

RESEARCH ARTICLE

Open Access

Sample size re-assessment leading to a raised sample size does not inflate type I error rate under mild conditions

Per Broberg

Abstract

Background: One major concern with adaptive designs, such as the sample size adjustable designs, has been the fear of inflating the type I error rate. In (Stat Med 23:1023-1038, 2004) it is however proven that when observations follow a normal distribution and the interim result show promise, meaning that the conditional power exceeds 50%, type I error rate is protected. This bound and the distributional assumptions may seem to impose undesirable restrictions on the use of these designs. In (Stat Med 30:3267-3284, 2011) the possibility of going below 50% is explored and a region that permits an increased sample size without inflation is defined in terms of the conditional power at the interim.

Methods: A criterion which is implicit in (Stat Med 30:3267-3284, 2011) is derived by elementary methods and expressed in terms of the test statistic at the interim to simplify practical use. Mathematical and computational details concerning this criterion are exhibited.

Results: Under very general conditions the type I error rate is preserved under sample size adjustable schemes that permit a raise. The main result states that for normally distributed observations raising the sample size when the result looks promising, where the definition of promising depends on the amount of knowledge gathered so far, guarantees the protection of the type I error rate. Also, in the many situations where the test statistic approximately follows a normal law, the deviation from the main result remains negligible. This article provides details regarding the Weibull and binomial distributions and indicates how one may approach these distributions within the current setting.

Conclusions: There is thus reason to consider such designs more often, since they offer a means of adjusting an important design feature at little or no cost in terms of error rate.

Keywords: Adaptive trial design, Type I error, Conditional power

Background

The last few years interest in various adaptive trial designs has surged [1]. A greater flexibility of clinical study design and conduct has followed from the application of these new ideas [2]. In [1] adaptive designs are defined as:

“...a clinical trial design that allows adaptations or modifications to aspects of the trial after its initiation without undermining the validity or integrity of the trial.”

More and more trials of this sort are being reported and regulatory bodies take an increasingly favourable view on them [3]. All stand to win if these designs come to optimal use [4]. However, some concerns have been raised. One of these involve the risk of inflating type I error. The current article will assess that risk in the context of sample size adjustable (SSA) designs that allow choosing between raising the sample size, continuing as originally planned, or closing the trial due to futility.

The following will recapitulate parts of [5], add detail and draw conclusions for trial procedures. In that reference the authors show that if the interim results look promising, no inflation of type I error rate occurs. Here ‘promising’ means that the conditional power at the current parameter estimate, i.e. the power updated by the

Correspondence: Per.Broberg@med.lu.se
Department of Oncology and Cancer Epidemiology, Clinical Sciences, Lund University and Skåne University Hospital, 221 85 Lund, Sweden

accumulated knowledge at interim, amounts to at least 50%. This article will show that a less strict bound applies, in agreement with [6], exhibit the bound in terms of a test statistic, and present mathematical as well as computational aspects of it.

Methods

Assumptions

Denote the planned final sample size by N_0 , the number of patients available at the pre-planned interim analysis by n , and the possible raise determined at the interim taking conditional power into account by r . Let us consider a one-sided test at level α based on observing $X_1, \dots, X_{N_{final}}$. Here $N_{final} = N_0$ or $N_{final} = N = N_0 + r$, depending on a decision taken during the course of the trial. The main result assumes normal distribution, but as will be outlined, it will still hold true for more general distributions. Further, assume the X_i to be independent normal with mean θ and variance 1. The null hypothesis states that $\theta = 0$. Define the normalised test statistic $Z^{(x)}$ by $Z^{(x)} = \sum_{i=1}^x X_i / \sqrt{x}$. The test rejects if $Z^{(N_{final})} > z_\alpha$, where z_α is the $100 \times (1 - \alpha)$ percentile of the standard normal distribution: $\Phi(z_\alpha) = 1 - \alpha$ (Φ being the cumulative distribution function of the standard normal distribution). The normalised test statistic $Z^{(n)} = \sum_{i=1}^n X_i / \sqrt{n}$ is observed when n patients have provided data, and the Data Monitoring Committee (DMC) will in part base its recommendations on the observed value. At this interim analysis an adaptation may lead to closing the study due to futility, continuing the study without changes or raising the sample size by recruiting an extra r subjects, yielding a total of $N = N_0 + r$ subjects. Closing the study due to futility may only decrease the type I error rate. So let us, for the sake of argument, disregard that possibility, and show that the type I error rate still remains protected.

The study protocol will specify n and N_0 , and at the interim we will consider raising the final sample size based on the conditional power evaluated at the current parameter estimate. Since the objective is to assess if the interim results are promising the current estimate of the parameter of interest gives the appropriate information [6]. As pointed out by Müller and Schäfer in [7], the over-all type I error can be preserved unconditionally under any general adaptive change, provided the conditional type I error that would have been obtained had there been no adaptation is preserved. This article however only considers the case of SSA. Unlike the situation in [8] the design does not permit sequential testing. Also, the article only considers the conventional hypothesis tests and p-values without adjustments.

We assess the conditional error rate as a function of r . By showing that the conditional type I error rate is bounded by the error rate which arises from the design without

adaptation the unconditional error rate is proven to be controlled at a pre-specified level α .

Derivation of the main result

We use the notation $X \sim N(\mu, \sigma^2)$ to signify that X follows a normal law with mean μ and variance σ^2 .

The change in type I error rate conditional on a sample size increase decided at the interim equals $G(r) = Pr(Z^{(N_0+r)} > z_\alpha | Z^{(n)} = z, \theta = 0) - Pr(Z^{(N_0)} > z_\alpha | Z^{(n)} = z, \theta = 0)$. The conditional distribution equals $(Z^{(N)} | Z^{(n)} = z, \theta = 0) \sim N(\rho z, 1 - \rho^2)$, where $\rho = Cov(\sum_{i=1}^N X_i / \sqrt{N}, \sum_{j=1}^n X_j / \sqrt{n}) = \sum_{i=1}^n 1 / \sqrt{nN} = \sqrt{n/N}$, and similarly for $Z^{(N_0)}$.

Expressing the difference in terms of normal distributions yields

$$G(r) = \Phi\left(\frac{\sqrt{\frac{n}{N_0+r}}z - z_\alpha}{\sqrt{1 - \frac{n}{N_0+r}}}\right) - \Phi\left(\frac{\sqrt{\frac{n}{N_0}}z - z_\alpha}{\sqrt{1 - \frac{n}{N_0}}}\right)$$

Now, in order to show this difference to be less than or equal to zero it may be equivalently shown that the difference of the arguments is negative (in the sense of non-positive), and denote this by $H(r)$. Obviously $H(0) = 0$.

To simplify notation put $q = n/(N_0 + r)$ and $V = (N_0 + r)/N_0$ for arbitrary n, N_0 and r , satisfying $N_0 > n > 0$, and $r > 0$. Please note $qV = n/N_0$. Then we aim to show

$$\frac{\sqrt{q}z - z_\alpha}{\sqrt{1 - q}} - \frac{\sqrt{qV}z - z_\alpha}{\sqrt{1 - qV}} \leq 0$$

This implies

$$z \left(\frac{\sqrt{q}}{\sqrt{1 - q}} - \frac{\sqrt{qV}}{\sqrt{1 - qV}} \right) \leq z_\alpha \left(\frac{1}{\sqrt{1 - q}} - \frac{1}{\sqrt{1 - qV}} \right)$$

And, then, multiply by -1 and divide by the multiplier of z , to obtain

$$z \geq z_\alpha \frac{\sqrt{1 - q} - \sqrt{1 - qV}}{\sqrt{1 - q}\sqrt{1 - qV}} / \frac{\sqrt{qV}\sqrt{1 - q} - \sqrt{q}\sqrt{1 - qV}}{\sqrt{1 - q}\sqrt{1 - qV}},$$

which after cancelling out $\sqrt{1 - q}\sqrt{1 - qV}$ becomes

$$\begin{aligned} z &\geq z_\alpha (\sqrt{1 - q} - \sqrt{1 - qV}) / (\sqrt{qV}\sqrt{1 - q} - \sqrt{q}\sqrt{1 - qV}) \\ &:= z_\alpha b(q, V) \end{aligned} \tag{1}$$

Now let us compare this bound to $z_\alpha \sqrt{n/N_0}$. Denote the bound in (1) by $z_\alpha b(q, V)$, and set out to prove $b(q, V) \leq \sqrt{qV}$. By subtracting \sqrt{qV} from both sides and equating denominators we have

$$\frac{\sqrt{1 - q} - \sqrt{1 - qV} - \sqrt{qV}(\sqrt{qV}\sqrt{1 - q} - \sqrt{q}\sqrt{1 - qV})}{\sqrt{qV}\sqrt{1 - q} - \sqrt{q}\sqrt{1 - qV}} \leq 0$$

But the denominator is positive under current assumptions. Thus, we may disregard it, and thus, we need to prove

$$\sqrt{1-q}(1-qV) + \sqrt{1-qV}(q\sqrt{V}-1) \leq 0$$

Division by $\sqrt{1-qV}$, moving the right term to the right hand side of the inequality symbol, and squaring both sides yields

$$(1-q)(1-qV) \leq 1-2q\sqrt{V}+q^2V$$

By expanding the left hand side product, eliminating terms and multiplying both sides by $-1/q$, we finally have

$$V+1 \geq 2\sqrt{V},$$

which is true for all positive V .

Now regard b as a function of r for n and N_0 fix. One may show that, $b(r) \nearrow \sqrt{n/N_0}$ asymptotically as $r \searrow 0$. Further, b decreases as r grows in a close to linear fashion. Also, $b(r) \searrow (1-\sqrt{1-qV})/\sqrt{qV}$ when r tends to infinity.

Please note that since the $b(q, V) \leq \sqrt{qV} = \sqrt{n/N_0}$ a sufficient but not necessary condition is $z \geq z_\alpha \sqrt{n/N_0}$, which will be seen to give the conditional power 50% (the simple criterion). Consequently, this new criterion is less restrictive than the one presented in [5], and, importantly, changes with r . The reference [6] provides an example where the type I error remains intact although the conditional power descends down to 36%.

To obtain the conditional power please note that $E[Z^{(N_0)}|Z^{(n)} = z, \theta = z/\sqrt{n}] = z/\sqrt{n/N_0}$, and furthermore $(Z^{(N_0)}|Z^{(n)} = z, \theta = z/\sqrt{n}) \sim N(z/\sqrt{n/N_0}, 1 - n/N_0)$. Then

$$\begin{aligned} Pr[Z^{(N_0)} > z_\alpha | Z^{(n)} = z, \theta = z/\sqrt{n}] \\ = 1 - \Phi\left(\frac{z_\alpha - z/\sqrt{n/N_0}}{\sqrt{1 - \frac{n}{N_0}}}\right) \end{aligned}$$

The minimum of this probability over $z > b(q, V)z_\alpha$ equals

$$\Phi\left(\frac{z_\alpha(b(q, V)/\sqrt{qV}-1)}{\sqrt{1-qV}}\right)$$

From the definition of $G(r)$ it follows that one cannot go further without increasing the conditional error rate. In this sense the bound is optimal.

Weibull ditributed survival time points

We will now study the situation where survival times follow a Weibull distribution and right censoring time points are exponentially distributed.

In [9] the details of an Edgeworth expansion of the product limit estimator are given $\Phi(x) - n^{-\frac{1}{2}}\phi(x)(\tilde{\kappa}_3(x^2 -$

$1) + 3\sigma_1)/6$. First some notation: $X =$ lifetime, $T =$ left truncation time point, $Y =$ right censoring time point, $Z = \min(X, Y)$, $\delta = I(X \leq Y)$. Further, put $C(z) := P(T \leq z \leq Z|T \leq Z)$. But since $T \equiv 0$ this probability equals $P(Z \geq z) = P(X \geq z, Y \geq z) = P(X \geq z)P(Y \geq z)$. Then $W_1(y) := P(Z \leq y, \delta = 1) = P(X \leq y, Y \geq X)$. With $\sigma_1(z) = \int_0^z \frac{dW_1(u)}{C^2(u)}$, the constant $\tilde{\kappa}_3$ in the Edgeworth expansion equals $\sigma_1^{-1}(-7.5\sigma_1^4 + \int_0^z C^{-3}(t)dW_1(t))$. As stated we assume $X \sim Weib(\lambda, \beta)$ and $Y \sim Exp(\mu)$. From this follows that $C(y) = exp(-\mu y - \lambda y^\beta)$, $W_1(y) = \int_0^y e^{-\mu x} \lambda \beta x^{\beta-1} e^{-\lambda x^\beta} dx$. Thus one may at the interim use parameter estimates to calculate a normal approximation to the conditional power. Alternatively, one may simulate the remainder of the trial. A third option is to base the procedure on the logrank test whose statistic converges to a normal distribution. Consider the situation where the time to some event is compared between patients in an active treatment group and those in a control group. Let r_i refer to the number of patients remaining at time i and o_i refer to the number of observed events. Further, let A refer to the active treatment group and C to the control group. If $T = \sum_{i=1}^k \frac{r_{iA}o_{iC} - r_{iC}o_{iA}}{r_i}$ and $V = \sum_{i=1}^k \frac{o_i(r_i - o_i)r_{iA}r_{iC}}{(r_i - 1)r_i^2}$, then $z = T/\sqrt{V}$ will asymptotically be standard normal, e.g. [10]. Hence one may apply the simple criterion to z observed at the interim.

Binomial proportion

For the sake of simplicity of exposition we focus attention to a single binomial proportion p and a one-sided test at the 5% level. Let the null hypothesis and alternative hypothesis be $H_0 : p = p_0, H_1 : p > p_0$. Please note that for $X_{N_0} \sim Bin(N_0, p_0)$ the conditional distribution given $\{X_n = k\}$ is the same as $X_{N_0-n} + k \sim Bin(N_0 - n, p_0)$, and similarly for X_{N_0+r} .

From this follows that we may obtain $G(r)$ exactly: in terms of R code [11]

```
#let the level of the test be 5%
qs.r <- qbinom(p = 0.95, size = N0+r,
prob = p0)
qs <- qbinom(p = 0.95, size = N0, prob = p0)
G.r <- pbinom(qs-k, size = N0-n, prob = p0) -
pbinom(qs.r - k, size = N0+r-n, prob = p0)
```

However, we will look at a normal approximation. In the binomial case several test statistics with close to normal distribution exist:

- the score test statistic:
 $z = \sqrt{n}(\hat{p} - p_0)/\sqrt{p_0(1 - p_0)}, \hat{p} = k/n$
- the log-odds:
 $z = \sqrt{n\hat{p}(1 - \hat{p})}(\log(\hat{p}/(1 - \hat{p})) - \log(p_0/(1 - p_0)))$
 [12]

The simple criterion would then say that if z as above exceeds $\sqrt{\frac{n}{N_0}}z_\alpha$, then the procedure protects the type I error rate (unconditionally). But we set out to find a more accurate approximation.

Now using the condition $\{X_n = k\}$

$$G(r) = P_{p_0}(X_{N_0+r-n} > q_{\alpha, N_0+r} - k) - P_{p_0}(X_{N_0-n} > q_{\alpha, N_0} - k),$$

where $q_{\alpha, m}$ is the $100 \times (1 - \alpha)$ percentile of $Bin(m, p_0)$.

Also the binomial distribution $X_n \sim Bin(n, p)$ admits a normal approximation of the pivotal statistic $U = (X_n - E[X_n])/SD(X_n)$, which coincides with the score test statistic above, such that

$$P(U > q) = 1 - \Phi\left(q - \frac{1}{6\sqrt{n}}(\kappa_3(q^2 - 1))\right) + O\left(\frac{1}{n}\right),$$

in terms of the third cumulant of U , which picks up the skewness. As a rule of thumb it is often said that the normal approximation is quite accurate when np and $n(1 - p)$ both exceed 5. But this statement holds even without the correction with respect to skewness.

In this case we may approximate the difference $G(r)$ defined above by

$$G(r) = P_{p_0}(X_{N_0+r} > q_{\alpha, N_0+r} | X_n = k) - P_{p_0}(X_{N_0} > q_{\alpha, N_0} | X_n = k),$$

where we insert the percentiles obtained from the inversion of the Cornish-Fisher expansion, cf. e.g. [13]:

$$q_{\alpha, n} = \mu_n + \sqrt{n}\sigma_0(z_\alpha + \frac{\gamma_0}{6\sqrt{n}}(z_\alpha^2 - 1)),$$

denoting $\mu_n = np_0$, $\sigma_0 = \sqrt{p_0(1 - p_0)}$, and the third cumulant $(1 - 2p_0)/\sigma_0$ by γ_0 . This quantity will deviate less than 1 from the true percentile for n from 20 to 200, and $\min\{np_0, n(1 - p_0)\} > 5$. Let us consider $G(r)$ through the pivotal quantities

$$U_N = \frac{q_{\alpha, N} - k - \mu_{N-n}}{\sigma_{N-n}}$$

We will be concerned with the difference

$$U_{N_0+r} - U_{N_0} = \frac{q_{\alpha, N_0+r} - k - \mu_{N_0+r-n}}{\sigma_{N_0+r-n}} - \frac{q_{\alpha, N_0} - k - \mu_{N_0-n}}{\sigma_{N_0-n}}$$

The task is now to identify when this difference is positive. To simplify notation denote by n_1 the larger of the two sample sizes and by n_2 the smaller. After equating the denominators and noting $\mu_{n_i} - \mu_{n_i-n} = np_0, i = 1, 2$ the difference equals:

$$\frac{\sqrt{n_2 - n}(np_0 + \sqrt{n_1}z_\alpha\sigma_0 + \gamma_0\sigma_0(z_\alpha^2 - 1)/6 - k)}{\sqrt{(n_1 - n)(n_2 - n)}\sigma_0} - \frac{\sqrt{n_1 - n}(np_0 + \sqrt{n_2}z_\alpha\sigma_0 + \gamma_0\sigma_0(z_\alpha^2 - 1)/6 - k)}{\sqrt{(n_1 - n)(n_2 - n)}\sigma_0}$$

Disregarding the positive denominator yields the condition:

$$(\sqrt{n_2 - n} - \sqrt{n_1 - n})(np_0 + \frac{\gamma_0\sigma_0(z_\alpha^2 - 1)}{6} - k) + (\sqrt{n_2 - n}\sqrt{n_1} - \sqrt{n_1 - n}\sqrt{n_2})z_\alpha\sigma_0 > 0$$

Some algebra will unearth the condition

$$k > np_0 + \frac{\gamma_0\sigma_0(z_\alpha^2 - 1)}{6} + \frac{(\sqrt{n_2 - n}\sqrt{n_1} - \sqrt{n_1 - n}\sqrt{n_2})z_\alpha\sigma_0}{\sqrt{n_2 - n} - \sqrt{n_1 - n}}$$

Please note that the first term corresponds to expectation of the null distribution. Further, the second term will be negative if $p_0 > 0.5$, and the third will always be negative under the conditions of this paper. From this follows that the normal approximation of $G(r)$ is non-positive for k satisfying the above condition. Finally, invoke the fast convergence of the binomial distribution towards a normal law, which means that already 20 observations will make the normal approximation quite accurate, provided $\min\{np, n(1 - p)\} > 5$. Simulations indicate that this decision rule is accurate already at an interim sample size n as low as 20. However, the rule does not guarantee preservation of the conditional type I error rate for all p . Thus the conclusion is that for the binomial distribution there is no inflation of the unconditional type I error rate under the above conditions. A total of 900000 simulations with n from 20 to 100, p_0 picked randomly in $[5/n, 1 - 5/n]$, k randomly generated from $Bin(n, p_0)$ and $N_0 = 2n$ and $r = n$ gave a median and mean of $G(r)$ equal to -0.004762 and -0.004574 , respectively, over the set defined by the inequality above. A set of similar simulations using the simple criterion ($k > n(p_0 + \sqrt{p_0(1 - p_0)})/N_0z_\alpha$) gave median and mean equal to -0.02429 and -0.02389 , respectively. Thus the simple criterion will be on the conservative side.

Results

Main result

In **Methods** the following result was derived.

A conditional power that guarantees preservation of nominal significance level

If the conditional power at the interim, which occurs after n out N_0 planned observations and leads to a raise of r , equals at least

$$\Phi\left(\frac{z_\alpha(b(q, V)/\sqrt{qV} - 1)}{\sqrt{1 - qV}}\right), \tag{2}$$

where $q = n/(N_0 + r)$, $V = (N_0 + r)/N_0$, and

$$b(q, V) = (\sqrt{1-q} - \sqrt{1-qV}) / (\sqrt{qV} \sqrt{1-q} - \sqrt{q} \sqrt{1-qV}), \quad (3)$$

then the type I error rate is preserved. The function b satisfies the inequalities $(1 - \sqrt{1-qV}) / \sqrt{qV} \leq b(q, V) \leq \sqrt{qV} = \sqrt{n/N_0}$.

A more practical criterion, or rule of thumb, may be to derive a test statistic z with close to a standard normal distribution under the null hypothesis, and check whether $z > \sqrt{\frac{n}{N_0}} z_\alpha$. This will be referred to as the simple criterion, and stems from [5]. More generally, the condition $z > b(q, V) z_\alpha$ suffices (cf. equations (11) and (12) in [6]). The conditional power bound in (2) decreases as r increases, but the lower bound on b implies a limit.

Example

Take the example of $n = 55, N_0 = 110, r = 40$ and $\alpha = 0.025, z_\alpha = 1.96$. Then the minimum conditional power equals 43%, see next subsection. Thus a conditional power of considerably less than 50% is permissible from the point of view of type I error rate preservation. This may be good to know if the original sample size calculation was grossly wrong. Then recruiting more subjects than planned may resolve the issue without jeopardising the type I error rate. On the other hand, in such a situation the validity of the scientific hypotheses on which the trial design rests may be questioned, and the sponsor will have to judge whether the updated hypotheses suggest a commercially viable route. Nevertheless, in some cases raising the sample size will make sense, and may save the trial from unnecessary disaster.

Above we assume the variance to be known. If it is not we may estimate it and use for instance a t -test statistic which quickly converges to a normal as the sample size increases.

Examination of the t -test has provided evidence of a small degree of inflation [14]. In [15] further details of when inflation occurs are given. However, already at a sample size of 30 the t -distribution and the normal distribution appear almost identical.

Calculations in R

In the statistical software environment R [11] one may easily define functions. Let us regard the bound $z_\alpha b(q, V)$ as a function of (n, N_0, r) instead. We may explore the bound $z_\alpha b(n, N_0, r)$ through the R function *B.func*

```
B.func <- function(n, N0, r) {q <- n/(N0+r);
V <- (N0+r)/N0; (sqrt(1-q) - sqrt(1-q*V)) /
(sqrt(q*V) * sqrt(1-q) - sqrt(q) * sqrt(1-q*V)) }
```

and, the minimum conditional power through

```
CP.min <- function(alpha1, n, N0, r) {
za <- qnorm(1-alpha1)
q <- n/(N0+r); V <- (N0+r)/N0
b <- B.func(n = n, N0 = N0, r = r)
pnorm(za * (b/sqrt(q*V) - 1) / sqrt(1-q*V)) }
```

So, for instance $CP.min(alpha1 = 0.025, n = 55, N0=110, r = 40) = 0.43$, and, $B.func(n = 55, N0 = 110, r = 0.01) = 0.7070907$, which approximately equals $\sqrt{n/N_0} = \sqrt{55/110} \approx 0.7071068$. Also $CP.min(alpha1 = 0.025, n = 55, N0 = 110, r = 110) = 0.3575873$.

Deviations from normal distribution

If we use non-normal data such as survival type of data, then it is often possible to approximate the test statistic by a normal variate. Many test statistics, e.g. those derived by the maximum likelihood method, converge quickly to a normal distribution when the sample size increases. This feature extends the relevance of the main result to measurements following other distributions than the normal.

In **Methods** we looked into the situation where a Kaplan-Meier (KM) estimate is used. The Edgeworth expansion of the distribution of the (standardised) KM estimator has the form $\Phi(x) - n^{-\frac{1}{2}} \phi(x) \tilde{\kappa}_3 (x^2 - 1) / 6$, where $\tilde{\kappa}_3$ is specified in **Methods** [9,16], Φ the cumulative distribution function of a standard normal variate and ϕ its frequency function. So if we express the change in conditional error rate ($G(r)$ below) in terms of this expansion the correction term to difference between normal distribution functions will approach zero as $1/\sqrt{n}$. Assuming some parametric distribution, such as the Weibull distribution, one may work out the details regarding this approximation. Or, one may assess the deviation from normality through a simulation procedure.

In the case of a single binomial proportion p and a one-sided test of the null hypothesis $H_0 : p = p_0$ versus the alternative hypothesis $p > p_0$, it holds that if we at the interim observe $X_n = k$ satisfying

$$k > np_0 + \frac{\gamma_0 \sigma_0 (z_\alpha^2 - 1)}{6} + \frac{(\sqrt{N_0 - n} \sqrt{n_1} - \sqrt{n_1 - n} \sqrt{N_0}) z_\alpha \sigma_0}{\sqrt{N_0 - n} - \sqrt{n_1 - n}},$$

with $\sigma_0 = p_0(1 - p_0)$, $\gamma_0 = (1 - 2p_0)/\sigma_0$ and $n_1 = N_0 + r$, then inflation of the type I error rate will not occur. More precisely put: on average, over all possible outcomes, the procedure will preserve the type I error rate. However, the conditional error rate will not always fall below the nominal one.

Discussion

There are operational issues with adaptive designs that must be addressed during the planning stage. In order to safeguard the integrity of the trial and avoid operational bias following an unblinded interim precautions need to be put in place to limit access to both the results and, even, the analysis plans. The latter will specify the output and decision rules, but will leave open the possibility of including other information, such as external factors in the final decision whether to stop for futility or to continue, and if so, whether or not to raise the sample size.

Further, a number of concerns have been raised involving the risk of violating statistical principles or lack of efficacy compared to group sequential designs, e.g. [17-19].

However valid these objections may be, more and more practitioners have felt that the challenges are tractable and have found SSA designs an attractive option. For small biotechnology companies this option gives the possibility of starting a trial with rather limited resources, followed by an additional investment conditional on the interim results being promising. Also, the SSA design makes a lot of sense whenever a fix size design would have to rely on quite limited amount of information regarding the primary variable.

Several references have argued the superiority of seamless phase II/III designs over the traditional phase II and III trials. Merging the two phases produces gains in valuable time [20], and, under reasonable conditions, saves sample size [21].

Earlier research has established that a conditional power at the interim analysis exceeding 50% implies that the conditional, and hence also the unconditional, type I error rate is preserved, cf. [5,7]. Further, the reference [6] builds on [8] and others to identify a more general region where this happens. The region is identified through equations (11) and (12) in [6]. The derivation of the region relies on results for Brownian motion. Together these two equations implicitly define a bound that coincides with b in (3) above.

Further, one cannot use a lower bound without risking inflation of the conditional error rate, and thus one may not rely on the Müller-Schäfer principle of conditional error functions [7] (new does not exceed the original) to prove preservation of unconditional error rate¹. By virtue of the Müller-Schäfer principle of conditional error functions any interim decision rule, pre-defined or not, that does not violate this fundamental requirement will permit a redesign of the trial. So from this perspective the SSA designs described here are well behaved and offer great flexibility.

Conclusions

This article has shown that the risk of compromising the nominal significance level of a statistical test by allowing a sample size increase during the course of a trial remains low and controllable. The conditional error rate and power provide key decision tools.

Endnotes

¹ Also, by reversing the order of terms in $G(r)$ and tracing the same line of thought one may conclude that a sample size decrease is permissible when results are discouraging. But then it may make more sense to discontinue the trial due to futility.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

Thanks are due to the referee who found an error in an earlier version of the main result and pointed my attention to crucial references I had overlooked.

Received: 25 November 2012 Accepted: 15 July 2013

Published: 19 July 2013

References

1. Chang M: *Adaptive Design Theory and Implementation Using SAS and R*. Boca Raton, Florida: Chapman and Hall/CRC; 2007.
2. Bretz F, Schmidli H, König F, Racine A, Maurer W: **Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts**. *Biom J* 2006, **48**(4):623-634.
3. FDA: *Guidance for industry: adaptive design clinical trials for drugs and biologics* 2010, [http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf]
4. PhRMA: **Working group on adaptive designs. Full white paper**. *Drug Inf J* 2006, **40**:421-484.
5. Chen YHJ, DeMets DL, Gordon Lan KK: **Increasing the sample size when the unblinded interim result is promising**. *Stat Med* 2004, **23**(7):1023-1038, [http://dx.doi.org/10.1002/sim.1688]
6. Mehta CR, Pocock SJ: **Adaptive increase in sample size when interim results are promising: A practical guide with examples**. *Stat Med* 2011, **30**(28):3267-3284, [http://dx.doi.org/10.1002/sim.4102]
7. Müller HH, Schäfer H: **A general statistical principle for changing a design any time during the course of a trial**. *Stat Med* 2004, **23**(16):2497-2508, [http://dx.doi.org/10.1002/sim.1852]
8. Gao P, Ware J, Mehta C: **Sample size re-estimation for adaptive sequential design in clinical trials**. *J Biopharmaceutical Stat* 2008, **18**(6):1184-1196, [http://www.tandfonline.com/doi/abs/10.1080/10543400802369053]
9. Wang Q, Jing BY: **Edgeworth expansion and bootstrap approximation for studentized product-limit estimator with truncated and censored data**. *Commun Stat - Theory and Methods* 2006, **35**(4):609-623, [http://www.informaworld.com/10.1080/03610920500498840]
10. Whitehead J: *The Design and Analysis of Sequential Clinical Trials*. Chichester, West Sussex: Ellis Horwood, second edition; 1992.
11. R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2010, [http://www.R-project.org] [ISBN 3-900051-07-0].
12. Zhou X, Li C, Yang Z: **Improving interval estimation of binomial proportions**. *Philos Trans R Society A: Math, Phys Eng Sci* 2008, **366**(1874):2405-2418, [http://rsta.royalsocietypublishing.org/content/366/1874/2405]
13. DiCiccio TJ, Efron B: **Bootstrap confidence intervals**. *Stat Sci* 1996, **11**(3):189-212, [http://www.jstor.org/stable/2246110]
14. Friede T, Kieser M: **Sample size recalculation in internal pilot study designs: a review**. *Biom J* 2006, **48**(4):537-555, [http://dx.doi.org/10.1002/bimj.200510238]

15. Graf AC, Bauer P: **Maximum inflation of the type 1 error rate when sample size and allocation rate are adapted in a pre-planned interim look.** *Stat Med* 2011, **30**(14):1637–1647, [<http://dx.doi.org/10.1002/sim.4230>]
16. Chang MN: **Edgeworth expansion for the kaplan-meier estimator.** *Commun Stat - Theory and Methods* 1991, **20**(8):2479–2494, [<http://www.informaworld.com/10.1080/03610929108830645>]
17. Burman CF, Sonesson C: **Are flexible designs sound?** *Biometrics* 2006, **62**(3):664–669, [<http://dx.doi.org/10.1111/j.1541-0420.2006.00626.x>]
18. Tsiatis AA, Mehta C: **On the inefficiency of the adaptive design for monitoring clinical trials.** *Biometrika* 2003, **90**(2):367–378, [<http://biomet.oxfordjournals.org/content/90/2/367.abstract>]
19. Jennison C, Turnbull BW: **Mid-course sample size modification in clinical trials based on the observed treatment effect.** *Stat Med* 2003, **22**(6):971–993, [<http://dx.doi.org/10.1002/sim.1457>]
20. Walton M: **Adaptive designs: opportunities, challenges and scope in drug development.** *PhRMA-FDA Workshop* 2006. [http://www.innovation.org/documents/File/Adaptive_Designs_Presentations/04_Walton_Adaptive_Designs_Trial_Issues_Goals_and_Needs.pdf]
21. Bischoff W, Miller F: **A seamless phase II/III design with sample-size re-estimation.** *J Biopharmaceutical Stat* 2009, **19**(4):595–609, [<http://www.tandfonline.com/doi/abs/10.1080/10543400902963193>]

doi:10.1186/1471-2288-13-94

Cite this article as: Broberg: Sample size re-assessment leading to a raised sample size does not inflate type I error rate under mild conditions. *BMC Medical Research Methodology* 2013 **13**:94.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

