



Reducing Sample Sizes in Phase II Trials based on Exact Binomial Tests by Shifting Design Parameters

Dirk Klingbiel*, Qiyu Li and Martin Bigler

Swiss Group for Clinical Cancer Research (SAKK), Switzerland

*Corresponding author: Dirk Klingbiel, SAKK Koordinationszentrum, Effingerstrasse 40, 3008 Bern, Switzerland, Tel: +41 (0) 31 389 92 05, E-mail: dirk.klingbiel@sakk.ch

Abstract

Background: Currently, a large number of new anticancer agents is accompanied by constrained resources, requiring prioritization. In addition, emerging molecular subtypes subdivide common cancer entities into rare diseases, making it harder to conduct phase II trials as a crucial step in drug development and therapy optimization. We extend a recent work by Khan et al. on the design of phase II clinical trial based on exact binomial tests.

Results/Methods: The effect of flexibility in design parameters on the required sample size and error rates is assessed. Changing p_0 and p_1 allows to reduce sample sizes in cases where altering α and β is not an option. When allowing flexibility in all four parameters, one can further reduce sample sizes.

Conclusion: In phase II trials based on exact binomial tests, minimal deviations from the original design parameters might ultimately reduce costs and trial duration.

Keywords

Phase II trial, Sample size, Type I error, Power

Introduction

Phase II clinical trials are designed to test whether new treatments show enough activity to warrant further investigation. In cancer research, a large number of new anticancer agents is at the same time accompanied by constrained resources, requiring prioritization [1,2]. With the emerging molecular subtypes even the “big” cancer entities become rarer diseases, making it harder to reach accrual targets.

The outcome of phase II trials is often a binary variable measuring response; examples from oncology include patients being progression-free at a given point in time or patients responding to treatment.

For single-arm trials, many designs test the null hypothesis $H_0: p \leq p_0$ that the true response probability p is less than some undesirable probability p_0 versus the alternative hypothesis $H_1: p \geq p_1$ that p is greater than or equal to some desirable probability p_1 . Examples of such single-arm designs are A'Hern's single-stage design [3] or Simon's two-stage design [4] which are still widely used. These and many other designs are based on the binomial distribution. When using A'Hern's design, the null hypothesis is rejected, if r or

less among n accrued patients show a response to the treatment. The values of r and n are chosen such that the type I error α , the probability of rejecting a true null hypothesis, and the type II error β , the probability of not rejecting a false null hypothesis, are not bigger than some pre-specified error boundaries (e.g. $\alpha \leq 0.05$ and $\beta \leq 0.20$).

Since these designs are based on the discrete binomial distribution, the boundaries on the type I and II errors are almost never met exactly. As an example consider a single-stage trial with $p_0=0.1$ and $p_1=0.2$, and specified with $\alpha=0.05$ and $\beta=0.20$. A'Hern [3] provides tables with sample sizes, yielding $r=12$ and $n=78$ (i.e. the null hypothesis is rejected if >12 patients out of 78 respond to treatment). The theoretical type I and II error rates in this example are 0.0453 and 0.1918, respectively.

In a recent article Khan et al. [5] proposed to relax the desired error rates to save patients in rare disease entities or for proof-of-concept studies [5]. We propose to allow flexibility in the other two design parameters, the probabilities p_0 and p_1 , keeping the effect size $\delta=p_0-p_1$ fix. Since a priori definition of these parameters is hard and almost never exact, a small deviation seems acceptable. Of note, our approach of varying p_0 and p_1 can be combined with the method of Khan et al. [5].

Materials and Methods

Throughout this paper we will use the notations given in table 1. Given the (un) promising probabilities p_0 and p_1 , and type I and II errors α and β an original design can be computed. We defined ranges $[p_{0,\min}, p_{0,\max}]$, $[\alpha, \alpha_{\max}]$ and $[\beta, \beta_{\max}]$ for the design parameters p_0 , α and β , respectively, which we considered acceptable, and discretized them using a small number $\tau > 0$, e.g. $\tau=0.001$:

$$P = \{p_{0,\min}, p_{0,\min} + \tau, p_{0,\min} + 2\tau, \dots, p_{0,\max}\} \quad (1)$$

$$A = \{\alpha, \alpha + \tau, \alpha + 2\tau, \dots, \alpha_{\max}\} \quad (2)$$

$$B = \{\beta, \beta + \tau, \beta + 2\tau, \dots, \beta_{\max}\}. \quad (3)$$

For each possible combination of parameters $p_0' \in P$, $p_1' = p_0' + \delta$, $\alpha' \in A$ and $\beta' \in B$ we computed a new design. From these designs, one can choose to minimize the minimal total sample size n , the minimal expected sample size EN (for Simon's optimal two-stage design), or any other criterion.

Citation: Klingbiel D, Li Q, Bigler M (2015) Reducing Sample Sizes in Phase II Trials based on Exact Binomial Tests by Shifting Design Parameters. Int J Cancer Clin Res 2:020. doi:10.23937/2378-3419/2/3/1020

Received: March 28, 2015; **Accepted:** August 18, 2015; **Published:** August 20, 2015

Copyright: © 2015 Klingbiel D. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Table 1: Notations used in the present paper

Notation	Parameter
ρ	(Unknown) true response probability of the treatment
ρ_0, ρ_1	Probabilities that are deemed unpromising (ρ_0) and promising (ρ_1)
δ	Effect size ($\delta = \rho_1 - \rho_0$)
α, β	Type I and II error probability
n, n_1	Total sample size and sample size after the first stage of a two-stage design
r, r_1	Rejection number at the end of a trial and after the first stage of a two-stage design
$B(\cdot; n, p)$	Binomial distribution function with parameters n and p
$\rho_{0,\min}, \rho_{0,\max}$	Lower and upper bound of the acceptable range of ρ_0
$\alpha_{\max}, \beta_{\max}$	Upper bounds for acceptable values of α and β
τ	Step size for the discretized ranges of acceptable values of ρ_0, α and β
EN	Expected sample size of a two-stage design
ϵ	Small value of misspecification of the true ρ_0
n_ϵ, r_ϵ	Sample size and rejection number of a design where ρ_0 is misspecified by ϵ
$\alpha_\epsilon(\rho_0)$	Probability of rejecting $H_0: \rho \leq \rho_0 + \epsilon$, given $\rho = \rho_0$
$\beta_\epsilon(\rho_0)$	Probability of not rejecting $H_0: \rho \leq \rho_0 + \epsilon$, given $\rho = \rho_1$

Table 2: 10 solutions with a total sample size of $n=65$

Solution	ρ_0	ρ_1	α	β	n	r	Exact α	Exact β
1	0.084	0.184	0.050	0.220	65	9	0.044	0.219
2	0.085	0.185	0.050	0.214	65	9	0.047	0.213
3	0.086	0.186	0.051	0.208	65	9	0.050	0.208
4	0.087	0.187	0.054	0.202	65	9	0.054	0.202
5	0.088	0.188	0.058	0.200	65	9	0.057	0.196
6	0.089	0.189	0.061	0.200	65	9	0.061	0.191
7	0.090	0.190	0.065	0.200	65	9	0.065	0.185
8	0.101	0.201	0.061	0.218	65	10	0.060	0.217
9	0.102	0.202	0.064	0.212	65	10	0.064	0.211
10	0.103	0.203	0.068	0.206	65	10	0.067	0.206

All results were obtained using R 3.0.0 [6] and the clinfun package version 1.0.5 [7].

Results

Example: single-stage trial Khan et al. [5] used the example of $\rho_0=0.1, \rho_1=0.2, \alpha=0.05$ and $\beta=0.2$, which yields a single-stage design with a sample size of $n=78$, and allowed a flexibility of +3% for the type I and II error rates (i.e. $\alpha \leq 0.08$ and $\beta \leq 0.23$). They found a design with only 65 patients and exact error rates $\alpha=0.0567$ and $\beta=0.2229$.

Thus, with α and β which are 0.0067 and 0.0229 higher than the originally desired ones, 13 (17%) patients can be saved in this example.

By, instead of allowing flexibility in α and β , varying the value of ρ_0 between $\rho_{0,\min}=0.09$ and $\rho_{0,\max}=0.11$ we find a solution requiring 70 patients with parameters $\rho_0=0.090$ and $\rho_1=0.190$. The exact values of the type I and II error rates are 0.048 and 0.199, respectively.

If we allow additional flexibility of 0.01 for α and β we find a design with parameters $\rho_0=0.090, \rho_1=0.190, \alpha=0.060$ and $\beta=0.201$ with a total sample size of only 64 patients. The exact type I and II error rates are 0.059 and 0.201, respectively. Thus with only a change of 0.01 for α and β instead of 0.03 but an additional variation of ρ_0 , one additional patient can be saved compared to the approach of Khan et al. [5] Compared to the original design according to A'Hern 14 patients less are required, a saving of 18%. ρ_0 and ρ_1 are 0.010 lower than in the original plan. α is only 0.009 higher than originally required and β only 0.001.

When we allow for 2 percent change in each parameter, the required sample size can be reduced to only 54 patients. The design parameters leading to this sample size are $\rho_0=0.081, \rho_1=0.181, \alpha=0.069$ and $\beta=0.215$. The exact type I and II error rates in this case are 0.068 and 0.215, respectively. The saving of 24 patients (35%) is achieved with ρ_0 and ρ_1 that are 0.019 lower and α and β that are 0.019 and 0.015 higher than desired.

Instead of looking for the solution with the lowest sample size one can also look for other configurations that give some sample

size above the minimal one. Table 2 shows 10 possible designs with a sample size of $n=65$, the minimal sample size of the approach according to Khan et al. [5].

Discussion

Phase II is a crucial phase in drug development and therapy optimization. Randomized designs which directly compare the treatment arms are preferred by some authors (see [1,8], for instance), but not always feasible. Often rare disease entities, or even distinct subtypes of a common disease, and limited resources prevent a comparative randomized design. In clinical trials, a reduction by even a few patients can make a substantial difference in terms of trial cost or duration. In single-arm trials our approach can be employed to achieve such reductions. In randomized phase II trials, a reduced sample size might prevent the exposure of patients to serious adverse effects of new agents, which often show to be ineffective in phase III trials, especially in oncology where only 37.5% of trials are positive [9].

Khan et al. [5] rightfully argue that a deviation from the conventional values for α or for β is useful because phase II trials only provide preliminary evidence and also the common values of $\alpha=5\%$ and $\beta=10\%$ are arbitrary themselves.

Likewise, there is little reason to believe that the "correct" unpromising rate ρ_0 , derived from historical data, is exactly 0.10, for example. Similarly, if a ρ_1 of 0.20 is deemed promising, why should one consider a rate of 0.19- or even 0.21- as not promising? Similarly, one could argue that the effect size could be varied as well. For instance, an increase of 9% in the response rate could be considered clinically meaningful if an increase of 10% is considered as such. In this case, however, a larger sample size will be required to gather evidence for this smaller effect size. And if 10% are deemed promising 11% should be even more so. This would allow further options and reductions. We kept the effect size fix, however, since we would be less comfortable with relaxing effect sizes which in many cases are already rather optimistic.

Authorities or grant providers sometimes do not like to use higher values for α or β in clinical trials. In this case and with limited resources it could be a viable option to search for other designs with smaller sample sizes using a slightly different value for ρ_0 and ρ_1 . The limitations of trading-off type I and type II errors are discussed in detail by Khan et al. [5]. One limitation is that final results should be seen in context of the α level used in the sample size calculation which might be different in there and our setting. Using only arbitrary thresholds for α and p-values, however, seems not sufficient either. In the decision taking to proceed to phase III or not, other factors such as other clinically relevant endpoints, safety, and accrual rates have to be considered.

Another limitation seems more serious to us: reducing the sample size too much might result in a loss of precision, making it difficult to judge treatment effects. Khan et al. [5] recommend to not

reduce sample sizes below 20 patients per treatment arm. We rather recommend to also consider p_0 instead of n only, since 20 patients can give enough precision for very small or large values of p_0 , but for $p_0 \approx 1/2$ even $n=30$ patients will result in a 95%-confidence interval width of 0.35 and more.

A limitation unique to our approach is that the misspecification of response rates has a serious impact on the type I error and power of a trial, as shown by Baey et al. [10] for two-stage designs. Shifting the (un-)promising response rates might deteriorate one or the other. But even if coming from larger phase III trials, there is always uncertainty attached to those parameters. With the proposed approach and an ϵ of one or two per cent, one would always be within the confidence intervals provided from historical data or within a reasonable range of the educated guess which was used to define the design parameters. Of course, there is the danger of increasing the type I or type II error because one moves further away from “the truth”. Since the truth is unknown, however, it could well be that one will end up closer to it.

In conclusion, it seems useful to assess a wider range of scenarios when designing a phase II trial based on exact binomial tests. By doing so, a smaller sample size may be chosen with only minimal deviations from the original design parameters. Ultimately, this could reduce costs and trial duration, which are crucial features of clinical trials. Projects in rare diseases or molecular subtypes of more frequent diseases might become feasible and more treatments might be tested.

Future Perspective

The costs of bringing a new medicine to the market are enormous. Clinical trials account for a good share of that. Phase II trials are a crucial step in the clinical development programs as here for the first time the medicine is given to a larger number of patients and efficacy is assessed to make a go or no-go decision to test the medicine in large, and hence expensive phase III trial.

With the emerging molecular subtypes even the “big” cancer entities become rarer diseases. While breast cancer is quite common it can already be subdivided into at least Luminal A, Luminal B, Triple-negative/basal-like, HER2 type and normal-like subentities. Likely, these subentities can be further subdivided when the understanding of tumor biology advances, driven by technologies like next-generation sequencing. Other cancer types like colon and lung cancer are moving into the same direction. Precision medicine seeks to attack specific molecular targets, often only present in one of those subtypes.

Well-defined patient groups, the target populations of precision medicine, will become smaller. Already now many trials are stopped early because of accrual problems. This problem will aggravate with smaller target populations for clinical trials testing targeted agents. A way out might be to conduct smaller trials, with a higher probability to finish accrual successfully without sacrificing too much power and significance as described in this article.

Executive Summary

- Many phase II trials have binary outcomes.
- Design parameters are somewhat arbitrary.
- Small changes in the design parameters seem acceptable
- Thirty or more percent of patients can be saved by this approach.

References

1. Ratain MJ, Sargent DJ (2009) Optimising the design of phase II oncology trials: the importance of randomisation. *Eur J Cancer* 45: 275-280.
2. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, et al. (2005) Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 23: 7199-7206.
3. A'Hern RP (2001) Sample size tables for exact single-stage phase II designs. *Stat Med* 20: 859-866.
4. Chen K, Shan M (2008) Optimal and minimax three-stage designs for phase II oncology clinical trials. *Contemp Clin Trials* 29: 32-41.
5. Khan I, Sarker SJ, Hackshaw A (2012) Smaller sample sizes for phase II trials based on exact tests with actual error rates by trading-off their nominal levels of significance and power. *Br J Cancer* 107: 1801-1809.
6. Dean CB, Nielsen JD (2007) Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal* 13: 497-512.
7. VE Seshan (2013) Clinfun: Clinical Trial Design and Data Analysis Functions, R package version 1.0.5.
8. Rubinstein L, Leblanc M, Smith MA (2011) More randomization in phase II trials: necessary but not sufficient. *J Natl Cancer Inst* 103: 1075-1077.
9. Gan HK, You B, Pond GR, Chen EX (2012) Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Natl Cancer Inst* 104: 590-598.
10. Baey C, Le Deley MC (2011) Effect of a misspecification of response rates on type I and type II errors, in a phase II Simon design. *Eur J Cancer* 47: 1647-1652.