ORIGINAL PAPER

BIOETHICS APPROACH OF BIOSTATISTICS IN CLINICAL TRIALS. AVOID THE USE OF EXCESSIVE OR INADEQUATE NUMBERS OF RESEARCH SUBJECTS

MIHAI MANOLACHE[1,2] , EMIN CADAR[3], DINU ANTONESCU[4], CONSTANTIN MIRCIOIU[3], IRINA PRASACU[5*], ROXANA SANDULOVICI[6]

Manuscript received: 15.11.2017; Accepted paper: 23.02.2018; Published online: 30.03.2018.

Abstract. *Assurance of a statistical and clinical significance to results of a clinical trial in the era of evidence based medicine is a complex analysis starting from establishing of end-points, hypothesis to be verified, risks α and β of type I and type II errors, and expected improvements of effects, measured by difference - Δ in selected endpoints.*

Ethic – nonethic in calculation of the number of subjects was discussed until now in context of "statistical risks" and in context of certitudes of conclusions from the point of view of inovator company, of investigator and finally of regulatory authorities. Patient is never entering in formulas. Paper put in evidence that critical factor in determining the size of experimental lots is the difference Δ, expected additional effect of the new treatment. Or, this factor is established mainly by Inovator Company and principal investigator and imply less the statisticians. So that long time passionate dispute between statisticians concernig the sample size remain mainly an academic problem. Actual decisions are established by company financial resources and regulatory authorities' rigid rules. Final conclusion is that the essential ethical aspect, connected with the risk of patients is ignored by all stakeholders of clinical trials.

Keywords: *calculation of the number of subjects, size of experimental lots, ethical aspect, risk of patients*

1. INTRODUCTION

Regulation (EU) no 536/2014 of the European Parliament and of the Council of 16 april 2014, establishing the rules for clinical trials beginning with 2018, starts with the assertion that **"**A clinical trial may be conducted only if: (a) the rights, safety, dignity and well-being of subjects are protected and prevail over all other interests; and (b) it is designed

[1] "Carol Davila" University of Medicine and Pharmacy, Doctoral School, 030167 Bucharest, Romania.
[2] CEBIS International, 060016 Bucharest, Romania.
[3] Ovidius University of Constanta, Department Pharmaceutical Science, 900527 Constanta, Romania.
E-mail: emincadar@gmail.com.
[4] Academy of Medical Sciences, National Commission Bioethics of Drugs & Medical Devices, 020125 Bucharest, Romania. E-mail constantin.mircioiu@yahoo.com.
[5] "Carol Davila" University of Medicine and Pharmacy, Department of Biostatistics, 030167 Bucharest, Romania. *Corresponding Author: irinaprasacu@yahoo.com.
[6] "Titu Maiorescu" University, Department of Biostatistics, 004051 Bucharest, Romania.
E-mail: roxana.sandulovici@yahoo.com.

to generate reliable and robust data". The second part of the fundamental rule referrs mainly to statistical and clinical significance of the data obtained in clinical trial.

Evidenced based medicine (EBM) is a desire old like medicine itself but the mathematical bases of the concept were established only some fifty years ago starting with book of American physician and mathematician Alvan Feinstein [1] and a book of statistician Archie Cochrane' [2]. Starting from Cochrane ideas it was established an international network for efficacy assessment in medicine – the *Cochrane Collaboration* which contributed decisivly to foundation and institutionalization of the EBM rules. The most important reason for practicing EBM is to improve quality of care through the identification and promotion of practices that effectively assure the accuracy and precision of diagnostic tests and the elimination of those that are ineffective or harmful. The concept was further extended to evidence-informed healthcare or evidence-based health care in diagnostic, treatment and quality of life evaluations [3-5].

Statistical techniques which are essential in assurance of the robustness of tests and reliability of conclusions include [6] as a first chapter the estimation of the number of sujects to be enrolled, sufficient to detect an estimated difference between treatment arms.

But, since a "sufficient number" could tend to infinite, it appears also the necessity of an ethical restriction of this number. Ethical aspects of statistical analysis plan for the protocol of a clinical study was a continuous concern of biostatisticians and their organizations even from the early beginning of clinical studies. Particularly "Avoid the use of excessive or inadequate numbers of research subjects for study size" was a main regard of the American Statistical Asociation Ethical Guidelines for Statistical Practice [7] as well as of The United Nations Statistical Commission [8].

The paper presents the statistical methods for estimation of the required number of patients as function of the objectives and type of the study, but also in conditions of ethical restrictions.

2. ESTIMATION OF THE NUMBER OF SUBJECTS IN CLINICAL TRIALS

The **fundamental contradiction** appears even in the mentioned definition of acceptable clinical trial. Safety, dignity and well-being of subjects imply in extremis a zero number of subjects. On the contrary, to generate reliable and robust data is a tendency toward to an infinite number of subjects. So that, between the regulatory bodies and biostatisticians on one part and ethical committees on the other part, is an irreducible contradiction.

Clinical trials are usual run on groups of subjects receiving different treatments. Estimation of the dimensions of group sizes is performed in practice using some mathematical formulas or, more "blind", some softwares, but the problem implies essentially in the same time medical, statistical and ethical aspects. Application of formulas requires preliminary data concerning the size of clinical significant difference, hypothesis, assumed risck of type I and type 2 errors, as well as estimation of variability of measured endpoint parameters.

In the beginning is the choice of the primary endpoint, which choice is further determinant for protocol and analysis of results. If this endpoint is a continuous or discrete distributed variable, if it is normal or binomial distributed, is first of all information which have to be provided to biostatistician.

From statistical reasons hypothesis to be verified (H_0) is usually the hypothesis of "equality", i.e. of non additional effect of tested treatment. Consequently the selected groups can be considered as coming from the same population which further allows pooling together the groups in order to better estimate the parameters of the entire population.

From ethical reasons, H_0 is the "guilty" hypothesis (presumption) : "effect of drug is equal to placebo", "effect is not superior to a standard treatment", "risk is greater" etc, . If the result of the analysis of study data does not allow the rejection of H_0 hypothesis, the new treatment is not accepted. In bioequivalence studies hypothesis H_0 is that drugs are not bioequivalent. Alternative hypothesis H_a is, as a rule, superiority of the tested treatment.

Risc of rejecting H_0 is the risk of rejecting the equality and accepting "superiority of the new treatment". If superiority is not true the patient pays for the error.
Usual notations are:

- probability of type I error: $\alpha = \Pr(reject\ H_0 / H_0\ true)$
- probability of type II error $\beta = \Pr(accept\ H_0 / H_0\ false)$
- $\pi = 1 - \beta$, called "power of the test" is the probability of rejecting a false hypothesis:

$$\pi = \Pr(reject\ H_0 / H_0\ false) \qquad \pi = 1 - \beta$$

β and consequently are at the choice of producer being its risk.

Decrease of the number of subjects can be made only by increasing the risk of producer. Risk of patient is usually limited by guidances. For example in bioequivalence studies $\alpha = 0.10$.

If there are compared the frequencies of an endpoint in placebo (p_1) and in treatment (p_2) arms, the number of subjects to be included $n_{pl} = n_T = n$ equals

$$n = \frac{\left[z_{\frac{\alpha}{2}} \sqrt{2p(1-p)} + z_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2}$$

where

$p = \frac{p_1 + p_2}{2}$, $z_{\frac{\alpha}{2}}$ and z_β are quantiles associated to normal standard distribution

$$P(Z < z_{\frac{\alpha}{2}}) = z_{\frac{\alpha}{2}} \text{ and } P(Z < z_\beta) = z_\beta$$

A "more ethical approach" is to take the placebo group only of fraction of the tested drug $n_T = n$ si $n_{PL} = kn_T = kn$ with k < 1 :

$$n = \frac{\left(z_{\frac{\alpha}{2}} \sqrt{\hat{p}\hat{q}(k+1)} + z_\beta \sqrt{\hat{p}_1\hat{q}_1 k + \hat{p}_2\hat{q}_2} \right)^2}{\Delta^2 k}$$

Unfortunately this approach is not enough studied and applied, since placebo is considered a treatment without adverse effects. But placebo group of patients are privated of treatment, which implies in most cases unacceptable risks connected with evolution of disease.

Practical example. Let us consider as practical example the estimation of number of subjects to prove an increased overall survival in first year after myocardial infarction following the aspirin treatment. Calculus is retrospectiv, starting from data obtained in one [10] of the numerous clinical studies verifying the platelet antiaggregant effect of low doses of aspirine.

Starting data:

- 10% rate of mortality in placebo group during the first year (p1),
- 20 % expected reduction of mortality in treatment arm,
- $\alpha = 0,05$ and a power of 0.9 (β=0,1)

So that we consider the reduction $10\% \xrightarrow{20\%} 8\%$

Putting into a formula proposed [11] for the sample size in case of comparison of two proportions:

$$n = \frac{\left[z_{\frac{\alpha}{2}} \sqrt{2p(1-p)} + z_{\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2}$$

the data :

$$p_1 = 0,10\ ,\ p_2 = 0,08\ ,\ and\ \ p = \frac{10\% + 8\%}{2} = \frac{18\%}{2} = 9\% \Rightarrow p = 0,09$$

$$\Delta = 0,10 - 0,08 = 0,02$$

it obtains the result

$$n = \left(\frac{1.96\sqrt{2 * 0.09 * 0.91} + 1,28\sqrt{0,1 * 0,9 + 0,08 * 0,92}}{0,02} \right)^2 \cong 4.300$$

It is to note that the critical parameter is the “clinical significance difference” which is tested $\Delta = p_1 - p_2\ \ 1/(p_1 - p_2)^2 = 1/0.02^2 = 2500$.

Since the proving of formula for the number of subjects is derived from the definition of power of the statistical tests (see Annex 1), the clinical trials (CT) with insufficient number of subjects are called “low power CT”.

3. OPPINIONS CONCERNING ETHICS OF LOW POWER CLINICAL TRIALS

Some forty years ago Chalmers and coworkers published a paper [12] which induced the idea that “negative trials”, i.e. trials where the power for detecting 25 % or 50 % therapeutic improvement is lower than 0.90, are unethical. Many of the therapies labeled as "no different from control" in clinical trials using inadequate samples have not received a fair rating. Since the paper was cited by 2016 at least 1600 times, this conclusion was widely accepted by biostatisticians and even some ethical committees.

Many clinicians and regulatory authorities regarded underpowered trials as unethical since these don’t offer a sufficient certitude concerning the presence of effect.

The desire to provide most precise information push people less informed on fundamental limits of mathematics to consider that approximations and estimations are bad science and consequently this type of research itself is unethic. In the same years Altman stated, “A study with a sample too small will be unable to detect clinically important effects. Such a study might, thus, be scientifically useless and hence unethical in its use of subjects and resources” [13].

Chalmers changed later his mind and considered that his paper from 1978 exerted a negative effect on the valuation of clinical studies. A very agresive paper against

underpowered CT was published later [14] which blamed the continuing unethical conduct of underpowered clinical trials. Other authors argued that underpowered" trials are not necessarily unethical [15] and the overemphasis on power analysis is not useful [16, 17, 23, 24].

In a more technical approach, Bachetti et all [18] concluded that "the average projected burden per participant remains constant as the sample size increases, but the projected study value does not increase as rapidly as the sample size … and smaller studies therefore have more favorable ratios of projected value to participant burden".

It is a false conclusion that an effect not demonstrated is an effect not present. Reports with low power may be better than no report.

Edwards and coworkers invoked a practical argument. Low power studies can be combined as "replicates" in meta-analysis in integrated studies with sufficient greater power than individual studies [19].

Beyond the scientific discussions about power, in case of rare diseases it is clear that there is no chance to increase the number of subjects.

4. OPPINIONS CONCERNING ETHICS OF OVERPOWERED CLINICAL TRIALS

Keeping in mind that the first concern in CT is that the rights, safety, dignity and well-being of subjects are protected and prevail over all other interests it is to observe that the safety of ***Overpowered*** studies implies risk to more subjects and additionally imply waste resources. When human or animal subjects are involved, running an overpowered study can be considered unethical.

Beyond the scientific dispute, the problem becomes acute and burning in case of impossibility to accrue the recruitment of patients in order to obtain the desired power, leading to premature closing of CTs. In a recent study it was shown that in 2011, a number of 481 studies terminated for failed accrual of expected enrolment [20],

5. RELATIVITY OF POWER ESTIMATION

In fact the power is estimated starting from the size of effect "Δ" which is not known apriori. As was observed in our example, for detection of a decrease of mortality from 10 to 8 %, the inverse of square of Δ introduced a factor of 2500. If the difference would be only 1%, the factor would be 10.000. If in a more optimistic approach we make calculus with 3 %, we obtain some 1100. So that effect of delta, as can be seen from the Fig. 1 is really critical.

Consequently the project is underpowered or overpowered, depending on our hopes it concerns the effect of the treatment.

Kenneth and Greems [21] disclosed that investigators beyond the "mandatory and mistical approach" sometimes perform a "sample size samba" to achieve adequate power, starting from the number of subjects and money available, for obtaining appropriate delta to introduce in formulas. This could be considered a trick but represents an acceptable solution to a real problem, an alternative to making nothing.

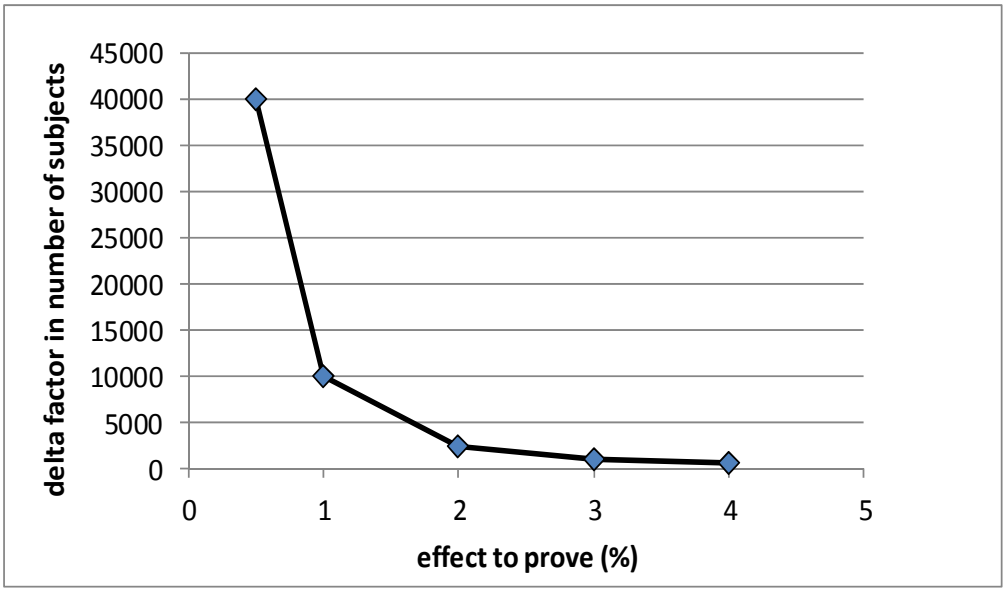

Figure 1. Dependence of the number of subjects on the expected size of difference in effects of treatments.

Regulatory approach American Statistical Association Ethical Guidelines for Statistical Practice (Approved by the Board of Directors, August 7, 1999) [7] asks "avoid the use of excessive or inadequate numbers of research subjects by making informed recommendations for study size". Additionally it is recommended to present to sponsors with choices are among valid alternative statistical approaches that may vary in scope, cost, or precision.

Unfortunately, in 2015 and 2016 approach, EMA groups for elaboration of rules for aprovals by ethics committees of CT, in proposals for a standardised document ignored completely the ethical aspects in projecting of sample sizes. Actually, it seems that many Ethics Committes are not evaluating the appropriateness of the calculus of the sample sizes of CTs.

In an evaluation of the information regarding sample size determinations in protocols submitted to UK research ethics committees published by Brittish Medical Journal in year 2009 [22] it was found that overall, only 42%, mainly coming from non-commercial sponsors, of protocols reported all of the information to accurately recalculate the sample size. Study size tended to be over-estimated rather than under-estimated.

6. CONCLUSIONS

Regulatory bodies and even ethical committeess ask the justification of the number of subjects but are less concerned on ethical aspects of the problem.

Biostatisticians were much concerned on this subjects which are analyzed by "ethical statistical codes". Usual approach was performed in terms of overpowered or underpowered CT. Long time, low numbers of research subjects were considered as non-ethical approaches following a low scientific value. More recently, following valuation of small studies by metanalysis, the oppinions on this subject became less and less aggressive and opposed conclusions being heared.

Since the models and formula for calculus use entering data concerning the size of clinical significant difference, hypothesis, assumed risck of type I and type 2 errors, as well as

estimation of variability of measured parameters calculus is equally problem of investigators and biostatisticians.

Paper put in evidence that the critical factor determining the number of subjects is the size of efffect $-\Delta$, which is desired to be proved in CT. A decrease of Δ from 3 % to 1 % leads to an increase of the number from thousands to tens of thousands. Choice of Δ determines practically all components of the CT and this is practically made outside statistical team.

It appears that there is a lack of cooperation between regulatory bodies, sponsors, investigators, biostatisticians and ethics committees for scientific based bioethics approach in clinical studies.

Last but not least all these approaches are connected with "reliable and robust data" safety of patients missing completely.

REFERENCES

[1] Alvan Feinstein, A.R., *Clinical Judgment*, Baltimore, Williams and Wilkins Co., 1967.
[2] Cochrane, A.L., *Effectiveness & Efficiency: Random Reflections on Health Service*s, CRC Press, New Ed edition CRC Press, New Ed edition, 1999.
[3] Greenhalgh, T., *How to Read a Paper: The Basics of Evidence-Based Medicine*, 4th Ed., John Wiley & Sons, 2010.
[4] Steven N.G., *Annals of Internal Medicine*, **130**(12), 995, 1999.
[5] Sackett, D. L., *Evidence-based Medicine. Encyclopedia of Biostatistics*, 3 UK Cochrane Centre, Oxford, UK, 2005.
[6] The International Statistical Institute's, *Declaration on Professional Ethics*, 1985.
[7] American Statistical Asociation, *Ethical Guidelines for Statistical Practice*, Approved by the Board of Directors, 1999.
[8] The United Nations Statistical Commission's, *Fundamental Principles of Official Statistics,* 1994.
[9] Hubert, L., Wainer, H., *Ethical aspects of statistical analysis plan for the protocol of clinical studies - A Statistical Guide for the Ethically Perplexed*, CRC Press Boca Raton, 2013.
[10] Yusuf, S., Wittes, J., Friedman, L., *Journal of the American Medical Association,* **260**(15), 2259, 1988.
[11] Chow, S.C., Shao, J. *Statistics in drug research*, CRC Press, Wang, H., Chow, S.C. *Sample Size for Comparing Proportions, in ethods and Applications of Statistics in Clinical Trials* 653, Duke Scholars, J Wiley, 2014.
[12] Freiman, J.A., Chalmers, T.C., Smith, H. Jr., Kuebler, R.R., *New England Journal of Medicine*, **299**, 690, 1978.
[13] Altman, D.G., *British Medical Journal*, **281**, 1336, 1980.
[14] Halpern, S.D., Karlawish, J.H., Berlin, J.A., *Journal of the American Medical Association*, **288**(3), 358, 2002.
[15] Edwards, S.J.L., Lilford, R.J., Braunholtz, D., et al., *Lancet*, **350**, 804, 1997.
[16] Knapp, T.R., *Journal of Nursing Research*, **45**, 379, 1996.
[17] Lilford, R., Stevens, A.J., *British Journal of Surgery*, **89**, 129, 2002.

[18] Bacchetti, P., Wolf, L.E., Segal, M.R., McCulloch, C.E., *American Journal of Epidemiology*, **161**(2), 105, 2005.
[19] Edwards, S.J.L., Lilford, R.J., Braunholtz, D., Jackson, J., *Lancet*, **350**, 804, 1997.
[20] Benjamin Carlisle, M.A., Kimmelman J., Ramsay, T., MacKinnon, N., *Clinical Trials*, **12**(1), 77, 2015.
[21] Schulz, K. F., Grimes, D. A., *Lancet*, **365**, 1348, 2005.
[22] Clark, T., Berger, U., Mansmann, U, *British Medical Journal*, **346**, 1135, 2013.
[23] Vatasescu, A., Enache, F., Mircioiu, C., Miron, D.S., Sandulovici, R., *Farmacia*, **59**(2), 161, 2011.
[24] Enache, F., Mircioiu, I., Corlan, G., Sandulovici, R., Mircioiu, C., *Farmacia*, **60**(2), 227, 2012.