Derivation and analysis of kidney HCE

Niels Jongs, Samvel B. Gasparyan

27 October, 2025

Contents

List of Tables

List of Figures

Introduction

In this supplementary material we will explain how to derive the analysis dataset for the kidney hierarchical composite endpoint (HCE), how to analyze it using win odds (Samvel B. Gasparyan et al. 2021), and how to visualize it using *maraca* plots (Karpefors, Lindholm, and Gasparyan 2023). Individual GFR slopes are calculated from the power-of-the-mean (POM) mixed-effects model (Vonesh et al. 2019). All calculations are done using the R software (R Core Team 2022).

- The win odds and its confidence interval are calculated using the `hce` package (Samvel B. Gasparyan 2025).
- The *maraca* plots are implemented in the package `maraca` (Martin Karpefors, Samvel B. Gasparyan, and Monika Huhn 2025).

In addition, we will be using the following packages:

- The package `readxl` for reading the data (Wickham and Bryan 2023).
- The package `nlme` for fitting a power-of-the-mean model for GFR slopes (Pinheiro and Bates 2000).
- The package `multcomp` to derive the total GFR slope from the power-of-the-mean model (Hothorn, Bretz, and Westfall 2008).
- The package `dplyr` for data manipulation (Wickham et al. 2023).
- The package `ggplot2` for customizing the maraca plots (Wickham 2016).

```
library(readxl)
library(ggplot2)
library(nlme)
library(multcomp)
library(dplyr)

library(hce)
library(maraca)
```

```
packageVersion("hce")
```

```
## [1] '0.8.5'
```

```
packageVersion("maraca")
```

```
## [1] '1.0.1'
```

Input datasets

We will use the following three input synthetic datasets that are structured according to clinical data standards for analysis datasets. The datasets are `ADSL` containing patient level data, `ADLB` containing GFR measurements for all patients, and `ADET` containing events of interest (death, chronic dialysis, sustained eGFR < 15, sustained eGFR declines of given threshold) and their study day of occurrence. Only one event per patient per type is kept in the `ADET` dataset.

```
ADSL <- read_excel("data/ADSL.xlsx")
ADET <- read_excel("data/ADET.xlsx")
ADLB <- read_excel("data/ADLB.xlsx")
```

```
head(ADSL)
```

Table 1: ADSL dataset structure

ID	TRTPN	EGFRBL	STRATAN
1	1	44	1
2	1	31	3
3	1	28	4
4	1	61	4
5	1	37	4
6	2	35	1

The dataset `ADSL` contains (see Table 1) the columns `ID` for patient ID, `TRTPN`, the planned treatment group, 1 for the active group, 2, for the control group. `STRATAN` contains randomization stratum for each patient (1-4), where a higher value means a higher risk of kidney progression.

```
head(ADET)
```

Table 2: ADET dataset structure

ID	AVAL	PARAM	PARAMCD	PARAMN	TRTPN
11	359	Sustained >=50% decline in eGFR	EGFR50	5	2
11	467	Sustained >=57% decline in eGFR	EGFR57	4	2
11	359	Sustained >=40% decline in eGFR	EGFR40	6	2
11	841	Sustained eGFR<15 (mL/min/1.73 m2)	EGFR15	3	2
15	962	Death (adj)	DTHADJ	1	1
21	737	Sustained >=40% decline in eGFR	EGFR40	6	1

The dataset `ADET` contains subject IDs, `AVAL` column specifies the study day of the event (days from randomization), `PARAM` specifies the type of the event, `PARAMCD` the coded type of the event, while `PARAMN` specifies the priority of the event, the numbers 1 to 6, where a higher value means a better (less severe) outcome.

```
ADET |> select(PARAM, PARAMCD, PARAMN) |> arrange(PARAMN) |> unique()
```

Table 3: Kidney HCE

PARAM	PARAMCD	PARAMN
Death (adj)	DTHADJ	1
Chronic dialysis (adj) >=90 days	DIAL90	2
Sustained eGFR<15 (mL/min/1.73 m2)	EGFR15	3
Sustained >=57% decline in eGFR	EGFR57	4
Sustained >=50% decline in eGFR	EGFR50	5
Sustained >=40% decline in eGFR	EGFR40	6

We need to derive the category 7 of the hierarchical composite endpoint (measurements), which is for patients not having any of the dichotomous outcomes 1-6 and who contribute to the analysis with their individual GFR slope. For this we will use the dataset `ADLB` of laboratory measurements.

```
head(ADLB)
```

Table 4: ADLB dataset structure

ID	TRTPN	AVAL	ADAY	AVISITN	PARAM	PARAMCD	PARAMN
1	1	40	357	7	eGFR measurements	eGFR	7
1	1	49	119	5	eGFR measurements	eGFR	7
1	1	41	238	6	eGFR measurements	eGFR	7
1	1	36	469	8	eGFR measurements	eGFR	7
1	1	40	721	10	eGFR measurements	eGFR	7
1	1	34	602	9	eGFR measurements	eGFR	7

In this dataset the column `AVAL` is the GFR measurement of patients done at the visit specified in the column `AVISITN`, while `ADAY` is the analysis day of the visit.

We will restrict the analysis to events and GFR measurements up to given cut-off of 3 years since randomization and we will consider the acute effect happening at the first visit, which is Day 14 (we divide by 360 to convert to years). We will derive also the coefficient for the chronic phase, which is the proportion of the length of the chronic phase (total follow-up minus the acute phase) divided by the total follow-up for annualizing the results. This will be used in deriving the total GFR slope from the two slope power-of-the-mean model.

```
CUTOFF <- 3
ACUTE <- 14/360
CHRONIC_coef <-  (CUTOFF - ACUTE)/CUTOFF
```

The following code annualizes the analysis days (by dividing them by 360) and restricts the GFR measurements and dichotomous outcomes to the cut-off timepoint, and only one event per patient, their most severe event, is selected (the event with the lowest `PARAMN` value). The resulting summary table shows the number of patients with a given most severe event.

```
ADLB$ADAY <- ADLB$ADAY/360
ADET$AVAL <- ADET$AVAL/360
ADLB <- ADLB[ADLB$ADAY <= CUTOFF, ]
EVNT <- ADET |> filter(AVAL <= CUTOFF) |>  arrange(ID, PARAMN)|> group_by(ID) |>
  mutate (n = row_number()) |> ungroup() |> filter(n == 1) |> select(- c("n"))
```

```
table(EVNT$PARAM, EVNT$TRTPN)
```

Table 5: Frequency of outcomes by treatment group

	Active	Control
Chronic dialysis (adj) >=90 days	17	29
Death (adj)	40	50
Sustained >=40% decline in eGFR	36	34
Sustained >=50% decline in eGFR	7	22
Sustained >=57% decline in eGFR	2	9
Sustained eGFR<15 (mL/min/1.73 m2)	16	28

Total GFR slope difference calculation

We will merge the datasets `ADLB` and `ADSL` in order to have `STRATAN` and `EGFRBL` columns present with the GFR measurements in the same dataset. Then we will fit a *linear mixed effect model (LME)* using the function `lme()` in the package `nlme`.

The model will include only measurements up to the cut-off day, will be adjusted for baseline GFR values, and will include in the model covariates for the stratification, the treatment group (converted to a 0, 1 variable

with 1 for the active group and 0 for the control), analysis day ADAY, the derived SPLINE variable which corresponds to time since the acute phase, treatment by analysis day, and treatment by spline interaction terms, without an intercept term.

```
ADLBO <- merge(ADLB, ADSL[, c("ID", "STRATAN", "EGFRBL")], by = "ID", all.x = T)
ADLBO$SPLINE <- ifelse(ADLBO$ADAY <= ACUTE, 0, ADLBO$ADAY - ACUTE)

ADLBO$TRT <- ifelse(ADLBO$TRTPN == 1, 1, 0)
fit <- lme(AVAL ~ EGFRBL + STRATAN + TRT + ADAY + SPLINE + ADAY*TRT + SPLINE*TRT - 1,
            random = list(ID = pdSymm(form = ~ 1 + ADAY)),
            weights = varComb(varIdent(form = ~ 1|TRT),
                              varPower(form = ~ 1 + ADAY)),
            na.action = na.omit,
            data = ADLBO,
            method  = "REML",
            control = lmeControl(maxIter = 1e8,
                                 msMaxIter = 1e8,
                                 opt = "optim",
                                 optimMethod = "L-BFGS-B",
                                 msVerbose = T))
```

```
## Warning in optim(c(oldPars), function(lmePars) -logLik(lmeSt, lmePars), :
## method L-BFGS-B uses 'factr' (and 'pgtol') instead of 'reltol' and 'abstol'
```

```
## final  value 92903.802147
## converged
```

Then, the function glht() from the package multcomp can be used to calculate the total GFR slope from a two-slope model accounting for the chronic phase through the CHRONIC_coef coefficient.

```
k <- ncol(coef(fit)) - 1
MTP <- glht(fit,
            linfct = rbind("Total: A - C" =   c(rep(0, k - 2), 1, CHRONIC_coef)))
SLP <- confint(summary(MTP))
Slope <- SLP$confint[1:3]
```

Therefore, the total GFR slope difference between (and its 95% confidence interval) active and control group per year is retained in the variable Slope and presented in Table 6. GFR slope in the active and control groups is calculated in the variables Slope0 and Slope1 respectively (using the same model) and presented in Table 7.

```
MTP0 <- glht(fit,
            linfct = rbind("Total: A - C" =   c(rep(0, k - 4), 1,
                                                CHRONIC_coef, 1, CHRONIC_coef)))
SLP0 <- confint(summary(MTP0))
Slope0 <- SLP0$confint[1:3]
MTP1 <- glht(fit,
             linfct = rbind("Total: A - C" =   c(rep(0, k - 4), 1, CHRONIC_coef, 0, 0)))
SLP1 <- confint(summary(MTP1))
Slope1 <- SLP1$confint[1:3]
```

Table 6: Total GFR slope difference between treatment groups

GFR Slope difference	Lower 95% CI	Upper 95% CI
1.1132	0.5888	1.6377

Table 7: Total GFR slope by treatment groups

	GFR Slope	Lower 95% CI	Upper 95% CI
Active	-3.6464	-4.0162	-3.2765
Control	-2.5332	-2.9049	-2.1614

We will use this model for predicting individual GFR slope values for all patients at the cut-off timepoint.

Derivation of individual GFR slopes

For predicting the annualized change from baseline in GFR values for all patients at the end of follow-up, we need to create a dataset containing the necessary (the ones included in the model above) baseline covariates of all patients.

Analysis day is selected as the length of follow-up (`ADAY`), the `SPLINE` variable for the years since the acute phase will be length of follow-up minus the acute phase. Then we can predicted the GFR values at the end of follow-up using the `predict()` function to obtain predicted GFR values for all patients at the end of follow-up, and subtract the corresponding baseline GFR values for each patient (dividing by the length of follow-up will provide the annualized individual GFR slopes for all patients).

```
ADLB1 <- ADLB0[ADLB0$ADAY == 0, ]
ADLB1$ADAY <- CUTOFF
ADLB1$SPLINE <- CUTOFF - ACUTE
ADLB1$AVALP <- (predict(fit, newdata = ADLB1) - ADLB1$EGFRBL)/CUTOFF
Desc <- tapply(ADLB1$AVALP, ADLB1$TRT, function(x) list(mean = mean(x), sd = sd(x)))
do.call(rbind, Desc)
```

```
##   mean         sd
## 0 -3.690067355 4.37486335
## 1 -2.545862177 4.328511519
```

```
ADLB1$AVALP <- round(ADLB1$AVALP, 2)
ADLB1 <- ADLB1[, c("ID", "AVALP")]
ADLB2 <- merge(ADLB1, EVNT, by = "ID", all.x = T)
ADLB2$AVAL <- ifelse(is.na(ADLB2$AVAL), ADLB1$AVALP, ADLB2$AVAL)
ADLB2$PARAMCD <- ifelse(is.na(ADLB2$PARAMCD), "eGFR", ADLB2$PARAMCD)
ADLB2$PARAM <- ifelse(is.na(ADLB2$PARAM), "eGFR slope", ADLB2$PARAM)
ADLB2$PARAMN <- ifelse(is.na(ADLB2$PARAMN), 7, ADLB2$PARAMN)
```

Note that although these individual GFR slope values are calculated for all patients, but they will be used only for those patients who did not get one of the dichotomous events described in Table 3. Then, we derive the `PARAM` and `PARAMCD` values for those patients, and the priority is set to 7, `PARAMN=7`. We will obtain the following dataset.

```
head(ADLB2)
```

Table 8: Predicted individual GFR slope values

ID	AVALP	AVAL	PARAM	PARAMCD	PARAMN	TRTPN
1	-3.03	-3.03	eGFR slope	eGFR	7	NA
2	1.75	1.75	eGFR slope	eGFR	7	NA
3	-1.90	-1.90	eGFR slope	eGFR	7	NA
4	-4.27	-4.27	eGFR slope	eGFR	7	NA
5	-2.29	-2.29	eGFR slope	eGFR	7	NA
6	-1.18	-1.18	eGFR slope	eGFR	7	NA

Kidney HCE dataset

Lastly, we derive the dataset for the kidney HCE based on predicted individual GFR values (if the patient did not experience a dichotomous outcome during the follow-up) or the study day of the most severe dichotomous event, if the patient experienced one of the outcomes 1-6 in Table 3.

```
HCE <- ADLB2[, c("ID", "AVAL", "PARAM", "PARAMCD", "PARAMN")]
names(HCE)[names(HCE) == "AVAL"] <- "AVAL0"
names(HCE)[names(HCE) == "PARAM"] <- "GROUP"
```

In the code below we derive the column `GROUPN` which is the prioritization of outcomes and is meant to introduce order between categories and within each category. Since within each category 1-6 the maximum value is the length of fixed follow-up (since within these categories the study day of the most severe events is used), then we can multiply the length of the follow-up with the priority number. We do this since in the next step we will be adding the study day of the events to the `GROUPN` values to introduce a ranking within each category. A particular care is needed for the category 7. The values in this category can be negative hence adding to the value of `GROUPN` we may get lower values than the values in the category 6. Hence we need to make sure that the value for `GROUPN` in this category is large enough.

```
M <- floor(abs(max(HCE$AVAL0[HCE$PARAMCD == "eGFR"]))) + 1
CUTOFF0 <- max(c(M, CUTOFF))
HCE$GROUPN <- CUTOFF*HCE$PARAMN
HCE$GROUPN[HCE$PARAMCD == "eGFR"] <- CUTOFF0*HCE$PARAMN[HCE$PARAMCD == "eGFR"]
HCE$AVAL <- HCE$AVAL0 + HCE$GROUPN

ADHCE <- merge(HCE, ADSL, by = "ID")
ADHCE$TRTP <- ifelse(ADHCE$TRTPN == 1, "A", "P")
```

```
head(ADHCE)
```

Table 9: Kidney HCE data structure

ID	AVAL0	GROUP	PARAMCD	PARAMN	GROUPN	AVAL	TRTPN	EGFRBL	STRATAN	TRTP
1	-3.03	eGFR slope	eGFR	7	217	213.97	1	44	1	A
2	1.75	eGFR slope	eGFR	7	217	218.75	1	31	3	A
3	-1.90	eGFR slope	eGFR	7	217	215.10	1	28	4	A
4	-4.27	eGFR slope	eGFR	7	217	212.73	1	61	4	A
5	-2.29	eGFR slope	eGFR	7	217	214.71	1	37	4	A
6	-1.18	eGFR slope	eGFR	7	217	215.82	2	35	1	P

Win odds and maraca plots

In this section we calculate the win odds and its confidence interval and plot the kidney HCE using the *maraca* plot (Karpefors, Lindholm, and Gasparyan 2023).

```
res <- calcWO(x = ADHCE, AVAL = "AVAL", TRTP = "TRTP", ref = "P")
res0 <- res[, c("WO", "LCL", "UCL", "Pvalue")]
```

```
res0
```

Table 10: Win odds and its confidence interval

WO	LCL	UCL	Pvalue
1.32	1.1733	1.485	0

And the maraca plot can be created as follows:

```r
hce_test <- maraca(
  data = ADHCE,
  step_outcomes = c("DTHADJ", "DIAL90", "EGFR15", "EGFR57", "EGFR50",  "EGFR40"),
  last_outcome = "eGFR",
  fixed_followup_days = CUTOFF,
  column_names = c(outcome = "PARAMCD", arm = "TRTP", value = "AVALO"),
  arm_levels = c(active = "A", control = "P"),
  compute_win_odds = FALSE
)
```

```r
plot(hce_test) + theme_classic() + xlab("") + ylab("") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 7),
        legend.position = "bottom")
```

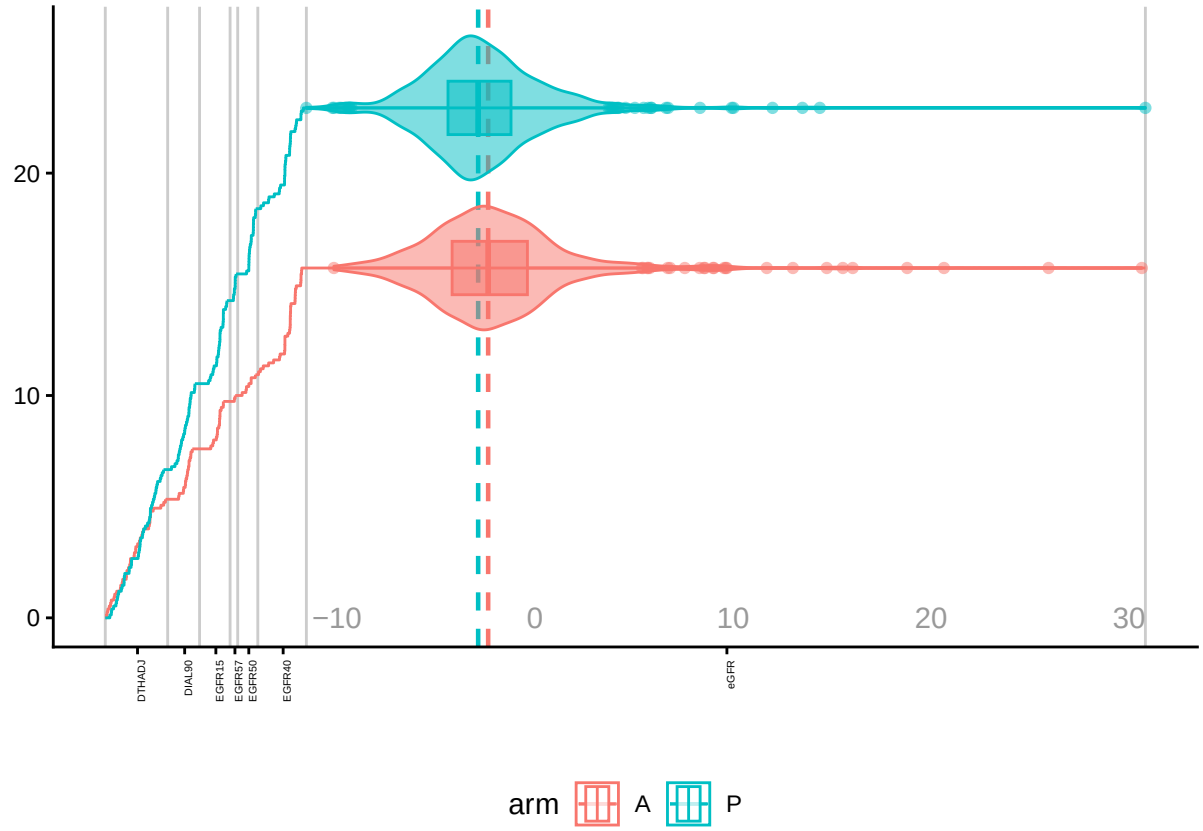

Figure 1: The maraca plot of kidney HCE

Acknowledgment

The synthetic datasets were created and kindly provided by the Analytics Data Preparation Team (ADAPT), Data Office, Data Science and AI, AstraZeneca.

Appendix

```
library(readxl)
library(ggplot2)
library(nlme)
library(multcomp)
library(hce)
library(maraca)
library(dplyr)

ADSL <- read_excel("data/ADSL.xlsx")
ADET <- read_excel("data/ADET.xlsx")
ADLB <- read_excel("data/ADLB.xlsx")
CUTOFF <- 3
ACUTE <- 14/360
CHRONIC_coef <-  (CUTOFF - ACUTE)/CUTOFF

ADLB$ADAY <- ADLB$ADAY/360
ADET$AVAL <- ADET$AVAL/360
ADLB <- ADLB[ADLB$ADAY <= CUTOFF, ]
EVNT <- ADET |> filter(AVAL <= CUTOFF) |>  arrange(ID, PARAMN)|> group_by(ID) |>
  mutate (n = row_number()) |> ungroup() |> filter(n == 1) |> select(- c("n"))

table(EVNT$PARAM, EVNT$TRTPN)
ADLB0 <- merge(ADLB, ADSL[, c("ID", "STRATAN", "EGFRBL")], by = "ID", all.x = T)
ADLB0$SPLINE <- ifelse(ADLB0$ADAY <= ACUTE, 0, ADLB0$ADAY - ACUTE)
ADLB0$TRT <- ifelse(ADLB0$TRTP == 1, 1, 0)

fit <- lme(AVAL ~ EGFRBL + STRATAN + TRT + ADAY + SPLINE + ADAY*TRT + SPLINE*TRT - 1,
           random = list(ID = pdSymm(form = ~ 1 + ADAY)),
           weights = varComb(varIdent(form = ~ 1|TRT),
                             varPower(form = ~ 1 + ADAY)),
           na.action = na.omit,
           data = ADLB0,
           method  = "REML",
           control = lmeControl(maxIter = 1e8,
                                msMaxIter = 1e8,
                                opt = "optim",
                                optimMethod = "L-BFGS-B",
                                msVerbose = T))
k <- ncol(coef(fit)) - 1
MTP <- glht(fit,
            linfct = rbind("Total: A - C" =   c(rep(0, k - 2), 1, CHRONIC_coef)))
SLP <- confint(summary(MTP))
Slope <- SLP$confint[1:3]
Slope

MTP0 <- glht(fit,
            linfct = rbind("Total: A - C" =   c(rep(0, k - 4), 1,
                                                CHRONIC_coef, 1, CHRONIC_coef)))
SLP0 <- confint(summary(MTP0))
Slope0 <- SLP0$confint[1:3]
Slope0
MTP1 <- glht(fit,
```

```
            linfct = rbind("Total: A - C" =   c(rep(0, k - 4), 1, CHRONIC_coef, 0, 0)))
SLP1 <- confint(summary(MTP1))
Slope1 <- SLP1$confint[1:3]
Slope1

ADLB1 <- ADLB0[ADLB0$ADAY == 0, ]
ADLB1$ADAY <- CUTOFF
ADLB1$SPLINE <- CUTOFF - ACUTE
ADLB1$AVALP <- (predict(fit, newdata = ADLB1) - ADLB1$EGFRBL)/CUTOFF
Desc <- tapply(ADLB1$AVALP, ADLB1$TRT,
               function(x) list(mean = mean(x), sd = sd(x)))
do.call(rbind, Desc)

ADLB1$AVALP <- round(ADLB1$AVALP, 2)
ADLB1 <- ADLB1[, c("ID", "AVALP")]
ADLB2 <- merge(ADLB1, EVNT, by = "ID", all.x = T)
ADLB2$AVAL <- ifelse(is.na(ADLB2$AVAL), ADLB1$AVALP, ADLB2$AVAL)
ADLB2$PARAMCD <- ifelse(is.na(ADLB2$PARAMCD), "eGFR", ADLB2$PARAMCD)
ADLB2$PARAM <- ifelse(is.na(ADLB2$PARAM), "eGFR slope", ADLB2$PARAM)
ADLB2$PARAMN <- ifelse(is.na(ADLB2$PARAMN), 7, ADLB2$PARAMN)

HCE <- ADLB2[, c("ID", "AVAL", "PARAM", "PARAMCD", "PARAMN")]
names(HCE)[names(HCE) == "AVAL"] <- "AVAL0"
names(HCE)[names(HCE) == "PARAM"] <- "GROUP"
M <- floor(abs(max(HCE$AVAL0[HCE$PARAMCD == "eGFR"]))) + 1
CUTOFF0 <- max(c(M, CUTOFF))

HCE$GROUPN <- CUTOFF*HCE$PARAMN
HCE$GROUPN[HCE$PARAMCD == "eGFR"] <- CUTOFF0*HCE$PARAMN[HCE$PARAMCD == "eGFR"]
HCE$AVAL <- HCE$AVAL0 + HCE$GROUPN
ADHCE <- merge(HCE, ADSL, by = "ID")
ADHCE$TRTP <- ifelse(ADHCE$TRTPN == 1, "A", "P")
head(ADHCE)

res <- calcWO(x = ADHCE, AVAL = "AVAL", TRTP = "TRTP", ref = "P")
res0 <- res[, c("WO", "LCL", "UCL", "Pvalue")]
res0

hce_test <- maraca(data = ADHCE,
                   tte_outcomes = c("DTHADJ", "DIAL90", "EGFR15",
                                    "EGFR57", "EGFR50",  "EGFR40"),
                   continuous_outcome = "eGFR",
                   fixed_followup_days = CUTOFF,
                   column_names = c(outcome = "PARAMCD", arm = "TRTP", value = "AVAL0"),
                   arm_levels = c(active = "A", control = "P"),
                   compute_win_odds = FALSE)

plot(hce_test) + theme_classic() + xlab("") + ylab("") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 7),
        legend.position = "bottom")
```

References

Gasparyan, Samvel B. 2025. *hce: Design and Analysis of Hierarchical Composite Endpoints*. https://CRAN.R-project.org/package=hce.

Gasparyan, Samvel B, Elaine K Kowalewski, Folke Folkvaljon, Olof Bengtsson, Joan Buenconsejo, John Adler, and Gary G Koch. 2021. "Power and Sample Size Calculation for the Win Odds Test: Application to an Ordinal Endpoint in COVID-19 Trials." *Journal of Biopharmaceutical Statistics* 31 (6): 765–87.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. "Simultaneous Inference in General Parametric Models." *Biometrical Journal* 50 (3): 346–63. https://CRAN.R-project.org/package=multcomp.

Karpefors, Martin, Daniel Lindholm, and Samvel B Gasparyan. 2023. "The Maraca Plot: A Novel Visualization of Hierarchical Composite Endpoints." *Clinical Trials* 20 (1): 84–88.

Martin Karpefors, Samvel B. Gasparyan, and Monika Huhn. 2025. *Maraca: The Maraca Plot: Visualization of Hierarchical Composite Endpoints in Clinical Trials*. https://CRAN.R-project.org/package=maraca.

Pinheiro, José C, and Douglas M Bates. 2000. "Fitting Nonlinear Mixed-Effects Models." *Mixed-Effects Models in S and S-PLUS*, 337–421. https://CRAN.R-project.org/package=nlme.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing, Version 4.2.2*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Vonesh, Edward, Hocine Tighiouart, Jian Ying, Hiddo L Heerspink, Julia Lewis, Natalie Staplin, Lesley Inker, and Tom Greene. 2019. "Mixed-Effects Models for Slope-Based Endpoints in Clinical Trials of Chronic Kidney Disease." *Statistics in Medicine* 38 (22): 4218–39.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, and Jennifer Bryan. 2023. *Readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.