Response to comments on “Properties of the weighted log-rank test in the design of confirmatory studies with delayed effects” by José L. Jiménez, Viktoriya Stalbovskaya and

Received: 6 March 2020 | Accepted: 13 March 2020

DOI: 10.1002/pst.2021

LETTER TO THE EDITOR

WILEY

Response to comments on "Properties of the weighted log-rank test in the design of confirmatory studies with delayed effects" by José L. Jiménez, Viktoriya Stalbovskaya and Byron Jones, Pharmaceutical Statistics, 2019; 18:287-303, DOI: 10.1002/pst.1923

We thank Dr Kundu for his comments[1] on our work,[2] where the primary objective was to study the impact of delayed effects in a standard confirmatory setting. This setting included the well-known O'Brien and Fleming rejection boundaries, which are extensively used in the design of confirmatory trials in oncology. As Dr Kundu mentions, we identified a slight type-I error rate inflation when using the standard O'Brien and Fleming rejection boundaries due to the use of the Fleming-Harrington class of weights.[3] In our article, we followed Golkowski *et al.*[4] who maintain the final type-I error rate at 2.5% by setting α to a slightly lower value than 2.5%.

Another alternative, in line with the commentaries made by Dr Kundu, is to calculate the actual correlation between the test statistics from the interim and final analysis, and numerically find the rejection boundary for the final analysis.

If we assume that t_1 and t_2 are the (unweighted) log-rank test statistics at the interim and final analyses, respectively, under the null hypothesis where the logarithm of the hazard ratio is equal to 0, t_1 and t_2 have the joint distribution:

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{i_1}{i_2} \\ \frac{i_1}{i_2} & 1 \end{pmatrix} \right), \tag{1}$$

where i_1 and i_2 are the square root of the information fraction after stages 1 (interim analysis) and 2 (final analysis) of the trial. For the unweighted log-rank test, under the null hypothesis, $i_1 = \sqrt{\frac{d_1}{4}}$ and $i_2 = \sqrt{\frac{d_2}{4}}$, where d_1 and d_2 are the number of events across both arms after stages 1 and 2 of the trial.

Now let $t' = \frac{i_2 t_2 - i_1 t_1}{\sqrt{i_2^2 - i_1^2}}$. Then by the standard formula for the multivariate normal distribution,[5] the joint distribution of $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ is:

$$\begin{pmatrix} 1 & 0 \\ -\frac{i_1}{\sqrt{i_2^2 - i_1^2}} & \frac{i_2}{\sqrt{i_2^2 - i_1^2}} \end{pmatrix} \times \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right). \tag{2}$$

Since

$$\begin{pmatrix} 1 & 0 \\ -\frac{i_1}{\sqrt{i_2^2 - i_1^2}} & \frac{i_2}{\sqrt{i_2^2 - i_1^2}} \end{pmatrix} \times \begin{pmatrix} 1 & \frac{i_1}{i_2} \\ \frac{i_1}{i_2} & 1 \end{pmatrix} \times \begin{pmatrix} 1 & -\frac{i_1}{\sqrt{i_2^2 - i_1^2}} \\ 0 & \frac{i_2}{\sqrt{i_2^2 - i_1^2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3}$$

t' and t_1 are stochastically independent and hence t' is referred as an "independent increment". However, as mentioned by Dr Kundu, if we approximate $\sqrt{\frac{i_1}{i_2}}$ by $\sqrt{\frac{d_1}{d_2}}$ but $corr(t_1, t_2) = \sqrt{\frac{i_1}{i_2}} \neq \sqrt{\frac{d_1}{d_2}}$, then t' and t_1 will not be

stochastically independent and hence t' will not be an independent increment. The Fleming-Harrington class of weights is defined as:

$$w(k) = \hat{S}(k)^{\rho}\left(1-\hat{S}(k)\right)^{\gamma}, \tag{4}$$

where $\hat{S}(k)$ represents the pooled survival function and $k = 1, ..., D$ represents the observed event times. In our article, we focused on the use of ($\rho = 0, \gamma = 1$), hence $w(k) = 1-\hat{S}(k)$.

Let ρ be equal to $corr(t_1, t_2) = \sqrt{\frac{i_1}{i_2}}$. Hence, to calculate ρ using the Fleming-Harrington class of weights, i_1 and i_2 are calculated as follows:

$$i_1 = \sqrt{\frac{\sum_{k=1}^{d_1} w^2(k)}{\sum_{k=1}^{d_2} w^2(k)}}, \tag{5}$$

$$i_2 = \sqrt{\frac{\sum_{k=1}^{d_2} w^2(k)}{\sum_{k=1}^{d_2} w^2(k)}} = 1, \tag{6}$$

assuming that the expected number of events at the final analysis corresponds to the actual number of events at the final analysis.

Assuming the same cumulative α spent we used in our article,[1] where at the interim analysis we spent 0.01 of the α, the second stage rejection boundary need to be numerically calculated using a bivariate normal probability density function of the form:

$$f(\mathbf{x}) = \frac{1}{2\pi}(\det\Sigma)^{\frac{1}{2}}\int_{-\infty}^{y}\int_{-\infty}^{x}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)d\mathbf{x}, \tag{7}$$

where $x = 1 - \Phi(1\text{-}0.01)$, Φ corresponds to the cumulative distribution function of the normal distribution, $\boldsymbol{\mu} = (0, 0)$ and $\boldsymbol{\Sigma}$ is a 2×2 positive definite covariance matrix with $y < \infty$.

We numerically find the value of y using the piece of **R** code presented below.

```
#Function that calculates the 2nd stage rejection boundary
crit_adj = function(y, alpha1, alpha, corr){
   x<-qnorm(1-alpha1)
   rejprob<-1-pmvnorm(upper=c(x,y),corr=corr)
   return(rejprob-alpha)
}

#All correlation values that are tested
rho_values = seq(0,by=0.001,1)
#Definition of output variables
z_stat = mat.or.vec(length(rho_values),1)
p_final = mat.or.vec(length(rho_values),1)

for(i in 1:length(rho_values)){
      rho = rho_values[i]
      corr = (1-rho)*diag(2)+rho
```

```
      alpha = 0.025 #final type-I error rate
      alpha1 = 0.01 #alpha spent at the interim

      #critical value at 2nd stage (z-scale)
      crit_final = uniroot(crit_adj, lower = 0, upper = 3,
      alpha1=alpha1, alpha=alpha, corr=corr, tol=1E-12)
      z_stat[i] = crit_final$root

      #p-value at which we test in the 2nd stage
      p_final[i] = 1-pnorm(crit_final$root)
}
```

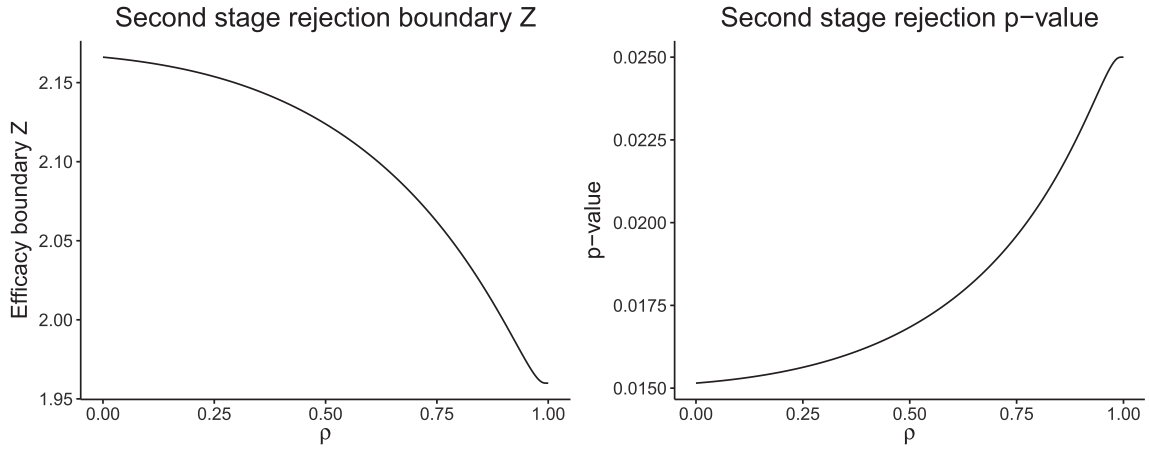

FIGURE 1 Second stage rejection boundaries and correspondent *p*-values for different ρ values

In this code, we look a value of y that is a root (ie, zero) of the function **`crit_adj(y, alpha1 = 0.01, alpha = 0.025, corr = ((1-ρ)*diag(2) + ρ))`** within the interval [0,3]. The value of y will correspond to the rejection boundary for the final analysis (or second stage). In Figure 1, we present the rejection boundaries in the Z-scale and also as a P value for correlation values that range from 0 to 1.

Therefore, at the final analysis (ie, analysis at the second stage of the trial), after calculating the correlation between the test statistics obtained at each stage, the new rejection boundary is calculated to guarantee type-I error control.

The arguments presented in the document can also be easily shown in a simple numerical example. Assume that that we have the survival data presented in Figure 2. This data was generated using the **R** code provided in Reference 1 with a delay of 5 months, median OS in the control group of 6 months, and median OS in the experimental group of 6 months (before the delay) and 9 months (after the delay). This dataset has 20% of censoring, and the number of events at the interim and final analyses is equal to 16 and 32, respectively. By calculating the survival function using the **R** function **`survfit`** and applying the Fleming and Harrington class of weights with $\rho = 0$ and $\gamma = 1$, we can easily how $corr(t_1, t_2) = \sqrt{\frac{i_1}{i_2}} \neq \sqrt{\frac{d_1}{d_2}}$ since $\sqrt{\frac{16}{32}} = 0.7071068$ and $\sqrt{\frac{3.4}{1}} = 0.503974$ (see Equations (5) and (6) to calculate i_1 and i_2).

ACKNOWLEDGEMENTS

We thank Dr Ekkehard Glimm (Novartis) for his support during the development of this work. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 633567.

	survival_time	censor	group
31	0.071651054601017	1	0
32	0.677906860681359	1	1
27	0.722500626929104	1	1
39	1.16623106469013	1	0
38	1.61353727476671	1	1
40	1.80849146022643	1	1
14	2.29729606821087	1	1
11	2.56071537128923	1	1
24	2.77684887889744	1	1
6	2.85299727972597	1	0
20	3.20778141085787	1	0
2	3.38166847033426	1	1
19	4.88705849647522	1	0
35	5.08672975446098	1	1
23	5.13365351165104	1	1
4	5.51792960551904	1	0
10	6.08883890130078	1	0
17	6.10617013316123	1	0
26	6.21738883201033	1	1
1	6.39813753543422	1	0
12	7.17822533092901	1	0
9	7.19119129315961	1	1
30	7.20232536459635	1	0
25	8.47713092827689	1	0
29	8.89840001100674	1	0
36	9.43487367313327	1	0
16	10.1624908517115	1	0
22	11.0938184307845	0	0
33	12.0938184307845	0	1
5	12.2192273664169	1	1
3	12.4934582217931	1	1
8	12.8335208315402	1	0
28	14.0938184307845	0	0
18	15.0938184307845	0	0
34	16.0938184307845	0	1
13	17.0938184307845	0	1
21	18.0938184307845	1	1
37	18.0938184307845	0	1
7	18.5955557872188	1	0
15	20.0938184307845	0	1

FIGURE 2 Simulated survival times

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

José L. Jiménez[1,2]
Viktoriya Stalbovskaya[3]
Byron Jones[2]

[1]*Dipartimento di Scienze Matematiche, Politecnico di Torino, Turin, Italy*
[2]*Novartis Pharma AG, Basel, Switzerland*
[3]*Merus, Utrecht, The Netherlands*

Correspondence
José L. Jiménez, Dipartimento di Scienze Matematiche, Politecnico di Torino, Turin, Italy.
Email: jose.jimenez@polito.it

ORCID

José L. Jiménez https://orcid.org/0000-0002-8809-2717

REFERENCES

1. Kundu, M.G. Comments on "Properties of the weighted log-rank test in the design of confirmatory studies with delayed effects" by Jose Jimenez, Viktoriya Stalbovskaya, and Byron Jones. *Pharm Stat*. 18:287-303, 2019, https://doi.org/10.1002/pst.1923. Pharm Stat. 2020. https://doi.org/10.1002/pst.2017.
2. Jiménez JL, Stalbovskaya V, Jones B. Properties of the weighted log-rank test in the design of confirmatory studies with delayed effects. *Pharm Stat*. 2019;18(3):287-303.
3. Fleming TR, Harrington DP. A class of hypothesis tests for one and two sample censored survival data. *Commun Stat Theory Methods*. 1981;10(8):763-794.
4. Golkowski D, Friede T, Kieser M. Blinded sample size re-estimation in crossover bioequivalence trials. *Pharm Stat*. 2014;13(3):157-162.
5. Jupp PE, Mardia KV. Maximum likelihood estimators for the matrix von mises-fisher and bingham distributions. *Ann Stat*. 1979;7(3):599-606.