Estimation in adaptive group sequential trials

Niklas Hack and Werner Brannath

January 29, 2015

Contents

1 Introduction

During the last years much research was spent on making mid-course corrections to the sample size of a clinical trial while the overall type I error rate of the test was preserved. Adaptive or flexible designs for clinical trials are attractive to clinical scientists and researchers since they provide a method to add flexibility to the frequentist paradigm. An important feature of adaptive designs is that the precise adaptation rule needs not to be pre-planed. [Müller and Schäfer, 2001] and [Müller and Schäfer, 2004] presented a general way to make adaptive changes to an on-going group sequential clinical trial while preserving the overall type I error rate. Their method allows to make data dependent changes to the sample size, the spending function and the number and spacing of interim looks at one or more time points. Adaptations can depend on the observed data up to the interim analysis and if no adaptation is performed the originally planned group sequential analysis can be applied. Only in the case of adaptations a modifed test statistic based on the conditional error rate has to be performed. In recent years there have been several approaches to calculate point estimates and confidence intervals following an adaptive change. [Mehta et al., 2006] proposed an approach for the calculation of repeated confidence intervals for adaptive group sequential trials. The [Müller and Schäfer, 2001] method is applied to the dual tests derived from the repeated confidence intervals (RCI) of [Jennison and Turnbull, 1989]. However, this method can only provide conservative coverage of the efficacy parameter δ. [Brannath et al., 2009] extended the stage-wise adjusted confidence intervals of [Tsiatis et al., 1984] to adaptive designs. Stage-wise adjusted confidence intervals provide exact coverage for classical group sequential designs. In the case of design adaptations it cannot be guaranteed that the stage-wise adjusted confidence interval provides exact coverage in general. The package *AGSDest* allows to compute repeated confidence intervals and p-values as well as confidence intervals and p-values based on the stage-wise ordering in groug sequential designs (GSD) and adaptive groug sequential designs (AGSD). The implemented principles allow us to perform data dependent changes to the sample size, the spending function, and the number and spacing of interim looks while preserving the overall type I error rate. Currently the procedures do not support the use of futility boundaries as well as more than one adaptive interim analysis. Furthermore, the package is currently restricted to the computation of one-sided confidence intervals.

2 Group Sequential Designs (GSD)

We consider a group sequential test (see for example, [Jennison and Turnbull, 2000]) for a comparative study of an experimental treatment E to a control treatment C, with a total of N normally distributed observations X_{il} , $i = E$ or $C, l = 1, 2, \ldots, N/2$, with known variance σ^2 . Let μ_E and μ_C denote the means based on a treatment E and a control C group and $\delta = \mu_E - \mu_C$ the difference of the population means. We focus on group sequential tests of the hypothesis

$$H_0 : \delta \leq 0$$

against the one-sided alternative $\delta > 0$. The trial is performed in K sequential

stages after observing the cumulative responses for $n_1, \ldots, n_K = N$ subjects. At stage j the data are summarized by the Wald statistics

$$Z_j = \hat{\delta}_j \sqrt{I_j}, j = 1, \ldots, K$$

where $\hat{\delta}_j$ is the maximum likelihood estimate of δ and $I_j \approx [se(\hat{\delta}_j)]^{-2} = n_j/(4\sigma^2)$ is the estimate of the Fisher information. We calculate sequentially for every interim analysis the Wald statistic $Z_1, \ldots, Z_K$. The trial stops at look j when the observed Wald statistic z_j is larger than the rejection boundary b_j. An α-spending function can be used to establish the boundaries $b_1, b_2, \ldots, b_K$ for each interim monitoring point, given the overall α. We denote by T the random variable which gives the stage where the trial stops.

2.1 GST object

Most of the functions for group sequential designs in this package need a `GST object` as input. A `GST object` is a collection of lists containing the design parameters of a group sequential design (GSD), namely:

GSD object:

K:	Number of stages
al:	Alpha (type I error rate)
a:	Lower critical bounds of group sequential design (are currently always set to -8)
b:	Upper critical bounds of group sequential design
t:	Vector with cumulative information fraction
SF:	Spending function (for details see help from `R-function bounds` (package: `ldbounds`))
phi:	Parameter of spending function when SF=3 or 4
alab:	Alpha-absorbing parameter values of group sequential design
als:	Alpha-values "spent" at each stage of group sequential design
Imax:	Maximum information number
delta:	Effect size used for planning the group sequential trial

Optionally, the object can also contain the group sequential design outcome (GSDo), which is necessary to calculate confidence bounds, p-values and point estimates (see next sections).

GSDo object:

T:	Stage where trial stops
z:	z-statistic at stage where trial stops

Furthermore, the package also provides the generic function `summary` (see next sections), which can be used to extend the `GST object` by the following quantities, e.g.:

cb.s	Confidence bound based on the stage-wise ordering
cb.r	Repeated confidence bound
pvalue.so	Stage-wise adjusted p-value
pvalue.r	Repeated p-value
est.ml	Maximum likelihood estimate
est.mu	Median unbiased point estimate
est.cons	Conservative point estimate

One of the basic R-functions of this package is the `plan.GST` function, which plans a GSD and creates a `GST object`.

2.1.1 Alpha spending function (SF)

Before we continue with the `plan.GST` function, we first describe the α-spending function, which is currently available in the package. Alpha spending functions establish α-values spent at each interim analysis given the overall α. The package supports the following spending functions, $\alpha(t)$:

O'Brien and Fleming type (1979)	$2(1-\Phi\left(\frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{t}}\right))$
Pocock type (1977)	$\alpha \cdot \log(1+(e-1)t)$
Kim and DeMats (1987)	$\alpha \cdot t^{\gamma}$
Hwang, Shih an DeCani (1990)	$\alpha\frac{1-e^{-\gamma t}}{1-e^{-\gamma}}$

2.2 plan.GST example

We consider a comparative study of an experimental treatment E to a control treatment C. Assume that the trial is planed as a three-look, one-sided group sequential design at level $\alpha = 0.025$. We initially want to test $H_0 : \delta \leq 0$ with 80% power to detect $\delta = 5$ with known standard deviation $\sigma = 15$. The stopping boundaries are derived from the γ -family proposed by Hwang, Shih and DeCani (1990) with $\gamma = -4$.

```
> library(AGSDest)
> GSD<-plan.GST(K=3,SF=4,phi=-4,alpha=0.025,delta=5,pow=0.8,compute.alab=TRUE,compute.als=
> GSD

3  stage group sequential design
 alpha :  0.025   SF:  4   phi:  -4   Imax:  0.32   delta:  5   cp:  0.8

Upper bounds          3.011  2.547  1.999
Lower bounds         -8.000 -8.000 -8.000
Information fraction  0.333  0.667  1.000


als  0.001 0.006 0.025
alab 3.222 1.194 0.000
```

The created `GST object` can now be plotted using the function `plot`.

```
> plot(GSD)
```

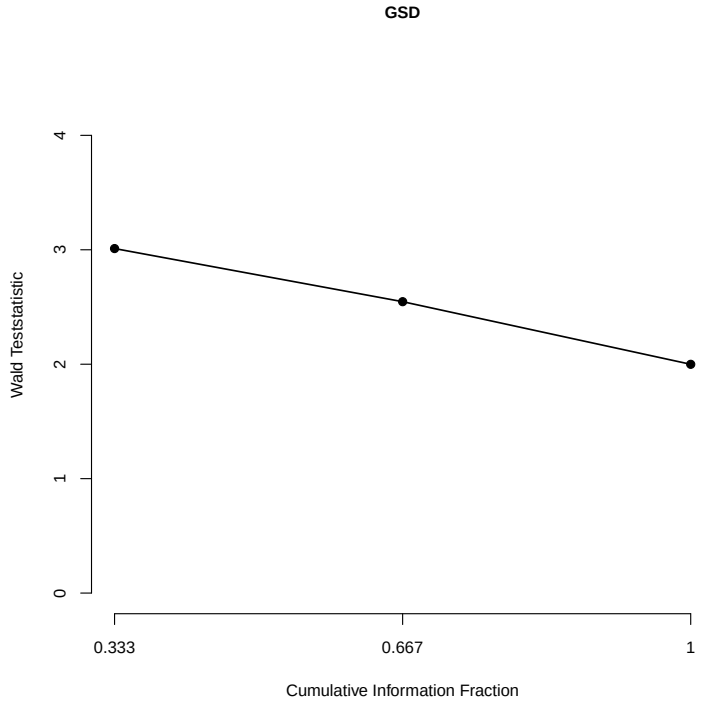


Figure 1: Group sequential design plot (Hwang, Shih and DeCani boundaries with $\gamma = -4$ at level 0.025) from example 2.2

2.3 Overall p-values for GSDs

An overall p-value can be defined via a family of nested hypotheses tests. A family of hypotheses tests is nested, if the rejection of the level-u test in the family implies the rejection of all level-u' tests where $u' > u$. An overall p-value q can be defined as the minimum of the levels of the tests which reject H_0 . In other words, we continue rejecting $H_0 : \delta \leq 0$ in a sequence of nested tests, with decreasing significant levels $0 < u < 1$, until we reach the level q, such that we cannot reject H_0. We are now introducing the repeated p-value and the p-value based on the stage-wise ordering.

2.3.1 Repeated p-values in a classical GSD

Repeated p-values have the advantage that they can be computed at any stage, whether the trial stops or not, but they are in general only conservative. For the repeated p-value the rejection boundaries of the trial can be specified via a spending function $g_u(t)$ that generates boundaries $b_{k,u}$ for all levels $0 < u < 1$ which are non-decreasing in u. We consider the GSD from section 2.1 in which the boundaries were determined from the γ-family spending function. The package determines from $g_u(t_j)$ the critical boundaries $b_{j,u}$ of a GSD at level u. This gives the family of nested hypotheses test. In order to obtain nested rejection regions we must have $b_{j,u} < b_{j,u'}$ for all $0 \leq u' < u \leq 1$. This requires a specific assumption on the spending function, which is satisfied for most spending functions including those of [Lan and DeMats, 1983], [Kim and DeMats, 1987] and [Hwang et al., 1990]. Now we can define the repeated p-value at stage k by

$$p_k = \inf\{u : z_k \geq b_{k,u}\} = \sup\{u : z_k < b_{k,u}\}$$

We consider the example from section 2.2 and calculate the repeated p-value at stage $T = 2$ assuming that $z = 1.088$. With `as.GST` we create a new object of `class GST`.

```
> GST<-as.GST(GSD=GSD,GSDo=list(T=2, z=1.088))
> GST

3  stage group sequential design
 alpha :  0.025   SF:  4   phi:  -4   Imax:  0.32   delta:  5   cp:  0.8

Upper bounds          3.011  2.547  1.999
Lower bounds         -8.000 -8.000 -8.000
Information fraction  0.333  0.667  1.000

als  0.001 0.006 0.025
alab 3.222 1.194 0.000


group sequential design outcome:

        T:  2   z:  1.088
```

Now we call the `pvalue` function with the new created GST-object and set `type` equal to 'r' to calculate the repeated p-value.

```
> pvalue(GST,type="r")

$pvalue.r
[1] 0.5834961
```

2.3.2 P-value based on the stage-wise ordering in a classical GSD

Stage-wise p-values have exact coverage probability, but they can only be calculated at the stage where the trial stops according to the prespecified stopping rule. Assume that the primary trial stops at look T. The stage wise ordering considers a sample point (j, z_j) as more extreme than the sample point (k, z_k), if either $j < k$ or $j = k$ and $z_j \geq z_k$. This ordering can be used to define an overall p-value p for H_0 as

$$p = P_0 \left(\bigcup_{j=1}^{T-1} \{Z_j \geq b_j\} \cup \{Z_T \geq z_T\} \right)$$

which is the probability under H_0 to get a more extreme sample point (in the sense of the stage wise ordering) than the one we have observed. We consider the same design as in section 2.2, but assume now that the trial stops at the stage $T = 2$ with the z-statistic $z = 2.63$ and hence can stop the trial and calculate the stage-wise adjusted p-value by setting the `type` to 'so'.

```
> pvalue(as.GST(GSD,list(T=2,z=2.63)),type='so')

$pvalue.so
[1] 0.005131236
```

2.4 Construction of one sided confidence intervals for GSDs

2.4.1 Classical repeated confidence bounds

The classical repeated confidence interval for a given group sequential design, was proposed by [Jennison and Turnbull, 1989]. It has the advantage that it can be computed at any stage, whether the trial stops or not, but it has only conservative coverage probability. This repeated confidence interval is defined by a family of dual significance tests for the hypothesis $H_h : \delta \leq h$ versus $\delta > h$ for all $h \in (-\infty, \infty)$. The confidence interval includes all values of h where the shifted null hypothesis H_h is not rejected. First, we sequentially compute the shifted Wald statistics $Z_j(h) = Z_j - h\sqrt{I_j}, j = 1, \ldots, K$, where I_j is the cumulated information until stage j . It is known that $Z_j(h)$ is $N(0, 1)$-distributed under H_h. Now we apply the same group sequential design to all h. At stage j we reject H_h if $Z_j - h\sqrt{I_j} \geq b_j$, i.e., we reject all $h \leq \frac{Z_j - b_j}{\sqrt{I_j}}$. Hence, the lower confidence bound at each step $j = 1, \ldots, K$ of the one-sided confidence interval is

$$(\underline{\delta}_j, \inf), j = 1, \ldots, K \quad \text{with } \underline{\delta}_j = \frac{Z_j - b_j}{\sqrt{I_j}}$$

We assume the example from section 2.3.1 and calculate the repeated confidence bound by setting the `type` to 'r'.

```
> seqconfint(GST,type='r')

$cb.r
[1] -3.162014
```

2.4.2 Classical stage-wise confidence bounds

Stage-wise confidence intervals have exact coverage probability, however they can only be calculated at the stage where the trial stops according to the pre-specified stopping rule. The stage-wise adjusted confidence interval also provides at level 0.5 a median unbiased point estimate for δ. The stage-wise ordering can be used to define an overall p-value for H_h as

$$p(h) = P_h\left(\bigcup_{j=1}^{T-1}\{Z_j \geq b_j\} \cup \{Z_T \geq z_T\}\right)$$

By definition $p(h)$ has an uniform distribution under H_h. Since $p(h)$ is strictly increasing in h, the equation $p(h) = \alpha$ has a unique solution. We perform a level-α test for H_h, if we reject H_h in the case that $p(h) \leq \alpha$, and otherwise accept H_h. We consider the example from section 2.2 and calculate the stage-wise confidence bound at stage $T = 2$ with the observed z-statistic $z = 2.63$ by setting the `type` to 'so'.

```
> seqconfint(as.GST(GSD,list(T=2,z=2.63)),type='so')

$cb.so
[1] 1.356988
```

2.5 Point estimates for GSDs

2.5.1 Median unbiased point estimate

Median unbiased point estimates are exact, but they can only be calculated at the stage where the trial stops according to the prespecified stopping rule. To calculate the point estimate $\underline{\delta}_{0.5}$ based on the stage-wise ordering we compute the lower stage-wise confidence bound at level 0.5. If the GSD stops at stage T, then $\underline{\delta}_{0.5}$ is the value of h that satisfies $p(h) = 0.5$. We assume the example from section 2.2 and calculate the median unbiased point estimate at stage $T = 2$ with the observed z-statistic $z = 2.63$ by setting the `type` to 'so' and the `level` to 0.5.

```
> seqconfint(as.GST(GSD,list(T=2,z=2.63)),type="so",level=0.5)

$est.mu
[1] 5.659091
```

2.5.2 Conservative point estimate

To calculate the conservative point estimate, we compute the lower repeated confidence bound at level 0.5. This point estimate is flexible, in the sense that it can be calculated at every stage of the trial and not only at the stage T where the trial stops. However, in general it's conservative in the sense that

its median can be below the true parameter value (but is assumed to be never above the true value). Hence we may overestimate the true value but only with a probability lower than 50%. We assume the example from section 2.2 and calculate the conservative unbiased point estimate at stage $T = 2$ with the observed z-statistic $z = 1.088$ by setting the `type` to 'r' and the `level` to 0.5.

```
> seqconfint(as.GST(GSD,list(T=2,z=1.088)),type='r',level=0.5)

$est.cons
[1] -0.2121496
```

2.6 Summary function for a GST object

As aforementioned the package also provides a generic summary function which takes as input a `GST object` and additional parameters. This summary function produces the results from the different functions for GSDs, e.g.: confidence bounds, p-values and point estimates. By specifying the `type` (`ctype`, `ptype` and `etype`), the user can define which values are calculated:

Confidence bounds and p-values:
- ctype: Confidence type
- ptype: P-value type

Possible value for these two parameters are:
- r: Repeated
- so: Stage-wise adjusted

Point estimates (etype):
- ml: Maximum likelihood estimate (ignoring the sequential nature of the design)
- mu: Median unbiased estimate (stage-wise lower confidence bound at level 0.5) for a classical GSD
- cons: Conservative estimate (repeated lower confidence bound at level 0.5) for a classical GSD

If no type is specified the summary function calculates by default all values.

We assume the example from section 2.2 where we stop at stage $T = 2$ with the observed z-statistic $z = 2.63$. With `as.GST` we create a new `GST object` and pass this object to the summary function. Now we want to calculate stage-wise adjusted confidence bound, the stage-wise adjusted p-values, but no point estimates. If we assign the output from the summary function to the new created `GST object`, the object gets extended by the calculated values.

```
> GSD1<-as.GST(GSD,list(T=2,z=2.63))
> GSD1<-summary(GSD1,ctype='so',ptype='so',etype=NULL)
> GSD1

stage-wise adjusted lower confidence bound:  1.357

stage-wise adjusted p-value:  0.005

3  stage group sequential design
 alpha :  0.025   SF:  4   phi:  -4   Imax:  0.32   delta:  5   cp:  0.8

Upper bounds          3.011  2.547  1.999
```

```
Lower bounds         -8.000 -8.000 -8.000
Information fraction  0.333  0.667  1.000

als  0.001 0.006 0.025
alab 3.222 1.194 0.000


group sequential design outcome:

        T:  2   z:  2.63
```

3 Adaptive group sequential Design (AGSD)

3.1 Müller and Schäfer method

[Müller and Schäfer, 2001] presented a general method for the full integration of the concept of adaptive interim analyses [Bauer and Kuehne, 1994] into group sequential testing. This method allows to change statistical design elements of a given group sequential design such as the α-spending function and the number of interim analyses, without effecting the type I error rate. The method is described by statistical decision functions and is based on the conditional rejection probability of a decision variable.

The conditional rejection probability gives the conditional probability to finally reject the null hypothesis given the interim data, assuming that the null hypothesis is true. To explain the method, consider as in the previous section the case of a comparative study of an experimental treatment E to a control treatment C with means μ_E and μ_C and common known variance σ^2 . As before assume a group sequential trial with $H_0 : \delta \leq 0$ against the one-sided alternative $H_A : \delta > 0$ and a maximum of K stages. Let us assume that the trial continues until stage $L < K$ without rejection, i.e., $z_j < b_j$ for all $j \leq L$, where z_j is the observed value of the Wald test statistic Z_j from stage j. Let us further assume that one decides to make data dependent changes to the study design at look L. Let R denote the event that H_0 will be rejected at any future analyses $j = L + 1, \ldots, K$. R can be written as the union of disjoint events

$$R = \bigcup_{i=L+1}^{K} R_i$$

where

$$R_i = \{Z_i \geq b_i \text{ and } Z_j < b_j \text{ for all } j < i\}$$

The conditional probability for H_0 of the event R given Z_j for $j \leq L$, is called conditional rejection probability. It can formally be written as

$$\epsilon(0) = P_0(R|Z_1 = z_1, \ldots, Z_L = z_L).$$

We now plan a new group sequential design at level $\epsilon(0)$. This trial starts at stage L and is based on a patient cohort which is independent from the cohort of patients recruited up to look L. This trial can be seen as a new, independent 'secondary' trial in which the sample size is initialized to zero and the type I error is equal to $\epsilon(0)$. The Wald z-statistics for the secondary trial are only based on the data observed after the stage of the adaptation L. We will distinguish the secondary trial from the original 'primary' trial by labeling the stages, sample sizes, stopping boundaries and test statistics by the superscript '(2)'. Assume that the secondary trial has a maximum number of $K^{(2)}$ stages, cumulated information numbers $I_j^{(2)}, j = 1, \ldots, K^{(2)}$ and rejection boundaries $b_j^{(2)}, j = 1, \ldots, K^{(2)}$. The boundaries for the secondary group sequential trial have to be chosen in such a way, that the resulting test procedure has type I error $\epsilon(0)$, i.e.,

$$\epsilon(0) = P_0\left(\bigcup_{j=L+1}^{K^{(2)}} \{Z_j^{(2)} \geq b_j^{(2)}\}|Z_1 = z_1, \ldots, Z_L = z_L\right)$$

Assume that the secondary trial terminates at look $T^{(2)} \leq K^{(2)}$ with the observed test-statistic $Z^{(2)}_{T^{(2)}} = z^{(2)}_{T^{(2)}}$. Now, the null hypothesis is rejected if and only if $z^{(2)}_{T^{(2)}} \geq b^{(2)}_{T^{(2)}}$. Note that the conditional rejection probability is the only information which is carried over to the secondary trial.

3.2 AGST object

Most of the functions for adaptive group sequential designs (AGSD) in this package need an `AGST object` as input. An `AGST object` is a collection of lists containing the design parameters of the primary trial (pT), the interim data (iD) and the design parameters of the secondary trial (sT), namely:

pT object:

K:	Number of stages
al:	Alpha (type I error rate)
a:	Lower critical bounds of group sequential design (are currently always set to -8)
b:	Upper critical bounds of group sequential design
t:	Vector with cumulative information fraction
SF:	Spending function (for details see help from `R-function bounds` (package: `ldbounds`))
phi:	Parameter of spending function when SF=3 or 4
alab:	Alpha-absorbing parameter values of group sequential design
als:	Alpha-values "spent" at each stage of group sequential design
Imax:	Maximum information number
delta:	Effect size used for planning the primary trial

iD object:

L:	Stage of the adaptation
z:	z-statistic at adaptive interim analysis

sT object:

K:	Number of stages
al:	Conditional rejection probability
a:	Lower critical bounds of secondary group sequential design (are currently always set to -8)
b:	Upper critical bounds of secondary group sequential design
t:	Vector with cumulative information fraction
SF:	Spending function (for details see help from `R-function bounds` (package: `ldbounds`))
phi:	Parameter of spending function when SF=3 or 4
Imax:	Maximum information number
delta:	Effect size used for planning the secondary trial

Optionally, the object can also contain the secondary trial outcome (sTo), which is necessary to calculate confidence bounds, p-values and point estimates.

sTo object:

T:	Stage where secondary trial stops
z:	z-statistic at stage where secondary trial stops

Furthermore, the package also provides the generic function summary (see next sections), which can be used to extend the `AGST object` by the following quantities, e.g.:

cb.s	Confidence bound based on the stage-wise ordering
cb.r	Repeated confidence bound
pvalue.so	Stage-wise adjusted p-value
pvalue.r	Repeated p-value
est.ml	Maximum likelihood estimate
est.mu	Median unbiased point estimate
est.cons	Conservative point estimate

The function `as.AGST` can be used to create an object of `class AGST`.

3.2.1 adapt example

We continue with the example from the section 2.2 and suppose that at the first interim analysis, after $n_1 = 95$ subjects in total (both groups together) have been evaluated, the estimate of δ is $\hat{\delta}_1 = 3$ with the estimated standard deviation $\hat{\sigma}_1 = 20$ which gives $z_1 = 0.731$. Since the observed $\hat{\delta}_1$ is below the anticipated δ and $\hat{\sigma}_1$ is higher, we decide to increase the sample size. As described above we set the significance level of the secondary trial equal $\epsilon(0)$ to control the type I error rate. The sample size is calculated on the bases of $\delta = 4$, which is the mean of the original δ_0 and the interim estimate $\delta_1 = 3$, with $\sigma = 20$ and a power of 90%. In order to calculate the conditional rejection probability $\epsilon(0)$ we first have to define the primary trial (`pT`), which is the originally planed GSD and the interim data(`iD`), which is the data we observed at the interim analyses. For the new secondary trial we changed the spending function from the Hwang-Shih-DeCani family (SF=4) to the O'Brien and Fleming type spending function (SF=1) to have a higher change for early rejection. Furthermore, we increased the power from 80% to 90%, based on the new effect size of $\delta = 4$.

```
> pT=GSD
> iD=list(T=1, z=0.731)
```

The function `cer` calculates the conditional rejection probability of `pT` given `iD`.

```
> cer(pT,iD)
```

```
[1] 0.02739815
```

The secondary trial can be planned with the function `adapt`. For safety reasons, we aim on a stage wise sample size of at most 200 patients in the secondary trial. This implies a maximum for the incremental information of the sequential steps, which can be calculated as:

```
> swImax=200/(4*20^2)
```

With `I2min` and `I2max`, we define the minimal and maximal total information for the secondary trial. These numbers can be determined by a minimum and maximum of steps and `swImax`. We aim on a minimum of 2 and a maximum of 5 stages for the secondary trial. If `I2max` is to small to reach the specified conditional power `cp` the functions returns a warning.

```
> I2min=2*swImax
> I2max=5*swImax
> sT=adapt(pT=pT,iD=iD,SF=1,phi=0,cp=0.9,theta=4,I2min=I2min,I2max=I2max,swImax=swImax)
> sT
```

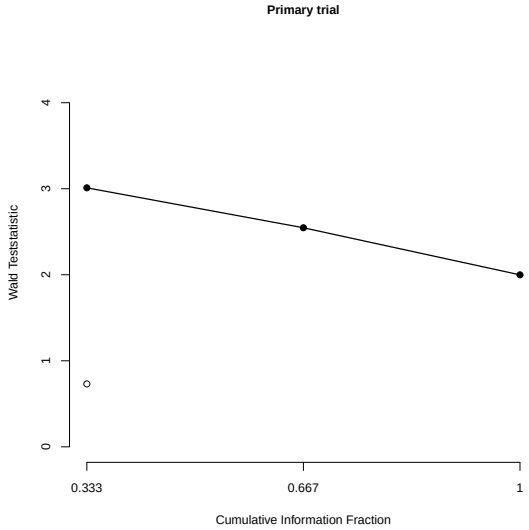
```
5  stage group sequential design
 alpha :  0.027   SF:  1   phi:  0   Imax:  0.62   delta:  4   cp:  0.89

Upper bounds          4.795  3.298  2.632  2.248  1.994
Lower bounds         -8.000 -8.000 -8.000 -8.000 -8.000
Information fraction  0.200  0.400  0.600  0.800  1.000
```

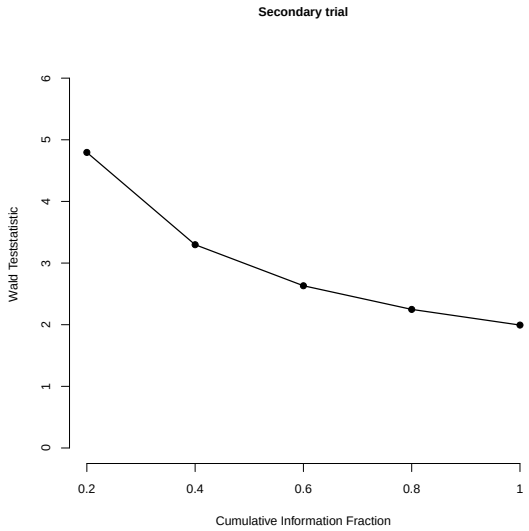
```
> AGSD<-as.AGST(pT,iD,sT)
```

The created `AGST object` can now be plotted using the function `plot`.

```
> plot(AGSD)
```



(a) Primary trial plot (Hwang, Shih and DeCani boundaries with $\gamma = -4$ at level 0.025 and the observed z-statistic $z = 0.731$ at stage $T = 1$) from example 3.2.1



(b) Secondary trial plot (O'Brien-Fleming boundaries at level 0.027) from example 3.2.1

Figure 2: Plots from example 3.2.1

3.3 Overall p-values with adaptations

3.3.1 Repeated p-values

Repeated p-values can be defined at every interim look j of an adaptive secondary trial and not just at the look $T^{(2)}$ where the trial terminates. However, they produce conservative tests. Let us assume that we perform some design adaptations at stage L. The conditional type I error rate for the test at level u is then given by

$$\epsilon_u = \begin{cases} 0 & \text{if } u \leq \alpha_L \\ P_0(\bigcup_{j=L+1}^{K} \{Z_j \geq b_{j,u}\} | Z_1 = z_1, \ldots, Z_L = z_L) & \text{if } u > \alpha_L \end{cases}$$

Let $p^{(2)}$ denote the repeated p-value of the secondary trial of the stage $T^{(2)}$ where the trial stops, i.e.,

$$p^{(2)} = \inf(u : z^{(2)}_{T^{(2)}} \geq b^{(2)}_{T^{(2)},u})$$

where $b^{(2)}_{k,u}$ is from the monotone family of boundaries from the spending function for the secondary trial. Now the overall p-value, considering the data from the primary and secondary trial, is defined by

$$q = \inf\{u : p^{(2)} \leq \epsilon_u\}$$

ϵ_u is increasing in u (if all $b_{j,u}$'s are decreasing in u) and hence the corresponding adaptive level-u tests are nested. Therefore the p-value can be computed as the solution of the equation $p^{(2)} = \epsilon_u$.
We continue with the example from section 3.2.1 and compute the repeated p-value for the adaptive design. We assume that we want to calculate the p-value at stage $T = 2$ with an observered test-statistic of $z = 1.532$. Before we can calculate the p-value we have to include in the `AGST object` a list containing the outcome from the secondary trial (sTo). With the now created `object` from class `AGST` we can calculte the repeated p-value after a design adaptation.

```
> sTo=list(T=2,z=1.532)
> AGSD<-as.AGST(pT,iD,sT,sTo)
> pvalue(AGSD,type='r')

$pvalue.r
[1] 0.1645508
```

3.3.2 P-values based on the stage-wise ordering with adaptations

Stage-wise p-values are exact, but they can only be calculated at the stage where the trial stops according to the prespecified stopping rule. In the case of a design adaptation at look L we compute the corresponding conditional error functions

$$\epsilon_u = \begin{cases} 0 & \text{if } u \leq \alpha_L \\ P_0(\bigcup_{j=L+1}^{k-1} \{Z_j \geq b_j\} \cup \{Z_k \geq b_{k,u}\} | Z_1 = z_1, \ldots, Z_L = z_L) & \text{if } \alpha_{k-1} < u < \alpha_k, k = L+1, \ldots, K \end{cases}$$

where $b_{k,u}$ defines the "threshold boundary" in such a way that it satisfies the relationship

$$P_0\left(\bigcup_{j=1}^{k-1}\{Z_j \geq b_j\} \cup \{Z_k \geq b_{k,u}\}\right) = u$$

Let $p^{(2)}$ denote the stage-wise adjusted p-value of the secondary trial at the stage $T^{(2)}$ where the trial stops, i.e.,

$$p^{(2)} = P_0\left(\bigcup_{j=1}^{T^{(2)}-1}\{Z_j^{(2)} \geq b_j^{(2)}\} \cup \{Z_{T^{(2)}}^{(2)} \geq z_{T^{(2)}}^{(2)}\}\right)$$

Now we can calculate the overall p-value by

$$q = \inf\{u : p^{(2)} \leq \epsilon_u\} = \sup\{u : p^{(2)} > \epsilon_u\}$$

We continue with the example from section 3.2.1 and calculate the stage wise adjusted p-value for a group sequential trial with design adaptations. We assume that the trial stops at stage $T = 3$ with the observered test-statistic $z = 2.73$.

```
> AGSD1<-as.AGST(pT,iD,sT,list(T=3,z=2.73))
> pvalue(AGSD1,type='so')

$pvalue.so
[1] 0.007435759
```

3.4 Construction of one-sided confidence intervals

3.4.1 Repeated confidence bounds with adaptations

Repeated confidence bounds have the advantage that they can be computed at any stage, whether the trial stops or not, but they have only conservative coverage probability. In the case of a design adaptation we apply the Müller and Schäfer principle to all dual tests. Collecting all h's where $H_h : \delta \leq h$ is accepted, gives the $1-\alpha$ confidence interval. To obtain this confidence interval we shift the observed test statistic of the primary trial to

$$z_j(h) = z_j - h\sqrt{I_j}, \quad j = 1, \ldots, L$$

and the test-statistic observed in the secondary trial is shifted to

$$z_j^{(2)}(h) = z_j^{(2)} - h\sqrt{I_j^{(2)}}, \quad j = 1, \ldots, T^{(2)}$$

Now, the conditional rejection probability can be calculated by

$$\epsilon(h) = P_0\left(\bigcup_{j=L+1}^{K}\{Z_j \leq b_j\} | Z_1 = z_1 - h\sqrt{I_1}, \ldots, Z_L = z_L - h\sqrt{I_L}\right)$$

With the [Müller and Schäfer, 2001] principle we can define the family of dual tests for H_h with the rejection rule

$$p^{(2)}(h) \leq \epsilon(h),$$

where $p^{(2)}(h)$ is a p-value of the secondary trial for the shifted test statistic $z^{(2)} - h\sqrt{I_j^{(2)}}$. To preserve the flexibility of the repeated confidence intervals we use the repeated p-value for $p^{(2)}(h)$. Applying the upper equation to all values of h gives the one-sided confidence interval $(\underline{\delta}, \infty)$ where $\underline{\delta}$ is the unique solution of $p^{(2)}(h) = \epsilon(h)$ in h. We assume the example from section 3.3.1 and calculate the repeated confidence bound.

```
> seqconfint(AGSD,type='r')
```

```
$cb.r
[1] -2.063108
```

3.4.2 Stage-wise confidence bounds with adaptations

Stage-wise confidence intervals have exact coverage probability, however they can only be calculated at the stage where the trial stops according to the prespecified stopping rule. Hence, with the stage-wise confidence intervals we cannot deviate from the pre-specified stopping rule. The stage-wise adjusted confidence intervals also provide a less conservative point estimate for δ. Let us now assume that we want to perform some design adaptations at look L. Recall that we have to compute the conditional type I error rate

$$\epsilon(0) = P_0\left(\bigcup_{j=L+1}^{K}\{Z_j \geq b_j\}|Z_1 = z_1, \ldots, Z_L = z_L\right).$$

In order to test $H_h : \delta \leq h$ at level α we apply the [Müller and Schäfer, 2001] principle for any given h by computing the conditional error function $\epsilon(h)$ of the test for $H_h : \delta \leq h$. The determination of $\epsilon(h)$ is now more complex and we refer to [Brannath et al., 2009] for details. We assume that the secondary trial stops at look $T^{(2)}$. Then we compute the p-value according to the stage-wise ordering of the secondary trial as

$$p^{(2)}(h) = P_h\left(\bigcup_{j=1}^{T^{(2)}-1}\{Z_j^{(2)} \geq b_j^{(2)}\} \cup \{Z_{T^{(2)}}^{(2)} \geq z_{T^{(2)}}^{(2)}\}\right)$$

With this new p-value we can define the dual test in such a way that
$H_h : \delta \leq h$ is rejected if and only if $p^{(2)}(h) \leq \epsilon(h)$. With the above adaptive tests for H_h it is now possible to compute the lower confidence bound δ in the case of an adaptive change at look L. We build the confidence set of all parameter values h that were accepted, i.e., $p^{(2)}(h) > \epsilon(h)$. We have the problem that $p^{(2)}(h) = \epsilon(h)$ may have more than one solution. The reason is the non-monotonicity of $\epsilon(h)$ (see [Brannath et al., 2009]). Thus we define δ as the smallest solution of $p^{(2)}(h) = \epsilon(h)$ which gives a conservative lower confidence bound. The conservatism was found to be natural in simulation studies. We consider the numerical example from section 3.3.2 and calculate the stage-wise adjusted confidence interval.

```
> seqconfint(AGSD1,type='so')
```

```
$cb.so
[1] 0.8017689
```

3.5 Point estimates with adaptations

3.5.1 Median unbiased point estimates with adaptations

Median unbiased point estimates with adaptations are almost exact, but they can only be calculated at the stage where the trial stops according to the pre-specified stopping rule. We consider the numerical example from section 3.3.2 and calculate the median unbiased point estimate.

```
> seqconfint(AGSD1,type="so",level=0.5)

$est.mu
[1] 3.799511
```

3.5.2 Conservative point estimates with adaptations

To calculate the conservative point estimate after an adaptation, we compute the lower repeated confidence bound at level 0.5 after an adaptation. This point estimate is flexible, in the sense that it can be calculated at every stage of the trial and not only at the stage T where the trial stops. Its median is never above the true parameter value, but can be below it. We consider the numerical example from section 3.3.1 and calculate the conservative point estimate.

```
> seqconfint(AGSD,type="r",level=0.5)

$est.cons
[1] 1.88595
```

3.6 Summary function for an AGST object

As aforementioned the package also provides a generic summary function which takes as input a `AGST object` and additional parameters. This summary function produces result summaries of the different functions for AGSDs, e.g.: confidence bounds, p-values and point estimates. By specifying the `type` (`ctype`, `ptype` and `etype`), the user can define which values are calculated:

Confidence bounds and p-values:
- ctype: Confidence type
- ptype: P-value type

Possible value for these two parameters are:
- r: Repeated
- so: Stage-wise adjusted

Point estimates (etype):
- ml: Maximum likelihood estimate (ignoring the sequential nature of the design)
- mu: Median unbiased estimate (stage-wise lower confidence bound at level 0.5) for an AGSD
- cons: Conservative estimate (repeated lower confidence bound at level 0.5) for an AGSD

If no type is specified the summary function calculates by default all values.

We assume the example from section 3.3.2. With `as.AGST` we create a new `AGST object` and pass this object to the `summary` function. Now we want to calculate the stage-wise adjusted confidence bound, the stage-wise adjusted p-value,

and all point etsimates (maximum likelihood, median unbiased and conservative point estimate). If we assign the output from the `summary` function to the new created `AGST object`, the object gets extended by the calculated values.

```
> AGSD1<-summary(AGSD1,ctype='so',ptype='so',etype=c('ml', 'mu', 'cons'))
> AGSD1

stage-wise adjusted lower confidence bound:  0.802

stage-wise adjusted p-value:  0.007

maximum likelihood estimate:  3.968

median unbiased estimate:  3.8

conservative estimate:  3.24

Primary trial:

3  stage group sequential design
 alpha :  0.025   SF:  4   phi:  -4   Imax:  0.32   delta:  5   cp:  0.8

Upper bounds          3.011  2.547  1.999
Lower bounds         -8.000 -8.000 -8.000
Information fraction  0.333  0.667  1.000


als  0.001 0.006 0.025
alab 3.222 1.194 0.000



interim data:

        T:  1   z:  0.731

Secondary trial:

5  stage group sequential design
 cer:  0.027   SF:  1   phi:  0   Imax:  0.62   delta:  4   cp:  0.89

Upper bounds          4.795  3.298  2.632  2.248  1.994
Lower bounds         -8.000 -8.000 -8.000 -8.000 -8.000
Information fraction  0.200  0.400  0.600  0.800  1.000


Secondary trial outcome:

        T:  3   z:  2.73
```

References

[Bauer and Kuehne, 1994] Bauer, P. and Kuehne, K. (1994). Evalution of experiments with adaptive interim analyses. *Biometrics*, 50(1):1029–1041.

[Brannath et al., 2009] Brannath, W., Mehta, C. R., and Posch, M. (2009). Exact confidence bounds following adaptive group sequential tests. *Biometrics*, 65(2):539–546.

[Hwang et al., 1990] Hwang, I., Shih, W., and DeCani, J. (1990). Group sequential designs using a family of type i error probability spending functions. *Statistics in Medizine*, 9(1):1439–1445.

[Jennison and Turnbull, 1989] Jennison, C. and Turnbull, B. (1989). Interim analyses: the repeated confidence interval approach (with discussion). *J.R. Statist. Soc. B*, 51(1):305–361.

[Jennison and Turnbull, 2000] Jennison, C. and Turnbull, B. (2000). *Group sequential methods with applications to clinical trials*. Chapman Hall,Boca Raton, London, New York, Washington, D.C.

[Kim and DeMats, 1987] Kim, K. and DeMats, D. (1987). Confidende intervals following group sequential tests in clinical trials. *Biometrics*, 43(1):857–864.

[Lan and DeMats, 1983] Lan, K. and DeMats, D. (1983). Discrete sequential boundaries for clinical trials. *Biometrics*, 70(1):659–663.

[Mehta et al., 2006] Mehta, C., Bauer, P., Posch, M., and Brannath, W. (2006). Repeated confidence intervals for adaptive group sequential trials. *Statistics in Medicine*, 26(1):5422–5433.

[Müller and Schäfer, 2001] Müller, H.-H. and Schäfer, H. (2001). Adaptive group sequential design for clinical trials: Combining the advantages of adaptive and of classic group sequential approaches. *Biometrics*, 57(1):886–891.

[Müller and Schäfer, 2004] Müller, H.-H. and Schäfer, H. (2004). A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine*, 23(1):2497–2508.

[Tsiatis et al., 1984] Tsiatis, A., Rosner, G., and Mehta, C. (1984). Exact confidence intervals following a group sequential test. *Biometrics*, 40(3):797–803.