Appendix: Software for Adaptive Designs

The availability of software is a necessary condition for the applicability and acceptance of a statistical methodology. Many of the procedures proposed for adaptive designs additionally require high levels of computational performance such that software should be able to perform complex computations in a relatively short time. This kind of software is available today, and we briefly review the available software in this chapter which is a bit more general review as the one provided in Bauer et al. (2016). Up to now, the reviews of software packages concentrated on packages specifically designed for group sequential methods (Emerson 1996; Wassmer 2006; Zhu et al. 2011), the reason simply being that software for adaptive designs was not available at that time. A review of software for adaptive designs is provided in Tymofyeyev (2014).

One essential core of many if not all packages available for group sequential design is the numerical computation of the multivariate normal integral as described in Chap. 1. For group sequential designs it turns out that due to the independent increment structure of the underlying stochastic process the multivariate integral can be computed through the successive computation of univariate integrals. This is a consequence of the well-known recursive integration formula described in Armitage et al. (1969), and makes the computation feasible. It is interesting to recognize that these authors were already able to provide accurate results for the problem for up to 100 dimensions, i.e., stages of the trial. Due to the enormous growth in the computational capacity many alternative algorithms are available today that make the computation feasible. For an overview, see Genz and Bretz (2009).

A wide range of computations necessary in the context of the assessment of group sequential designs is possible with the use of software programs freely available on the homepage of Christopher Jennison: www.bath.ac.uk/~mascj. He provides the Fortran code for all the tabulated results of the Jennison and Turnbull monograph on group sequential designs in clinical trials (Jennison and Turnbull 2000). This might
serve as a very valuable tool to find the source code for algorithms to be used in group sequential designs.

Fortran programs for the computation of the use function approach are available from the University of Wisconsin School of Medicine and Public Health site [www.biostat.wisc.edu/content/lan-demets-method-statistical-programs-clinical-trials](http://www.biostat.wisc.edu/content/lan-demets-method-statistical-programs-clinical-trials) programs for Computing Group Sequential Boundaries Using the Lan-DeMets Method, Version 2.1. It comes with a Microsoft Windows graphical user interface and hence additionally provides a convenient way to perform the calculation. The last update is from 11/2003. So this tool was not further developed, and it is restricted to the use function approach. However, an R tool is available now (see below). R is free, and it also compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS. This might be advantageous, and reason for its widespread use. We checked CRAN (Comprehensive R Archive Network) cran.rstudio.com on January 20, 2015, and list the available packages which are available, together with a short description and its potential use in adaptive designs. We hope to provide a more or less complete list though it is emphasized that this is a dynamic development and we expect a lot of more packages in the near future. We also note that we concentrate on tools for confirmatory adaptive designs and not on tools especially developed for early phase dose-finding trials.

- **adaptTest**: Adaptive two-stage tests (Vandemeulebroecke 2009). The functions defined in this program serve for implementing adaptive two-stage adaptive tests that are based on the combination testing principle.
- **AGSDest**: Estimation in adaptive group sequential trials (Hack et al. 2013). This module enables the calculation of confidence intervals in adaptive group sequential trials.
- **asd**: Simulations for adaptive seamless designs (Parsons 2013). This package runs simulations for adaptive seamless designs with and without early outcomes for treatment selection and population enrichment type designs.
- **gMCP**: Graph Based Multiple Comparison Procedures (Rohmeyer and Klinglmüller 2014). This package provides functions and a graphical user interface for adaptive (Klinglmüller et al. 2014) and non-adaptive (Bretz et al. 2009b) graph-based multiple comparison procedures.
- **GroupSeq**: A GUI-based program to compute probabilities regarding group sequential designs (Pahl 2014). This program can be used for assessing the test characteristics of group sequential design and providing the boundaries for a group sequential approach or an inverse normal combination test approach.
- **gsDesign**: Group Sequential Design (Anderson 2014). gsDesign is a comprehensive package that derives group sequential designs and describes their properties. A graphical user interface gsDesignExplorer is available as well. The resulting boundaries can be used for adaptive settings.
- **interAdapt** (Fisher et al. 2014). This is an interactive tool for designing and evaluating certain types of adaptive enrichment designs.
- **ldbounds** Lan-DeMets method for group sequential boundaries (Casper and Perez 2014) is based on the Fortran from the University of Wisconsin and can also be used to provide the test characteristics of the use function approach.
– PwrGSD: Power in a Group Sequential Design (Izmirlian 2014). This program evaluates analysis plans for sequentially monitored trials on a survival endpoint. It can also be used to perform power calculations in a group sequential setting.

– seqmon: Sequential Monitoring of Clinical Trials (Schoenfeld 2012). This program elementarily computes the probability of crossing sequential boundaries in a clinical trial and uses a method described by the author (Schoenfeld 2001).

There is also an R package called RCTDesign: Methods and Software for Clinical Trials. This package builds on the formerly available S-Plus module S+SeqTrial. RCTdesign is currently not available at CRAN but is freely available to users through a joint agreement between Tibco, Inc. (the owners of the S-Plus software system and the S-Plus code in the module S+SeqTrial) and Scott S. Emerson (the developer of the C code that serves as the engine for S+SeqTrial). RCTdesign makes the computation and evaluation of a wide range of commonly used designs possible. It also comes with an add-on for adaptive methods. Furthermore, the book (Chang 2014) contains R programs for adaptive designs. These are elementary programs for performing sample size reassessment procedures and some basic adaptive randomization designs. The book also comes with SAS macros, most of them performing simulations for the adaptive design described in the book.

Since version 6, SAS comes with some function calls in SAS/IML for doing groups sequential tests (SAS Institute Inc. 1995). Currently available are the SEQ, SEQSCALE, and SEQSHIF calls. These procedures provide accurate results for computing decision regions, maximum and expected sample sizes, group sequential densities, etc. Examples can be found in Wassmer (1999c), SAS Institute Inc. (1995), Wassmer and Biller (1998), Dmitrienko et al. (2005). Within SAS/IML it is straightforward to produce results for group sequential designs although the calculation of, for example, bias adjusted estimates might become cumbersome. New in SAS 9.2 are procedures for doing group sequential designs in a more comfortable way (SAS Institute Inc. 2009). Specifically, the SEQDESIGN procedure designs interim analyses for clinical trials. It directly computes the boundary values and required sample sizes for the trial within a wide range of possible designs. The SEQTEST procedure performs the interim analyses (tests and confidence intervals) based on design information produced by the SEQDESIGN procedure. SAS currently does not provide any direct capabilities for doing confirmatory adaptive designs as considered in this monograph.

Since the very beginning of adaptive designs the software ADDPLAN was designed for doing confirmatory adaptive designs (www.addplan.com). It is commercially available since 2002 as a tool for designing, simulating, and performing analysis for group sequential designs with an emphasis on the confirmatory adaptive technique. The MC module provides additional multiple comparison features for more than two treatment arms in simulation and analysis, and the PE module additional features for patient enrichment designs in simulation and analysis. There is also the new DF module with capabilities for adaptive dose-finding designs (MCPMod, CRM, etc.).
East from Cytel (www.cytel.com) is a comprehensive tool for design, simulation, and analysis of trials with interim analyses. In the current release, adaptive extensions are provided with the EastAdapt and the EastSurv module. Recently, the modules EastMultiarm and EastEndpoint provide extensions to multi-arm designs and designs with multiple endpoints. An extension to dose finding trials comes with EastEscalate and Cytel’s Compass.

We also mention the nQuery module for calculating designs for the group sequential case in the nQuery package (www.statsols.com/products/nquery-advisor-interim) as well as corresponding capabilities in PASS from NCSS (www.ncss.com). Both do not provide any adaptive extensions but can be used for performing interim decisions and assessing group sequential designs, for example, with respect to maximum and expected sample size.

To summarize, some software is free and hence attractive for statistical research. This is particularly true for the increasing number of available R packages. Simulation-based evaluation of operating characteristics of adaptive designs is becoming increasingly important, some of the available adaptive R packages typically address this issue. The R and SAS packages are available only within the programming environment, whereas the ADDPLAN, EaSt, nQuery, and PASS programs come with a user-friendly graphical user interface (GUI). We note that a free GUI is also available for gsDesign and some other R packages. Within commercially available packages, only ADDPLAN, EastAdapt, and EastSurv address the specific requirements for confirmatory adaptive designs.
References


References


FDA. (2015). *Draft guidance for industry and food and drug administration staff. Adaptive designs for medical device clinical studies*. Food and Drug Administration. Center for Devices and Radiological Health (CDRH) and Center for Biologics Evaluation and Research (CBER), Rockville, MD.


Graf, A. C., & Bauer, P. (2011). Maximum inflation of the type I error rate when sample size and allocation rate are adapted in a pre-planned interim look. Statistics in Medicine, 30, 1637–1647.


Rohmeyer, K., & Klinglmüller, F. (2014). gMCP: Graph based multiple comparison procedures. [http://cran.r-project.org/web/packages/gMCP](http://cran.r-project.org/web/packages/gMCP). R package version 0.8-7.


References


Index

A priori hierarchical intersection test, 247, 265
Adaptive change of hypothesis, 271
Adaptive choice of test statistic, 272
Adaptive closed test, 235
  example, 235, 237, 238
  increase in power, 243
  with combination test, 235, 243
  with CRP principle, 235
Adaptive design, 133
  basic principle, 134
  Bauer and Köhne method, 137, 138, 140, 144, 168
  combination test, 135
  conditional error function approach, 151
  conditional invariance principle, 134
  CRP principle, 155
  Fisher’s product test, 137, 138, 140, 141, 144, 168
  futility stop, 136, 163
  iDMC, 133, 275
  inverse chi-squared method, 137
  inverse normal combination test, 143, 144, 149, 167
  Müller & Schäfer method, 155
  multi-stage, 167
  Proschan and Hunsberger method, 151
  recursive combination test, 169
  regulatory view, 274
  reverse engineering, 276
  sample size recalculation, 146
  two-sided adaptive test, 164
  Type I error maximization method, 159
Adaptive Dunnett test, 252–256
  conditional second stage test, 254
derivation, 253
  power comparison, 256
  unconditional second stage test, 255
Adaptive endpoint selection
  case study, 272–274
Adaptive enrichment design, 261–271
  binary response, 267
  case study, 268–271
  effect specification, 266
  repeated confidence interval, 266
  repeated p-value, 266
  selection rule, 269
  survival data, 268
Adaptive treatment selection, 237, 241–261
  binary response, 258
  case studies, 259–261
  effect size pattern, 256
  example, 249–252
  power definition, 256
  repeated confidence interval, 247
  repeated p-value, 247
  selection rule, 256
  shortcut, 251
  survival data, 258
Adaptive treatment switch, 238
Adaptively weighted ML estimate, 213
Adjusted nominal significance level, 27
ADVENT trial, 260
α-spending function approach, 74–81
  approximation of Pocock’s design, 78
  calculation of sample size, 79
  example, 76
  random over- and underrunning, 78
Bauer and Köhne method, 137, 138, 140, 144, 168
Bayesian predictive power, 181–184
β-spending function approach, 81
Bias adjusted point estimation, 98
Bias in adaptive design
  adaptively weighted ML estimate, 213
  fixed weighted ML estimate, 211
  ML estimate, 209
Binary response, 110–121
  confidence interval, 114, 121
  in adaptive enrichment design, 267
  in multiple treatments, 258
  non-inferiority testing, 121
  power calculation, 113
  sample size calculation, 112
  testing a single rate, 110
  two-sample comparison, 116
Bivariate t intersection test, 264
Blinded sample size recalculation, v, 134
Bonferroni correction, 11
Bonferroni intersection test, 246, 265
Case studies, 259–261, 268–274
Circular conditional error function
definition, 152
derivation, 162
Classical group sequential design, 25–62
Closed testing in adaptive design, 235–239
Closure principle, 233
Combination test, 135
  Bauer and Köhne method, 137, 138, 140, 144, 168
  circular conditional error function, 152
decision rule, 136
  Fisher’s product test, 137, 138, 140, 141, 144, 168
  inverse normal, 143
  linear conditional error function, 154
  relationship to conditional error function, 153
  weighted inverse normal, 143
Conditional bias, 98
Conditional error function
calculation with conditional error function, 172
effect specification, 151
example, 154
optimal, 155
Conditional error function approach, 151
relationship to combination test, 153
Conditional invariance principle, 134
Conditional power, 171
Bayesian predictive power, 181–184
Critical values
  α-spending function approach, 74–81
  DeMets and Ware test, 51
  O’Brien and Fleming test, 12, 26
  Pampallona and Tsiatis test, 45
  Pocock test, 12, 26
  Wang and Tsiatis test, 36
CRP principle, 155
  adaptive Dunnett test, 252–256
  advantage, 158
general formulation, 157
  relationship to inverse normal, 157
Data Monitoring Committee, v, 35, 133, 275
Delayed responses, 225
Δ-class critical values, 36–42
DeMets and Ware test, 50–55
Dunnett intersection test, 245
Effect specification
  for adaptive enrichment design, 266
  for multiple treatments, 256
Exact confidence bound for combination test, 201
Expected reduction in sample size, 32
Familywise error rate, 233
Fisher’s product test, 137
  choice of boundaries, 140
Conditional Type I error rate, 156
Confidence interval
  binary response, 114, 121
definition, 85
  exact confidence bound for combination test, 201
  example, 206
  for adaptive design, 200–209
  normal response, 105
  survival data, 227
Covariance of test statistic, 9
Conditional power, 171
comparison with inverse normal combination test, 144
example, 138, 140
multi-stage, 168
non-stochastic curtailment, 138
weighted, 141
Fixed sample test, 7
Fixed weighted ML estimate, 211
Futility stop, 14, 52
\(\beta\)-spending function approach, 81
binding and non-binding, 54, 150
in adaptive design, 136, 163
in treatment arm selection design, 243
with conditional power, 175

Group sequential design
\(\alpha\)-spending function approach, 74–81
and inverse normal combination test, 143
delayed responses, 225
\(\Delta\)-class critical values, 36–42
DeMets and Ware test, 50–55
equal stage sizes, 25–62
expected reduction in sample size, 32
Haybittle and Peto boundaries, 42
iDMC, 35
inflation factor, 31
maximum number of stages, 35
O’Brien and Fleming test, 12, 26
one-sided, 50–62
optimum design in \(\Delta\)-class, 40
optimum information rates, 71
Pampallona and Tsiatis test, 43–50
Pocock test, 12, 26
power and average sample size, 15–19,
30–35
power at given sample size, 34
power family, 36–42
sample size calculation, 19
sample size recalculation, 146
stopping probability, 34
symmetric design, 43–50
unequal stage sizes, 63–81
Wang and Tsiatis test, 36–42

Haybittle and Peto boundaries, 42
Hazard ratio
definition, 123
estimation, 227

I-SPY 2 TRIAL, 268
iDMC, 35, 133, 275
Independent increments, 126, 222
Inflation factor, 31
Information rates, 64, 74
INHANCE trial, 260
Intersection hypothesis, 234, 243
Intersection test, 245–247, 264–266
a priori hierarchical intersection test, 247,
265
bivariate \(t\) test, 264
Bonferroni test, 246, 265
Dunnett test, 245
Šidák test, 246, 265
Simes test, 246, 265
Inverse chi-squared method, 137
Inverse normal combination test, 143
advantage, 150, 186
comparison with Fisher’s product test, 144
criticism, 149
multi-stage, 167
sample size calculation, 143
use of group sequential boundaries, 143

Left truncation of survival data, 223
Linear conditional error function, 154
Log-rank test, 123, 222
Loss in power due to over- or underweighting,
148

Müller & Schäfer method, 155
Many-to-one-comparison, 242
Maximum number of stages, 35
Median unbiased point estimation, 212
MSE reduction, 100, 209
Multi-stage adaptive test, 167
Multiplicity
closed testing in adaptive design, 235–239
closure principle, 233
dexample, 231
familywise error rate, 233
intersection hypothesis, 234
sources of, 232
Multivariate normality, 9
MUSEC trial, 187

Non-inferiority testing
Fieller’s theorem, 109
normal response, 106
Normal response, 102–110
cross-over design, 110
non-inferiority testing, 106
paired comparison, 106
power calculation, 104
sample size allocation ratio, 108
sample size calculation, 103
significance level approach, 103
two-sample comparison, 106
Welch approximation, 107

Numerical coincidence of critical values, 51

Observation times
estimation of, 127

One-sided design, 50–62
Optimum design in $\Delta$-class, 40
Optimum information rates, 71
example, 73

Overall exact $p$-value, 85, 191
example, 198
for Fisher’s product test, 192
for inverse normal method, 193

Paired comparison, 106
Pampallona and Tsiatis test, 43–50
example, 45, 57
one-sided, 55

Point estimation
bias of ML estimate, 95
bias reduction, 98
conditional bias, 98
definition, 86
in adaptive design, 209–218
MSE reduction, 100, 209
Rao-Blackwell, 99

Power and average sample size, 15–19, 30–35
Power at given sample size, 34

Power calculation
binary response, 113
normal response, 104
survival data, 129

Power definition for multiple treatments, 256
Power family, 36–42
Power of two-sided tests, 62
Promising zone approach, 176
Proschan and Hunsberger method, 151

$r$-selection rule, 257
Random over- and underrunning, 78
Rao-Blackwell, 99
Recursive combination test, 169
Recursive integration formula, 20–23
Regulatory view, 274
Repeated confidence interval

Adaptive enrichment design, 266
definition, 91
example, 207
for binary response, 218
for inverse normal method, 204
for unknown variance, 218
in adaptive design, 204
inconsistency in multiple testing, 251
monitoring property, 91, 205
multi-arm testing situation, 247
normal mean, 92, 207
survival data, 227
with futility stop, 208

Repeated $p$-value
consistency, 94

Repeated $p$-value
adaptive enrichment design, 266
definition, 93
example, 199
for combination test, 193
for Fisher’s product test, 196
monitoring property, 195
multi-arm testing situation, 247
Repeated significance test, 3, 27
Reverse engineering, 276

Sample mean ordering, 90
Sample size calculation, 19
binary response, 112
normal response, 102, 103
survival data, 124

Sample size recalculation, 146
based on conditional power, 176
case study, 187
critism, 179, 186
decreasing the sample size, 188
loss in power, 148
optimum rule, 176
with promising zone approach, 176
with survival data, 226

Sample space ordering, 84–91
likelihood ratio ordering, 89
sample mean ordering, 90
score test ordering, 90
stage-wise ordering, 86

Sampling scheme
group sequential, 8
sequential, 4

Score test ordering, 90
Selection rule
for enrichment design, 269
for multiple treatments, 256
$r$-selection rule, 257
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential probability ratio test</td>
<td>4</td>
</tr>
<tr>
<td>Short-term endpoint</td>
<td>259</td>
</tr>
<tr>
<td>Šidák intersection test</td>
<td>246, 265</td>
</tr>
<tr>
<td>Significance level approach</td>
<td>103</td>
</tr>
<tr>
<td>Šimes intersection test</td>
<td>246, 265</td>
</tr>
<tr>
<td>Single rate testing</td>
<td>110</td>
</tr>
<tr>
<td>Software</td>
<td>277–280</td>
</tr>
<tr>
<td>ADDPLAN</td>
<td>279</td>
</tr>
<tr>
<td>EaSt</td>
<td>279</td>
</tr>
<tr>
<td>Fortran</td>
<td>278</td>
</tr>
<tr>
<td>R</td>
<td>278</td>
</tr>
<tr>
<td>SAS</td>
<td>279</td>
</tr>
<tr>
<td>Stage-wise ordering</td>
<td>86, 89, 190</td>
</tr>
<tr>
<td>for two-stage combination test</td>
<td>190</td>
</tr>
<tr>
<td>in adaptive design</td>
<td>190, 201</td>
</tr>
<tr>
<td>Stage-wise test statistic</td>
<td>9</td>
</tr>
<tr>
<td>Stopping probability</td>
<td>34</td>
</tr>
<tr>
<td>Survival data</td>
<td>122–130, 221–228</td>
</tr>
<tr>
<td>accrual pattern</td>
<td>122</td>
</tr>
<tr>
<td>adaptation with short-term endpoint</td>
<td>259</td>
</tr>
<tr>
<td>adaptive design</td>
<td>222</td>
</tr>
<tr>
<td>calculation of event probability</td>
<td>124</td>
</tr>
<tr>
<td>combination test</td>
<td>222</td>
</tr>
<tr>
<td>confidence interval</td>
<td>227</td>
</tr>
<tr>
<td>Cox proportional hazards regression model</td>
<td>130</td>
</tr>
<tr>
<td>estimation of hazard ratio</td>
<td>227</td>
</tr>
<tr>
<td>estimation of observation times</td>
<td>127</td>
</tr>
<tr>
<td>hazard function</td>
<td>123</td>
</tr>
<tr>
<td>in adaptive enrichment design</td>
<td>268</td>
</tr>
<tr>
<td>in multiple treatments</td>
<td>258</td>
</tr>
<tr>
<td>independent increments property</td>
<td>126, 222</td>
</tr>
<tr>
<td>left truncation</td>
<td>223</td>
</tr>
<tr>
<td>log-rank test</td>
<td>123, 222</td>
</tr>
<tr>
<td>power calculation</td>
<td>129</td>
</tr>
<tr>
<td>restriction of information used in adaptation</td>
<td>224, 259</td>
</tr>
<tr>
<td>sample size allocation ratio</td>
<td>124</td>
</tr>
<tr>
<td>sample size calculation</td>
<td>124</td>
</tr>
<tr>
<td>sample size recalculation</td>
<td>226</td>
</tr>
<tr>
<td>Triangular plans</td>
<td>4</td>
</tr>
<tr>
<td>Two one-sided tests</td>
<td>62, 165</td>
</tr>
<tr>
<td>Two-sample comparison</td>
<td></td>
</tr>
<tr>
<td>binary response</td>
<td>116</td>
</tr>
<tr>
<td>normal response</td>
<td>106</td>
</tr>
<tr>
<td>survival data</td>
<td>123</td>
</tr>
<tr>
<td>Two-sided adaptive test</td>
<td>164</td>
</tr>
<tr>
<td>Two-sided confidence interval</td>
<td></td>
</tr>
<tr>
<td>example</td>
<td>206</td>
</tr>
<tr>
<td>in adaptive design</td>
<td>205</td>
</tr>
<tr>
<td>Type I error maximization method</td>
<td>159</td>
</tr>
<tr>
<td>Unequal stage sizes</td>
<td>63–81</td>
</tr>
<tr>
<td>adjusting critical values</td>
<td>67</td>
</tr>
<tr>
<td>( \alpha )-spending function approach</td>
<td>74–81</td>
</tr>
<tr>
<td>calculation of sample sizes</td>
<td>71</td>
</tr>
<tr>
<td>effect on Type I error rate</td>
<td>64</td>
</tr>
<tr>
<td>information rates</td>
<td>64</td>
</tr>
<tr>
<td>maximum increase on Type I error rate</td>
<td>66</td>
</tr>
<tr>
<td>O’Brien and Fleming design</td>
<td>70</td>
</tr>
<tr>
<td>optimum information rates</td>
<td>71</td>
</tr>
<tr>
<td>overall test statistic</td>
<td>64</td>
</tr>
<tr>
<td>random over- and underrunning</td>
<td>78</td>
</tr>
<tr>
<td>sample sizes fixed in advance</td>
<td>69</td>
</tr>
<tr>
<td>Unlinded sample size recalculation</td>
<td>133</td>
</tr>
<tr>
<td>Wang and Tsiatis test</td>
<td>36–42</td>
</tr>
<tr>
<td>Weighted Fisher’s product test</td>
<td>141</td>
</tr>
<tr>
<td>Weighted inverse normal combination test</td>
<td>143</td>
</tr>
<tr>
<td>Welch approximation</td>
<td>107</td>
</tr>
</tbody>
</table>