

## Additional Material

This is the R code used to generate the first three lines of Table 2, i.e. to do simulations for the first three Scenarios in the base-case configuration.

```
# Load libraries
library(splines)
library(boot)
library(flux)
library(mfp)
# Set seed
set.seed(1)
#Define Simulation Scenarios (for example Scenarios 1-3)
# Scenario 1: Logistic growth model
formula1.1<-function(x) {
  return(0.05+0.90/(1+exp(-x*2+25)))
}
# Scenario 2: Gompertz model, b<1:
formula2.1<-function(x) {
  return(0.9*exp(-1*exp(-0.5*(x-11))))
}
# Scenario 3: Gompertz model, b=1:
formula3.1<-function(x) {
  return(0.9*exp(-1*exp(-1*(x-11))))
}
#Some parameter initializations
n.sim<-1000
n.obj<-500
n.arms<-7
scenarios<-c(1:3)
n.scen<-length(scenarios)
min.dur<-10
max.dur<-20
x.dur<-seq(min.dur,max.dur, length.out = 100)
# Some matrices initializations:
max.err.fp<-matrix(NA,n.sim,n.scen)
where.max.err.fp<-matrix(NA,n.sim,n.scen)
diff.auc.fp<-matrix(NA,n.sim,n.scen)
reldiff.auc.fp<-matrix(NA,n.sim,n.scen)
cover.auc.fp<-matrix(NA,n.sim,n.scen)
# Start simulations looping over different scenarios
for (s in 1:n.scen) {
  formula.1<-get(paste("formula",scenarios[s],".1", sep=""))
```

```

#Calculate y from formula to calculate AUC
y.dur<-formula.1(x.dur)
#Calculate true area under the curve
true.auc<-auc(x.dur,y.dur)
# Define formula to calculate root in prediction curve for fractional polynomial model:
formula.3<-function(x) {
  return(predict(fit.fp.i,data.frame(durlong=x),type="resp"))
}
# Run n.sim simulations:
for (i in 1:n.sim) {
  # We choose equidistant duration arms, between the minimum and the maximum
  durations<-seq(min.dur,max.dur,length.out = n.arms)
  # number of patients in each duration group, aiming for a total of n.obj patients
  npergroup<-ceiling(n.obj/n.arms)
  # total number of patients
  n<-npergroup*n.arms
  # Durations in long format
  durlong<-rep(durations, each=npergroup)
  # Calculate probabilities
  pro<-formula.1(durlong)
  # Generate events from binomial distribution
  y<-rbinom(n,1,pro)
  # Fit Fractional Polynomial regression model
  fit.fp.i<-mfp(y~fp(durlong), family="binomial")
  #calculate area between true and estimated curve:
  y.dur.est<-formula.3(x.dur)
  y.dur.ses<-predict(fit.fp.i,data.frame(durlong=x.dur),type="resp", se.fit=T)$se.fit
  y.fit<-formula.3(x.dur)-y.dur
  est.auc<-auc(x.dur,abs(y.fit))
  #Calculate maximum absolute error and its position along the curve
  max.err<-max(abs(y.fit))
  where.max.err<-which.max(abs(y.fit))
  #calculate coverage level
  cov<-0
  for (j in 1:100) {
    if (y.dur.est[j]-qnorm(0.975)*y.dur.ses[j]<y.dur[j]&y.dur.est[j]+qnorm(0.975)*y.dur.ses[j]>y.dur[j])
      cov<-cov+1
  }
}
#results
max.err.fp[i,s]<-max.err

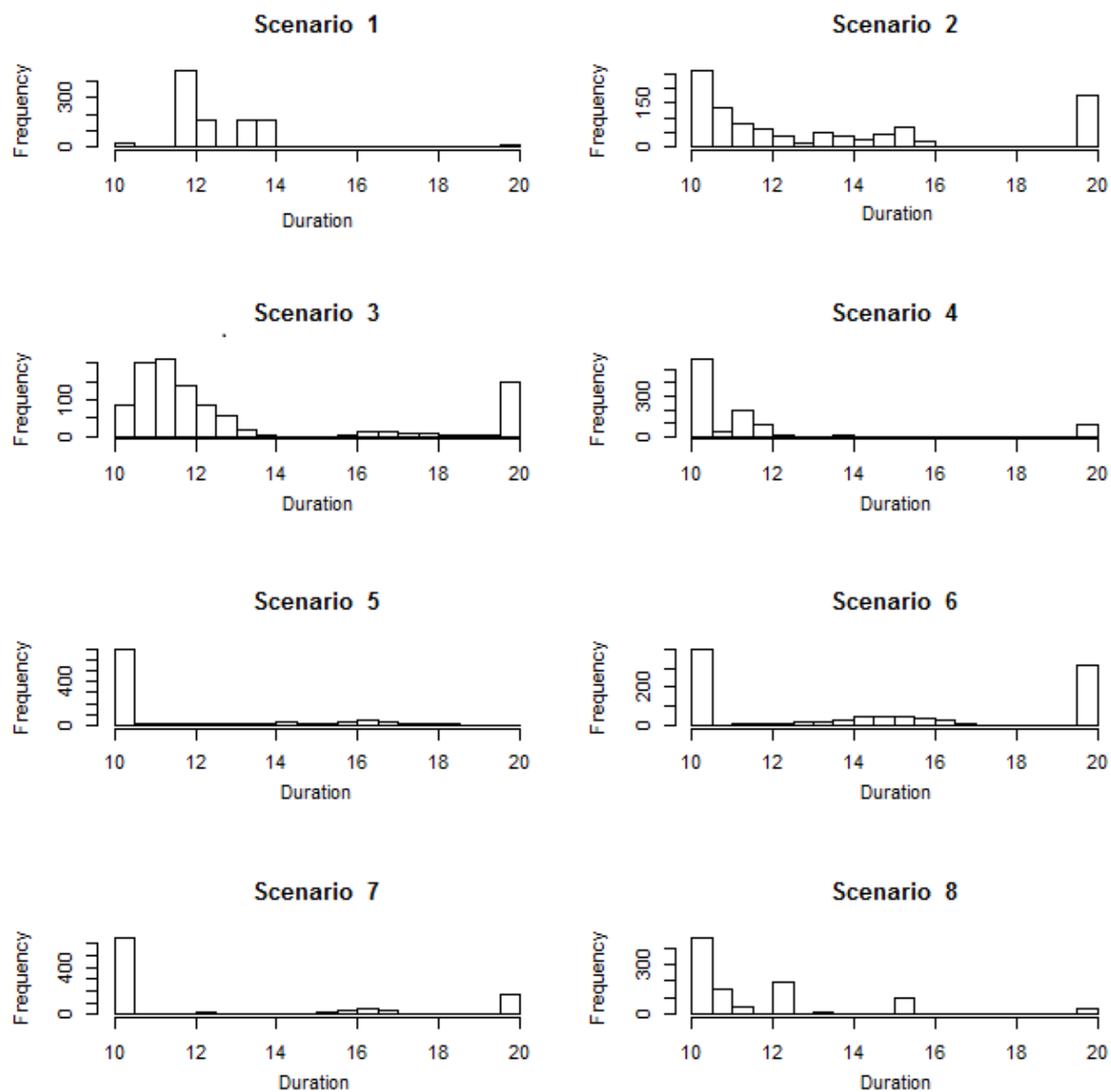
```

```

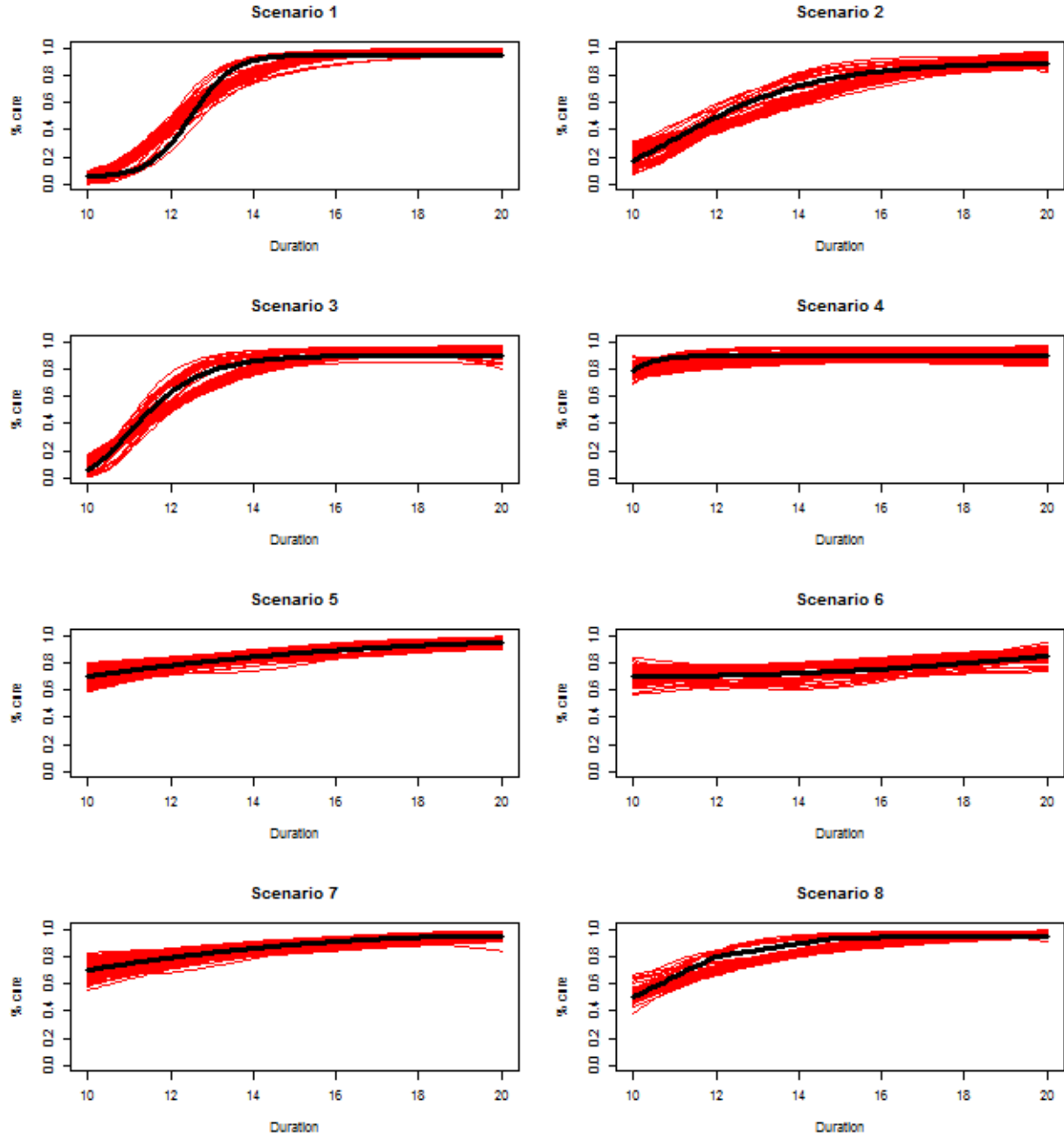
where.max.err.fp[i,s]<-x.dur[where.max.err]
diff.auc.fp[i,s]<-est.auc
reldiff.auc.fp[i,s]<-est.auc/(max.dur-min.dur)
cover.auc.fp[i,s]<-cov

  if (i%%10==0) cat("Simulation number ", i, " completed\n")
}
cat("Scenario number ", s, " completed\n")
}
Table<-matrix(NA,n.scen+1,8)
for (s in 1:n.scen) {
  Table[s,1:5]<-quantile(reldiff.auc.fp[,s], c(0, 0.05, .50, .95, 1))
  Table[s,6:7]<-quantile(max.err.fp[,s], c(.50, .95))
  Table[s,8]<-mean(cover.auc.fp[,s])
}
Table[n.scen+1,1:5]<-quantile(as.vector(reldiff.auc.fp), c(0, 0.05, .50, .95, 1))
Table[n.scen+1,6:7]<-quantile(as.vector(max.err.fp), c(.50, .95))
Table[n.scen+1,8]<-mean(as.vector(cover.auc.fp))
rownames(Table)<-c("Scenario 1", "Scenario 2", "Scenario 3", "Scenario 4", "Scenario 5", "Scenario 6",
  "Scenario 7", "Scenario 8", "Overall")

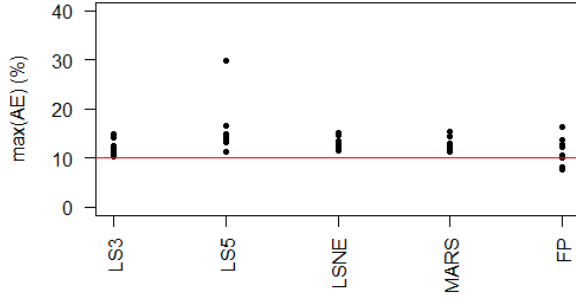
```



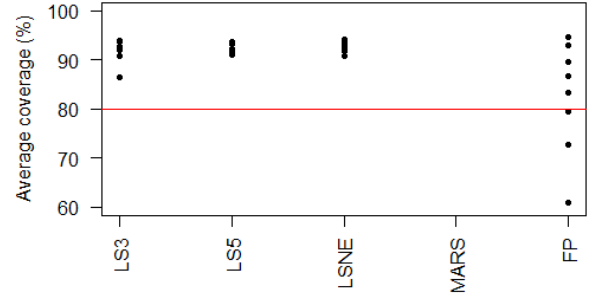
**Figure 5.** Histogram of the durations corresponding to the maximum absolute error along the duration-response curve for all the eight scenarios under the base-case configuration; this is assuming a sample size of 504 patients, randomised to 7 equidistant arms, and fitting a fractional polynomial model to estimate the duration-response curve.



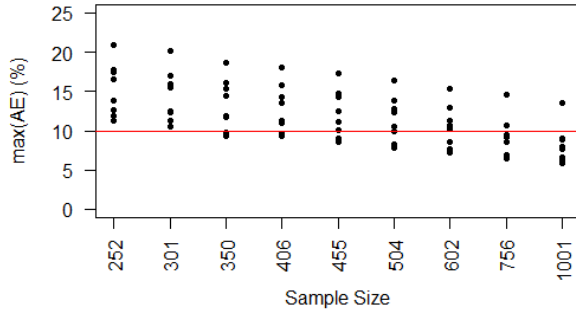
**Figure 6.** Prediction curves (red) from the worst 100 simulations in terms of coverage level against the true data generating curve (black) for all the eight scenarios under the base-case configuration; this is assuming a sample size of 504 patients, randomised to 7 equidistant arms, and fitting a fractional polynomial model to estimate the duration-response curve.



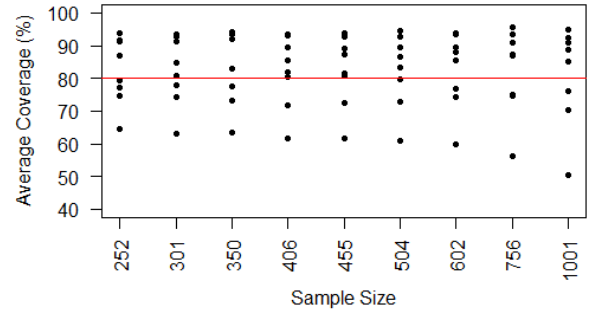
(a) Comparison flexible regression methods: 95<sup>th</sup> percentiles  $\max_d AE(d)$



(b) Comparison flexible regression methods: mean coverage

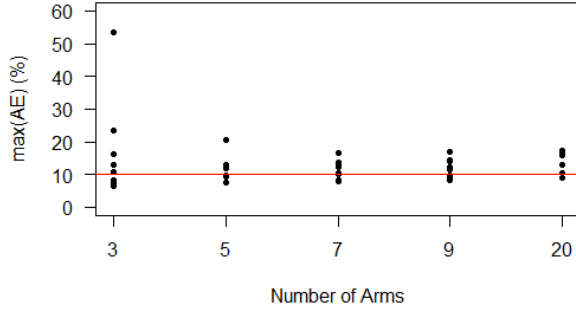


(c) Sensitivity to sample size: 95<sup>th</sup> percentiles  $\max_d AE(d)$

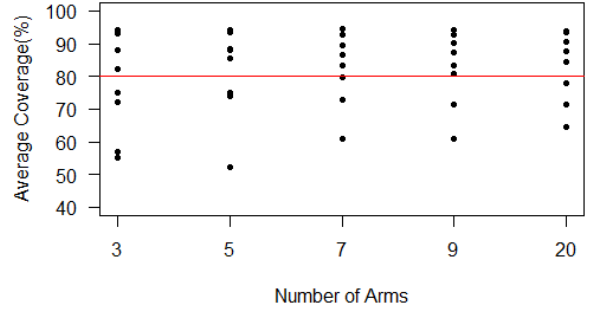


(d) Sensitivity to sample size: mean coverage

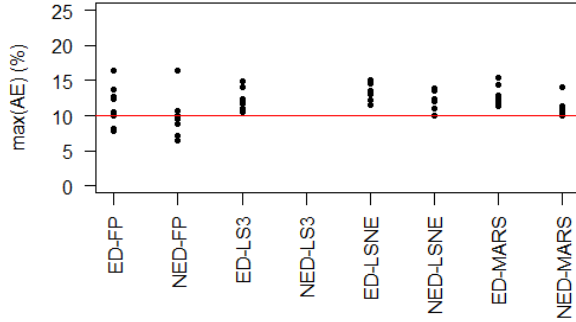
**Figure 7.** Comparison of results of trial simulations from the eight scenarios varying either (i) the flexible regression method used (LS3, LS5, LSNE, MARS, FP), with total sample size of 504 patients (panels (a) and (b)), or (ii) the total sample size between 250 and 1000 patients, using FP (panel (c) and (d)). Patients are divided in 7 equidistant duration arms. In the left panels we show the 95<sup>th</sup> percentiles of maximum absolute error ( $\max_d AE(d)$ ) from the eight scenarios, while in the right panels we compare mean coverage along the curve. The red horizontal lines indicate 10%  $\max_d AE(d)$  or 80% mean coverage. Points related to MARS mean coverage are missing as always lower than 60%. LS3-5: Linear Spline with 3-5 knots. LSNE: Linear Spline with Non-Equidistant knots. MARS: Multivariable Adaptive Regression Splines. FP: Fractional Polynomials.



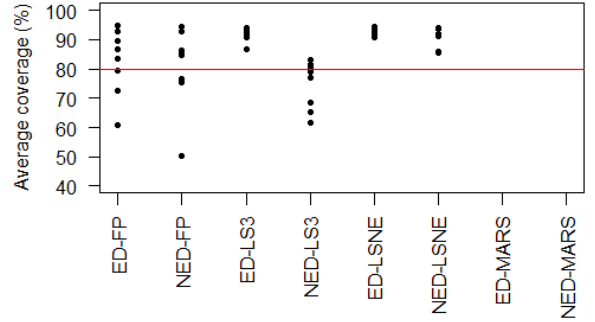
(a) Sensitivity to number of arms: 95<sup>th</sup> percentiles  $\max_d AE(d)$



(b) Sensitivity to number of arms: mean coverage



(c) Sensitivity to placement of arms: 95<sup>th</sup> percentiles  $\max_d AE(d)$



(d) Sensitivity to placement of arms: mean coverage

**Figure 8.** Comparison of results of trial simulations from the eight scenarios either varying the number of equidistant arms (panels (a) and (b)) between 3 and 20, using Fractional Polynomials (FP), or using different designs, Equidistant (ED) or Not Equidistant (NED), comparing four different regression methods (panels (c) and (d)). The total sample size is always 504 patients. In the left panels we show the 95<sup>th</sup> percentiles of maximum absolute error ( $\max_d AE(d)$ ) from the eight scenarios, while in the right panels we compare mean coverage along the curve. The red horizontal lines indicate 10%  $\max_d AE(d)$  or 80% mean coverage. Points related to MARS mean coverage are missing as always lower than 40%. LS3: Linear Spline with 3 knots. LSNE: Linear Spline with Non-Equidistant knots. MARS: Multivariable Adaptive Regression Splines. FP: Fractional Polynomials.