Appendix

We developed the package “cbpti” to explore the effects of continuous biomarkers in patient-treatment interaction. This is a first version which uses a linear relationship between the biomarker and a continuous response. Future versions of the package will implement non-linear relationships, binary response and survival data. The package includes the data of Gadbury & Iyer 2000 [1] and the data published by Schwenke 1990 [2].

We use the data of Gadbury & Iyer to demonstrate the functions of this package below. This data consists of counts of epileptic seizures during four time-periods, two treatment groups randomly assigned to placebo or the antiepileptic drug progabide, and two continuous baseline measures (baseline number of seizures and age). The continuous baseline measurements can be considered as two specific biomarkers. Here we present the results of the two biomarkers for comparison. The data can be loaded from the package by applying the function `data()`, specifying the package “cbpti” to be chosen.

```r
data(gadbury, package="cbpti")
```

The treatment outcome is taken as the asinh-transformed total number of seizures under treatment. The transformation is motivated by the fact that the seizure counts are over-dispersed and the proposed transformation is suitable for variance stabilization [3]. The outcome and the baseline number of seizures are transformed with the function `asinh()` in R.

```r
gadbury$Y<-asinh(gadbury$Y1+gadbury$Y2+gadbury$Y3+gadbury$Y4)
gadbury$Base_tr <- asinh(gadbury$Base)
```

In the following we present our methods to produce (1) interaction plots, (2) contrast plots, (3) proportion of unfavorable treatment effect plots, (4) ROC curves, and (5) prediction curves for the two biomarkers, baseline number of seizures and age. In our R package, we calculate pointwise confidence intervals instead of simultaneous confidence bands since they are much easier to calculate than simultaneous confidence bands and commonly used in clinical reports. Moreover, previous study has shown that pointwise confidence intervals work well in practice [4]. For the sake of simplicity, we assume that the outcome measurements of the two treatment groups are not correlated and the covariance is thus assumed to be zero. We give the functions written in R accompanied by the produced plots. Note that the functions were written and tested and the plots produced with R version 3.1.2 (2015-10-08) -- "Pumpkin Helmet", Copyright (C) 2014 The R Foundation for Statistical Computing, platform: x86_64-w64-mingw32/x64 (64-bit).

**Interaction plot**

We produced the interaction plots with 95% pointwise confidence intervals by applying the function `bmplot.interaction()`. The function draws a fitted regression line for each treatment arm. The dots show the response of population distribution relative to biomarker value. The argument `arms` is the list of treatment arms to be plotted. The following examples are interaction plots for the baseline number of seizures in the left panel and age in the right panel:
The blue and red colors denote the placebo group and the antiepileptic drug progabide group, respectively. The individuals receiving the progabide treatment have a smaller total number of seizures than those receiving the placebo treatment when the Asinh-transformed baseline number of seizures is below five. As the Asinh-transformed baseline number of seizures moves above five, the placebo treatment becomes favorable compared to the progabide treatment. When employing age as a biomarker, the progabide treatment is superior to the placebo treatment among individuals aged above 22.

**Contrast plot**

The contrast plot showing the difference of treatment effects between treatment arms with 95% pointwise confidence interval is given by the function `bmplot.constrast()`. The horizontal benchmark line is calculated by the mean of predicted difference between treatment arms and
displayed as a long dashed line in the contrast plot. The argument `arms` is the list of treatment arms and the first element specifies the reference group. The following R codes are the contrast plots for the baseline number of seizures in the left panel and age in the right panel:

```r
p2.base <- bmplot.contrast(data=gadbury, biomarker="Base_tr", treatment="Trt", arms=c(0,1), outcome="Y", base_font="Times", base_size=14, xlab="Asinh-transformed baseline number of seizures", ylab="Difference of asinh-transformed total number of seizures")

p2.age <- bmplot.contrast(data=gadbury, biomarker="Age", treatment="Trt", arms=c(0,1), outcome="Y", base_font="Times", base_size=14, xlab="Age", ylab="Difference of asinh-transformed total number of seizures")

multi.bmplot(plotlist=list(p2.base,p2.age),ncol=2,nrow=1)
```

The results show that the difference of asinh-transformed total number of seizures becomes higher with the increase of the asinh-transformed baseline number of seizures when comparing the antiepileptic drug progabide treatment group with the placebo group. In contrast, it is the opposite direction when using age as a biomarker.

**Proportion of unfavorable treatment effect plot**

The proportion of unfavorable treatment effect plot shows the estimated proportion of patients who would possibly suffer an unfavorable outcome from treatment and is given by the function `bmplot.proportion()`. A 95% pointwise confidence interval is also presented in the proportion of
unfavorable treatment effect plot. The argument \texttt{ncoef} is the number of produced regression coefficients from the specified multivariate normal distribution and used for the estimation of confidence intervals. The following R codes are the proportion of unfavorable treatment effect plots for the baseline number of seizures in the left panel and age in the right panel:

\begin{verbatim}
p3.base <- bmplot.proportion (data=gadbury, biomarker="Base_tr", treatment="Trt", arms=c(0,1), outcome="Y", ncoef = 1000, xlab="Asinh-transformed baseline number of seizures", ylab="Proportion")

p3.age <- bmplot.proportion (data=gadbury, biomarker="Age", treatment="Trt", arms=c(0,1), outcome="Y", ncoef = 1000, xlab="Age", ylab="Proportion")

multi.bmplot(plotlist=list(p3.base,p3.age),ncol=2,nrow=1)
\end{verbatim}

In our package, we do not have the function to add a vertical line displaying the choice of threshold automatically. It is recommended to add a vertical line as a threshold where the curve reaches 0.5 in the left panel; the implication is that the progabide treatment is suggested for the individuals whose baseline number of seizures falls below this threshold. When employing age as a biomarker, the progabide treatment is suggested for individuals who age falls above this threshold.

\textbf{ROC curve}

In regard to the ROC curve in the cbpti package, we do not employ Huang’s methodology since their approach uses ROC curves in the potential outcome framework under severe restriction. We
generalize this idea to the standard two parallel armed RCT setting with continuous outcome and use a simple algorithm for estimation of the ROC.

The ROC displays true positive fractions and false positive fractions and can be plotted using the function `bmplot.ROC()`. The ROC analysis is constructed on a basis of contrast plot. For the treatment differences (delta) between two groups given each biomarker value, a randomly generated sample of deltas from a normal distribution is created. A certain biomarker value is then specified as a cut-off point. If an individual’s biomarker value is above this cut-off, this individual is considered a good responder, otherwise the individual is considered a bad responder. In this randomly generated population, we can categorize all individuals into four types: true positive, false positive, true negative and false negative individuals. The true positive fraction (TPF) is defined as the probability of correctly identifying a good responder. The false positive fraction (FPF) is defined as the probability of incorrectly classifying a bad responder as a good responder. Given different cut-off points, a ROC curve can be created and the area under the curve can be estimated. The argument `n` is the size of the generated random sample from the normal distribution and used for the improved estimation of sensitivity and specificity. The calculation of these estimates was done by applying the package ROCR [5].

```r
p4.base <- bmplot.ROC(data=gadbury, biomarker="Base_tr", treatment="Trt", arms=c(1,0), outcome="Y", ncoef=100, n = 100, xlab="False positive fraction", ylab="True positive fraction")
p4.age <- bmplot.ROC(data=gadbury, biomarker="Age", treatment="Trt", arms=c(0,1), outcome="Y", ncoef=100, n = 100, xlab="False positive fraction", ylab="True positive fraction")
multi.bmplot(plotlist=list(p4.base,p4.age),ncol=2,nrow=1)
```

The ROC curve may lie below the chance diagonal (implying the area under the curve (AUC) of the ROC curve < 0.5) depending on the direction of treatment effect across the range of biomarker value. In this case, we can reverse the classifier by switching the treatment arms in the argument `arms` and recalculate the AUC for the corrected classifier (corrected AUC=1-AUC), in order to enable the
comparison between the performance of different biomarkers. The interpretation of AUC for the baseline number of seizures in the left panel is that a bad responder will have a higher biomarker value than 62.4% (95%CI=52.3%-70.2%) of the good responders, while for age in the right panel is that a good responder will have a higher biomarker value than 60.8% (95%CI=53.2%-67.3%) of the bad responders.

**Prediction curves**

Since the true positive fraction and the false positive fraction are of no practical use in helping clinicians estimate the probability of good response in individual patients, we calculate the proportion of good responders and bad responders based on our ROC analysis. Given each biomarker value, a randomly generated sample of deltas from a normal distribution is created. The proportion of good responders (delta<0) and bad responders (delta>=0) can be estimated given each randomly generated sample for each biomarker value. The following R codes create the plots for the baseline number of seizures in the left panel and age in the right panel:

```r
p5.base <- bmplot.prediction(data=gadbury, biomarker="Base_tr", treatment="Trt", arms=c(0,1), outcome="Y", ncoef=1000, n=100, xlab="Asinh-transformed baseline number of seizures", ylab="Probability")
p5.age <- bmplot.prediction(data=gadbury, biomarker="Age", treatment="Trt", arms=c(0,1), outcome="Y", ncoef=1000, n=100, xlab="Age", ylab="Probability")
multi.bmplot(plotlist=list(p5.base,p5.age),ncol=2,nrow=1)
```
The blue and red colors denote the proportion of good responders and the proportion of bad responders under the progabide treatment, respectively. The proportion of good responders under the progabide treatment is higher than bad responders for smaller Asinh-transformed baseline number of seizures. Once the Asinh-transformed baseline number of seizures moves above five, the proportion of bad responders exceeds that of good responders. When employing age as a biomarker, the proportion of bad responders exceeds that of good responders for age below 22.

References