QOL ANALYSIS METHODOLOGY

Descriptive information of patient and tumor characteristics is provided for each Arm. Since this is a randomised study, no p-values were reported for the comparison of baseline variables between both groups (Roberts and Torgerson, 1999).

Following the definitions presented by Annota et al (2017), the time until definitive deterioration (TUDD) was defined as the time from inclusion in the study (date of randomisation) to a first deterioration of at least 10 points in Global Health Status (GHS) score as compared to the baseline GHS score, with no further improvement of more than 10 points as compared to the score qualifying the deterioration (i.e., the score at the time of the first deterioration observed), or with no GHS data after the deterioration was observed. Further, all-cause death was also considered as an event if the patient did not experience deterioration before death. In this way, TUDD or death was redefined as "GHS deterioration-free survival". A limitation of the study is the potentially large interval between the last questionnaire and the last follow up (or death). This implies that for some patients at some moment in time, the event is only defined by the occurrence of death and not by a change in score. Patients without event were censored at the time of last follow-up.

Kaplan-Meier curves were used to describe the GHS deterioration-free survival, as well as the overall survival (OS) and the progression-free survival (PFS). Comparison of the curves was done using logrank tests. Since the randomisation was stratified on centre (more specifically, a minimisation procedure with a random element (Altman and Bland, 2005) involving ECOG performance status, stage and location of the tumor was applied within each centre separately), stratified versions of the tests were performed as well and are available upon request.

An alternative approach was considered using 20 points difference instead of 10 points in the definition of the TUDD.

For the analysis of the primary outcome, i.e. the deterioration-free survival at three months, the survival estimates at three months were compared between both groups using a Z-test on the log-log transformed Kaplan-Meier estimates. A non-parametric bootstrap procedure was used to compare the estimates at three months after stratification on centre. A P-value smaller than 0.05 (or an estimate outside the 95% percentile-based confidence interval in the bootstrap procedure) was considered significant.

Definitive deterioration, not including death in its definition, was used as an irreversible binary time-varying (TV) covariate in Cox regression models for OS and PFS (Therneau and Grambsch, 2000). The aim was to link the GHS trajectory and the risk. This approach has several disadvantages. First, at each point in time the whole trajectory of GHS values is reduced to a dichotomous value: has definitive deterioration already occurred or not. Second, at a point in time a relation between GHS and risk is established using information from the future (this disadvantage is inherent in the definition of the TUDD). Last but not least, the Cox model, extended for time-varying covariates, is only appropriate for so-called exogenous covariates (Rizopoulos, 2012). The path of such exogenous or external covariate can never be influenced by the event time or status. Typical examples are age, time since randomisation, season of the year, air pollution, etc. An internal or endogenous covariate as the quality of life clearly does not fit within this

definition. Treating endogenous covariates as exogenous may produce spurious results in the Cox regression. A more appropriate approach to evaluate the relation between the GHS trajectory and the risk for dying (or progression) is the use of a joint model (Rizopoulos, 2010). In the joint model, the risk for death was allowed to depend on the current level of GHS. A linear mixed model using splines was used to model the evolution of GHS (the use of splines allows a nonlinear relation between time since randomisation and GHS). A spline-approximated baseline risk function was used for the relative risk model. The average evolution of GHS in each group was plotted (with a pointwise 95% confidence interval). The predictions for time until death curve as a function of GHS were visualised for four hypothetical patients having the following GHS trajectories during the first 6 months

- (1) constant rather low GHS (GHS=50)
- (2) constant rather high GHS (GHS=83.3)
- (3) decrease in GHS (from 91.7 to 16.7 over 6-months period)
- (4) increase in GHS (from 16.7 to 91.7 over 6-months period)

Analyses have been performed using SAS software, version 9.4 of the SAS System for Windows and with R software version 3.6.0 (2019 The R foundation for Statistical Computing), using the package Jmfit (Rizopoulos, 2010) for fitting the joint model.

Detailed data available upon request.

References:

- Altman DG, Bland JM (2005). Treatment allocation by minimisation. BMJ. 330(7495):843.
- A. Anota, M.Savina Bergonie, C. Bascoul-Mollevi (2017). QoLR: An R Package for the Longitudinal Analysis of Health-Related Quality of Life in Oncology. Journal of Statistical Software, Volume 77, Issue 12. doi:10.18637/jss.v077.i12
- F.E. Harrell (2001). Regression Modeling Strategies. New York: Springer
- C. Roberts and D.J. Torgerson (1999). Understanding controlled trials. Baseline imbalance in randomised controlled trials. BMJ. 17;319(7203):185.
- D. Rizopoulos (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. Journal of Statistical Software, July 2010, Volume 35, Issue 9.
- Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Boca Raton: Chapman and Hall/CRC
- Therneau TM, Grambsch PM (2000). Modeling survival data: extending the Cox model. New York, NY: Springer-Verlag.