A portfolio-based approach to optimize proof-of-concept clinical trials

Peer-reviewed author version

Mallinckrodt, Craig H.; MOLENBERGHS, Geert; Persinger, Charles; Ruberg, Stephen J.; Sashegyi, Andreas & Lindborg, Stacy R. (2012) A portfolio-based approach to optimize proof-of-concept clinical trials. In: Journal of Biopharmaceutical Statistics, 22 (3), p. 596-607.

DOI: 10.1080/10543406.2011.564340
Handle: http://hdl.handle.net/1942/14897

A portfolio-based approach to optimize proof-of-concept clinical trials.

C.H. Mallinckrodt, G. Molenberghs, S.J. Rubergh, C. Persiinger, A. Sashegy, S.R. Lindborg

Abstract

Improving proof-of-concept (PoC) studies is a primary lever for improving drug development. Since drug development is often done by institutions that work on multiple drugs simultaneously, the present work focused on optimum choices for rates of false positive (α) and false negative (β) results across a portfolio of PoC studies. Simple examples and a newly derived equation provided conceptual understanding of basic principles regarding optimum choices of α and β in PoC trials. In examples that incorporated realistic development costs and constraints, the levels of α and β that maximized the number of approved drugs and portfolio value varied by scenario. Optimum choices were sensitive to the probability the drug was effective and to the proportion of total investment cost prior to establishing PoC. Results of the present investigation agree with previous research in that it is important to assess optimum levels of α and β. However, the present work also highlighted the need to consider cost structure using realistic input parameters relevant to the question of interest.

Key words: Power, Type I error, Proof-of-concept

Introduction

Only 11% of the drugs that enter phase I testing receive regulatory approval (Kola and Landis, 2004). Much of this attrition comes late in development, after significant investment. For example, the attrition rate in phase II is 62%. Despite this high culling rate, 45% of compounds fail in phase III (Kola and Landis, 2004). Approximately 50% of compound attrition is due to not demonstrating efficacy, which is a 15% increase compared with the previous decade (Hurko and Ryan, 2005). The probability of success in phase II is the most important factor in determining the overall cost per approval of new molecular entities (Paul et al, 2010). And, improving the quality of Proof of Concept (PoC) trials where the efficacy of a drug is first evaluated is an important factor in improving drug development (Gelenberg et al, 2008; Kola and Landis, 2004).

A key aspect of PoC studies is the rate of false positive and false negative results. False negative results in PoC studies may terminate development of effective therapies. False positive results may mistakenly trigger large, expensive, but futile phase III programs. Statistical issues are key considerations regarding false negative and false positive results. Therefore, it is not surprising that many manuscripts and text books have been devoted to the design and analysis of clinical trials. See for example Piantadosi (2005). However, developing a drug involves a series of studies, and optimizing each individual trial in that series does not necessarily optimize the series (Mallinckrodt et al, 2010). Moreover, most drug approvals come from large companies that have many compounds in development (Munos, 2009). Optimizing each individual compound in a portfolio does not necessarily optimize the

portfolio. Thus, when seeking to optimize any individual clinical trial, consideration must be given to individual trial-level factors, project-level (program-level) factors, and portfolio-level (franchise-level) factors (Chen and Beckman, 2009ab).

Chen and Beckman (2009ab) considered cost effective designs for PoC studies in oncology, focus on the rate of false positive (α) and false negative (β) results. The current research also focused on optimum choices of α and β in PoC studies, but the problem was not limited to oncology and development costs included varying levels of fixed costs plus the variable costs due to differing sample sizes whereas the previous research defined efficiency based solely on patient numbers. In considering optimum choices for α and β, it was assumed that the individual trials were designed efficiently, using such things as futility analyses or adaptive designs to deliver the chosen levels of α and β in the most efficient manner. The focus here was on what levels of α and β should be chosen.

In the next section motivating scenarios are used to introduce key questions regarding choice of α and β in PoC studies. Subsequently, basic principles of how α and β influence portfolio outcomes are illustrated using a simple example and with an equation whose derivation arises from similar problems in statistics. With these basic ideas fixed, the impact of α and β on portfolio outcomes is evaluated in realistic situations, incorporating realistic cost structures and investment constraints. Results from these quantitative evaluations are put into context and summarized in the discussion section and concluding remarks.

Motivating Scenarios

To illustrate the need to consider the portfolio as a whole when optimizing choices of α and β in an individual PoC trial consider a portfolio of 100 drugs to be tested in PoC studies. The PoC study is by definition the first chance to evaluate efficacy and the successful development of each drug hinges on making the correct decision in the PoC study. Therefore, intuition suggests that for each individual drug, more power and more rigorous control of Type I error for the PoC study is ideal.

However, the formula for sample size determination $N = 2(Z_{1-\alpha} + Z_{\beta})^2 / \Delta^2$ (Piantadosi, 2005 p 267) shows the diminishing marginal return from increasing sample size. Increasing sample size from 50 to 100 increases power more than increasing sample size from 200 to 250. For an individual compound perspective, at some point, the cost from increasing sample size is prohibitive relative to the gain in α and/or β.

Moreover, resource constraints may dictate that not all 100 drugs can be developed at the level of α and β desired for each compound. If so, would it be better to rigorously test only those drugs that the budget allows, or should α and β be relaxed to ensure all drugs can be tested? It is further interesting to note from the sample size formula that many choices of α and β lead to the same sample size. Put another way, if the budget (i.e., sample size) is fixed, latitude still exists to choose from a wide array of combinations of α and β in order to optimize the trial conditional on cost.

Additional consideration may be given to the disease state and magnitude of unmet medical need. For example, is it appropriate to use the same α and β in PoC trials for drugs to treat erectile dysfunction and pancreatic cancer? If not, in which case is need greater for control of false positive results and in which case is need greater for control of false negative results? Similarly, within a particular disease state where existing treatments are well-established, say major depressive disorder, would it be appropriate to use the same α and β to test a drug with a mechanism of action similar to the approved drugs as for novel, potentially first-in-class treatments?

Basic Principles

The data in Table 1 summarizes how choices of α and β in PoC trials influence portfolio outcomes when there are no constraints and the portfolio is fixed. In this illustration, 100 drugs entering phase II are modeled, with 20/100 or 30/100 of the drugs being effective. The number of true positives and false positives at the end of phase II are calculated based on the number of effective drugs, power, and the false positive rate. The phase III success rate is also calculated, assuming 100% success of true positives in phase III and 0% success of false positives in Phase III. The rate of false positive results is ½ the type I error rate (when using the traditional two-tailed test), and the false negative rate is 1 - power. It is assumed that there is no attrition in phase III such that the number and percent of true positive PoC results is the number and percent of compounds that are successful in phase III and receive regulatory approval. Although some phase III attrition is inevitable due to safety, efficacy, and/or financial considerations, basic principles are best illustrated without additional complicating factors.

Comparing results within blocks of 3 in Table 1 (rows 1-3, 4-6, 7-9) illustrates that portfolio efficiency is influenced by choice of α, with lower rates of false positive PoC studies leading to higher success rates in phase III. Also, the false positive rate has a greater influence on efficiency as the percentage of effective compounds [p(E)] decreases because the need for protection against false positive results increases as the percent of ineffective drugs increases. Comparing the following sets of rows (1, 4, 7) (2, 5, 8) (3, 6, 9) illustrates that portfolio effectiveness is influenced by choice of β, with greater power in PoC studies leading to more phase III successes and launches. Also, the false negative rate has a greater influence on effectiveness as p(E) increases because the need for protection against false negative results increases as the percent of effective drugs increases. Comparing differences in results from the scenarios with 20 vs. 30 effective drugs illustrates that both portfolio efficiency and effectiveness increase when a greater percentage of drugs are effective.

These basic relationships can be quantified by drawing analogy to other optimization problems in statistics. For example, choosing the optimum level of α and β for PoC studies with various levels of p(E) is similar to optimizing sensitivity and specificity of a diagnostic test at varying levels of disease prevalence. In Appendix A this analogy is used to derive equation 1.

$$\frac{\beta}{\alpha} \times \frac{1-p(E)}{p(E)} \times \frac{C\alpha}{C\beta} = 1 \qquad [1]$$

Where:

α = false positive rate (one-tailed)

β = false negative rate

p(E) = probability the drug is effective, which is also the opportunity for false negative results

1–p(E) = probability the drug is not effective, which is also the opportunity for false positive results

Cα = cost of a false positive result

Cβ = cost of a false negative result

The equation can be rearranged as follows to solve for the optimum ratio of α and β,

$$\frac{1-p(E)}{p(E)} \times \frac{C\alpha}{C\beta} = \frac{\alpha}{\beta} \quad [2]$$

Similarly, the equation can be rearranged to solve for specific values of α given a choice for β, and vice verse,

The cost terms in equation 1 and 2 deserve careful consideration. For example, Cα describes the cost (i.e., consequence) of a false positive result in a PoC study. It is assumed that a false positive PoC triggers phase III development that fails. Hence, the economic cost of a false positive PoC is the phase III development cost. The cost (i.e., consequences) of a false negative PoC (Cβ) can be defined, using similar economic logic, as the profit that would have been earned had the drug become a marketed medicine.

It can be argued that profit is a less than perfect indicator of value. However, given the difficulties in quantifying the true value in prolonging the life of a cancer patient, reducing the pain caused by arthritis, or maintaining glycemic control in diabetics, profit is a usable proxy.

The p(E) term can also involve complexity. For example, p(E) may be the probability associated with being superior to placebo or to a standard of care (SoC). The placebo comparison may be used for novel therapies and in uncrowded therapeutic areas, whereas the SoC comparison may be more useful in crowded therapeutic areas. The key aspect is that p(E) be defined based on the same criteria and comparison as $C\beta$.

If p(E) = 0.50, and if $C\alpha = C\beta$, the two terms left of the equal sign in equation 2 simplify to 1 and the optimum ratio of α and $\beta = 1$; for example, $\alpha = 0.10$ (two-tailed $\alpha = 0.20$) with power = 90% ($\beta = 0.10$). Since the circumstances under which a false positive or false negative error can occur are equally likely and the errors are equally costly, the optimum is to have equal protection against them. Given the same costs, but p(E) = 0.20 the optimum ratio = 4/1 since the circumstances under which a false positive result can occur are now 4-fold more likely.

If the anticipated profit is 10-fold greater than the phase III development cost (e.g., $C\alpha = 1$ and $C\beta = 10$), with p(E) = 0.20 the optimum ratio of α and $\beta = 4/10 = 1/2.5$; for example if $\alpha=0.025$ (two-tailed $\alpha = 0.05$) then the optimum $\beta = 6.25\%$ (power = 93.75%).

Alternatively, if p(E) = 0.33 the optimum ratio of α and β is now 1/5; for example, if α=0.025 then the optimum β = 0.5% (power = 99.5%). Alternatively, if power is 95% the corresponding α = 0.25 (two-tailed α = 0.50).

The impact of α and β on portfolio outcomes in realistic situations.

With basic ideas fixed in the previous section, the impact of α and β in PoC studies on portfolio outcomes can be evaluated in realistic situations by incorporating realistic cost structures and investment constraints. For this exercise, industry average costs of $40 million for phase II and $150 for phase III were taken from Paul et al (2010). Two development paradigms were considered: The so-called fast to PoC and fast to registration archetypes discussed by Mallinckrodt et al (2010). In the fast to PoC approach, investments are minimized prior to establishing proof of concept. For example, the toxicology studies needed to support longer phase III studies, dose ranging studies, carcinogicity studies, etc are only begin after establishing PoC. In the fast to registration paradigm all phase II investments are undertaken in parallel in order to accelerate timelines. The fast to PoC approach terminates ineffective drugs at lower cost because it has lower fixed cost prior to PoC; however, this approach takes longer to get effective drugs to market. The fast to registration approach gets effective drugs to market sooner, but is less efficient at terminating ineffective drugs. However, total development costs for compounds making it to market are assumed to be the same for both approaches.

Specifically, the cost for each PoC trial in the fast to registration approach was modeled as a fixed cost of $35,000,000 plus $25,000 per patient, with post PoC (phase III) cost of $150,000,000. For the fast to PoC approach, each trial was modeled as a fixed cost of $5,000,000 plus $25,000 per patient, with the $30,000,000 in fixed cost deferred to phase III, yielding a total post PoC cost of $180,000,000. Development was constrained to a total investment of $5 billion dollars for phase II and Phase III combined.

The outcome variables of interest were number of launches and total portfolio value. The potential value, such as lifetime profit, of a drug in development is particularly hard to estimate reliably (Munos et al, 2009). Therefore, a simplistic approach was adopted where portfolio value was a function of the number of launches. In the fast to PoC paradigm each launch (compound making it to market) was given a value of 1.0. In the fast to registration paradigm each launch was valued at 1.2, assuming that the compound got to market faster, thereby having more time prior to patent expiration such that total profit increased by 20%.

Using these two development paradigms, portfolio outcomes were investigated based on α (one-sided) at 0.025, 0.05, and 0.10, with power $(1 - \beta)$ at 90%, 80%, and 65%. These levels of α and β were applied to scenarios where the probability the drug was effective was either 30% or 60%, with a 20% phase III attrition rate assumed for safety outcomes in all scenarios, and assuming the true effect size for the effective drugs was either 0.3, 0.4, or 0.6. Therefore, this investigation can be viewed as a 3 x 3 factorial arrangement of decision variables (α and β) applied to a 2 x 2 x 3 factorial arrangement of design variables (2 development paradigms, 2 levels of

probability of efficacy, and 3 magnitudes of treatment effect). Results from the scenario using an effect size of 0.4 in effective drugs are summarized in Table 2.

When the probability the drug was effective = 0.3, the number of launches ranged from 10 to 17 across the various scenarios. The fast to PoC archetype yielded more launches and greater portfolio value as its lower costs allowed more drugs to be developed given the fixed budget. In the fast to PoC archetype, the three scenarios with the greatest number of launches all had $\alpha = .025$. In contrast, for the fast to registration archetype, the three scenarios with the greatest number of launches and greatest portfolio value all had 90% power.

When the probability the drug was effective = 0.6, development scenario and choice of α and β resulted in smaller differences in portfolio outcomes as across the scenarios portfolio value ranged from 18.5 to 20.9. With power of 80% or 90% the fast to registration archetype yielded greater portfolio value than the fast to PoC archetype. In the fast to PoC archetype, $\alpha = .025$ or .05 yielded more launches and greater portfolio value than $\alpha = .10$. In the fast to registration archetype the three scenarios with the greatest number of launches all had 90% power.

When the effect size was 0.3 or 0.6 (results not shown) for the effective compounds, results were qualitatively similar to the previously described results where the effect size was 0.4. After completing these pre-planned comparisons a post hoc comparison

was done to investigate how the cost of phase III influenced the optimum levels of α and β (results not shown). As phase III cost increased, the benefit from lower false positive rates increased; and, as phase III costs decreased the benefit from greater power increased.

Discussion

Optimizing PoC studies is important because the probability of success in phase II and early, accurate assessment of PoC are the most important drivers of drug development productivity (Paul et al, 2010; Gelenberg et al, 2008, Kola and Landis, 2004). Since most drugs are developed by companies that work on multiple drugs simultaneously (Munos, 2009), the present work focused on optimizing the portfolio. Specifically, it was assumed that individual trials were designed to efficiently deliver the chosen levels of α and β, and the focus here was on determining what levels of α and β these trials should seek to deliver.

Basic principles of how α and β influence portfolio outcomes were illustrated with a simplistic example and with an equation whose derivation arises from similar problems in statistics. With these basic ideas fixed, the impact of α and β on portfolio outcomes was evaluated in realistic situations, incorporating realistic cost structures and investment constraints. These examples further illustrated that the optimum choice for α and β can vary depending on compound attributes, such as the probability the drug is effective, and on the development approach (fast to PoC vs. fast to registration).

Although cost estimates that were representative of average costs were used, these costs were not intended to represent any specific situation. Moreover, a simple model was used to determine how costs varied across the scenarios and to determine portfolio value from number of launches. These simplistic approaches were intended to show general trends, but lacked the detail that is probably necessary to evaluate specific circumstances. Therefore, these results illustrate that levels of α and β can have an important impact on portfolio productivity and can vary by scenario; but they do not illustrate what is optimum for any specific situation.

Other limitations of the present work include that consideration was not given to situations where the outcome in the PoC study was different from the outcome used in confirmatory trials, which if not perfectly correlated would result in phase III success rates being different from what would be predicted by the α and β used in the PoC study. In addition, the examples considered portfolios of drugs as if all were at the same stage of development. In practice, research enterprises typically have drugs at all stages of development with decisions made one at a time,

Similar to the present research, Chen and Beckman (2009ab) noted that optimum levels of α and β varied by scenario. These authors used oncology specific examples and assessed portfolio outcome based on an efficiency score function driven by the number of patients per approval. Unlike the previous research, the present investigation did not uncover scenarios where low power was optimal. This difference likely arises from the cost modeling assumptions. When patient number is the only cost consideration, the benefit from smaller sample sizes will be greater than when also incorporating fixed costs into the scenarios. Therefore, results of the present

investigation agree with previous research in that it is important to assess optimum levels of α and β. However, the present work also highlights the need to consider cost structure using realistic input parameters for the situation and the question of interest. For example, in a deadly illness such as cancer, it may be appropriate to describe efficiency in terms of numbers of patients studied whereas is other disease states a purely economic cost may be most useful.

Conclusion

Simple examples and a newly derived equation provided conceptual understanding of basic principles regarding optimum choices of α and β in PoC trials. In examples incorporating realistic development costs and constraints, the optimum choice of α and β varied by scenario, and was particularly sensitive to cost structure and the probability the drug was effective Results of the present investigation agree with previous research in that it is important to assess optimum levels of α and β. However, the present work also highlights the need to consider cost structure using realistic input parameters relevant to the question of interest.

References

Chen C and Beckman RA. (2009) Optimal Cost-Effective Designs of Phase II Proof of Concept Trials and Associated Go-No Go Decisions', Journal of Biopharmaceutical Statistics, 19: 3, 424 - 436

Gelenberg AJ, Thase ME, Meyer RE, Goodwin FK, Katz MM, Kraemer HC, Potter WZ, Shelton RC, Fava M, Kahn A, Trivedi MH,

Ninan PT, Mann JJ, Bergeson S, Endicott J, Kocsis JH, Leon AC; Manji HK, Rosenbaum JF. The history and current state of antidepressant clinical trial design: a call to action for proof–of–concept studies. *J Clin Psychiatry*. 2008;69(10):1513-1528

Hurko O and Ryan JL. (2005). Translational research in central nervous system drug discovery. *NeuroRx,* **2**, 671-682.

Kola I and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*, **3**, 711-715.

Johnson, R.A. and Wichern, D.W. (1992) Applied Multivariate Statistical Analysis (3rd ed.). En-
glewood Cliffs, NJ: Prentice-Hall.

Mallinckrodt CH, Detke MJ, Prucka WR, Ruberg SJ, and Molenberghs G. (2010).
Design Archetypes for Phase II Clinical Trials in Central Nervous System Disorders.
Drug Information Journal. In Press.

Munos, BH. (2009). Lessons learned from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov*, **8**, 959-968.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR and Schacht AL. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov, 9*, 203-214.

Piantadosi, S. (2005) Clinical Trials: A Methodological Perspective (2nd ed.). New York: John Wiley & Sons.

Table 1. Outcomes from proof of concept trials with varying levels of power and false positive rate results and percentages of effective drugs.

	Phase II Decisions			Phase II Results[3] (20/100 drugs effective)			Phase II Results[3] (30/100 drugs effective)		
Scenario	pts/Arm[1]	Power	False+ rate[2]	True +s	False +s	Ph III success rate	True +s	False +s	Ph III success rate
1	35	65	10	13	8	61.9	19.5	7	73.6
2	52	65	5	13	4	76.5	19.5	3.5	84.8
3	69	65	2.5	13	2	86.7	19.5	1.75	91.8
4	57	80	10	16	8	66.7	24	7	77.4
5	78	80	5	16	4	80	24	3.5	87.3
6	99	80	2.5	16	2	88.9	24	1.75	93.2
7	83	90	10	18	8	69.2	27	7	79.4
8	108	90	5	18	4	81.8	27	3.5	88.5
9	132	90	2.5	18	2	90	27	1.75	93.9

1. All trials assume a standardized effect size of 0.4 and that is the true effect size for the effective drugs
2. The rate of false positive results in PoC trials, whis is ½ α in two-tailed testing
3. True +s = power multiplied by number of effective drugs, False +s = False + rate multiplied by the number of ineffective drugs

Ph III success rate = True +s / (True +s plus False +s)

Table 2. Portfolio outcomes for various choices of α and β across different development scenarios.

Power	Alpha	Prob[1] N per Arm	Prob[2] Enter PhIII	Eff PHIII	Apprvl[3] Prob	Fast to PoC archetype: Cost Per[4] Compd	Compds[5] Studied	Compds[6] Apprvd	Portfolio[7] Value	Fast to registration archetype: Cost Per[4] Compd	Compds[5] Studied	Compds[6] Apprvd	Portfolio[7] Value
Probability drug is effective = 0.3. Effect size = 0.4													
.65	.1	35	0.27	0.74	0.16	54.5	91.8	14.3	14.3	76.5	65.4	10.2	12.2
.65	.05	52	0.23	0.85	0.16	49.0	102.0	15.9	15.9	72.1	69.3	10.8	13.0
.65	.025	69	0.21	0.92	0.16	46.7	107.1	16.7	16.7	70.3	71.1	11.1	13.3
.8	.1	57	0.31	0.77	0.19	63.7	78.6	15.1	15.1	84.4	59.3	11.4	13.7
.8	.05	78	0.28	0.87	0.19	58.4	85.6	16.4	16.4	80.2	62.4	12.0	14.4
.8	.025	99	0.26	0.93	0.19	56.3	88.8	17.1	17.1	78.6	63.6	12.2	14.7
.9	.1	83	0.34	0.79	0.22	70.4	71.1	15.4	15.4	90.2	55.5	12.0	14.4
.9	.05	105	0.31	0.89	0.22	65.2	76.7	16.6	16.6	86.0	58.1	12.6	15.1
.9	.025	132	0.29	0.94	0.22	63.4	78.9	17.0	17.0	84.7	59.0	12.7	15.3
Probability drug is effective = 0.6. Effect size = 0.4													
.65	.1	35	0.43	0.91	0.31	84.2	59.4	18.5	18.5	101.3	49.4	15.4	18.5
.65	.05	52	0.41	0.95	0.31	81.4	61.4	19.2	19.2	99.1	50.5	15.7	18.8
.65	.025	69	0.40	0.98	0.31	80.5	62.2	19.4	19.4	98.5	50.8	15.8	19.0
.8	.1	57	0.52	0.92	0.38	101.5	49.3	18.9	18.9	115.9	43.2	16.6	19.9
.8	.05	78	0.50	0.96	0.38	98.9	50.6	19.4	19.4	113.9	43.9	16.9	20.2
.8	.025	99	0.49	0.98	0.38	98.2	50.9	19.6	19.6	113.5	44.1	16.9	20.2
.9	.1	83	0.58	0.93	0.43	113.6	44.0	19.0	19.0	126.2	39.6	17.1	20.5
.9	.05	105	0.56	0.96	0.43	111.1	45.0	19.5	19.5	124.3	40.2	17.4	20.9
.9	.025	132	0.55	0.98	0.43	110.6	45.2	19.5	19.5	124.1	40.3	17.4	20.9

1. Probability a compound will enter phase III. Includes true positive and false positive phase II compounds

2. Probability that a phase III compound is effective.
3. Probability a compound will be approved. Includes the 20% attrition in phase III due to safety and other reasons not due to efficacy
4. Development cost per compound. Total phase II cost + cost phase III cost for compounds entering phase III divided by number of compounds entering phase II.
5. Number of compounds studied. The $5 billion budget divided by average cost per compound.
6. Number of compounds approved. Number of compounds studies multiplied by approval probability.
7. Portfolio value = number of launches for fast to PoC and = number of launches x 1.2 for fast to registration.