

Statistical Analysis Plan (SAP) for NARLAL II

This SAP has been created using a layout recommended by "Liverpool clinical Trial centre - LCTC":

<https://www.lctc.org.uk/Content/SAP%20Statement%20Elaboration%20Document%20v1.0.pdf>

The form is also linked to from the Equator network:

<https://www.equator-network.org/reporting-guidelines/guidelines-for-the-content-of-statistical-analysis-plans-in-clinical-trials/>

Section 1: Administrative information

Part 1: Title and Trial registration

Statistical analysis plan for the NARLAL II trial (Novel Approach to Radiotherapy in Locally Advanced Lung cancer - Heterogeneous FDG-guided dose escalation with concomitant Navelbine), clinicaltrials.gov NCT02354274.

Part 2: SAP Version

The initial aim of the study and power analysis is defined in the NARLAL II study protocol. The current document is the first version of the final statistical analysis plan, created before the end of patient accrual. The date of the current version (version 1.0) is the 13th of February, 2023

Part 3: Protocol Version

The basis for this SAP is the English translation of the Danish protocol version (version 6) that the Danish research ethics committee approved on the 15th of August 2022 (that version was a small addendum to the protocol regarding an extension of the follow-up time to ten years after radiotherapy and inclusion beyond patient number 350)

Part 4: SAP Revisions – revision history, with justification and timing

The current SAP is the first version of the SAP; thus, there are no revisions.

Part 5 Roles and Responsibility – non-signatory names and contribution

Names are listed alphabetically by first name:

Ane Appelt, Member of the working group drafting the SAP

Carsten Brink, Member of the working group drafting the SAP, and author of the R-package for data analysis

Charlotte Kristiansen, Discussion partner during the development of the SAP

Christina Maria Lutz, Member of the working group drafting the SAP

Ditte Sloth Møller, Member of the working group drafting the SAP

Gitte Fredberg Persson, Discussion partner during the development of the SAP

Lone Hoffmann, Discussion partner during the development of the SAP

Marianne Marquard Knap, Discussion partner during the development of the SAP

Mette Pøhl, Discussion partner during the development of the SAP

Mikkel Drøgemüller Lund, Discussion partner during the development of the SAP

Rune Slot Thing, Discussion partner during the development of the SAP

Tine Bjørn Nielsen, Discussion partner during the development of the SAP

Torben Schjødt Hansen, Discussion partner during the development of the SAP

Part 6: Roles and Responsibility – signatures

Tine Schytte, Primary investigator

Signature:



Section 2: Introduction

Part 7: Background and rationale

The background and rationale for the study are described in section one (*Background*) of the clinical protocol

Part 8: Objectives

The objective of this study is to examine the effect of inhomogeneous, FDG-PET-driven escalation of radiation dose to the primary tumour and involved lymph nodes, compared to standard uniform dose, in definitive chemo-radiation treatment of inoperable locally advanced NSCLC (stage IIB-IIIB).

The null hypothesis of the primary endpoint is that there is no difference in locoregional control rate between the two study arms (standard uniform dose versus inhomogeneous dose escalation). The alternative hypothesis is that there is a difference between the two groups.

Secondary study objectives include evaluation of acute and late toxicity, overall and recurrence-free survival, and correlation of radiation dose with tumour and nodal control.

Section 3: Trial Methods

Part 9: Trial design – description of trial design

The trial is a phase III multi-centre trial, randomising between standard and inhomogeneous dose escalation during radiotherapy of patients with LA-NSCLC. The randomisation ratio is 1:1.

a) Treatment planning and randomization



b) Detailed study timeline

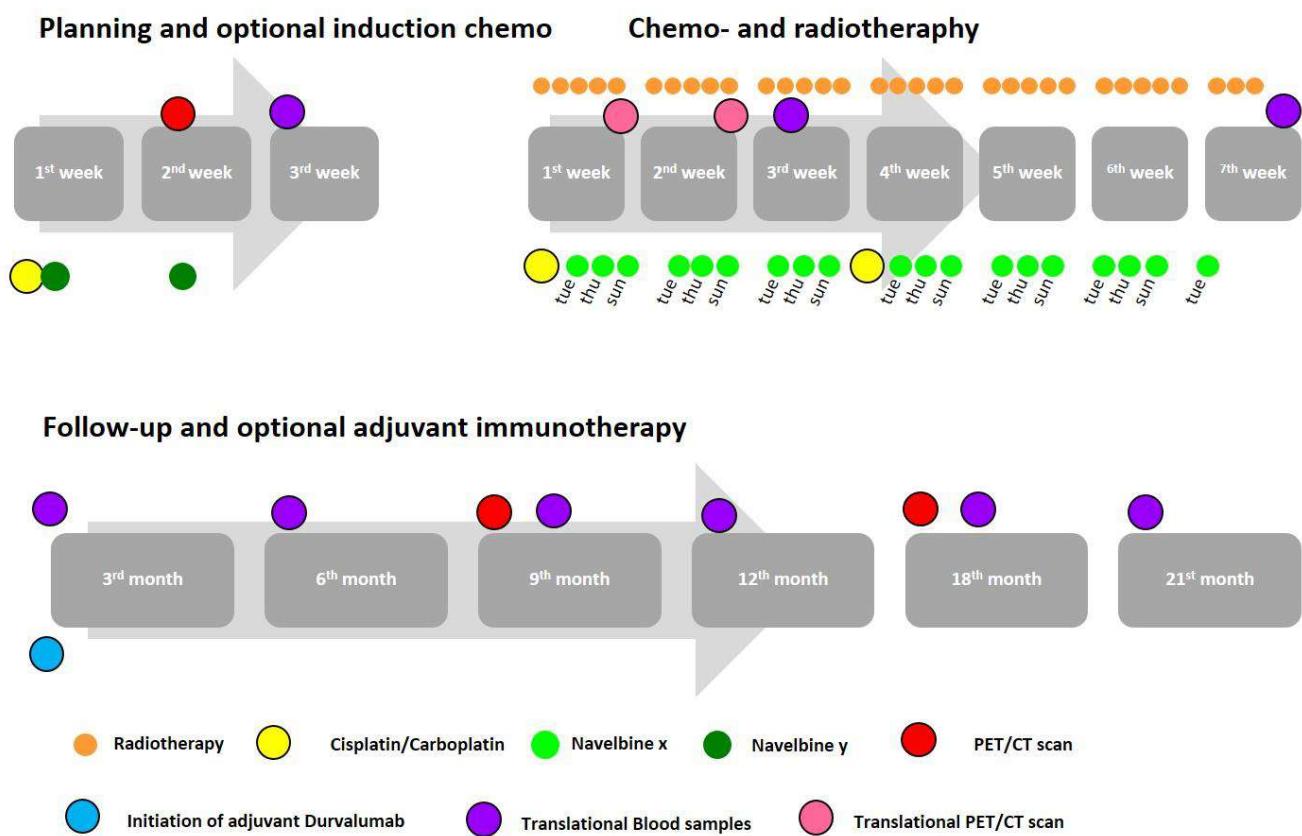


Figure 1: Study schedule for patients in the NARLAL2 trial.

a) Overall study schedule leading up to randomization.

b) Detailed time schedule of chemo- and radiation therapy for a patient who starts on Monday and the subsequent follow-up. Standard follow-up CTs are not shown. In 2019, adjuvant treatment with durvalumab was approved as part of the standard treatment of advanced lung cancer. The implementation in the NARLAL2 study schedule is indicated (blue). Translational PET/CT scans during RT as well as translational blood samples before, during and after RT are indicated.

Part 10: Randomisation

The randomisation is performed within blocks and is stratified on treatment institution and histology (squamous vs non-squamous cell carcinoma). The randomisation lists have been created by the "Open Patient data Explorative Network" (OPEN – Region of Southern Denmark) and implemented in the related

Redcap database. The size of the stratification blocks is unknown to any of the investigators. The study is open-label for the patient and the treating physician, but the allocation to each treatment arm is blinded. For each centre, the randomisation is performed within the Redcap database, typically accessed by the local clinical research unit. The ability to deliver a standard arm treatment plan is a prerequisite for study enrolment and randomisation. Radiotherapy treatment plans for both study arms were finalised by the local dose planner and approved by the treating physician before randomisation was communicated to the treating team.

Part 11: Sample size

The primary endpoint is cause-specific locoregional control, i.e. patients are censored on death and loss to follow-up for locoregional recurrence. If the assessment of locoregional control status is not possible physically or ethically (e.g. due to disseminated metastatic disease), the patient will be censored at the date of the last available imaging that can assess the locoregional control status. The study sample size was based on an expected effect size of HR=0.57 for locoregional control between the two treatment arms. This corresponds to a difference of 16 percentage points in locoregional control at 30 months or a change in median locoregional control from 36 to 63 months (assuming exponentially distributed event times and constant hazards in the two arms). Further, the modelling assumptions included censoring due to death (prior to locoregional failure) based on a median survival time of 42 months.

The sample size calculation is based on a power of 80% to resolve the expected hazard ratio with a 5% significance level. At the time of protocol initiation, the original sample size estimate assumed a simple log-rank test for the difference between the two arms. However, the introduction of Durvalumab (as consolidation treatment post chemoradiotherapy for responders) during the study period resulted in a protocol amendment, including an update of the sample size calculation. At the time of the protocol amendment, it was unclear whether the routine use of Durvalumab would significantly change local control rates. Therefore, the primary endpoint analysis was amended to use a stratified log-rank test, with stratification on whether or not the patient received Durvalumab. At the same time, new sample size calculations were performed (documented in the protocol). Following the amendment, the final number of planned patients in the study was 350 (175 per study arm).

As mentioned in part 3, the study will continue recruiting after the inclusion of patient number 350 and until the group is ready to publish the primary endpoint, to increase the power of detecting potential differences in toxicity (see part 27). However, the primary endpoint will be evaluated on the initial 350 patients alone.

Part 12: Framework

The study has been initiated with the expectation that the intervention arm will result in a superior local control rate. However, due to previous dose-escalation studies, which indicated a risk of reduced control in the dose-escalation arm, the study is designed as a two-sided test to show differences between the treatment arms.

Part 13: Statistical Interim analyses and stopping guidance

Interim analyses have been carried out during the inclusion period following the specifications in the clinical protocol. The main purpose of the interim analyses was to detect and pause the trial in case of unexpected and unacceptable acute or semi-acute/late toxicity or reduced overall survival in the dose escalation arm.

The interim toxicity analyses used the O'Brien-Fleming [1] method to define the stopping rules for acute and semi-acute/late toxicity, respectively, with an overall significance level of 5% for each set of analyses. The acute toxicity interims were performed after 3 months of follow-up of the initial 20, 40, and 60 patients included in the dose-escalation arm. The semi-acute/late analyses were performed for the same number of patients after one year of follow-up. The aim was to stop the study if radiation-related toxicity of grade 4+ was observed for more than 10% of the patients at three months and more than 5% at one year in the dose escalation arm.

An interim analysis of overall survival was performed halfway through the study (i.e. after randomisation of patient 165, as per the original sample size of 330 patients), using a log-rank test with a significance level of 1%. This analysis was not performed directly on the primary endpoint; thus, the significance level for the primary endpoint is not adjusted due to these interim analyses (no adjustment for alpha-spending).

Further details of the interim analyses are provided in the protocol. Due to legal requirements, the principal investigator must evaluate all serious adverse events (SAE) during the study. Thus, the principal investigator had access to all SAE information and performed the toxicity-related interim analyses. The survival interim was performed by an independent expert review (Professor Per Pfeiffer, Department of Oncology, Odense University Hospital). At the time of writing the SAP, all interim analyses were passed without pausing the study.

Part 14 Timing of final analysis

The primary endpoint analysis is planned to be performed one year after the randomisation of the last patient. This analysis depends on complete information for the primary endpoint at the trial management centre. Full information includes:

- Key baseline patient characteristics;
- Information on treatment delivered, including any protocol violations;
- Radiotherapy plans transferred to the national treatment plan bank (DcmCollab) and ready for evaluation;
- Complete information on local control. Complete information is defined as information on the local-control status of all patients at all preplanned imaging sessions until one year after the inclusion of the last patient. Imaging cancelled or delayed beyond the one-year limit should not delay the primary endpoint analysis;
- Information on the administration of Durvalumab as consolidation treatment (as the primary endpoint analysis is stratified on the use of Durvalumab);
- Survival status;
- Toxicity information for toxicities that are planned to be reported in combination with the primary endpoint.

The trial database will, at the end of inclusion, indicate the date for the latest update of survival, toxicity, and prescription of durvalumab for each patient. This date information will be used to ensure that all the requested information is available before analysis.

The database will be locked for primary endpoint analysis at the time when the trial management centre has received the minimum required information for all patients for the primary endpoint analysis. As mentioned, the study enrollment may be extended beyond the 350 planned patients, but the "last patient" for the primary endpoint analysis is patient number 350. The first date after a year since the randomisation of patient number 350, and where the minimum data is available in the trial management centre, will be

the date of database locking for the primary endpoint. Adding additional status updates on other patients is acceptable until the minimum required data is available in the trial centre.

No data analysis for primary or secondary endpoints related to local control status, survival or toxicity may be performed before the inclusion of the last patient in the study. Initial reporting of toxicity is allowed immediately after the inclusion of the last patient. Any data relating to the primary endpoint (disease control) should not be reported before the publication of the primary endpoint.

After reporting the primary endpoint, all patients will be followed for 10 years after randomisation or until death, as per the follow-up plan. This additional follow-up does not impact the timing of the primary endpoint, but ensures that potential toxicity differences between the two arms can be fully evaluated at a later stage.

Part 15: Timing of outcome assessments

All follow-up times are defined in the protocol (section 8.4). Specifically, disease control and recurrence are evaluated by CT every 3 months for the first two years, then half yearly until 5 years, and yearly until 10 years after randomisation. Each CT evaluation is combined with a follow-up visit in which the patient's toxicity level is recorded. Besides CT evaluation, two scheduled PET scans are performed 9 and 18 months after randomisation.

Section 4: Statistical Principles

Part 16 Confidence intervals and p-values

All applicable statistical tests will be 2-sided and performed using a 5% significance level; reported confidence intervals will be 95%.

Part 17: Description of any planned adjustment for multiplicity, and if so, including how the type 1 error is to be controlled

The primary endpoint is a single test; thus, there is no need to adjust for multiple testing. Neither will there be any adjustment for multiplicity in the analyses of toxicity differences between the two treatment arms.

Part 18: Confidence intervals (CI) to be reported

Confidence intervals for the Kaplan-Meier estimator and hazard ratios will be based on bootstrapping. The bootstrapping will use 2000 bootstraps, and the confidence intervals will be based on the 95% most centrally located bootstrap values (2.5% of the bootstraps will be below the confidence interval, and similar 2.5% will be above).

Part 19. Adherence and Protocol Deviations

Major protocol deviations are defined as the following:

- Patients treated according to another arm than the one they were allocated to by randomisation
- Patients for which the randomisation result was known before the finalisation of both treatment plans (escalated and standard arm)
- Patients treated with a plan that did not obey the high-priority OAR constraints (spinal cord, bronchi, oesophagus, lung, and heart)

Lack of protocol treatment adherence is defined as:

- Patients that did not finish the intended radiotherapy treatment

A consort diagram based on the recommendation from the CONSORT group will be provided and will include the number of deviations defined above (see part 21, section 5)

Part 20: Analysis populations

All analyses will be based on the intention-to-treat population, which consists of all patients randomised (i.e. not patients who provided initial consent but were deemed ineligible prior to randomisation).

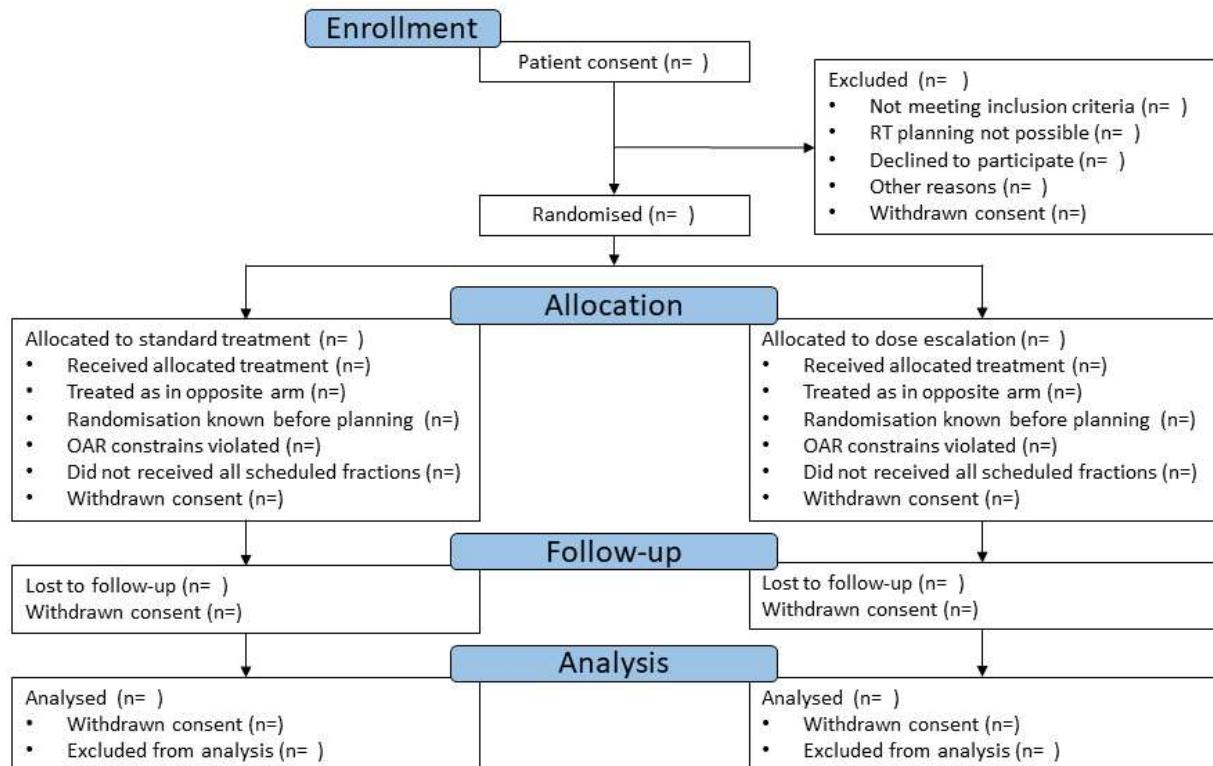
Comparisons will be based on the intention-to-treat arm, i.e. the arm selected by randomisation, independent of actually delivered treatment. Therefore, the population might include patients that accidentally have been included and randomised, but were subsequently deemed ineligible for the study or did not start treatment after randomisation. In this event, patients most likely will be censored before or during radiotherapy due to missing follow-up.

Section 5: Trial Population

Part 21:

All patients treated in the participating centres and fulfilling the inclusion criteria were candidates for the trial. The trial remained open during the COVID-19 pandemic, during which some centres had to reduce the trial capacity. Thus not all potential candidates may have been offered to participate in the trial. Systematic screening logs were not maintained at all centres; therefore, the number of patients screened for the trial will not be reported.

Figure 2: Example showing the number of patients included in the trial. The layout is based on the recommendations from the CONSORT group <http://www.consort-statement.org/consort-statement/flow-diagram>, with modifications to fit the current trial. Major protocol deviations (defined in part 19) should be included in the "allocation" section, number of consent withdrawals during follow-up should be mentioned in the "follow-up" section. For all relevant entries, the reason for a given deviation should, if possible, be documented in the caption of the figure or in the appendix.



Part 22: Eligibility

The number of patients randomised without fulfilling the inclusion criteria will be reported for each trial arm

Part 23: Recruitment

See part 21

Part 24: Withdrawal/Follow-up – level of withdrawal

The number of patients that withdrew their consent will be reported separately for the two arms, at each relevant time point.

Part 25: Baseline patient characteristics

All patient characteristics will be reported separately for the two treatment arms. The layout and variables included are shown in the example below, based on simulated data. Categorical and ordinal variables will be presented as numbers and percentages. Continuous values will be presented by their medians and interquartile ranges. Furthermore, cumulative and differential plots of all continuous variables will be plotted separately for the two treatment arms

Since this is a randomised trial, no test is performed to evaluate whether the baseline variables of the two treatment arms are sampled from the same population. All baseline differences will be assessed within the study group to identify parameters which accidentally have considerable differences. This evaluation is agreed to be performed before the primary endpoint is revealed to the study group.

The database includes the absolute value of FEV1 spirometry measures. For publication, the percentage of expected FEV1 will be reported; the reference values will be calculated using the equation in Løkke et al. [2]

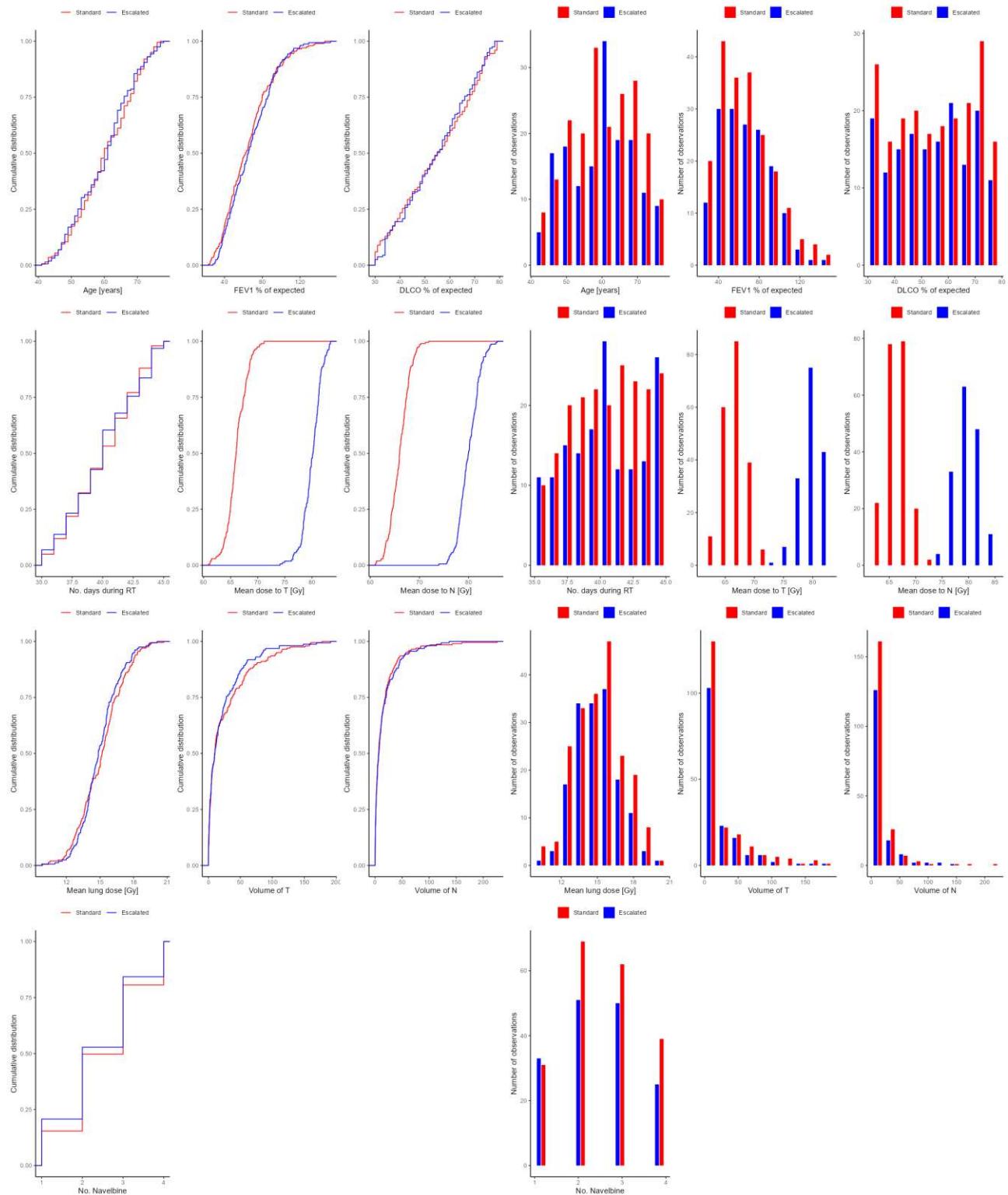
Table 1: Example of the patient characteristics table based on simulated data and summarised using the statistical software developed as part of writing the current SAP. The simulated data are not sampled from a previous cohort; thus, the data shown might be very different from an actual cohort. In particular, the simulated randomisation did not use blocks and stratification, and so appears considerably unbalanced. Some levels of the categorical variables might not be relevant to the actual study or are missing; these will be adapted automatically by the software package for the clinical data. Treatment detail variables (such as tumour volumes and doses) have not been finalised and may be updated prior to publication. Further, variables included in the table are the minimum set of variables for the publication, but others might be added.

| Patient characteristics and treatment details. | | | |
|--|---------------------|--------------------|-------------------|
| Variable | Overall | Standard | Escalated |
| n | 360 | 201 | 159 |
| Age [years] (median [IQR]) | 60.5 [53.0, 67.0] | 60.0 [54.0, 68.0] | 61.0 [52.5, 66.0] |
| Sex = Male/Female (%) | 175/185 (48.6/51.4) | 94/107 (46.8/53.2) | 81/78 (50.9/49.1) |
| Histology (%) | | | |
| Squamous carcinoma | 81 (22.5) | 47 (23.4) | 34 (21.4) |
| Adeno carcinoma | 75 (20.8) | 40 (19.9) | 35 (22.0) |
| Adenosquamous carcinoma | 111 (30.8) | 63 (31.3) | 48 (30.2) |
| NOS | 93 (25.8) | 51 (25.4) | 42 (26.4) |
| Stage (%) | | | |
| IB | 54 (15.0) | 30 (14.9) | 24 (15.1) |
| IIA | 58 (16.1) | 33 (16.4) | 25 (15.7) |

Patient characteristics and treatment details.

| Variable | Overall | Standard | Escalated |
|------------------------------------|---------------------|--------------------|--------------------|
| IIB | 66 (18.3) | 36 (17.9) | 30 (18.9) |
| IIIA | 71 (19.7) | 40 (19.9) | 31 (19.5) |
| IIIB | 63 (17.5) | 33 (16.4) | 30 (18.9) |
| IV | 48 (13.3) | 29 (14.4) | 19 (11.9) |
| Performance status = 0/1 (%) | 184/176 (51.1/48.9) | 111/90 (55.2/44.8) | 73/86 (45.9/54.1) |
| FEV1 % of expected (median [IQR]) | 64.1 [46.1, 82.9] | 62.3 [45.7, 79.9] | 65.2 [47.4, 85.7] |
| DLCO % of expected (median [IQR]) | 55.0 [42.0, 68.0] | 55.0 [41.0, 68.0] | 55.0 [42.0, 66.0] |
| Smoking at start RT = No/Yes (%) | 237/123 (65.8/34.2) | 128/73 (63.7/36.3) | 109/50 (68.6/31.4) |
| No. RT fractions (%) | | | |
| <30 | 1 (0.3) | 1 (0.5) | 0 (0.0) |
| 30-32 | 1 (0.3) | 1 (0.5) | 0 (0.0) |
| 33 | 358 (99.4) | 199 (99.0) | 159 (100.0) |
| No. days during RT (median [IQR]) | 40.0 [38.0, 42.0] | 40.0 [38.0, 42.0] | 40.0 [38.0, 42.0] |
| Mean dose to T [Gy] (median [IQR]) | 68.5 [65.8, 79.7] | 66.0 [65.0, 67.4] | 80.2 [78.8, 81.2] |
| Mean dose to N [Gy] (median [IQR]) | 68.6 [65.8, 79.5] | 66.0 [64.5, 67.4] | 80.0 [78.4, 81.7] |
| Mean lung dose [Gy] (median [IQR]) | 15.1 [13.7, 16.3] | 15.2 [13.7, 16.4] | 14.8 [13.9, 16.0] |
| Volume of T (median [IQR]) | 10.0 [2.3, 34.6] | 9.6 [2.5, 36.9] | 10.1 [2.0, 28.1] |
| Volume of N (median [IQR]) | 6.6 [1.5, 19.1] | 6.4 [1.4, 19.0] | 6.8 [1.5, 19.7] |
| No. Navelbine (median [IQR]) | 2.0 [2.0, 3.0] | 3.0 [2.0, 3.0] | 2.0 [2.0, 3.0] |
| No. platin during RT (%) | | | |
| 1 | 94 (26.1) | 46 (22.9) | 48 (30.2) |
| 2 | 172 (47.8) | 99 (49.3) | 73 (45.9) |
| 3 | 94 (26.1) | 56 (27.9) | 38 (23.9) |

Figure 3: Example plot of cumulative and differential distribution of continuous baseline variables, plotted using the statistical software developed as part of writing the current SAP.



Section 6: Analysis

Part 26: Outcome definitions

Locoregional recurrence is the event for the primary endpoint. Any recurrence in the lung region that is not classified as locoregional recurrence will be classified as a distant recurrence. Below is a list defining which recurrences are locoregional and which are not.

A locoregional recurrence is defined as any of the following events:

- 1) Recurrence in the same lung lobe as the primary tumour
- 2) Recurrence within mediastinum
- 3) Recurrence in the ipsilateral hilus (relative to the primary tumour)
- 4) Recurrence in the contralateral hilus (relative to the primary tumour)
- 5) Recurrence in the ipsilateral supraclavicular region (relative to the primary tumour)

While the following events are not defined as a locoregional recurrence:

- 1) Recurrence of clearly different histology than the primary tumour (nor if the recurrence fulfils any of the items above defining a local recurrence)
- 2) Recurrence in a lobe without primary tumour involvement
- 3) Recurrence in the opposite lung than that of the primary tumour
- 4) Recurrence in the opposite supraclavicular region
- 5) Multiple lung metastases
- 6) Malignant pleural effusion

Identification of a locoregional event is based on imaging and subsequent biopsy verification. The event date is always the imaging date; also, if the imaging is based on a prior clinical suspicion of recurrence. In the case of image/biopsy confirmation of the local recurrence, no backdating to previous imaging sessions should occur, even in the case of posthoc imaging evidence for recurrence on the previous images.

A biopsy should, as a standard, follow locoregional-control events identified by imaging. For some patients, the biopsy might not have been ethical or practically possible to perform. The events that are only identified on imaging without biopsy will be reviewed by a national response evaluation committee consisting of at least two radiologists and two oncologists to ensure a standard practice for recurrence evaluation based only on images. The same board should also be consulted if there is doubt about whether a given recurrence should be defined as locoregional or distant based on the above-provided definition.

A clinical suspicion of recurrence that, for some reason, is not followed by imaging or biopsy is not an event.

Patients are censored at the time of death; in other words, death is not an event for the primary endpoint. Patients are censored in case of disseminated metastatic disease (without local recurrence) or similar clinical events/situations that prohibit follow-up for locoregional control. In particular, note that metastasis is not an intercurrent (censoring) event for the primary endpoint. In case of no primary or no other censoring events, patients are censored at the last available image session that can be used to evaluate recurrence.

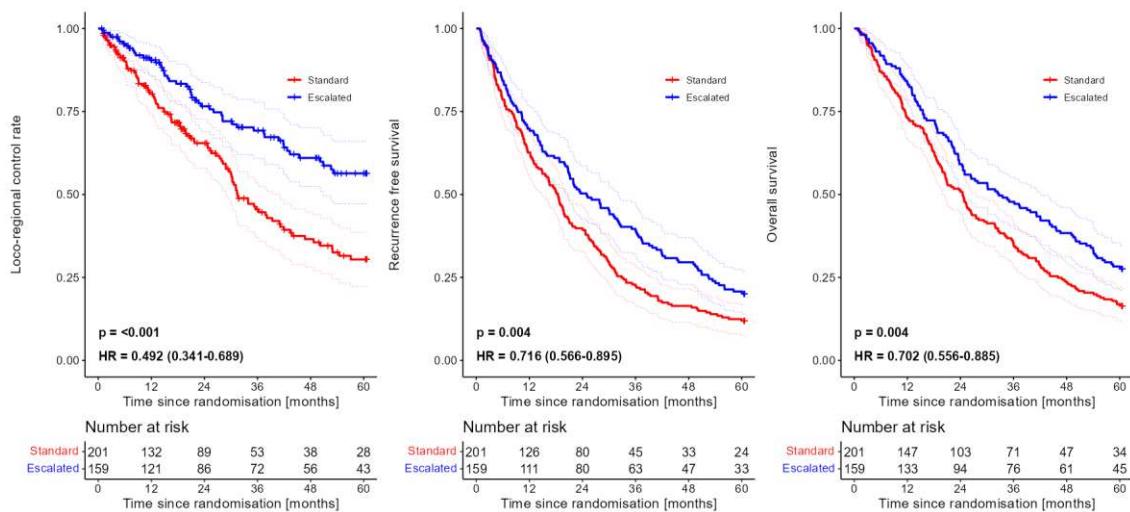
Time zero for the survival analysis is the day of randomisation.

Part 27: Analysis methods

The median potential follow-up time for the primary endpoint will be calculated using the reverse Kaplan-Meier estimator (the event variable for the reverse Kaplan-Meier is unity for all censoring events and zero for all primary endpoint events). The locoregional control rate, the primary endpoint, will be presented as Kaplan-Meier curves for the two treatment arms, including numbers at risk at regular time points. The test for the primary study hypothesis will be a stratified log-rank test between the arms, with the result presented as a test p-value. As a secondary endpoint, recurrence-free survival (with recurrence and death as an event) and overall survival will be presented similarly to the primary endpoint - locoregional control rate.

The hazard ratio between the study arms will be estimated based on a Cox proportional hazard model stratified on durvalumab. Confidence intervals for Kaplan-Meier estimates and hazard ratios will be based on bootstrapping, using 2000 bootstraps (as described above).

Figure 4: Example of Kaplan-Meier curves based on simulated data and the statistical software developed as part of writing the current SAP.



In the unlikely case of clear violation of proportional hazards (e.g. if Kaplan-Meier curves cross or split at late time points) such that a Cox model cannot fit the data, a secondary analysis utilising time-dependent hazard variation will be performed. This analysis will follow the method used by Brink et al. [3]. The evaluation of whether a time-dependent Cox model is needed will be based on an evaluation of log(-log(KM)) of the two treatment arms. If the curves deviate significantly from each other, even after a translation along the y-axis, time-dependent hazard analysis might be warranted.

A competing risk model with the cumulative risk of having the first event as death, locoregional failure, distant failure, and simultaneous locoregional and distant failure will be reported as a secondary endpoint. The reporting will be made in two versions, one in which the individual endpoint are compared per arm within separate plots and one in which the endpoints are shown together as stacked plots per arm within two subplots (at least one of the versions will be in the appendix). The aim is to show the potential difference between these events over time for each treatment arm. Similar to the above, the confidence intervals will be calculated based on 2000 bootstraps (confidence intervals can not be visualised for the stacked plots).

Example of competing risk figure based on simulated data and the statical software developed as part of writing the current SAP (in the simulated data, there were no simultaneous distant and local failures, thus effectively only three groups in the plot):

Figure 5: Example of competing risk figure based on simulated data and the statical software developed as part of writing the current SAP (in the simulated data, there were no simultaneous distant and local failures, thus effectively only three groups in the plot).

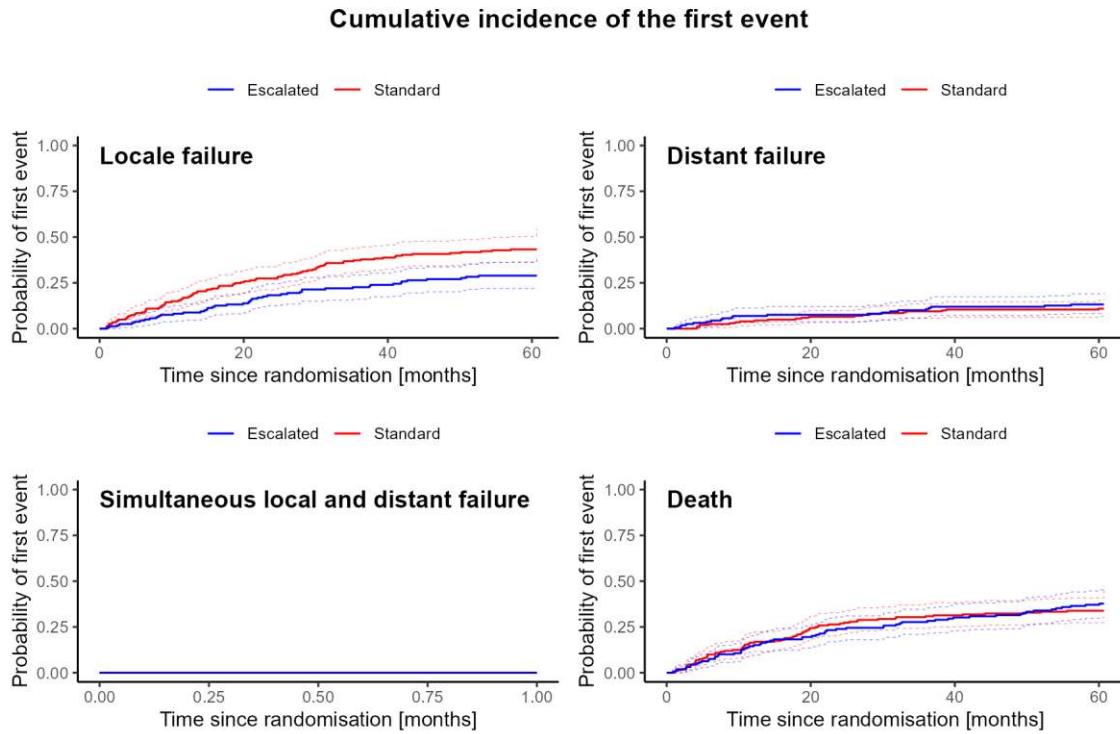
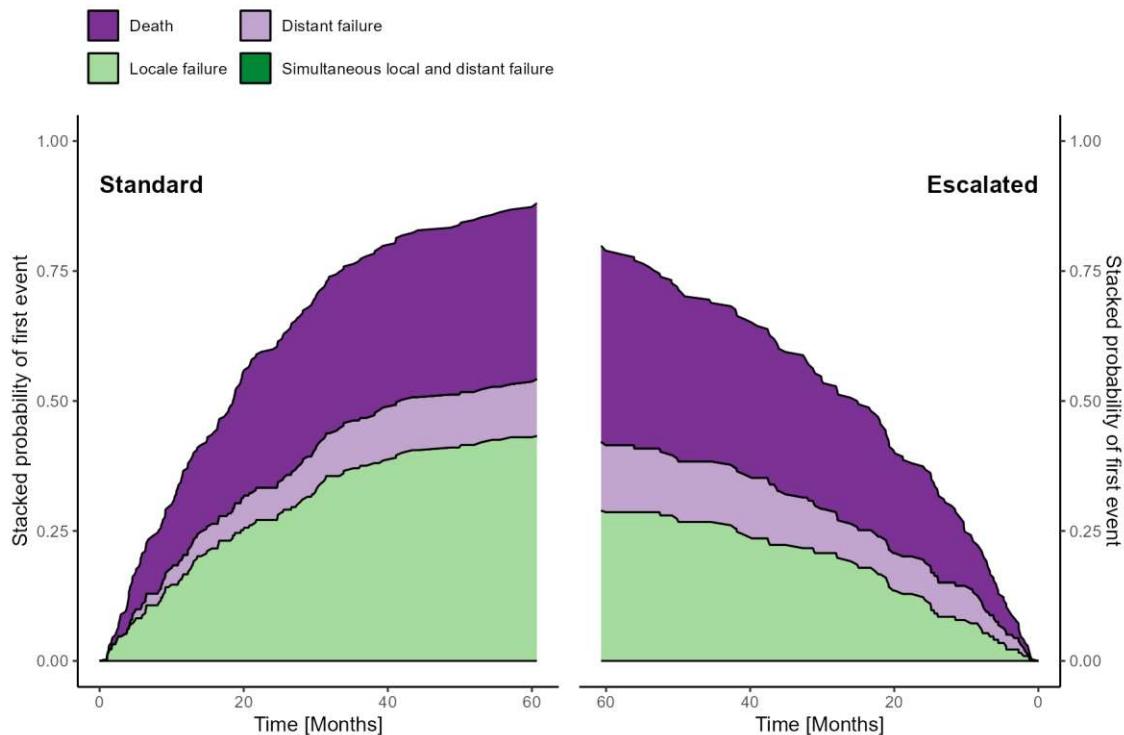


Figure 6: Example of stacked competing risk figure based on simulated data and the statical software developed as part of writing the current SAP (in the simulated data, there were no simultaneous distant and local failures, thus effectively only three groups in the plot).



As a planned subgroup analysis, all primary endpoint analyses will be repeated separately for the key stratification factor - squamous and non-squamous cell carcinoma. The squamous group consist of the histologies: squamous carcinoma and adenosquamous carcinoma. All other histologies are classified as non-squamous. There are no plans to conduct a similar subgroup analysis for receipt of consolidation Durvalumab; as this was not a randomisation stratification factor, and the number of patients receiving Durvalumab is estimated to be small (no more than 10% of the entire study cohort). There is no planned sensitivity testing related to centre size (e.g. inclusion of only centres contributing many versus few patients) nor to test the impact of any missing data.

A comprehensive report on all early toxicity is planned prior to the primary endpoint analysis. The report of the primary endpoint will include the following early toxicities: Fatigue, Cough, Dysphagia, Infection, and Pneumonitis. Furthermore, should the early toxicity report indicate large differences in other toxicities, those will also be included in the primary endpoint manuscript. The reported early toxicity for each patient is the maximum toxicity scored up to 6 months after start of radiotherapy (until and including the 6-month follow-up visit, irrespective of date) and before any recurrence (local or distant). Toxicity after any recurrence will not be included, as that toxicity is likely more related to the recurrence and the supporting treatment and only to a lesser extent to the radiotherapy. This toxicity reporting is in line with the study design, as this does not include detailed follow-up of patients after they reach the main event locoregional recurrence. In other words, only limited toxicity data have been systematically collected after locoregional recurrence. Early toxicity subdivided into toxicity reported during and after the radiotherapy course will be reported as supplementary material.

Test for difference in early toxicity between the two treatment arms will be based on Fisher's exact test. Early toxicity will be reported with grades 0 to 2 combined into one level since differences between grades

0-2 are not deemed clinically relevant (Fisher's test will be used after combining grades). If the study results in a large difference in early recurrence (before 6 months) for the two arms, the above test could be biased since the patients with a lower recurrence rate will have more toxicity assessments. Thus if Fisher's exact test shows differences between the arms, it should be further evaluated whether the result might be related to differences in recurrence rate. Such an evaluation may be based on actuarial analysis (with censoring after recurrence).

The article related to the primary endpoint will also include an evaluation of late toxicity (all toxicity scored after, but not including, the follow-up visit at 6 months). The reported late toxicities are those related to the heart, thoracal column (compression), oesophagus (perforation, stricture, fistula), airways (bleeding, stricture, fistula), and clinically relevant fibrosis. The evaluation of late toxicity will follow the same method described above for the early toxicity; however, the toxicity will be reported in two parts, one that only scores toxicity until recurrence and one that scores all observed toxicity independent of recurrence status. The toxicity before recurrence will be evaluated with the grades 0 to 2 combined since differences among these are deemed of no clinical relevance. For toxicity independent of recurrence, high-severity toxicity is of primary concern; thus, analysis of toxicity independent of recurrence will be performed with grades 0 to 3 combined.

Table 2: Example of a table showing early toxicity. The simulated data are not sampled from a previous cohort; thus, the data shown might be very different from an actual cohort. In particular, the simulated randomisation did not use blocks and stratification and appears considerably unbalanced. Some levels of the categorical variables might not be relevant to the actual study or are missing; these will be adapted automatically by the software package for the clinical data. Except for pneumonitis, which was not present in the simulated data, the shown variables are those planned to be included in the article on the primary endpoint.

| Toxicity: | | | | |
|------------------|------------|------------|------------|-------|
| Variable | Overall | Standard | Escalated | p |
| n | 360 | 201 | 159 | |
| Fatigue (%) | | | | 0.963 |
| 0-2 | 233 (64.7) | 131 (65.2) | 102 (64.2) | |
| 3 | 70 (19.4) | 38 (18.9) | 32 (20.1) | |
| 4 | 57 (15.8) | 32 (15.9) | 25 (15.7) | |
| Cough (%) | | | | 0.446 |
| 0-2 | 243 (67.5) | 138 (68.7) | 105 (66.0) | |
| 3 | 67 (18.6) | 33 (16.4) | 34 (21.4) | |
| 4 | 50 (13.9) | 30 (14.9) | 20 (12.6) | |
| Dysphagia (%) | | | | 0.223 |
| 0-2 | 263 (73.1) | 150 (74.6) | 113 (71.1) | |
| 3 | 51 (14.2) | 23 (11.4) | 28 (17.6) | |
| 4 | 46 (12.8) | 28 (13.9) | 18 (11.3) | |
| Infection (%) | | | | 0.212 |
| 0-2 | 243 (67.5) | 128 (63.7) | 115 (72.3) | |
| 3 | 56 (15.6) | 34 (16.9) | 22 (13.8) | |
| 4 | 61 (16.9) | 39 (19.4) | 22 (13.8) | |

Part 28: Missing data

For the primary endpoint, information about locoregional control, survival and administration of durvalumab is needed for all patients. These data items will be available for all patients: All patients will have been evaluated for locoregional control (and patient records will be reviewed by local investigators where any data are missing); survival data will be available for all patients (using national registries); information about the administration of durvalumab can be found within the patient record system for all patients. If a patient is lost to follow-up, the patient will be censored on the day of the last information about recurrence status. Thus no data imputation will be needed for the primary endpoint. For the toxicity, missing toxicity scores will be seen as not available, and no data imputation will be performed. Similarly, missing planned medical images will be seen as not available, and no attempt to perform data imputation will be performed.

Part 29: Additional Analyses

There are no further analyses planned related to the primary endpoint.

Part 30: Harms

As described above in the analysis section, the toxicity is reported as early (up to and including the follow-up visit 6 months after the start of radiotherapy) and Late (beyond 6 months). Within each period, the reported values are the highest grade that the patient has experienced.

Part 31: Statistical Software

A statistical package written in R has been developed as part of this analysis plan. The package is publically available and can be accessed at <https://github.com/oncology-ouh/NarlaI2>. All source code is available as individual files and as a tar file. The package can also create simulated data, which have been used to create the example figures in the current SAP. The repository also includes a signed version of the current SAP.

The module has the following functionalities:

- 1) Create simulated data that are used for the plots in the current SAP;
- 2) Load of all the raw data exported from the clinical database;
- 3) Extraction of patient characteristics, toxicity and survival data in data.frame that easily can be exported for analysis in other statistical packages (SAS, SPSS, Stata);
- 4) Analyses (including bootstrapping) and plotting of patient characteristics, toxicity data, and survival data.

The final analysis will be conducted in R and based on the package; but as stated, the package also facilitates easy export to other systems. All the analysis software was developed before access to the clinical data. The package will be used as designed prior to the release of the data; however, it is likely that small corrections e.g. small layout changes of tables and plots will be needed after the release of the clinical data. The version used to generate the plots for the current SAP is version 1.0.0 of the software, uploaded to GitHub on 13th of February 2023.

Part 32: References

- 1) Peter C. O'Brien TRF. A Multiple Testing Procedure for Clinical Trials. *Biometrics*. 1979;35:549-56.

- 2) Lokke A, Marott JL, Mortensen J, Nordestgaard BG, Dahl M, Lange P. New Danish reference values for spirometry Clin Respir J. 2013;7(2):153-67.
- 3) Brink C, Bernchou U, Bertelsen A, Hansen O, Schytte T, Hjelmborg JVB, et al. Causal relation between heart irradiation and survival of lung cancer patients after radiotherapy. Radiother Oncol. 2022;172:126-33.