



Is it time for the weighted log-rank test to play a more important role in confirmatory trials?



The log-rank test is frequently used to detect a potential treatment effect in randomized clinical trials with time-to-event endpoints. It is asymptotically the most powerful test under the proportional hazards setting, but it has been shown to markedly lose power when the proportional hazards assumption is violated [1]. Weighted log-rank tests with various fixed and adaptive weight functions have been proposed in the literature to increase the power of a trial when non-proportional hazards are expected e.g. [2–4]. In particular, the $G^{\rho,\gamma}$ family of weights proposed by Fleming and Harrington [5] allows the flexibility to assign greater weights to either early or late failure times as controlled by the two parameters, and one option to assign greater weights to late failure times is to set $\rho = 0$ and $\gamma > 0$. The most powerful weighted log-rank test assigns the weights proportionally to the magnitude of the log hazard ratio [1].

There are some examples in the literature that demonstrate the potential value of weighted log-rank tests. The ovarian cancer screening trial UKCTOCS failed to meet its primary endpoint with the log-rank test. As the authors stated in the article: “The main limitation of this trial was our failure to anticipate the late effect of screening in our statistical design. Had we done so, the weighted log-rank test could have been planned in line with many other large cancer screening trials” [6]. Similarly, the Phase 3 trial of the epidermal growth factor vaccine CIMAvax-EGF as a switch maintenance therapy in advanced non-small cell lung cancer failed to meet its primary endpoint with the log-rank test. As a delayed separation of the survival curves was observed the Fleming-Harrington weighted log-rank test was performed in a post hoc analysis [7]. In cases where there's a strong scientific rationale behind a delayed treatment effect, a weighted log-rank test may be considered. For example, in a trial to assess the effect of calcium and Vitamin D supplementation on the risk of colorectal cancer, a weighted log-rank test was specified in the protocol with weight increasing linearly from 0 at randomization to a maximum of 1 at 10 years to enhance the statistical power of the trial [8].

Delayed separation of survival curves has been frequently observed in clinical trials with time-to-event endpoints but the adoption of weighted log-rank tests has been limited. In cases where there are both a strong biological rationale and some clinical evidence to support a delayed treatment effect, we would encourage the consideration of pre-specifying a weighted log-rank test for the primary analysis with the log-rank test being a sensitivity analysis. Clinical evidence may be generated from either trials of other treatments with a similar mechanism of action or completed trials with the same treatment. For example, the pivotal Phase 3 KEYNOTE-040 trial investigating pembrolizumab, an anti-PD1 therapy, in previously treated patients with recurrent or metastatic head and neck squamous cell carcinoma, did not meet its pre-specified primary endpoint of overall survival with a p-value of 0.06 [9]. Nivolumab, another anti-PD1 therapy, was investigated in a similar patient population earlier and the survival curves showed little separation during the first 3 months of treatment [10]. As an immuno-oncology therapy, it may be reasonably expected to have a delayed treatment effect. As the result of being supported by both the mechanism of action and competitor data, a weighted log-rank test could have been pre-specified as the primary analysis for the KEYNOTE-040 trial.

Another example where the clinical evidence generated from earlier trials may support the use of a weighted log-rank test is vonapanitase in patients with chronic kidney disease. The first Phase 3 PATENCY-1 trial did not meet the primary endpoint of improved primary unassisted patency compared to placebo, and the secondary patency endpoint demonstrated promising results with a hazard ratio of 0.66 and a p-value of 0.048 [11]. Based on the results of the PATENCY-1 trial the sample size of the ongoing PATENCY-2 trial has been increased to provide sufficient power for the secondary patency endpoint, which has been elevated to be one of the two co-primary endpoints [12]. As the Kaplan-Meier curves for secondary patency showed little separation for the first 3 months in the PATENCY-1 trial a weighted log-rank test may be considered for the PATENCY-2 trial if there is a sufficient biological rationale to support a delayed treatment effect. Table 1 shows the power of the trial with both the log-rank test and the Fleming-Harrington weighted log-rank test under the assumptions of exponential and piecewise exponential distributions, where $S_1(t)$ and $S_2(t)$ are the survival functions for the vonapanitase and placebo arms, respectively. Assuming that the respective 12-month event free rates are 0.74 and 0.61 for the two treatment arms, the log-rank test will have 90.4% power under the proportional hazards assumption, which is reduced to 81.8% if there's a delayed treatment effect with no treatment benefit for the first 3 months. The Fleming-Harrington test with $\rho = 0$ and $\gamma = 0.25$ has a minimal loss of power under the proportional hazards assumption and has a substantial gain in power when a delayed treatment effect is present. The value of the weighted log-rank test is particularly meaningful when the true treatment effect is smaller than that observed in the PATENCY-1 trial. For example, if the 12-month event free rates are 0.72 and 0.63, respectively, the log-rank test will have less than 50% power with a delayed treatment effect, and as a comparison the weighted log-rank test will maintain over 60% power irrespective of the proportional hazards assumption.

Various statistical methodologies have been developed for time-to-event endpoints with non-proportional hazards but their usage in confirmatory clinical trials has been limited. A new approach to complementing the weighted log-rank test by a Cox model with a time-varying treatment effect

<https://doi.org/10.1016/j.conctc.2017.09.007>

Available online 18 September 2017

2451-8654/ © 2017 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1Power (%) of the standard and weighted log-rank tests based on $n = 10,000$ simulations.

Distribution of Survival Time	Standard Log-rank	Fleming-Harrington (0, 0.25)	Fleming-Harrington (0, 0.5)	Fleming-Harrington (0, 1)
Exponential distribution				
$S_1(12) = 0.74, S_2(12) = 0.61$	90.4	89.2	87.0	80.8
$S_1(12) = 0.73, S_2(12) = 0.62$	78.9	77.4	73.7	66.6
$S_1(12) = 0.72, S_2(12) = 0.63$	61.8	60.2	57.0	50.2
Piecewise exponential distribution with a change point at 3 months and $S_1(3) = S_2(3) = 0.8$.				
$S_1(12) = 0.74, S_2(12) = 0.61$	81.8	90.6	94.3	97.3
$S_1(12) = 0.73, S_2(12) = 0.62$	66.8	78.9	85.0	94.7
$S_1(12) = 0.72, S_2(12) = 0.63$	48.2	60.1	68.1	83.2

was recently proposed in the Journal, which helps translate the weighted log-rank test to quantitative estimates that can facilitate the evaluation of a potential treatment effect in terms of its clinical meaningfulness [13]. We would like to encourage clinical trial practitioners to consider a weighted log-rank test when both the mechanism of action and existing clinical evidence point to a potential delayed treatment effect. With an appropriately chosen weight function the loss of power should be fairly minimal under the proportional hazards setting, and the gain in power can be substantial in the presence of non-proportional hazards. Even given the preference to follow the precedence of using the log-rank test in confirmatory trials, a weighted log-rank test may be pre-specified as an important sensitivity analysis to help better characterize the potential benefit of a new treatment.

References

- [1] D. Schoenfeld, The asymptotic properties of nonparametric tests for comparing survival distributions, *Biometrika* 68 (1981) 316–319.
- [2] D.M. Zucker, E. Lakatos, Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment, *Biometrika* 77 (4) (1990) 853–864.
- [3] Z. Xu, et al., Designing therapeutic cancer vaccine trials with delayed treatment effect *Stat, Med* 36 (4) (2017) 592–605.
- [4] S. Yang, R. Prentice, Improved logrank-type tests for survival data using adaptive weights, *Biometrics* 66 (2010) 30–38.
- [5] D.P. Harrington, T.R. Fleming, A class of rank test procedures for censored survival data, *Biometrika* 69 (1982) 553–566.
- [6] I.J. Jacobs, et al., Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial, *Lancet* 387 (2016) 5–11.
- [7] P.C. Rodriguez, et al., A phase III clinical trial of the epidermal growth factor vaccine CIMAvax-EGF as switch maintenance therapy in advanced non-small cell lung cancer patients, *Clin. Cancer Res.* 22 (2016) 3782–3790.
- [8] J. Wactawski-Wende, et al., Calcium plus vitamin D supplementation and the risk of colorectal cancer, *N. Engl. J. Med.* 354 (2006) 684–696.
- [9] Merck provides update on phase 3 study of KEYTRUDA® (pembrolizumab) monotherapy in patients with previously treated recurrent or metastatic head and neck squamous cell carcinoma (HNSCC), Available from: <http://www.mrknewsroom.com/news-release/prescription-medicine-news/merck-provides-update-phase-3-study-keytruda-pembrolizumab-m>, .
- [10] R.L. Ferris, et al., Nivolumab for recurrent squamous-cell carcinoma of the head and neck, *N. Engl. J. Med.* 375 (2006) 1856–1867.
- [11] Proteon therapeutics announces top-line data from phase 3 PATENCY-1 clinical trial of investigational vonapanitase in patients with CKD, Available from: <https://globenewswire.com/news-release/2016/12/13/897168/0/en/Proteon-Therapeutics-Announces-Top-Line-Data-from-Phase-3-PATENCY-1-Clinical-Trial-of-Investigational-Vonapanitase-in-Patients-with-CKD.html>, .
- [12] Proteon therapeutics announces increase to enrollment of ongoing phase 3 PATENCY-2 clinical trial, Available from: <https://globenewswire.com/news-release/2017/05/02/975853/0/en/Proteon-Therapeutics-Announces-Increase-to-Enrollment-of-Ongoing-Phase-3-PATENCY-2-Clinical-Trial.html>, .
- [13] R. Lin, L. Leon, Estimation of treatment effects in weighted log-rank tests, *Contemp. Clin. Trials Commun.* 8 (2017) 147–155.

Zheng Su*
Deerfield Institute, New York NY, 10017, USA
E-mail address: zhengsucctc@gmail.com

Ming Zhu
Sanofi Pasteur, Swiftwater PA, 18370, USA

* Corresponding author.