatment group, depending on the mixture proportion. Separate simulations were run for $N_0 = N_1 = 15$ or 500. Under each setting, 1,000 independent data sets were simulated. We considered the following two specifications of the $h(\cdot)$ function in the ETM:

- 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 Step 3, which always contains one. Thus, the fore mentioned procedure can always covert the true value of $\lambda$ with probability 0.95.

**Table 1: Simulation scenarios.**

| Scenario | $F_0$ | $F_1$ | $\lambda$ |
|----------|-------|-------|-----------|
| (i)      | LN(3.2, 0.9$^2$) | LN(3.7, 0.2$^2$) | 0.25 |
| (ii)     | LN(3.2, 0.9$^2$) | LN(3.7, 0.2$^2$) | 0.50 |
| (iii)    | LN(3.2, 0.9$^2$) | LN(3.7, 0.2$^2$) | 0.75 |
| (iv)     | LN(3.2, 0.9$^2$) | LN(3.7, 0.2$^2$) | —    |

The second set of simulations evaluates the performance of the testing statistic $LRT_\lambda(\lambda)$ proposed in Section 2.3. We considered the above three scenarios as in the first set of simulations. In addition, we considered a null scenario where both $F_0(\cdot)$ and $F_1(\cdot) \sim LN\left(3.2, 0.9^2\right)$ had the same distribution. 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).

The power appears to be higher for larger $\lambda$. For example, in 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 ($N_0 = N_1 = 150$), 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.

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

Table 2: Simulation results for estimating $\hat{\lambda}$.

| Scenario | N0=N1 | Bias | SSE  | SEE  | CP   | Bias | SSE  | SEE  | CP   |
|----------|-------|------|------|------|------|------|------|------|------|
| (i)      | 150   | -0.002 | 0.061 | 0.056 | 0.94 | -0.012 | 0.063 | 0.058 | 0.95 |
| (ii)     | 500   | -0.002 | 0.031 | 0.031 | 0.95 | -0.005 | 0.031 | 0.031 | 0.96 |
| (iii)    | 150   | -0.014 | 0.093 | 0.078 | 0.94 | -0.035 | 0.112 | 0.083 | 0.94 |
| (iv)     | 500   | -0.002 | 0.041 | 0.042 | 0.95 | -0.006 | 0.043 | 0.043 | 0.95 |

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

| Scenario | $N_0=N_1$ | $LRT_{\lambda}(0)$ | $LRT_{\lambda}(0.25)$ | $LRT_{\lambda}(0.5)$ | $LRT_{\lambda}(0.75)$ | $LRT_{\lambda}(0.9)$ |
|----------|-----------|---------------------|----------------------|---------------------|---------------------|---------------------|
| (i)      | 150       | 1.00                | 1.00                 | 1.00                | 1.00                | 1.00                |
| (ii)     | 500       | 0.86                | 0.94                 | 0.99                | 1.00                | 1.00                |
| (iii)    | 150       | 0.28                | 0.32                 | 0.52                | 0.77                | 0.70                |
| (iv)     | 500       | 0.04                | 0.05                 | 0.05                | 0.09                | 0.15                |
| (v)      | 500       | 0.05                | 0.05                 | 0.05                | 0.06                | 0.11                |

Table 4: Analysis results for the BCNU data.

| ETM                  | $\hat{\lambda}$ | 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      |

SE is the estimated error of $\hat{\lambda}$.

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 ($\beta=0$) using the test statistic $LRT_{\lambda}(\lambda)$ described in Proposition 3. Fixing $\lambda$ 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 $\lambda$ 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 $\lambda$ 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.

Conclusion

In this paper, we investigate the treatment effect heterogeneity problem in randomized clinical trials, where the survival time for only a subgroup of the population is affected by the treatment. To deal with this problem, we extend the ETMM to time-to-event outcome. We also develop a testing procedure for the existence of treatment effect. To characterize the magnitude of heterogeneity, we proposed a unified confidence interval for the mixture proportion.
One limitation of our method is that it assumes a sufficiently long follow-up period. In case of a short follow-up period with patients censored by the end of the study, a new parameter to characterize the survival fraction by the end of the study can be introduced. Statistical inference procedure and theoretical proofs will need to be modified correspondingly. Similar approach has been taken in Wang et al. [11,21]. Another limitation of our method is that the censoring time is assumed to be discrete. Extending our method to continuous censoring time is challenging due to the technical complexity in studying asymptotic properties of model parameters. More advanced theoretical tools, such as the modern empirical process, may be required. But the derivations are primarily of mathematical interest because censoring time can be approximated by a sufficiently fine set of discrete points in practice.

The power of $LRT_\lambda(\lambda)$ depends on the choice of $\lambda$. Based on simulations, we recommend to use $\lambda = 0.50$, which balances the sensitivity and stability. More powerful tests may be derived by taking the supreme over $\lambda$. For binomial mixture model, Lemdani and Pons [15] showed that the supreme test $\sup_{\lambda \in [0,1]} LRT_\lambda(\lambda)$, $0<\epsilon<1$ has the same asymptotic distribution as $LRT_\lambda(\lambda)$ for a given $\lambda$. The asymptotic distribution of the supreme test under the ETMM is more complicated. But the empirical distribution may be obtained using a re-sampling method.

Acknowledgement
No acknowledgement for this article.

Conflict of Interest
There is no conflict of interest.

References
1. Brem H, Piantadosi S, Burger PC, Walker M, Selker R, et al. (1995) Placebo-controlled trial of safety and efficacy of intra operative controlled delivery by bio degradable polymers of chemotherapy for recurrent gliomas. The Polymer-brain Tumor Treatment Group. Lancet 345(8956): 1008-1012.
2. Pearson K (1894) Contributions to the mathematical theory of evolution. Phil Trans R Soc Lond A 185: 71-110.
3. Larson MG, Dinse GE (1985) A mixture model for the regression analysis of competing risks data. Appl Statist 34(3): 201-211.
4. McLachlan GJ (1996) Mixture models and applications. In: Balakrishnan N & Basu AP (Eds.), The Exponential Distribution: Theory, Methods and Applications. CRC Press, USA, pp. 664.
5. McLachlan GJ, McGiffin DC (1994) On the role of finite mixture models in survival analysis. Statistics in Medical Research 3(3): 211-226.
6. Rosen O, Tanner M (1999) Mixtures of proportional hazards regression models. Stat Med 18(9): 1119-1131.
7. Chen HY, Little RJA (1999) Proportional hazards regression with missing covariates. Journal of the American Statistical Association 94(447): 896-908.
8. Qin J (1999) Empirical likelihood ratio based confidence intervals for mixture proportions. The Annals of Statistics 27(4): 1368-1384.
9. Zou F, Fine JP, Yandell BS (2002) On empirical likelihood for a semi parametric mixture model. Biometrika 89(1): 61-75.
10. Shen Y, Qin J, Costantino JP (2007) Inference of tamoxifen’s effects on prevention of breast cancer from a randomized controlled trial. J Am Stat Assoc 102(480): 1235-1244.
11. Wang C, Tan Z, Louis TA (2011) Exponential tilt models for two-group comparison with censored data. Journal of Statistical Planning and Inference 141(5): 1102-1117.
12. Hartigan JA (1985) A failure of likelihood a symptotics for normal mixtures. In: Le Cam LM & Olshen RA (Eds.), Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Volume 2). Wadsworth Publishing Co Inc., pp. 950.
13. Chen H, Chen J (2001) Large sample distribution of the likelihood ratio test for normal mixtures. Statistics & Probability Letters 52(2): 125-133.
14. Chen J (1998) Penalized likelihood-ratio test for finite mixture models with multinomial observations. The Canadian Journal of Statistics 26(4): 583-599.
15. Lemdani M, Pons O (1995) Tests for genetic linkage and homogeneity. Biometrics 51(3): 1033-1041.
16. Liang KY, Rathouz PJ (1999) Hypothesis testing using mixture models: Application to genetic linkage analysis. Biometrics 55(1): 65-74.
17. Vardi Y (1985) Empirical distributions in selection bias models (Com: P204-205). Am Statist 13(1): 178-203.
18. Owen AB (2001) Empirical Likelihood. Chapman & Hall Ltd.
19. Day NE (1969) Estimating the components of a mixture of normal distributions. Biometrika 56(3): 463-474.
20. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from in completed data via the EM Algorithm. Journal of the Royal Statistical Society, Series B: Methodological 39(1): 1-38.
21. Wang C, Tan Z, Louis TA (2014) An exponential tilt model for quantitative trait loci mapping with time-to-event data. Journal of Bioinformatics Research Studies, (in press).