Optimal Ethical Balance for Phase III Trials Planning

Bilancio Etico Ottimale nella Pianificazione della Fase III delle Prove Cliniche

Lucio De Capitani and Daniele De Martini

Abstract The need of an ethical evaluation is mandatory for every clinical trial, Ethics Committees have been built for that. The distinction between individual and collective ethics has been introduced in a seminal work by Lellouch and Schwartz in 1971, where individual ethics regard concerns related to the patients enrolled in the trial, and collective ethics those of the patients not enrolled who would benefit of a positive trial result. In this paper, a metrization of individual and collective ethics is proposed in order to evaluate their balance in a confirmatory clinical trial. The ethical balance evaluation, among these two aspects of ethics, can be performed before trial starting in order to address sample size determination. The metrization is based, among other parameters, on the drug effect size, on the quality of life of patients under therapy or placebo, and of that induced by adverse reactions. Some numerical examples show that the optimal ethical balance can be provided by sample sizes far from those computed by adopting the usual paradigm based on the prefixed power of 80-90%.

Abstract *La valutazione etica nelle prove cliniche è assolutamente necessaria, e infatti ogni studio clinico viene sottoposto ad adeguato comitato etico. La distinzione tra etica individuale e collettiva è stata introdotta in un lavoro pionieristico da Lellouch e Schwartz nel 1971, in cui i possibili danni subiti dai pazienti coinvolti nello studio vengono valutati dall'etica individuale, mentre l'etica collettiva considera quelli dei pazienti non coinvolti, che beneficerebbero di un positivo risultato dello studio clinico. In questo lavoro, si propone una quantificazione dell'etica individuale e collettiva al fine di valutare il bilancio etico in uno studio clinico confermativo. La valutazione del bilancio tra questi due aspetti di etica può essere eseguita prima dell'inizio dello studio clinico al fine di calcolare la dimensione campionaria*

Lucio De Capitani
Dipartimento di Statistica e Metodi Quantitativi, Universitá degli Studi di Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milano, e-mail: lucio.decapitani1@unimib.it

Daniele De Martini
Dipartimento di Statistica e Metodi Quantitativi, Universitá degli Studi di Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milano, e-mail: daniele.demartini@unimib.it

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

ottimale. La proposta quantificazione dell'etica dipende dall'entità dell'effetto del farmaco, dalla qualità della vita dei pazienti in terapia o placebo, e da quella indotta dai possibili effetti indesiderati. Attraverso degli esempi numerici si mostra che il bilancio etico ottimale pu essere raggiunto in corrispondenza di dimensioni del campione lontane da quelle calcolate adottando il paradigma usuale, in cui la numerosit campionaria viene scelta al fine di assicurare il raggiungimento di valori di potenza del test pari all'80-90%.

Key words: Individual Ethics, Collective Ethics, Global Ethics, Assurance

1 Introduction

It is a common habit to prefix the power of phase III clinical trials at 80-90%. These power settings are suggested by the most relevant books on clinical trials methodology (see, for example, [8] or [2]). Unfortunately, the practical choice of power is seldom motivated as to concern its clinical, and therefore ethical, impact.

In fact, on the one hand 80% looks high enough to guarantee that if the new treatment is effective the trial will succeed - and the drug will be available for the ill population, with high probability. On the other hand, the power is often set not higher than 90% in order to minimize, for the enrolled sample of patients, risks and wastings due to potential harm and lack/or of efficacy of the new treatment. In other words, the power threshold of 80-90% seems accomplishing the need to affect on medical practices, being the drug effective, and that of preserve the safety of enrolled patients.

These two concerns remind to the concept of individual and collective ethics, introduced by [7]. Originally, collective ethics (CE) concerned maximizing total group benefit, and individual ethics (IE) concerned maximizing the benefit of each person to be treated.

However, it is not clear when adopting either 80% or 90%, and why. FDA and [12] encourage to set the power at 90%, but not for explicit ethical reasons: they argue to adopt this power to decrease the rate of unsuccessful trials. In fact, in the literature there are not precise indications about which power threshold should be adopted in different ethical situations.

In the seminal paper [7] the concepts of IE and CE were introduced, together with some mathematical formulations of them under both fixed and sequential designs. [6] provide a critical history of IE and CE, and remark that "very little follow-up research in the lineage of the 1971 paper considers mathematical models".

The aim of this work is that of providing a model to quantify the ethical balance in fixed designs, which are the most widely adopted in phase II and phase III trials (see www.clinicaltrials.gov). Our perspective is in agreement with the view on the ethic of the trial proposed in [11], which "is dictated by the type of evidence sought and by balancing various costs of aggregate harm and benefit" (see [6]).

In [4] it is recalled that Nuremberg Codes point 2 offers strong support of the existing connection between power and ethics, and they add: "the underlying principle is that for any given outlay of human risk or resources, there is an obligation to maximize the power and efficiency of the experimental design". Here, we invert the point of view, by modeling ethics as a function of the sample size, and so of the statistical power. Then, we suggest to adopt the sample size that maximizes experimental ethics.

2 Theoretical framework

A two-arm parallel design with balanced sampling is considered for the phase III trial. A sample of size n is collected for each arm (i.e. new drug and standard treatment/placebo). The true, and unknown, proportions of healing are p_t and p_c. The statistical hypotheses are $H_0 : p_t = p_c$ vs $H_1 : p_t > p_c$, and α is the type I error probability. We assume that the proportions represent the responder rate under the two arms.

Given that $\hat{p}_{t,n}$ and $\hat{p}_{c,n}$ are the sample proportions, the test statistic is $T_n = (\hat{p}_{t,n} - \hat{p}_{c,n})/\sqrt{(\hat{p}_{t,n}(1-\hat{p}_{t,n}) + \hat{p}_{c,n}(1-\hat{p}_{c,n}))/2}$. The power function $\pi(n) = P(T_n > z_{1-\alpha})$ is approximated by $\Phi(ES\sqrt{n/2} - z_{1-\alpha})$, where ES is the standardized effect size: $ES = (p_t - p_c)/\sqrt{(p_t(1-p_t) + p_c(1-p_c))/2}$.

3 Modeling ethics

Ethics is the sum of individual and collective ethical contributions. Usually, individual ethics concern the sample of patients involved in the experiment, where collective ethics consider the remaining population.

Basically, ethics are computed through Benefit/risk indicators times the duration of periods of interest. Quality of Life measures (QoL) are adopted together with the probabilities of responders and of harms, which are related to the "Number Needed to Treat" (NNT) and "Number Needed to Harm" (NNH), all being classical benefit/risk indexes (see, for example, [5]). In particular, the ethical contribution of a group is given by the group size times the quality of life indicator times the duration of the period where such a QoL persists. Since the new treatment is available to the population if the trial succeeds, collective ethics are also multiplied by the power of the experiment. Thus, in general, Ethics are:

$$E = \text{group size} \times \text{probability} \times \text{quality of life} \times \text{duration}$$

Individual ethics (IE) concerns ethics of the population sample enrolled in the trial, and considers the placebo group just during the trial. IE depends on: the size of each sample (n), the rate of responder under new therapy (p_t), the harm probability under new therapy (hp_t), the rate of responder under control treatment (p_c), the quality of life during the disease and before treatments (QoL_d), the benefit (quality of life)

after the new treatment (QoL_t), the harm (risk) during the new treatment (QoL_r), the benefit after control treatment (QoL_c), life expectancy (D_L), duration of the therapy (D_{th}) and accrual rate (A_r, which is assumed to be uniform during enrollment). The duration of phase III trial is: $D_{P3} = 2n/A_r + D_{th}$.

Under the new treatment, the ethical contribution of (eventual) QoL improvement of responders and non responders is the sum of the two, resulting:

$$IE_T(n) = n \times \Big(p_t \times (QoL_t \times (D_L - (D_{P3}(n) - D_{th})/2) + \\ QoL_d \times (D_{P3}(n) - D_{th})/2) + (1 - p_t) \times QoL_d \times D_L \Big) \ .$$

Note that the duration of the benefit of responders is given by life expectancy minus the average of the time elapsed from the beginning of the trial and the end of the therapy. For non responders, the quality of life remains that of the disease during all life.

The ethical contribution of harm due to the new drug is:

$$IE_{TH}(n) = n \times hp_t \times QoL_r \times D_{th} \ .$$

Under the placebo control there is no harm. The (eventual) QoL improvement of responder and non responder is evaluated just during the trial (at the end of phase III this group could be treated with the new therapy if the trial succeeds) and it results:

$$IE_C(n) = n \times (p_c \times (QoL_c \times D_{th} + QoL_d \times (D_{P3}(n) - D_{th})) + \\ +(1 - p_c) \times QoL_d \times D_{P3}(n)) \ .$$

Finally, IE results: $IE(n) = IE_T(n) + IE_{TH}(n) + IE_C(n)$.

Collective ethics (CE) concerns ethics of the population not involved in the trial, and that enrolled in the trial under the control treatment once the trial has been completed. Besides the quantities already introduced to define IE, CE depends on: the population size (N), the incidence in the illness ($Prev_i$), the power of the experiment $\pi(n)$. The size of the ill population is $N_{ill} = N \times Prev_i - n$, that is the ill population minus the group that tested the new treatment in the trial.

First, the "during trial" ethical balance of the population not involved in the experiment is:

$$CE_{DT}(n) = (N \times Prev_i - 2n) \times QoL_d \times D_{P3} \ .$$

When the trial succeeds, the ethical contribution related to the benefit of the population due to treatment, for responder and non responder, is:

$$CE_{ST}(n) = N_{ill} \times (p_t \times QoL_t + (1 - p_t) \times QoL_d) \times (D_L - D_{P3}(n) - D_{th}) \times \pi(n) \ .$$

Note that the duration of the quality of life (benefit or not) is given by life expectancy minus the duration of the trial minus that of the therapy. In other words, it is assumed that the ill population adopts the new treatment whenever it is available, that is, just after the end of the successful trial.

The ethical contribution of harm due to the new drug in the ill population also accounts for the power of the experiment, resulting:

$$CE_{TH}(n) = N_{ill} \times hp_t \times QoL_r \times D_{th} \times \pi(n) \ .$$

In case the trial is unsuccessful the harm is due to the loss of benefit, and the quality of life of ill popuplation remains the same:

$$CE_{UT}(n) = N_{ill} \times QoL_d \times (D_L - D_{P3}(n)) \times (1 - \pi(n)) \ .$$

The total collective ethical contribution results:

$$CE(n) = CE_{DT}(n) + CE_{ST}(n) + CE_{TH}(n) + CE_{UT}(n) \ .$$

The global ethical contribution of the experiment is given by the sum of individual and collective ethics:

$$GE(n) = IE(n) + CE(n) \ . \tag{1}$$

It is of interest to compute the sample size providing the best ethical balance, and then to account for the power this optimal sample size provides, which could not be in the range of the classical range of power adopted for planning phase III trials, i.e. $[80\%, 90\%]$.

4 Examples

Two numerical examples are reported here, based on the ethical model in (1).

Let us consider first a situation where the new drug works well, and where side effects are low. We expect that the power at the phase III sample size giving the best ethical balance is high.

We set the parameters as follows: $\alpha = 2.5\%$, $p_t = 0.5$, $p_c = 0.1$, $Prev_i = 10\%$, $N = 1M$, $hp_t = 0.05$, $QoL_d = -2$, $QoL_t = 5$, $QoL_r = -5$, $QoL_c = 0.5$, $D_L = 20$, $D_{th} = 0.5$, $A_r = 200$. In this situation, if α, p_t and p_c were considered only, standard sample size computation would give group samples of size 17 to achieve a power of 80%, and of size 23 with power 90%. However, the sample size giving the optimal Ethical Balance, that is providing the maximum of $GE(n)$, is $argmax(GE(n)) = n_{opt} = 55$, per group. The subsequent power is: $\pi(55) = 0.9956$, meaning that the optimal power is quite higher than those usually adopted, viz. 80-90%.

Now, consider a situation where the effect of the new drug is just moderate and the side effects are quite remarkable. Some of the above parameters are modified as follows: $p_t = 0.3$, $Prev_i = 0.1\%$, $hp_t = 0.5$, $QoL_t = 3$, $QoL_r = -10$, $A_r = 50$. In this second situation, standard sample sizes per group would be of 59 and 79, with prefixed power of 80% and 90%, respectively. The optimal Ethical Balance is obtained here with samples of size $argmax(GE(n)) = n_{opt} = 50$. The power function has now changed, since $p_t = 0.3$. Consequently, the optimal power under the ethical

perspective is $\pi(50) = 0.7054$. This means that when the effect size is low and there are considerable side effects, the optimal power can be quite lower than 80-90%.

Often, the information on the parameters involved in the Ethical Balance is weak. To account for the possible deviation between the prefixed values of the parameters and their true value, a statistical distributions on parameters is usually introduced: this technique is called assurance (see [9]). We performed a sensitivity analysis of model (1) with assurance obtaining that, also in this case, the optimal power can be quite lower than 80-90%. These results are not reported here for the sake of brevity.

5 Conclusions

We have shown, through the model on individual and collective ethics we introduced, and a couples of appropriate examples, that defining the power within the range 80-90% can lead to poor choices in terms of the impact a new drug might have on the ill population.

To conclude, we would like to merge ethical models, like the one here introduced, and economical models, such as those developed in [10], [1], and in [3], since we believe that even profit and cost have an ethical impact on our society.

References

1. Antonijevic Z., Kimber M., Manner D., Burman C-F., Pinheiro J., and Bergenheim K.: Optimizing Drug Development Programs: Type 2 Diabetes Case Study. Therapeutic Innovation & Reg.Sci. **47**(3), 363–374 (2013)
2. Chow S.C., Shao J., Wang H.: Sample Size Calculations in Clinical Research. Chapman and Hall/CRC, 2nd ed. Boca Raton (2008)
3. De Martini D.: Profit Evaluations when Adaptation by Design is Applied. Therapeutic Innovation and Regulatory Science, **50**(2), 213–220 (2016)
4. Gelfond J.A., Heitman E., Pollock B.H., Klugman C.H.: Power, ethics, and obligation. Stat. Med. **31**(29), 4140–4141 (2012)
5. Guo J.J., Pandey S., Doyle J., Bian B., Lis Y., Raisch D.W.: A Review of Quantitative RiskBenet Methodologies for Assessing Drug Safety and EfcacyReport of the ISPOR RiskBenet Management Working Group. Value in Health **13**(5), 657–666 (2010)
6. Heilig C.M., Weijer C.: A critical history of individual and collective ethics in the lineage of Lellouch and Schwartz. Clinical Trails **2**, 244–253 (2005)
7. Lellouch J., Schwartz D.: L'essai therapeutique: ethique individuelle ou ethique collective? Revue de l'Institut International de Statistique **39** 127–136 (1971)
8. Meinert C.L.: Clinical Trials - Design, Conduct and Analysis. Oxford University Press, New York, 1986.
9. O'Hagan A., Stevens J.W., Campbell M.J.: Assurance in clinical trial design. Pharmaceutical Statistics, **4**, 187–201 (2005)
10. Patel N., Bolognese J., Chuang-Stein C., Hewitt D., Gammaitoni A., Pinheiro J.: Designing PhII Trials Based on Program-Level Considerations: A Case Study for Neuropathic Pain. Drug Inf.J. **46**(4), 439–454 (2012)
11. Schwartz D., Flamant R., Lellouch J.: Clinical Trials. Academic Press, New York, 1980.
12. Wang S.J., Hung H.M.J., O'Neill R.T.: Adapting the sample size planning of a phase III trial based on phase II data. Pharmaceutical Statistics **5**, 85–97 (2006)