Group sequential designs for clinical trials

Alex Asher

5 December, 2023

STATA 18

Intro: clinical trials

- Clinical trials
  - studies examining the effects of medical treatments on living beings (usually humans)
  - participants are usually accrued over time
  - ethical considerations

image: Journal of Pharmacology & Clinical Research, licensed under CC BY 4.0

Intro: clinical trial jargon

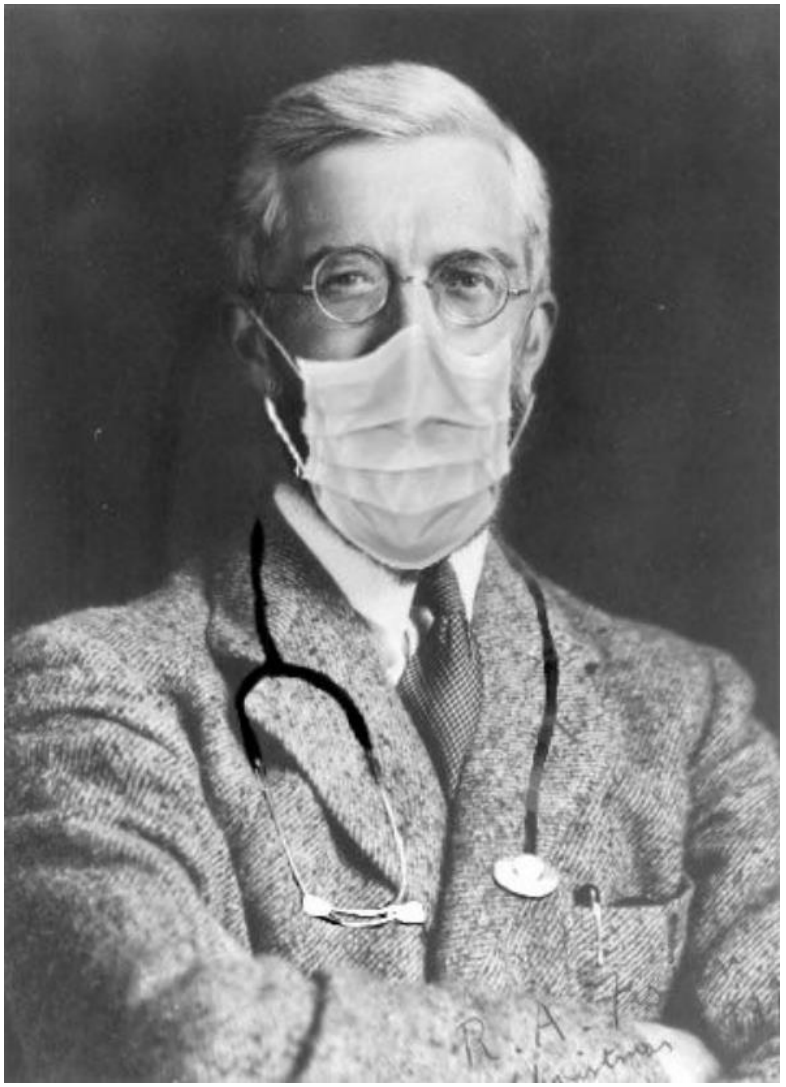
- Endpoint
  - the outcome variable
- Arm
  - treatment group
- Randomization
  - random allocation of participants to arms
- RCT
  - randomized controlled trial

Intro: fixed-sample designs (FSD)

- Predetermined sample size
- No analysis until all data collected

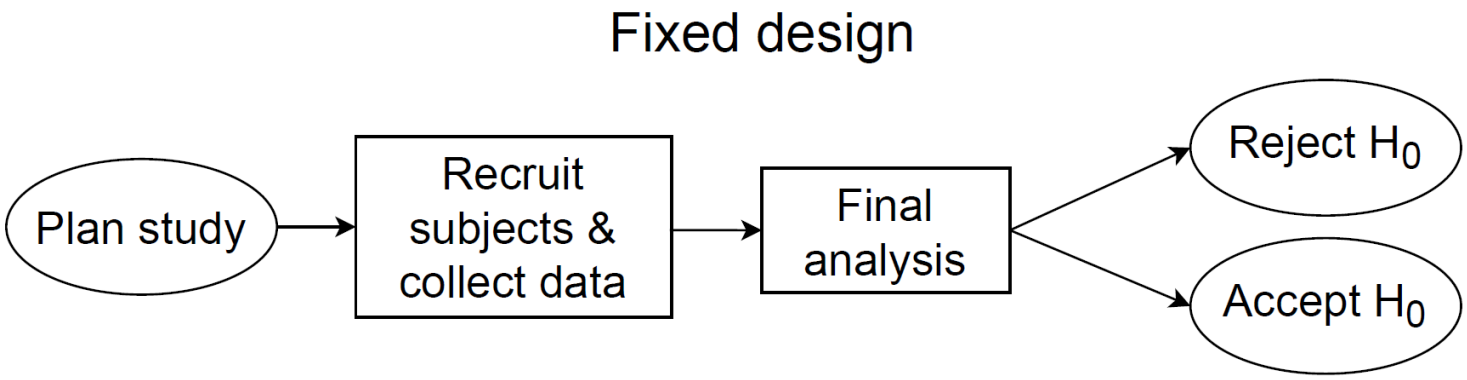
Intro: group sequential designs (GSD)

- Analyze data from an ongoing trial
- Early stopping for efficacy or futility
- Sample size depends on the data



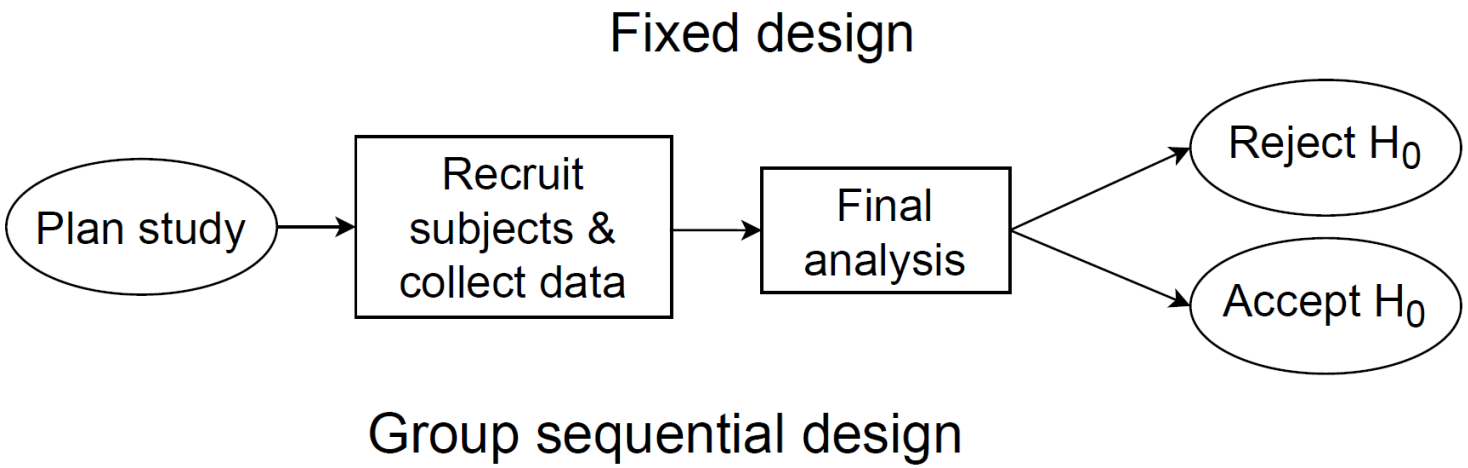
Intro: FSD vs GSD

or: how I learned to stop worrying and accept the null



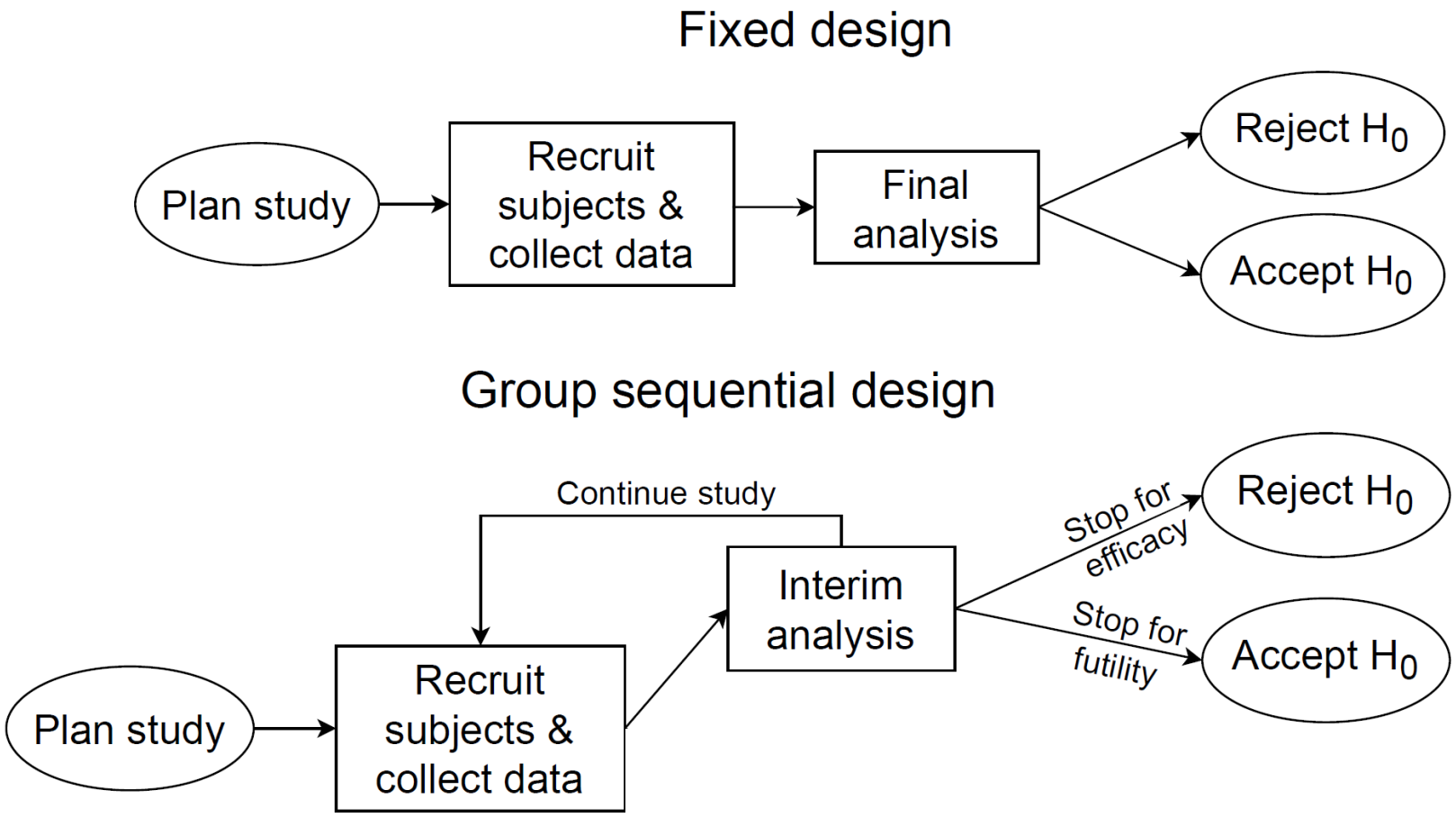
Intro: FSD vs GSD

or: how I learned to stop worrying and accept the null



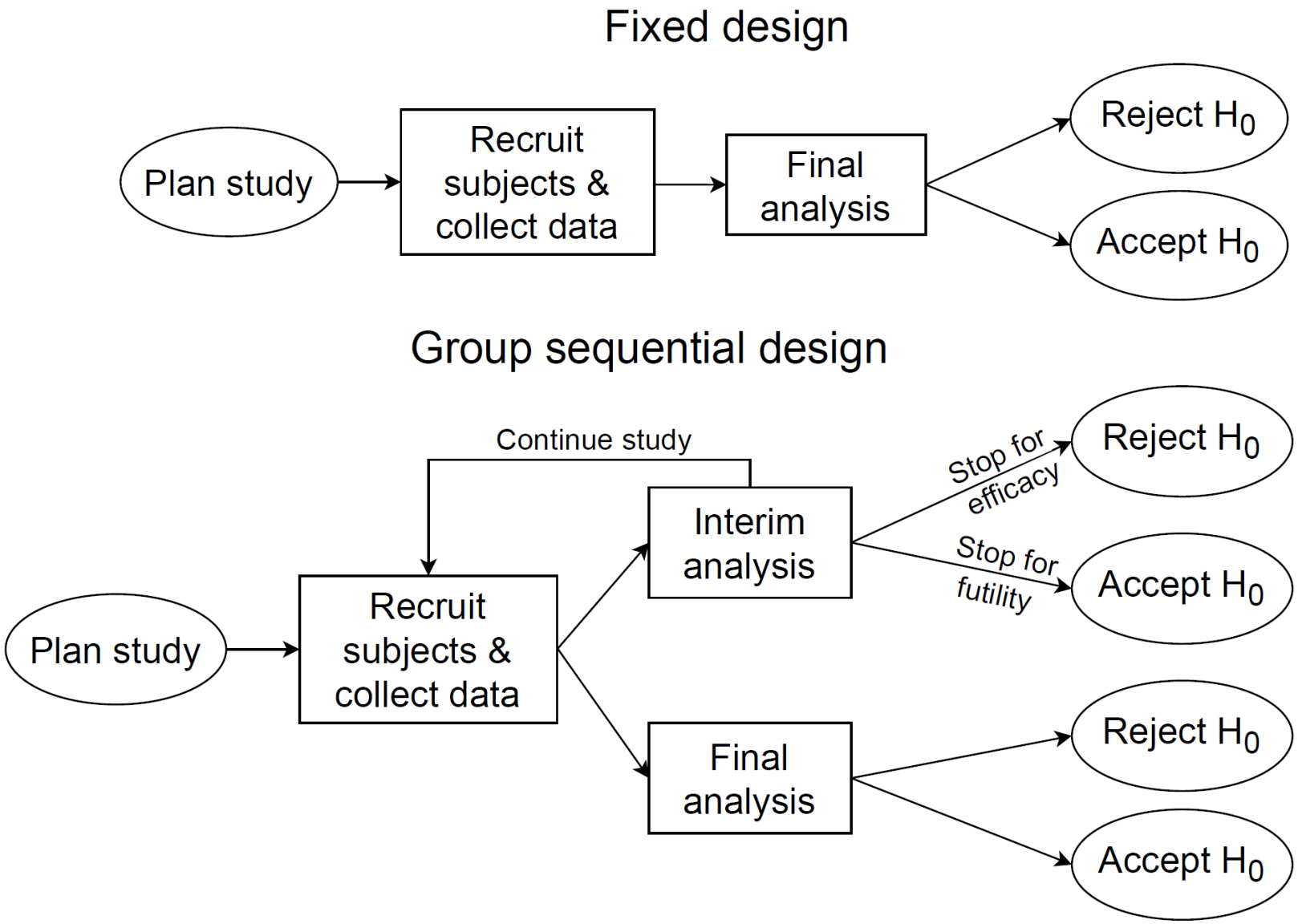
Intro: FSD vs GSD

or: how I learned to stop worrying and accept the null



Intro: FSD vs GSD

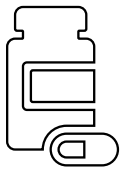
or: how I learned to stop worrying and accept the null

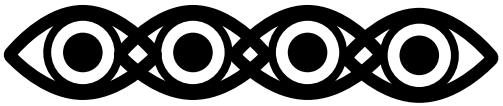


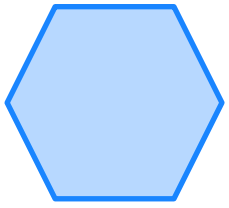
Intro: GSD jargon

- Look
  - analysis of the data to-date
- Expected sample size (ESS)
  - average sample size if the trial were to be repeated
- Maximum information
  - Fisher information of the parameter estimated for the test, calculated at the maximum sample size
- Information fraction
  - fraction of the maximum information that has been collected when a look is performed; proportional to sample size
- Information ratio
  - ratio of the maximum information (sample size) of a GSD to the information (sample size) of an equivalent FSD

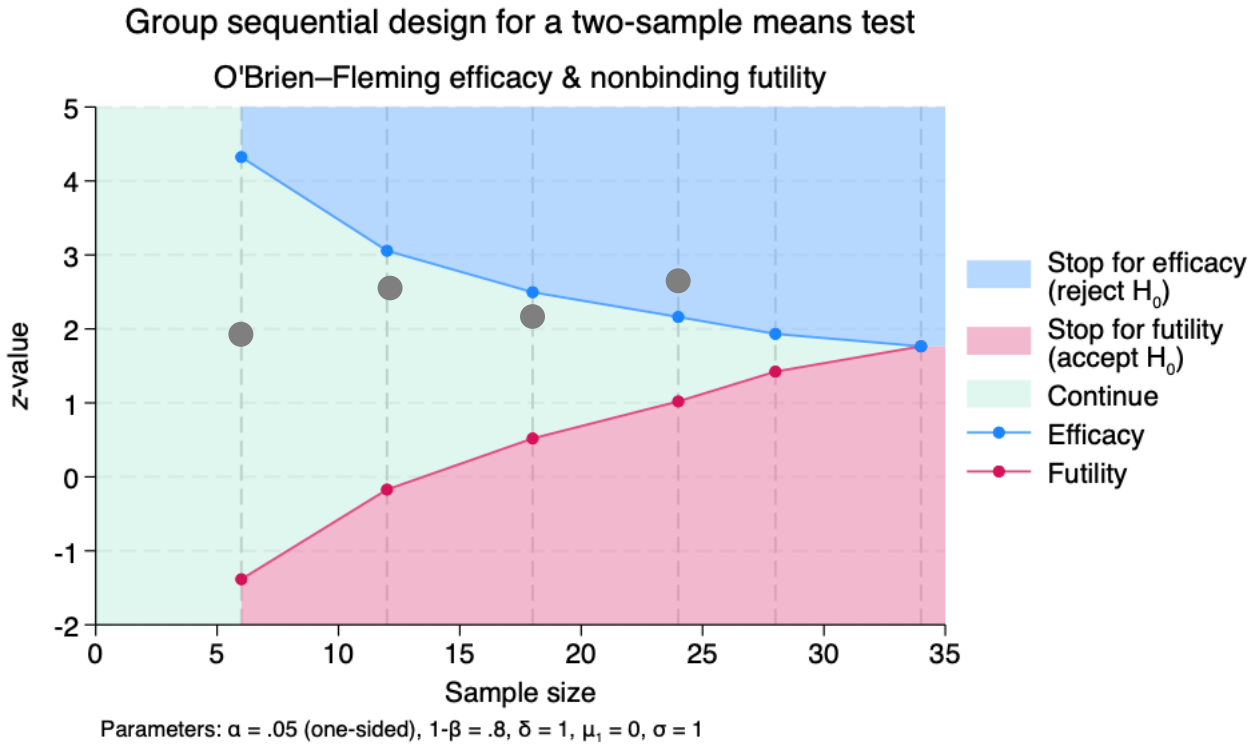
Intro: GSD in action







Treatment is effective



Intro: GSD theory

Test statistic Z_k is calculated using all observations through look k

$(Z_1, Z_2, \ldots, Z_K)$ is multivariate normal with

$$\mathrm{Cov}(Z_j, Z_k) = \sqrt{\frac{I_j}{I_k}}$$

where information fraction $I_k = n_k/n_K$

gs commands

- **gsbounds**
  - calculate stopping boundaries & information ratio
- **gsdesign**
  - calculate stopping boundaries & sample sizes
  - calls **power** and **gsbounds** under the hood

gsbounds: syntax

Calculate efficacy stopping boundaries

`gsbounds` [`,` `efficacy(`*boundary*`)` *options*]

Calculate futility stopping boundaries

`gsbounds, futility(`*boundary*[`,` `binding`]`)` [*options*]

Calculate efficacy and futility stopping boundaries

`gsbounds, efficacy(`*boundary*`)` `futility(`*boundary*[`,` `binding`]`)` [*options*]

boundary	Description
`obfleming`	classical O'Brien–Fleming bound
`pocock`	classical Pocock bound
`wtsiatis(#)`	classical Wang–Tsiatis bound with specified parameter value
`errpocock`	error-spending Pocock-style bound
`errobfleming`	error-spending O'Brien–Fleming-style bound
`kdemets(#)`	error-spending Kim–DeMets bound with specified parameter value
`hsdecani(#)`	error-spending Hwang–Shih–de Cani bound with specified parameter value

gsbounds: syntax

options	Description
Main	
`efficacy(`*boundary*`)`	boundary for efficacy stopping; if neither `efficacy()` nor `futility()` is specified, the default is `efficacy(obfleming)`
`futility(`*boundary*`[ , binding ])`	boundary for futility stopping; use `binding` to request binding futility bounds (default is nonbinding)
`nlooks(#)`	total number of analyses (`nlooks()` − 1 interim analyses and one final analysis)
`information(`*numlist*`)`	sequence of information levels for analyses; default is evenly spaced
`nopvalues`	suppress p-values
`alpha(#)`	overall significance level for all tests; default is `alpha(0.05)`
`power(#)`	overall power for all tests; default is `power(0.8)`
`beta(#)`	overall probability of type II error for all tests; default is `beta(0.2)`
`upper`	upper one-sided test; default is two-sided
`lower`	lower one-sided test; default is two-sided
`onesided`	synonym for `upper`
Graph	
`graphbounds[ (`*graphopts*`) ]`	graph boundaries
`matlistopts(`*general_options*`)`	control the display of boundaries; seldom used
optimopts	optimization options for boundary calculations; seldom used

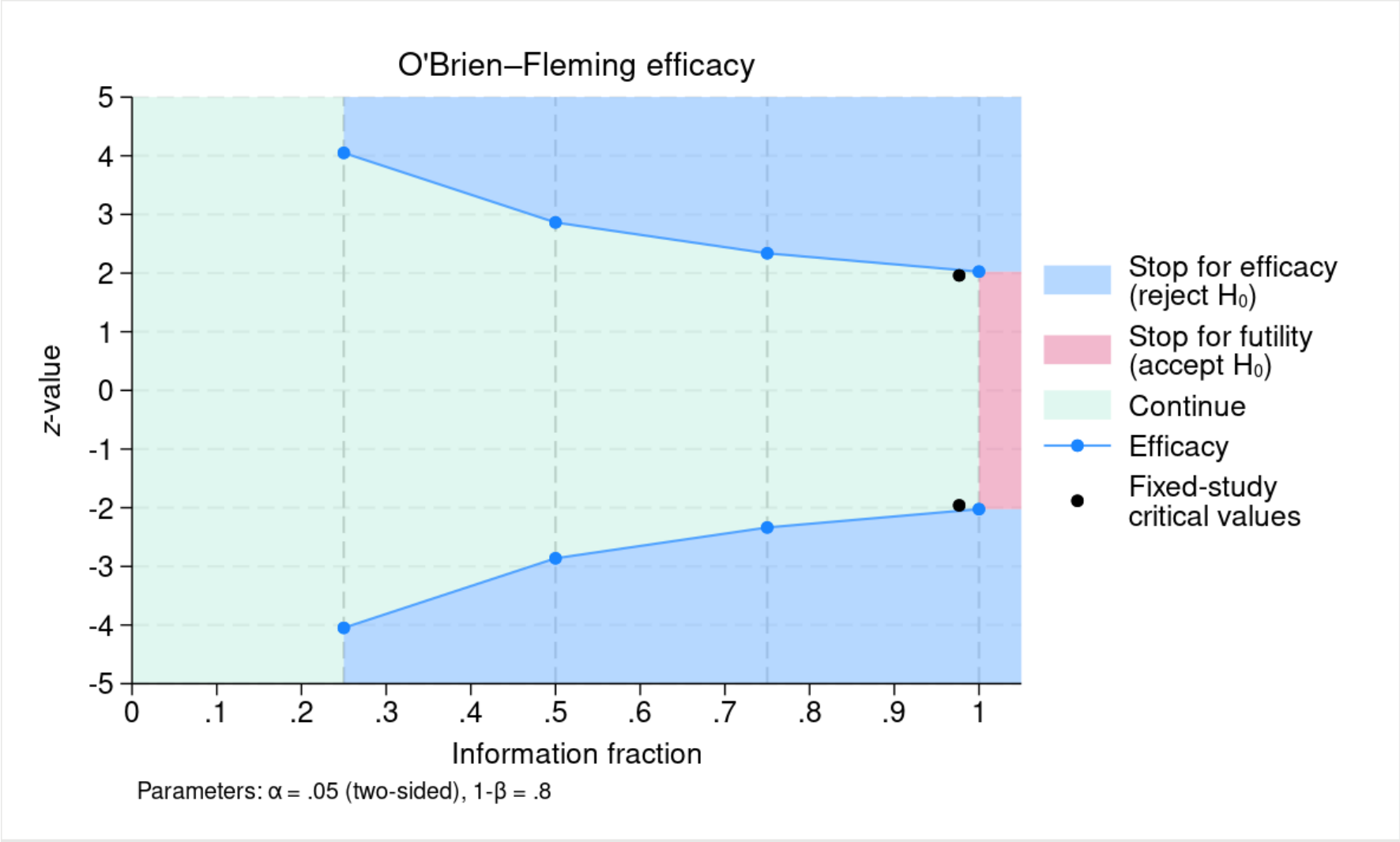
gsbounds: output

```
. gsbounds, efficacy(obfleming) nlooks(4)

Group sequential boundaries

Efficacy: O'Brien-Fleming

Study parameters:
      alpha = 0.0500  (two-sided)
      power = 0.8000

Info. ratio = 1.0238

Fixed-study crit. values = ±1.9600

Critical values and p-values for a group sequential design
```

Look	Info. frac.	Efficacy Lower	Efficacy Upper	p-value
1	0.25	-4.0486	4.0486	0.0001
2	0.50	-2.8628	2.8628	0.0042
3	0.75	-2.3375	2.3375	0.0194
4	1.00	-2.0243	2.0243	0.0429

```
Note: Critical values are for z statistics;
      otherwise, use p-value boundaries.
```

gsbounds: graph



gsdesign: syntax

`gsdesign` *method* ... [, *designopts boundopts*]

where *method* ... refers to a `power` *method* that is used for sample-size calculation, *designopts* are options controlling the sample-size calculation, and *boundopts* are options controlling the calculation of the stopping boundaries.

method	Description
`onemean`	GSD for one-sample mean test
`twomeans`	GSD for two-sample means test
`oneproportion`	GSD for one-sample proportion test
`twoproportions`	GSD for two-sample proportions test
`logrank`	GSD for a log-rank test
usermethod	user-defined sample-size calculation

`gsdesign` supports the above methods when they are used to calculate sample size with simple random sampling. To use an unsupported method, specify option `methodok`.

gsdesign: syntax

designopts	Description
Main	
methodopts	method-specific options
`alpha(#)`	overall significance level for all tests; default is `alpha(0.05)`
`power(#)`	overall power for all tests; default is `power(0.8)`
`beta(#)`	overall probability of type II error for all tests; default is `beta(0.2)`
`onesided`	request a one-sided test; default is two-sided
`nfractional`	report fractional sample size
`force`	allow calculation with unsupported *methodopts*
`methodok`	allow calculation with unsupported *method*
`poweriteration(`*powiteropts*`)`	iteration options for the calculation of fixed-study sample size; not available with *method* `logrank`; seldom used

gsdesign: output

```
. gsdesign twomeans 50 80, sd1(30) sd2(45) alpha(0.025) power(0.9) onesided efficacy(kdemets(2))
> futility(errobfleming) nlooks(4) graphbounds

Group sequential design for a two-sample means test
Satterthwaite's t test assuming unequal variances
H0: m2 = m1 versus Ha: m2 > m1

Efficacy: Error-spending Kim-DeMets, rho = 2.0000
Futility: Error-spending O'Brien-Fleming style, nonbinding

Study parameters:
      alpha =  0.0250  (upper one-sided)
      power =  0.9000
      delta = 30.0000
         m1 = 50.0000
         m2 = 80.0000
        sd1 = 30.0000
        sd2 = 45.0000

Expected sample size:
         H0 =   46.96
         Ha =   53.01

Info. ratio =  1.1134
    N fixed =      72
      N max =      80
     N1 max =      40
     N2 max =      40

Fixed-study crit. value = 1.9600

Critical values, p-values, and sample sizes for a group sequential design
```

	Info.	Efficacy		Futility		Sample size		
Look	frac.	Upper	p-value	Lower	p-value	N1	N2	N
1	0.25	2.9552	0.0016	-1.3792	0.9161	10	10	20
2	0.50	2.5594	0.0052	0.3580	0.3602	20	20	40
3	0.75	2.3009	0.0107	1.3336	0.0912	30	30	60
4	1.00	2.0920	0.0182	2.0920	0.0182	40	40	80

```
Note: Critical values are for z statistics; otherwise, use p-value boundaries.
```


STATA
18

Resources

- **Overview of features**
  - https://www.stata.com/new-in-stata/group-sequential-designs/
- **Video overview**
  - https://youtu.be/hO2qW1NLrMk?si=zu1-02rrzzkOw41E
- **Documentation**
  - https://www.stata.com/manuals/adapt.pdf
- **Power and sample size calculations**
  - https://www.stata.com/features/power-and-sample-size/