

EXACT CALCULATION OF TARGETED CLINICAL TRIALS EFFICIENCY IN THE PARAMETRIC CASE

Aboubakar Maitournam¹

*University Abdou Moumouni of Niamey, Faculty of Sciences and Techniques
Department of Mathematics and Computer Sciences
PB 10662, Niamey, Niger*

ABSTRACT. Previously we introduced the phase III efficiency of targeted clinical trials relative to standard untargeted clinical trials. It was defined as the ratio of required sample sizes of two designs. We established that efficiency by assuming that in the case of untargeted clinical trials, both the control and treatment groups are mixtures of two genotypes noted respectively R+ and R- with a common outcome standard deviation for these subgroups. R+ patients are assumed to be more likely to respond to a molecularly targeted therapy than R- patients. Hence, targeted clinical trials designs randomize only R+ patients. Moreover, we made the assumption that the outcome of control and treatment groups can be modeled by the mixture of two normal distributions due to the presence of two above mentioned genotypes in the overall population. Here, we consider distinct outcome standard deviations for genotypes R+ and R- in the control and treatment groups. The obtained efficiency expression still shows that the untargeted design will require many more patients than the targeted design. In addition, the established expression of efficiency can allow one to evaluate several scenarios in order to determine the appropriate sample size for targeted clinical trials design in function of the treatment effect and the prevalence of two genotypes in the overall population.

Résumé. Précédemment, nous avons défini l'efficacité des essais cliniques moléculaires ciblés par rapport aux essais cliniques conventionnels; comme étant le ratio des tailles d'échantillons requises dans la conception de la phase III des deux types d'essais susmentionnés. Nous avons établi l'expression analytique de cette efficacité relative en supposant que dans le cas des essais cliniques conventionnels, les groupes contrôle et traitement, sont composés de deux génotypes notés respectivement R+ et R- ayant la même variance en termes de réponse. Les patients R+ sont supposés être des bénéficiaires potentiels de la thérapie moléculaire ciblée. Par conséquent, les essais cliniques moléculaires ciblés randomisent seulement les patients de type R+. De plus, nous avons supposé que les variables aléatoires représentant les réponses à la thérapie dans les groupes contrôle et traitement peuvent être modélisées par des mélanges de deux lois gaussiennes à cause de la présence des génotypes R+ et R- dans la population mère. Dans cet article nous supposons que les réponses des génotypes R+ et R- dans les groupes contrôle et traitement ont des variances pouvant être distinctes. La formule explicite de l'efficacité relative que nous obtenons, prouve encore une fois que le modèle conventionnel non ciblé exigera beaucoup plus de patients que le modèle ciblé. D'autre part, cette expression analytique de l'efficacité relative peut permettre de déterminer la taille d'échantillon appropriée pour la conception de la phase III des essais cliniques moléculaires ciblés en fonction de l'effet du traitement et de la prévalence des génotypes R+ et R- dans la population mère.

Keywords: Sample size, targeted clinical trials, efficiency, Gaussian mixture, stratified and personalized medicines.

Mots clés: Calcul de tailles d'échantillons, mélange de lois normales, essais cliniques moléculaires ciblés, médecines stratifiée et personnalisée

AMS 2010 Subject Classification: 62P10.

Copyright © 2014, African Journal of Applied Statistics. All rights reserved

Article history: Received 2014/01/01; Accepted 2014/10/30; Published Online 2014/11/30.

1. INTRODUCTION

Our previous papers [Maitournam and Simon (2004) and Simon and Maitournam (2004)] dealt with the development of an expression for the phase III efficiency of targeted clinical trials. Targeted designs are of interest in the evaluation of a molecularly targeted therapy and involve comparing control and treatment groups, but with enrolling based on a molecular assay only patients considered likely to benefit from the new treatment. In contrast, conventional untargeted clinical trials have broad eligibility which increases macroscopic and

molecular heterogeneity. Targeted clinical trials are particularly suited to the current genomic era featured by the massive generation of molecular data like biomarkers which may be potential targets for molecularly targeted drugs [see Simon (2009)]. Targeted Clinical Trials certainly will be the cornerstone of stratified and personalized medicines [see Maitournam (2011)]. Phase III efficiency of targeted clinical trials versus untargeted clinical trials was calculated as the ratio of required sample sizes of two designs. We established an expression for the efficiency of targeted clinical trials by assuming that in the case

¹Corresponding author: (A. Maitournam): maitournai@gmail.com
maitournai@gmail.com, address : Department of Mathematics and Computer Sciences, PB 10662, Niamey, Niger.

of untargeted clinical trials, both the control and treatment groups are mixtures of two genotypes noted respectively R+ and R- with common outcome standard deviation for these genotypes in the control and treatment groups. R+ patients are assumed more likely to respond to the molecularly targeted therapy than R- patients. Only R+ patients are enrolled in the targeted clinical trials design. Moreover, we made the assumption that the outcome of control and the treatment groups can be modeled by the mixture of two normal distributions due to the presence of two mentioned genotypes in the overall population. Here we consider the less restrictive hypothesis of distinct outcome standard deviations for genotypes R+ and R- in the control and treatment groups, and consequently we develop an expression of targeted clinical trials efficiency in that setting. The paper is organized as follows. First, the exact expression of targeted clinical trials efficiency is established in section 2 and is proved in section 3, and then some remarks are drawn in section 4.

2. EXACT EXPRESSION OF TARGETED CLINICAL TRIALS EFFICIENCY

Assume that the random variables X and Y which represent respectively the overall outcome for the control and treatment groups have their densities defined as follows:

$$f_X = \gamma f_{\mu_0, \sigma_0^2} + (1 - \gamma) f_{\mu_1, \sigma_1^2} \tag{2.1}$$

$$f_Y = \gamma f_{\mu_{0T}, \sigma_0^2} + (1 - \gamma) f_{\mu_{1T}, \sigma_1^2} \tag{2.2}$$

where

- f_{μ, σ^2} is the density of the normal distribution with mean μ and variance σ^2 ,
- γ denotes in the studied population the proportion of genotype R- patients,
- $1 - \gamma$ is the proportion of genotype R+ patients,
- μ_0 is the mean value of outcome for genotype R- patients in the control group,
- μ_1 is the mean value of outcome for genotype R+ patients in the control group,
- μ_{0T} is the molecularly targeted therapy effect mean value for genotype R- patients in the treatment group,
- μ_{1T} is the molecularly targeted therapy effect mean value for genotype R+ patients in the treatment group,
- σ_0 is the outcome standard deviation for genotype R- patients in the control and treatment groups,
- σ_1 is the outcome standard deviation for genotype R+ patients in the control and treatment groups.

Let $\delta = \mu_{0T} - \mu_0$ and $\mu_{1T} - \mu_1$ respectively be the potential negligible therapy benefit for R- patients and the full benefit for R+ responders depending on the scenario or design of clinical trials with $\mu_{1T} - \mu_1 = r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0}$; $r > 1$; $\Delta > 0$, and $\delta \in [0, \epsilon]$, $\epsilon > 0$ where I_A is one if A is true and 0 otherwise.

Then given

- the power $1 - \beta$, $0 < \beta < 1$
- and the type I error α , $0 < \alpha < 1$,

if $2n$ and $2n_T$ are the required samples sizes respectively for the untargeted and targeted clinical trials designs in order to detect the molecularly targeted therapy mean effect, the efficiency e of targeted clinical trials versus untargeted clinical trials is obtained as follows:

$$e = \frac{2n}{2n_T} = \frac{n}{n_T} \tag{2.3}$$

$$= \frac{1 + \gamma(\frac{\sigma_0^2}{\sigma_1^2} - 1) + \frac{\gamma(1-\gamma)}{2\sigma_1^2} [(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2]}{[1 - \gamma(1 - \frac{\mu_{0T} - \mu_0}{\mu_{1T} - \mu_1})]^2} \tag{2.4}$$

$$= \frac{1 + \gamma(\frac{\sigma_0^2}{\sigma_1^2} - 1) + \frac{1}{2\sigma_1^2} [\frac{V(X)}{2} + \frac{V(Y)}{2} - \gamma(\sigma_0^2 - \sigma_1^2) - \sigma_1^2]}{[1 - \gamma(1 - \frac{\delta}{r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0}})]^2} \tag{2.5}$$

where $V(X)$ and $V(Y)$ are respectively the variances of X and Y .

3. PROOF

By using the formulae for the calculation of mixture means and variances established by Pearson (1894), the means and variances of random variables X and Y which represent respectively the control and treatment outcomes and which densities are represented in equations (2.1) and (2.2), are defined as follows:

$$E(X) = \gamma\mu_0 + (1 - \gamma)\mu_1$$

$$E(Y) = \gamma\mu_{0T} + (1 - \gamma)\mu_{1T}$$

$$V(X) = \gamma\sigma_0^2 + (1 - \gamma)\sigma_1^2 + \gamma(1 - \gamma)(\mu_1 - \mu_0)^2 = \sigma_C^2$$

$$V(Y) = \gamma\sigma_0^2 + (1 - \gamma)\sigma_1^2 + \gamma(1 - \gamma)(\mu_{1T} - \mu_{0T})^2 = \sigma_T^2. \tag{3.6}$$

The difference of means or effect size between the control and treatment outcomes, which represents the therapy effect is $d = E(Y) - E(X) = \gamma(\mu_{0T} - \mu_0) + (1 - \gamma)(\mu_{1T} - \mu_1) = \gamma\delta + (1 - \gamma)(r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0})$ where $\mu_{0T} - \mu_0 = \delta$ and $\mu_{1T} - \mu_1 = r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0}$

If we denote by \bar{X} and \bar{Y} the random variables which describe the estimated response means respectively for the control and treatment groups, $m_C = E(X)$ and $m_T = E(Y)$ the theoretical means or mathematical expectations respectively of X and Y ; then the usual central limit theorem implies that $\bar{X} \sim N(m_C, \frac{\sigma_C^2}{n})$, $\bar{Y} \sim N(m_T, \frac{\sigma_T^2}{n})$ where n is the sample size for control group assumed to be the same for that of treatment group.

Thus for a given type I error α , the null hypothesis H_0 of no difference of means between the control and treatment groups is rejected if $n^{\frac{1}{2}} \frac{\bar{Y} - \bar{X}}{\sqrt{\sigma_C^2 + \sigma_T^2}} > Z_{1 - \frac{\alpha}{2}}$ where $Z_{1 - \frac{\alpha}{2}}$ is the standard normal distribution $\frac{\alpha}{2}$ percentile.

For a given power $1 - \beta$, the calculation of the required sample size for the design without screening that is for the untargeted design, is obtained by using classical methods established for normal distributions [See Desu and Raghavarao (1990); Adcock (1997)], as follows:

$$P \left[n^{\frac{1}{2}} \frac{\bar{Y} - \bar{X}}{\sqrt{\sigma_C^2 + \sigma_T^2}} + \left(\frac{\gamma(\mu_{0T} - \mu_0) + (1 - \gamma)(\mu_{1T} - \mu_1)}{\sqrt{\sigma_C^2 + \sigma_T^2}} \right) n^{\frac{1}{2}} > Z_{1 - \frac{\alpha}{2}} \right] = 1 - \beta = 1 - \Phi(-Z_{1 - \beta})$$

where Φ is the cumulative distribution

function of the standard normal distribution.

So

$$1 - \Phi\left(Z_{1-\frac{\alpha}{2}} - \left(\frac{\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)}{\sqrt{\sigma_C^2 + \sigma_T^2}}\right)n^{\frac{1}{2}}\right) = 1 - \Phi(-Z_{1-\beta})$$

$$\text{Thus } Z_{1-\frac{\alpha}{2}} - \left(\frac{\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)}{\sqrt{\sigma_C^2 + \sigma_T^2}}\right)n^{\frac{1}{2}} = -Z_{1-\beta}$$

Hence we obtain the required sample size without screening that is for the untargeted design

$$n = \frac{A}{[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)]^2}$$

where

$$A = (Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2(2\gamma\sigma_0^2 + 2(1-\gamma)\sigma_1^2 + \gamma(1-\gamma)[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2]$$

and with screening ($\gamma = 0$) that is for targeted design, we get the following sample size

$$n_T = \frac{2\sigma_1^2(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{(\mu_{1T} - \mu_1)^2}$$

Then by denoting

$$B = 2\gamma\sigma_0^2 + 2(1-\gamma)\sigma_1^2 + \gamma(1-\gamma)[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2](\mu_{1T} - \mu_1)^2$$

and

$$C = \gamma\frac{\sigma_0^2}{\sigma_1^2} + (1-\gamma) + \frac{\gamma(1-\gamma)}{2\sigma_1^2}[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2](\mu_{1T} - \mu_1)^2,$$

we get the following equation noted (2.3) in section 2 which represents the efficiency of the targeted design versus the untargeted design:

$$e = \frac{2n}{2n_T} = \frac{n}{n_T} = \frac{B}{2\sigma_1^2[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)]^2} = \frac{C}{[\gamma(\mu_{0T} - \mu_0) + (1-\gamma)(\mu_{1T} - \mu_1)]^2}$$

and then, from (2.4), e is equal to

$$\frac{\gamma\frac{\sigma_0^2}{\sigma_1^2} + (1-\gamma) + \frac{\gamma(1-\gamma)}{2\sigma_1^2}[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2](\mu_{1T} - \mu_1)^2}{(\mu_{1T} - \mu_1)^2 [1 - \gamma + \gamma\frac{\mu_{0T} - \mu_0}{\mu_{1T} - \mu_1}]^2}$$

We finally obtain equation (2.5) by replacing $\gamma(1-\gamma)[(\mu_1 - \mu_0)^2 + (\mu_{1T} - \mu_{0T})^2]$ using (3.6) and hypothesis $\mu_{0T} - \mu_0 = \delta, \mu_{1T} - \mu_1 = r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0}$.

$$e = \frac{1 + \gamma\left(\frac{\sigma_0^2}{\sigma_1^2} - 1\right) + \frac{1}{2\sigma_1^2}\left[\frac{V(X)}{2} + \frac{V(Y)}{2} - \gamma(\sigma_0^2 - \sigma_1^2) - \sigma_1^2\right]}{\left[1 - \gamma\left(1 - \frac{\delta}{r\delta I_{\delta \neq 0} + \Delta I_{\delta = 0}}\right)\right]^2}$$

4. DISCUSSION

The restrictive assumption of common variance for the outcomes of two genotypes R+ and R- in the untargeted and treatment groups is released in that paper, because it is likely that polymorphism and other complex mutations and interactions may induce distinct variability [See Betensky *et al* (2002)] between the two subgroups. The expressions of efficiency (see equations 2.4 and 2.5) clearly show that the untargeted design will require more patients than the targeted design. Indeed, the equations (2.4 and 2.5) show that efficiency e is always greater than one as the numerator is always greater than one because σ_0^2 reasonably is greater than σ_1^2 whereas the denominator is always less than one whenever the treatment effect for R- patients is no greater than the treatment effect for R+ patients. Consequently, the obtained formula of efficiency (see equation 2.3) generalizes that established in Maitournam and Simon (2004) and formulas (2.4 and 2.5) can allow one to make several scenarios in order to determine the appropriate sample size for targeted clinical trials designs in function of the treatment effect and the prevalence of the two genotypes in the overall population by varying the parameters particularly r, δ or Δ .

5. CONCLUSION

The development here makes it clear that targeted clinical trials are more efficient than the conventional untargeted clinical trials. The expression of efficiency for targeted clinical trials that we have established in the case of distinct standard deviations between subgroups proves once again this decisive advantage of the former design over conventional untargeted clinical trials in which the efficiency is blurred by molecular heterogeneity. Targeted clinical trials are an important step toward stratified and personalized medicines which will bring huge opportunities for mathematical modelling [See Maitournam (2011)].

REFERENCES

A. Maitournam, R. Simon, 2004. On the efficiency of targeted clinical trials. *Statistics In Medicine.*, **24** (3), pp-pp 329-339
 R. Simon and A. Maitournam, 2004. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clinical Cancer Research.*, **10**, 6759-6763
 R. Simon, 2009. Advances in Clinical Trial Designs for Predictive Biomarker Discovery and Validation. *Current Breast Cancer Reports.* **1**, 216-221
 A. Maitournam, 2011. Statistics, Targeted Clinical Trials, Stratified and Personalized Medicines. In : *International Encyclopedia of Statistical Science.* Miodrag Lovric (ed.). Springer-Verlag Berlin Heidelberg.
 K. Pearson, 1894. Contribution to the mathematical theory of evolution. *Phil. Trans. Roy.* **185**, 77-110.
 M.M. Desu and D. Raghavarao, 1990. *Sample size methodology.* Academic Press, Boston.
 J. Adcock, 1997. Sample size determination: a review. *The statistician*, **46** (2), 261-263.
 R.A. Betensky, D.N. and J.G. and Cairncross, 2002. Influence of unrecognized molecular heterogeneity on randomized trials. *Journal of Clinical Oncology.*, **20** (10), 2495-2499.