ABSTRACT: **301146**

Sample Size Determination in Group-Sequential Trials Assessing Interim Futility by Intermediate Composite Endpoints

Shogo Nomura[1]

[1]Department of Biostatistics and Bioinformatics, Graduate School of Medicine, The University of Tokyo, Japan

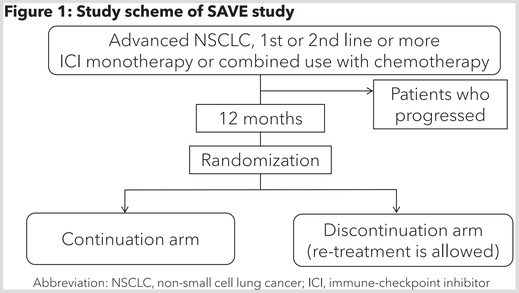
ABSTRACT

- In oncology phase 3 trials, non-inferiority concepts have sometimes been evaluated without preliminary knowledge of an experimental treatment being non-inferior to the standard treatment, and thus, a prompt consideration for futility stop is vital.
- Such concepts are usually examined in a patient population with a good prognosis, and thus, it is often the case that the number of observed events regarding primary endpoint is quite small at earlier interim analyses, which would miss an opportunity for a harmful treatment to be recommended a futility stop. To address these issues, the incorporation of an aggressive futility monitoring rule using an earlier-available intermediate endpoint is attractive.
- When the primary endpoint is overall survival (OS), typical intermediate endpoints include progression-free survival, and so on. Past studies have shown from simulation and case studies that the monitoring rule accelerated the timing of futility recommendation [1, 2]. However, they missed an issue of using an early aggressive rule in that the study power is sometimes dropped to a nonnegligible level.
- In this study, we numerically assess the power reduction of when an aggressive inefficacy monitoring based on an earlier-available intermediate endpoint is used. We then propose a novel sample size determination method to achieve power at a predetermined level.

A MOTIVATING EXAMPLE

- The SAVE study [3] is an ongoing phase 3 non-inferiority trial evaluating the value of breaking ICI agents (e.g., pembrolizumab and nivolumab) for NSCLC patients (primary endpoint: OS).
- Because there is no promising data supporting the utility of the discontinuation, and the number of deaths at the first interim analysis is expected to be quite small, the study investigators decided to add a strict inefficacy rule based on an intermediate endpoint, time to treatment failure of strategy (TFS).
- The required number of patients is 216 (106 OS events) supposing that 2-year OS rate is 70%, the expected hazard ratio (HR) for OS is one, and the non-inferiority margin is 1.53 (one-sided alpha: 5%, power: 70%, enrollment and follow-up period: 2.5 and 3 years).

Figure 1: Study scheme of SAVE study



Abbreviation: NSCLC, non-small cell lung cancer; ICI, immune-checkpoint inhibitor

A STATISTICAL CHALLENGE

Planned interim analyses in the SAVE study

- Total of two interim analyses (IAs) is planned.
- A stop for efficacy is recommended using the boundary calculated from the O'Brien-Fleming type alpha spending function.
- A stop for futility is considered if $\widehat{\text{HR}}$ for OS exceeds the non-inferiority margin and/or discontinuation of ICI is not less toxic.
- Only in the 1st IA, an aggressive inefficacy monitoring rule based on TFS is incorporated.

Statistical challenge

- Incorporation of the aggressive inefficacy monitoring rule by TFS may drop the study power would be dropped to a nonnegligible level.

EXISTING METHODS

Design and basic notations

- Suppose that, within τ_a yrs, total of N patients are uniformly enrolled and randomized into control (C) or test (T) arm in a $r:(1-r)$ ratio.
- θ_{kl}: HR for endpoint k ($k=1$: TFS, $k=2$: OS) at lth analysis, defined as a ratio of the hazard function of arm T to that of arm C.
- We assume that the proportionality of hazards assumption holds for both endpoints throughout all analyses, i.e., $\hat{\theta}_{kl}(t) \equiv \theta_k$.
- $\phi_{Ck}(t)$ and $\phi_{Tk}(t)$: distribution functions for endpoint k in each arm
- $\mathcal{H}_0$: $\log\theta_2 \geq \log\delta_2$, and $\mathcal{H}_1$: $\log\theta_2 < \log\delta_2$, where δ_k is a non-inferiority margin for endpoint k.
- With a large sample size, $\log\hat{\theta}_{kl}$ is normally distributed as $N(\log\theta_{kl}, \sigma^2_{kl})$, where

$$\sigma^2_{kl} = \frac{1}{Nr\phi_{Ck}(t_l)} + \frac{1}{N(1-r)\phi_{Tk}(t_l)}.$$

- Z statistic of endpoint k at analysis l is $Z_{kl} = (\log\hat{\theta}_{kl} - \log\delta_k)\sigma_{kl}^{-1} = (\log\hat{\theta}_{kl} - \log\delta_k)\sqrt{\mathcal{I}_{kl}}$
- With the nomial one-sided significance level of alpha and power as α and $1-\beta$, the required number of OS events e_0^* and sample size N_0^* can be calculated as:

$$e_0^* = \frac{(z_{1-\alpha}+z_{1-\beta})^2}{r(1-r)(\log\delta_2)^2},\ N_0^* = e_0^* \times \sqrt{\frac{1}{r\phi_{Ck}(t_L)} + \frac{1}{(1-r)\phi_{Tk}(t_L)}},$$

where $t_L = \tau_a + \tau_f$ (follow-up period [yrs]).

- c^E_{kl} and c^F_{kl}: Efficacy or futility boundaries for Z_{kl}.
- In this study, for an illustrative purpose, c^E_{2l} is calculated using the O'Brien-Fleming type cumulative error spending function, and, for c^F_{2l}, a harm look approach and linear inefficacy boundary (LIB)[4] are used for OS. For c^F_{11}, a more aggressive boundary for TFS assessment in the 1st IA. Note that our approach works for any decision rules that can be transformed to rules for Z_{kl}.

Decision making rules

At the lth interim analysis ($l = 1, \cdots, L-1$):

- If $Z_{1l} < c^F_{1l}$ and $Z_{2l} < c^E_{2l}$, reject $\mathcal{H}_0$ (stop for efficacy)
- If $Z_{1l} \geq c^F_{1l}$ or $Z_{2l} \geq c^F_{2l}$ is satisfied at least one of k, accept $\mathcal{H}_0$ (stop for futility)
- Otherwise, continue to the next interim analysis

At the final analysis ($l = L$):

- If $Z_{2L} < c^E_{2L}$, reject $\mathcal{H}_0$
- Otherwise, do not reject $\mathcal{H}_0$

PROPOSED METHOD

Power functions

- Suppose that the sequence of all the test statistics $\{Z_{kl}\}$ has a multivariate normal distribution. There are three types of correlation parameters: $\text{Corr}[Z_{kl}, Z_{kl'}] = \sqrt{\frac{e_{kl}}{e_{kl'}}}$, $\text{Corr}[Z_{1l}, Z_{2l}] = \rho_l$, and $\text{Corr}[Z_{1l}, Z_{2l'}] = \rho_{ll'}$, where $l < l'$ and e_{kl} denotes the total number of events for endpoint k at the lth analysis.
- In the simplest case of $L = 2$, we consider $\{Z_{11}, Z_{21}, Z_{22}\}$ having a three-dimensional multivariate normal distribution with

$$\boldsymbol{\mu} = \begin{bmatrix} \{\log\theta^*_{11} - \log\delta_1\}\,\sigma_{11}^{-1} \\ \{\log\theta^*_{21} - \log\delta_2\}\,\sigma_{21}^{-1} \\ \{\log\theta^*_{22} - \log\delta_2\}\,\sigma_{22}^{-1} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_1 & \rho_{12} \\ & 1 & \sqrt{e_{21}/e_{22}} \\ & & 1 \end{bmatrix}.$$

θ^*_{kl} is the expected hazard ratio under an assumed scenario.

- Based on the above-mentioned decision rules, the probability of observing events can then be obtained as:

$$\pi^E_1 = \Pr(\mathcal{B}_{11} \cap \mathcal{A}_{21}) = \int_{-\infty}^{c^F_{11}}\int_{-\infty}^{c^E_{21}} f_2(z_{11}, z_{21})\mathrm{d}z_{11}\mathrm{d}z_{21}$$

$$\pi^F_1 = 1 - \Pr(\mathcal{B}_{11} \cap \mathcal{B}_{21}) = 1 - \int_{-\infty}^{c^F_{11}}\int_{-\infty}^{c^F_{21}} f_2(z_{11}, z_{21})\mathrm{d}z_{11}\mathrm{d}z_{21}$$

$$\pi^F_2 = \Pr(\{\mathcal{B}_{11} \cap \mathcal{C}_{21}\} \cap \mathcal{A}_{22}) = \int_{-\infty}^{c^F_{11}}\int_{c^F_{21}}^{c^E_{21}}\int_{-\infty}^{c^E_{22}} f_3(z_{11}, z_{21}, z_{22})\mathrm{d}z_{11}\mathrm{d}z_{21}\mathrm{d}z_{22}$$

$$\pi^F_2 = \Pr(\{\mathcal{B}_{11} \cap \mathcal{C}_{21}\} \cap \bar{\mathcal{A}}_{22}) = \int_{-\infty}^{c^F_{11}}\int_{c^F_{21}}^{c^E_{21}}\int_{c^E_{22}}^{\infty} f_3(z_{11}, z_{21}, z_{22})\mathrm{d}z_{11}\mathrm{d}z_{21}\mathrm{d}z_{22}$$

where π^E_l and π^F_l denote the probability of declaring efficacy and futility at analysis l, f_2 and f_3 denote bivariate and trivariate normal distribution functions, $\mathcal{A}_{kl} = \{Z_{kl} < c^E_{kl}\}$, $\mathcal{B}_{kl} = \{Z_{kl} < c^F_{kl}\}$, and $\mathcal{C}_{kl} = \{c^E_{kl} < Z_{kl} < c^F_{kl}\}$.

- For general L, π^E_l ($l \geq 2$) is calculated as

$$\pi^E_l = \Pr\left(\left\{\bigcap_{l_0=1}^{l-1} \mathcal{B}_{1l_0} \cap \mathcal{C}_{2l_0}\right\} \cap \mathcal{A}_{2l}\right),$$

and the power function for trials with L analyses is $\sum_{l=1}^{L}\pi^E_l$.

Sample size determination

- Given τ_a, τ_f,α, $1-\beta$ (target power), δ_k, η_k, ξ_{kl}, L, t_l, $\phi_{C_k}(t)$, and $\phi_{T_k}(t)$, the procedure in case of $r = 0.5$ is as follows.
  1. Use N_0^* as an initial value for the sample size $N_{(0)}$. Run a simulation to estimate the expected values of ρ_l and $\rho_{ll'}$. This can be performed through many repetitions (here, 10,000 times) of the generation of a hypothetical trial dataset of size $N_{(0)}$.
  2. Set $N_{(i+1)} = N_{(i)} + 2$ ($i = 0, 1, \cdots$), and calculate e_{kl}, σ_{kl}, $\boldsymbol{\mu}_{2L-1}$, $\boldsymbol{\Sigma}_{2L-1}$, and efficacy/inefficacy boundaries c^E_l/c^F_l for all l.
  3. Calculate π^F_l for all l, and calculate the power, denoted as $\Psi(N = N_{(i+1)})$. Proceed to Step 2 when $\Psi(N = N_{(i+1)})$ does not reach the target value of $1-\beta$. When proceeding to Step 2, one sets the new i as $i+1$. Otherwise, finish the iteration steps. The resulting N is termed as N_1^*.
- We considered exponential model and its more complicated version termed three-component OS model for $\phi_{C_k}(t)$ and $\phi_{T_k}(t)$.

NUMERICAL AND SIMULATION STUDIES

Scenarios and evaluations

Under the following scenarios in the SAVE study, π^E_l and π^F_l were calculated from numerical integrations and simulations.

- Simple null ($\theta^*_1 = 1.53, \theta^*_2 = 1.53$);
- Complex null ($\theta^*_1 = 1.24, \theta^*_2 = 1.53$);
- Simple null ($\theta^*_1 = 1, \theta^*_2 = 1$);
- Complex alternative 1 ($\theta^*_1 = 1.24, \theta^*_2 = 1.00$);
- Complex alternative 2 ($\theta^*_1 = 0.84, \theta^*_2 = 1.00$).

Results (TFS: aggressive monitoring rule, OS: harm look)

- The calculation of π^E_l and π^F_l was successful, resulting in a good performance of the proposed N_1^*. The ratio N_1^*/N_0^* was 1.6-1.7.

Table : Numerically calculated π^E_l, π^F_l, and power (sample size: $N_0 = 216$)

	Efficacy monitoring			Type I error	Futility monitoring		
Scenario	t_1	t_2	t_3	/Power	t_1	t_2	t_3
TFS: exponential, OS: three-component model							
Simple null	0.0%	0.7%	1.4%	**2.1%**	74.9%	2.6%	20.4%
Complex null	0.0%	1.1%	2.8%	**3.9%**	43.4%	2.5%	50.2%
Simple alternative	0.1%	18.5%	45.3%	**63.8%**	15.3%	0.1%	20.8%
Complex alternative 1	0.1%	15.9%	28.8%	**44.8%**	42.0%	0.5%	12.7%
Complex alternative 2	0.0%	17.9%	53.3%	**71.2%**	4.2%	0.1%	24.5%

Table : Performance of N_1^* (target level: 80%)

Non-inferiority margin					OS: three-component		
$S(2)$	2-year %OS	Hazard ratio	N_0^*	N_1^*	N_1^*/N_0^*	$\Psi(N = N_1^*)$	Simulated power
0.7	6%	1.25	934	1501	1.61	80.0%	**80.0%**
	9%	1.39	442	709	1.60	80.0%	**80.6%**
	12%	1.53	262	421	1.61	80.0%	**78.9%**

CONCLUSION

- Our proposed method can return all the stopping probabilities precisely when an inefficacy due to an intermediate composite endpoint is incorporated in a typical group sequential design framework with single primary endpoint.
- When the reduction is noteworthy, one should consider applying our proposed sample size determination method.

For more details, please see the accepted article in *Statistics in Biopharmaceutical Research*:
https://doi.org/10.1080/19466315.2020.1799852

MAIN REFERENCES

[1] Goldman B, LeBlanc M, and Crowley J.
Interim futility analysis with intermediate endpoints.
Clinical Trials 2008; Vol. 5, No. 1, pp. 14–22.

[2] Wang M, Dignam JJ, Zhang QE, DeGroot JF, Mehta MP, and Hunsberger S.
Integrated phase II/III clinical trials in oncology: A case study.
Clinical Trials 2012; Vol. 9, No. 6, pp. 741–747.

[3] Nomura S, Goto Y, Mizutani T, Kataoka T, Kawai S, Okuma Y, Murakami H, Tanaka K, and Ohe Y.
A randomized phase III study comparing continuation and discontinuation of PD-1 pathway inhibitors for patients with advanced non-small-cell lung cancer (JCOG1701, SAVE study).
Japanese Journal of Clinical Oncology 2020.

[4] Freidlin B, Korn EL, and Gray R.
A general inefficacy interim monitoring rule for randomized clinical trials.
Clinical Trials 2010; Vol. 7, No. 3, pp. 197–208.

Presented at the 2020 ASA Biopharmaceutical Section Statistics Workshop (virtual).
September 22-25, 2020

snomura@m.u-tokyo.ac.jp