1. ADMINISTRATIVE INFORMATION

Title	Neoadjuvant chemoradiotherapy plus surgery versus active surveillance for oesophageal cancer
SAP ID	SANO
Short title	Surgery as Needed for Oesophageal cancer
Trial registration	Netherlands Trial Register (NTR 6803)
SAP version and date	v1.2, 01-02-2023
Corresponding protocol version	This document has been written based on information contained in the study protocol v1.7, 25-11-2020.
SAP revisions	-
Roles and responsibility	SAP author: - B.M. Eyck, Dept. of Surgery, Erasmus MC SAP coauthors: - dr. B.J. van der Wilk, Dept. of Surgery, Erasmus MC - M.J. Valkema, Dept. of Surgery, Erasmus MC - dr. B.J. Noordman, Dept, of Surgery, Erasmus MC Supervising senior statistician: - prof. dr. E.W. Steyerberg. Dept of Biomedical Data Sciences, Leiden University Medical Center Supervising senior epidemiologist for CEA/BIA: - prof. dr. H. Lingsma. Dept of Public Health, Erasmus MC Principal investigators: - prof. dr. J.J.B. van Lanschot - prof. dr. B.P.L. Wijnhoven - dr. S.M. Lagarde
Data Centre and Monitoring	Clinical Trial Center – Erasmus MC Cancer Institute Erasmus MC Rotterdam , room EE21-16 Dr. Molewaterplein 40, 3015 GD, Rotterdam, The Netherlands P/O box 2040, 3000CA Rotterdam, The Netherlands

2. SUMMARY

Design: The Surgery As Needed for Oesophageal cancer (SANO) trial is a phase-III, non-inferiority, stepped-wedge, cluster randomised controlled trial.

Research population: patients with esophageal cancer who have a clinically complete response (cCR) at twelve weeks after neoadjuvant chemoradiotherapy (nCRT) assessed by two clinical response evaluations CREs.

Study arms: active surveillance (intervention) will be compared with standard esophagectomy (control).

Framework: primary analysis will be a modified intention-to-treat analysis, secondary analysis will be per protocol.

Primary end point: overall survival, defined as the interval between group allocation (moment of achieving cCR) and death or last follow-up.

Secondary end points: HRQoL, surgical and pathological characteristics, postoperative characteristics, distant dissemination rate, disease-free survival.

Analysis of primary end point: the primary end point will be reported in Kaplan-Meier curves and with hazard ratios from mixed-effects Cox regression analyses, taking into account the cluster effect and calendar time.

3. INTRODUCTION

3.1. Background and rationale

See general SANO protocol (§1).

3.2. Objectives

See general SANO protocol (§2).

3.3. Research population

The research population is defined as: patients with locally advanced squamous cell carcinoma or adenocarcinoma of the esophagus or esophagogastric junction who have a clinically complete response (cCR) after neoadjuvant chemoradiotherapy according to the CROSS regimen and who have agreed to undergo the experimental or the control treatment.

To assess which patients have a cCR (and thus are eligible for inclusion in the study), patients will undergo a diagnostic work-up after completion of nCRT consisting of two clinical response evaluations (CREs) (**Appendix A**). The first CRE (CRE-1) is performed at six weeks after nCRT and consists of endoscopy with ≥4 bite-on-bite biopsies of the original tumor location. Patients without histological evidence of residual disease will continue to the second CRE (CRE-2) which is performed at twelve weeks after nCRT and consists of an FDG-PET/CT scan, endoscopy with ≥4 bite-on-bite biopsies and endoscopic ultrasonography with fine-needle aspiration. Only those patients without (cyto)histological evidence of residual disease, highly suspected locoregional disease, or distant metastases are considered to have a cCR and will be eligible for study participation. For more information on the diagnostic work-up until cCR (*i.e.* CRE-1, CRE-2 and CRE-2b), see the general SANO protocol (§3.3, 5.1 and 8.3).

3.4. Study arms

The research population (patients with a cCR) will be assigned to one of the two treatment arms:

- Active surveillance
- Standard surgery

3.5. Primary hypothesis

The primary alternative hypothesis is that overall survival in the active surveillance arm is non-inferior to the standard surgery arm.

3.6. Secondary hypotheses

The secondary alternative hypotheses are:

- Health-related quality of life of patients who undergo active surveillance is clinically relevantly better than that of patient who undergo standard surgery. No clinically relevant difference exists in concerns about developing cancer again and about the need for (second) surgery, as well as the impact of these concerns on mood and daily functioning (Cancer Worry Scale) between patients who undergo active surveillance and patient who undergo standard surgery.

- The proportion of all patients in the active surveillance arm with an unresectable or incurable (T4b or R2) regrowth in the absence of distant metastases, does not exceed the standard surgery arm.

- The rate of all patients in the active surveillance arm with a microscopically irradical (R1) resection does not exceed the standard surgery arm.

- The rate of all patients in the active surveillance arm with distant dissemination does not exceed the standard surgery arm.

- The frequency of postoperative complications in the active surveillance arm is statistically significantly lower than the frequency of postoperative complications in the standard surgery group

- Disease-free survival in the active surveillance arm is non-inferior to disease-free survival in the standard surgery arm.

For definitions of the outcome parameters see §6.1.

4. STUDY METHODS

4.1. Trial design

Phase III multi-center, stepped-wedge, cluster randomized controlled trial.

4.2. Randomization

Patients who achieved a cCR are randomized at institutional level in a 4:5 ratio. This ratio was chosen based on availability of 1/5th of sample size for surgery arm from the preSANO trial. All centers will start the trial by offering the control arm of the trial (standard surgery). Clusters of participating centers will be randomized one by one to offering the intervention arm (active surveillance) until all clusters have crossed over into the active surveillance arm. At the end of the trial, all centers will be including into the active surveillance arm until the total sample size is reached. Clusters will be randomized approximately every 3 months, but the time periods can be adjusted if the number of included patients will differ substantially from the expectations. Given this cluster randomized design, patients already know which intervention they will potentially receive at the moment of providing informed consent.

Based on 12 participating centers, 6 clusters of 2 centers with comparable estimated inclusion rates are formed. Clusters will be determined by randomization, but always consist of a center with high and low expected total inclusion. To ensure maximal patient safety, the first cluster offering the intervention will consist of two centers with extensive experience in response evaluations gained in the preSANO trial. This first cluster will hence consist of the Erasmus MC and either Zuyderland Medical Center or Catharina Cancer Center, the latter of which will be determined by randomization.

The stepped wedge design allows for control of underlying time trends. To reduce the possible effect of selection bias due to patients' preferences in the period in which both control and experimental treatment are provided simultaneously, there will be an extended period during the start-up phase and the end phase of the trial in which all centers are providing standard surgery and active surveillance, respectively.

For more information on randomization, see the general study protocol (§8.2).

4.3. Sample size

With a predetermined power of 80%, significance level of 0.05, non-inferiority margin of 15%, and intra-center correlation coefficient of 0.02, a total of 224 patients (*i.e.* 112 patients in each arm) with cCR will have to be enrolled in the trial. With 20% of patients not adhering to the allocated treatment at the moment of achieving cCR, the total number of required inclusions will be 280 (= 224/0.8) patients with cCR. Taking into account that 29 patients with cCR can be included from the preSANO trial and a cCR rate of 34%, this will translate into approximately 740 patients required at baseline.

For more information on sample size calculation, see the general study protocol (§4.4).

4.4. Framework

The primary hypothesis assumes comparability of the arms in terms of overall survival. In addition, advantages in HRQoL are expected for patients who undergo active surveillance. Therefore, a non-inferiority framework will be used for the primary analysis of the trial. Non-inferiority is defined as a two-year survival rate that is no more than 15 percentage points below the expected 75% two-year survival rate among patients in the standard surgery group.

The secondary hypotheses regarding HRQoL assumes superiority in patients who undergo active surveillance. Therefore, this hypothesis will be tested using a superiority framework rather than a non-inferiority framework.

4.5. Statistical interim analyses and stopping guidance

No interim analyses will be performed. For information on stopping guidance, see the general study protocol (§8.7).

4.6. Timing of final analysis

Outcomes will be analyzed after a minimum follow-up of two years after the last enrolled patient reached a cCR. Two years is a commonly used minimum follow-up time for comparable oncological trials, and is expected to capture the most relevant data for the short-term analysis. Moreover, results may lead to implementation of active surveillance as an alternative treatment strategy already and thus may reduce unnecessary delay of providing organ sparing treatment for patients with locally advanced esophageal cancer.

4.7. Timing of outcome assessments

The study algorithm and the schedule of the study procedures are provided in the general protocol (§3.3 and §3.4, resp.).

5. STATISTICAL PRINCIPLES

5.1. Analysis population

For non-inferiority trials, the context of the trial and the expected effect of non-adherence should be carefully considered when determining the analysis population and analysis principles (*i.e.* intention-to-treat, per protocol, and/or as treated).[1]

In the SANO trial, all patients undergo the same diagnostic work-up after nCRT to assess whether a patient has a cCR (**Appendix A**). To undergo these early study procedures, patients have provided informed consent before or shortly after nCRT. Given the stepped-wedge cluster randomized design of the trial, patients already know which intervention they will receive at the moment they provide informed consent. Once patients have a cCR and qualify for the treatment phase, they may still refuse to undergo the allocated treatment at this point (*e.g.* a patient with a cCR has gone through the diagnostic work-up at a center that offers standard surgery but the patient refuses to undergo surgery and prefers active surveillance). In non-inferiority trials, such non-adherence to the allocated treatment typically dilutes the observed treatment effects in respective allocation arms, and results in a higher probability of claiming non-inferiority.[1] To reduce the probability of falsely claiming non-inferiority, the analysis population will be defined as patients who agree to undergo either active surveillance or standard surgery at the moment of achieving cCR, regardless the allocated treatment. According to this modified intention to treat analysis, patients who cross-over to the other treatment arm at the moment of achieving cCR will be analyzed as treated (e.g. a patient with cCR who is allocated to standard surgery, but refuses surgery and undergoes active surveillance, will be analyzed in the active surveillance arm).

Active surveillance includes ten CREs over a five-year period. According to the protocol patients are offered esophagectomy in case of a locoregional relapse in the absence of distant metastases. However, protocol deviations by physicians or patients may negatively affect survival in the active surveillance arm (*e.g.* insufficient biopsies taken by the endoscopist, or patient's refusal of surgery in case of a locoregional relapse). Since this is inherent to an active surveillance strategy, our analysis population will primarily be analyzed per intention-to-treat. In this way, we further reduce the probability of falsely claiming non-inferiority. This intention-to-treat analysis thus also includes patients with a cCR who intended to, but are unable to undergo the assigned treatment (*e.g.* due to a poor condition at that time). Also, patients who do not adhere to the protocol, or who have a major protocol deviation remain in the analysis. To reduce the risk of confounding bias due to exclusion of patients who refuse to undergo the allocated treatment at the moment of achieving cCR, our analysis will be adjusted for predefined confounding factors (further specified in §7).[1]

To further support our hypothesis and provide information about a best-case scenario, a secondary per-protocol analysis will be performed. For the per-protocol analysis, patients who do not adhere to the protocol or who have a major protocol deviation will be excluded from the analysis. Patients with a minor protocol deviation will remain in the analysis. For definitions of adherence and protocol deviations see §5.2. Lastly, we will perform an intention-to-treat analysis as well.

5.2. Adherence and protocol deviations

Definitions for adherence and protocol deviations are different and will be handled differently in the diagnostic phase (*i.e.* period between completion of nCRT and achieving cCR in which CRE-1 and CRE-2 are performed) and the treatment phase (*i.e.* period where the patients undergo one of the treatments of the study arm).

Patients with non-adherence or a major protocol deviation in the diagnostic phase will be excluded. The overall numbers and reasons of non-adherence and major protocol deviations will be reported in the CONSORT flow chart. In the treatment phase, patients with non-adherence, major and minor protocol deviation will remain in the trial (as per intention-to-treat). The numbers of non-adherence and protocol deviations will be reported in the CONSORT flow chart and will be provided by treatment arm.

In the diagnostic phase:

- Non-adherence will be defined as:
    - Patients who refuse to undergo all diagnostic modalities that are required according to the protocol to assess whether a patient has a cCR;
- Major protocol deviations will be defined as:
    - At CRE-1, a combination of no bite-on-bite technique used and less than 4 biopsies taken
    - At CRE-2, either no bite-on-bite technique used or less than 4 biopsies taken;
    - At CRE-2, suspicious lymph nodes at EUS of which FNA is non-representative but the CRE is not repeated (will define the repeated CRE-2 in the rest of the protocol as CRE-2b);
    - At CRE-2, uncertain outcome at the pathological assessment but CRE-2b is not performed;
    - At CRE-2b, uncertain outcome at pathological assessment (*i.e.* non-representative FNA of suspected lymph nodes or uncertain outcome of biopsy pathology).
- Minor protocol deviations will be defined as:
    - At CRE-1, either no bite-on-bite technique used or less than 4 biopsies taken;
    - At CRE-2, the FGD-PET/CT scan is made after EGD and EUS;
    - CRE-2b not performed within 2 weeks after CRE-2.

In the treatment phase:

- Non-adherence will be defined as:
    - Patients in the standard surgery arm who agreed to undergoing surgery when achieving cCR, but do not want to undergo surgery any more at a later stage
    - Patients in the active surveillance arm who agreed to undergoing CREs when achieving cCR, but do not want to undergo CREs any more at a later stage
    - Patients in the active surveillance arm who are undergoing CREs after achieving cCR and have a local relapse at one of the CREs, but who refuse to undergo surgery.
- Major protocol deviations will be defined as:
    - At any CRE in the active surveillance arm, no bite-on-bite technique used;

  - At any CRE in the active surveillance arm, suspected lymph nodes at EUS of which FNA is non-representative but EUS is not repeated prior to the next planned CRE, or the next planned CRE is not expedited;
  - At any CRE in the active surveillance arm, uncertain outcome at the pathological assessment but the EGD with biopsies is not repeated prior to the next planned CRE, or the next planned CRE is not expedited;
- Minor protocol deviations will be defined as:
  - At any CRE in the active surveillance arm, less than 4 biopsies taken;
  - At any CRE in the active surveillance arm, the FDG-PET/CT scan is made after EGD and EUS;
  - At any CRE in the active surveillance arm, a required follow-up EGD or EUS is performed more than 2 weeks after the CRE in question.

5.3. Confidence intervals and P values

All applicable statistical tests will be one-sided and will be performed using a 5% significance level. All confidence intervals presented will be 95% and two-sided.

Subgroup analyses of the primary endpoint and secondary endpoints will be considered as exploratory and will not be adjusted for multiplicity. Presentation will be limited to point estimates of effects with 95% confidence intervals.

6. TRIAL POPULATION

6.1. Screening data

The screening population is defined as: all patients with locally advanced esophageal carcinoma who have been discussed by the upper GI multidisciplinary team (MDT) after complete staging and for whom the primary advice is to undergo nCRT according to CROSS followed by esophagectomy.

For each participating center, the screening population is assessed from the moment the SANO center opened for inclusion (*i.e.* after local approval and local trial initiation meeting) until the last inclusion. Centers will register all patients that are recruited as well as those that are not recruited via the trial's website. The sum of recruited patients and non-recruited patients is the total number of patients screened. If a center did not consistently report their non-recruited patients, the MDT registries will be assessed after finishing the accrual phase, to assess the screening population in that particular center.

For the preSANO phase II (bite-on-bite phase) trial, the screenings population will also be assessed (for those centers whose patients will be included in SANO).

The following summaries will be provided overall, by cluster and by study center:

- The number of days recruiting;
- The number of patients screened;
- The number of patients recruited;
- The number of screened patients not recruited, and the reason for non-recruitment.

6.2. Eligibility

The trial's eligibility criteria are specified in the general study protocol (§4.2 and §4.3).

6.3. Recruitment

The CONSORT flowchart for randomized controlled trials including the extension for non-inferiority will be used to summarize the recruitment of patients (**Appendix B**: example). Also, a flowchart of clusters according to the CONSORT extension for stepped-wedge cluster randomized trials will be included **(Appendix C**: example).[3]

For the CONSORT flowchart for randomized controlled trials (**Appendix B**), the following definitions will be used:

Excluded after assessment for eligibility:

- Patients who did not meet the inclusion criteria (*e.g.* tumor not FDG-avid, bulk of tumor in stomach, eventually did not undergo CROSS);
- Patients who refused participation;
- Patients who have not been asked to participate (*e.g.* other conflicting studies in center, missed or unknown reason).

Excluded before CRE-1:

- Patients who cannot undergo CREs (*e.g.* toxicity of nCRT);
- Disease progression (*e.g.* disseminated disease, local unresectable progression);
- Death;
- Withdrawal.

Exclusion at CRE-1:

- Patients with no-pass of EGD;
- Patients with a major protocol deviation;
- Patients with positive bite-on-bite biopsy;

Exclusion before CRE-2:

- Patients who cannot undergo CRE-2;
- Disease progression between CRE-1 and CRE-2;
- Death;
- Withdrawal.

Exclusion at CRE-2:

- Metastases;
- No-pass of EGD;
- Positive bite-on-bite biopsy;
- No-pass EUS;

- Positive FNA;
- Highly suspected disease without possibility of pathological confirmation;
- Major protocol deviation.

Exclusion at CRE-2b:

- Metastases;
- No-pass of EGD;
- Positive bite-on-bite biopsy;
- No-pass EUS;
- Positive FNA;
- Highly suspected disease without possibility of pathological confirmation;
- Major protocol violation.

Exclusion before start of the assigned treatment:

- Patients with cCR who could have undergone the assigned treatment but refused it at the moment of achieving cCR (*e.g.* patients included in surgery arm who refuse to undergo surgery, patients included in active surveillance arm who refuse to undergo active surveillance).

Population in the treatment phase (cCR):

- Surgery arm:
    - Patients with cCR who underwent surgery (inclusion);
    - Patients with cCR who opted for surgery but were unable to undergo it (*e.g.* due to a poor physical status at that time) (inclusion);
    - Patients with cCR where the operation was discontinued peroperatively (inclusion, discontinued intervention);
    - Lost to follow-up (*e.g.* patients in the surgery arm who died of non-disease-related cause and therefore failed to undergo the 16 months PET/CT which is used for comparing the distant dissemination rates). These patients will be listed in the flowchart but will not be excluded.
- Active surveillance:
    - Patients with cCR who started active surveillance (inclusion);
    - Patients with cCR who opted for undergo active surveillance but were unable to undergo this (inclusion);
    - Patients with cCR where active surveillance was discontinued (inclusion, discontinued intervention);
    - Lost to follow up.
- Analyzed in surgery arm
    - For the intention-to-treat analysis, all patients included in the analysis population will be included in the analysis.
        - For the per protocol analysis, patients with non-adherence and major protocol deviations will be excluded.
- Analyzed in active surveillance arm

- For the intention-to-treat analysis, all patients included in the analysis population will be included in the analysis.
    - For the per-protocol analysis, patients with non-adherence and major protocol deviations will be excluded.

The CONSORT flowchart of clusters for stepped-wedge cluster randomized trials (**Appendix C**) will be structured as follows:

Eligiblity:

- Number of clusters that have been assessed for eligibility (*e.g.* any other hospitals besides the participating hospitals in The Netherlands assessed for eligibility).

Excluded before start of trial:

- Centers that did not meet inclusion criteria to participate in the trial (*e.g.* could not abide to protocol because of participation in other study with overlapping treatments or endpoints);
- Centers that declined to participate;
- Other reasons to not participate in the trial.

Population after start of intervention (cCR):

- Number of clusters and participants allocated to each treatment arm per sequence.

For each allocated sequence:

- Average size and variance of clusters;
- The numbers of clusters and participants who were assessed for eligibility (*i.e.* size of the screenings population per time period);
- The numbers of clusters and participants who were randomly assigned;
- The numbers of clusters and participants who received intended treatments;
- The numbers of clusters and participants who did not received intended treatments, including reasons
- The numbers of clusters and participants who were analyzed for the primary outcome.

6.4. Withdrawal/follow-up

Consent withdrawals and losses to follow-up will be presented in the CONSORT flow chart by treatment arm, with numbers and reasons for withdrawal and/or exclusion from analysis.

Subjects can leave the study at any time for any reason if they wish to do so without any consequences. The investigator can decide to withdraw a subject from the study for urgent medical reasons.

The level of consent withdrawal will be tabulated by:

- Withdrawal from the intervention but continue with follow-up;

- Withdrawal from follow-up but allow data collected to date to be used;
- Withdrawal from follow-up and withdrawal of consent for data collected to date to be used; or be lost to contact/follow-up.

6.5. Baseline patient characteristics

Baseline characteristics of the analysis population (*i.e.* patients with cCR after nCRT who are allocated to one of the treatment arms) will be presented in a table. Baseline characteristics of the baseline population (*i.e.* all patients who provided informed consent and met the inclusion criteria before any treatment) will be presented in a separate table. Variables will be described overall and per intervention arm. NB: this also includes pre-SANO patients.

The following variables will be described:

- Age (years);
- Sex (male, female);
- Histology (adenocarcinoma, squamous cell carcinoma, other);
- Tumor differentiation grade (good-moderate, poor, undifferentiated)
- cT (cT1, cT2, cT3, cT4a, cT4b, cTx);
- cN (cN0, cN1, cN3, cNx);
- WHO performance status (0 – 5);
- Tumor location (upper, middle, lower esophagus);

If needed, Karnofsky performance status will be transformed to WHO performance status by the following conversion table:

Karnofsky grade	**WHO grade**
90-100	0
70-80	1
50-60	2
30-40	3
10-20	4
0	5

Continuous variables will be presented as means with standard deviations (SD) or as medians with interquartile ranges (IQR) for non-normally distributed values, as assessed by visual inspection of histograms and, if unclear, Shapiro-Wilk's test. Categorical variables will be presented as proportions with percentages. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

7. ANALYSIS

7.1. Outcome definitions and analysis method

Primary and secondary outcomes:

Primary outcome is overall survival. Secondary outcomes consist of health-related quality of life, postoperative complications, distant dissemination rate, disease-free survival and cost-effectiveness analysis.

7.1.1. Overall survival

Overall survival (the primary endpoint) will be defined as the interval between the moment of achieving cCR (*i.e.* the moment the patient will start the intervention) and death or last follow-up. It will be reported primarily in Kaplan-Meier curves and with hazard ratios from Cox regression analyses, with adjusted treatment strategy effect as primary effect parameter. Descriptive statistics will include median survival with interquartile range and survival rates per year of follow-up. Differences between treatment arms will be reported as hazard ratios with 95% confidence interval and *P*-values. Survival will be presented in:

- Manuscript text;
- Figure: Kaplan-Meier curves for overall survival according to treatment allocations;
- Table: adjusted hazard ratios for overall survival including hazard ratios for each subgroup.

Required variables:

- Date of death;
- Date of last follow-up;
- Optional: Cause of death.

Analysis:

- The proportional hazards assumption will be assessed; if the assumptions appear to be inconsistent with data, log-rank tests will be primarily used to test for differences in survival times.
- Mixed-effects Cox regression models will be used:
  - Use of a mixed regression model – including an institution-level frailty effect – is required to capture the potential between-institutional variation in survival.[5] The cluster effect by center and effect of calendar time will also be taken into account.
  - To adjust for potential selection bias, the treatment strategy effect will be estimated with adjustment for prognostic factors for survival, *i.e.* age, sex, histologic subtype of tumor, histological grading, clinical T stage, clinical N stage, and WHO performance score.
  - We will also use the mixed-effects Cox regression model to explore potential differences in treatment effect between subgroups of patients and we will test for interaction with subgroups. Subgroups are predefined according to age, sex, histologic subtype of tumor, histological grading, clinical T category, clinical N category, and WHO performance score.

7.1.2. HRQoL

HRQoL is captured with the EORTC QLQ-C30 and QLQ-OG25 questionnaires, the Cancer Worry Scale and the EuroQol EQ5D-5L questionnaire.[6-8] HRQoL is measured at baseline, and 3 (moment of achieving cCR), 6,9,12, 16, 20, 24, 30, 36, 48 and 60 months after baseline.

For the primary analysis of the SANO trial, global health related quality of life at two years will be reported. In a separate analysis, more extensive reporting of HRQoL will be done based on clinical relevance and hypothesized association with active surveillance or surgery.

The following points in time have been selected for consensus discussion:

- Baseline;
- 3 months (moment of achieving cCR);
- 6 months
- 9 months
- 1 year;
- 2 year;
- 5 year.

Analysis:

- QLQ-C30 and QLQ-OG25 items will be analyzed according to the EORTC scoring manual;[7, 9]
- Cancer Worry Scale items will be analyzed according to the Cancer Worry Scale scoring manual;[6]
- EQ5D-5L items will be analyzed according to the EuroQol scoring manual;[10]
- Repeated measurement analysis will be used to evaluate within and between group differences, as we have described before.[11]Generalized estimated equations (GEE) will be used to compare trends over time between both treatments arms and to assesses follow-up measurements longitudinally. GEE with an independence working correlation allows for missed assessments, since patients with different numbers of completed measurements can be included (*e.g.* patients who missed one or more questionnaire, or patients who dropped out during the trial (in particular due to death)) (12 + nieuwe ref?). Cohen's d (CD) effect sizes will be calculated to evaluate the clinical relevance of the effects and to allow for standardized comparison between different end points. Respective CD values of 0.2, 0.5, and 0.8 reflect small, medium and large effects.[13] A CD >0.5 is considered a clinically relevant effect.[14]

7.1.3. Surgical and pathological characteristics

Pathological staging will be defined according to the 7th ed. of the Union for International Cancer Control TNM staging manual.[15] Microscopically radical resection (R0) will be defined as a tumour-free at the proximal, distal and circumferential resection margin (margin >1mm not required). Tumour regression grade (TRG) will be defined according to the adjusted Mandard's classification by Chirieac (TRG 1 to 4).[16]

Continuous variables will be reported as means with standard deviations (SD) or as medians with interquartile ranges (IQR) for non-normally distributed values, as assessed by visual inspection of

histograms and, if unclear, Shapiro-Wilk's test. Categorical variables will be reported as proportions with percentages.

Surgical and pathological characteristics will be presented in:

- Table**:** Surgical and pathological characteristics according to treatment allocations.

Required variables:

Surgical:

- Number of patients proceeding to surgery;
- Resection (yes, no);
- Number of patients with cT4b;
- Number of patients with intraoperative metastases;
- Macroscopic radicality (yes, no);
- Surgical approach (transhiatal, transthoracic);
- Surgical technique (totally minimally invasive, hybrid minimally invasive, open);
- Conversion form minimally invasive to open (yes, no);
- Location of anastomosis (cervical, thoracic);
- Duration (minutes);
- Blood loss (ml);
- Number of blood transfusions during and <48 hours after surgery.

Pathological:

- Microscopic radicality (yes, no);
- Histological grading (good-moderate, poor, undifferentiated);
- TRG;
- Number of lymph nodes examined;
- Number of positive lymph nodes;
- ypTNM.

Analysis:

- The clinical unresectablity (cT4b) and R0 resection rates between both study arms will be compared by using multivariable binary logistic regression, adjusted for age, sex, histologic subtype of tumor, histological grading, clinical T stage, clinical N stage, and WHO performance score.
- Other outcomes will not be statistically compared.

7.1.4. postoperative characteristics

Postoperative complications will be defined according to the Esophagectomy Complications Consensus Group (ECCG) definitions.[17] Postoperative complications will be reported as proportions with percentages and will be presented in:

- Table: postoperative characteristics according to treatment allocations.

- Supplementary table: all postoperative characteristics according to treatment allocations.

Required variables (bold for primary table only, others in supplementary table):

Totals:

- **Total number of patients with complication;**
- **Number of patients with major complications (Clavien-Dindo ≥ 3b).**

Gastrointestinal:

- **Esophagogastric leak from anastomosis, staple line or localized conduit necrosis (Type I-III);**
- **Conduit necrosis/failure requiring surgery (Type I-III);**
- Ileus requiring intervention;
- Liver dysfunction.

Pulmonary:

- **Total respiratory complications;**
- **Pneumonia;**
- **Respiratory failure requiring reintubation;**
- Pleural effusion requiring additional drainage procedure;
- Pneumothorax requiring intervention;
- Atelectasis mucous plugging requiring bronchoscopy;
- **Acute respiratory distress syndrome.**

Cardiac:

- **Total cardiac complications;**
- **Dysrhythmia (atrial and/or ventricular) requiring intervention;**
- Myocardial infarction;
- Cardiac arrest requiring CPR;
- Congestive heart failure requiring intervention;
- Pericarditis requiring intervention.

Neurologic / psychiatric:

- **Recurrent laryngeal nerve injury (Type I-III, A-B);**
- Acute delirium.

Urologic:

- Urinary tract infection requiring antibiotics;
- Acute renal insufficiency;
- Acute renal failure requiring dialysis.

Thromboembolic:

- **DVT (deep venous thrombosis) or PE (pulmonary embolus);**

- Stroke (CVA).

Infection:

- **Wound Infection requiring opening wound or antibiotics;**
- Generalized sepsis – CDC definition;
- Other infections requiring antibiotics.

Wound/diaphragm:

- Thoracic wound dehiscence;
- Acute abdominal wall dehiscence / hernia.

Other complications

- **Chyle leak (Type I-III, A-B);**
- **ICU stay;**
- **Length of hospital stay**
- **90-day and/or in-hospital mortality;**
- **Reoperation (all reasons);**
- **Multisystem organ failure.**

7.1.5. Distant dissemination rate

Distant dissemination rates will be defined as:

- The proportion of metastases detected at FDG-PET/CT at 16 months and 30 months after nCRT (i.e. 13 and 27 after achieving cCR);
- In case the presence of metastases is FDG-PET/CT unclear, histological evidence and/or a second diagnostic modality will be used for confirmation. If metastatic disease remains unclear, patients undergo additional imaging studies over time. If the lesion grows or if more lesions will come to light, the patient will be considered to have metastatic disease;
- The definition of a distant metastasis will be according to the 8th edition of the AJCC TNM staging manual, and will include: 1) metastases in organs other than the esophagus, 2) lymph node metastases below the level of the celiac trunc, 3) lymph node metastases in cervical regions I-V (which includes the supraclavicular fossa (level V)). Regional nodes are defined as: all lymph nodes in continuity with the esophagus, extending from peri-esophageal cervical nodes (incl. cervical peri-esophageal level VI and VII) to the celiac nodes.[18]

Distant dissemination rates will be reported with hazard ratios and 95% confidence intervals in the manuscript text.

Required variables:

- Number of metastases at PET/CT at 16 months and 30 months after nCRT;
- Location of metastases.

Analysis:

- The distant dissemination rates in both study arms will be compared with cause-specific hazard models (Cox regression), adjusted for age, sex, histologic subtype of tumor, clinical N stage, and WHO performance score. A cause-specific hazard ratio is the appropriate statistic for answering etiologic questions.[19]

7.1.6. Disease-free survival

In accordance with the Assessment of Time-to-event End-points in CANcer trials (DATECAN) guidelines, disease-free survival will be defined as the interval between the day of achieving cCR and the date of:

- Primary (or per-operative) unresectable disease;
- Locoregional relapse (after completion of therapy, including surgery in active surveillance group);
- Distant metastasis (during or after completion of treatment);
- A second esophageal tumor;
- Death;

whichever occurs first.

Hence, in both arms:

- Clinically resectable relapse before resection ≠ event
- Clinically resectable relapse after resection = event
- Clinically resectable relapse but patient refuses to undergo surgery = event
- Unresectable relapse or M1 = event

Disease-free survival will be reported as median survival with interquartile ranges and as survival rates per year of follow-up. Differences between treatment arms will be reported as hazard ratios with 95% confidence interval and p-values. Survival will be presented in:

- Manuscript text;
- Figure: Kaplan-Meier curves for disease-free survival according to treatment allocations;
- Table: adjusted hazard ratios for disease-free survival including hazard ratios for each subgroup.

Required variables:

- Date of death;
- Date of relapse;
- Location of relapse;
- Date of last follow-up.

Analysis:

- The proportional hazards assumption will be assessed;
- Mixed-effects Cox regression model, adjusted for age, sex, histologic subtype of tumor, histological grading, clinical T stage, clinical N stage, and WHO performance score, taking into account the cluster effect by cente;
- Subgroups are predefined according to age, sex, histologic subtype of tumor, clinical N stage, and WHO performance score.

7.1.7. Other results

The diagnostic findings in the active surveillance arm at each CRE will be described and will be presented in stacked bar charts with 95% confidence interval (Appendix D: example).

Required variables at each CRE:

- Distant metastases;
- No-pass of EGD/EUS;
- Positive bite-on-bite biopsy;
- Positive FNA;
- Highly suspected disease without pathological confirmation;

7.2. Missing data

The number of participants with missing data for baseline variables will be reported in the table with baseline characteristics. Missing data for variables used in analyses will also be reported, either in tables (*e.g.* surgical, pathological and postoperative characteristics) or in the manuscript text (*e.g.* for distant dissemination rate).

For clinical unresectability (cT4b), the R0 resection rates and the distant dissemination rates at 16 and 30 months after completion of nCRT, multiple imputation at the individual person level will be used for missing data. Only participants who will be included in the analysis will be included in the imputation model. Imputation will be performed separately within each treatment arm.

The imputation model will include the outcome parameter, as well as site of recruitment, age, gender, baseline cT, baseline cN and baseline tumor length. Missing data in any of the covariates to be adjusted for in the analysis will be accounted for using the same multiple imputation model as above. We will perform sensitivity analyses to assess the robustness of the results to other methods to account for missing data.

7.3. Additional analyses

A sensitivity analysis will be performed for overall and disease-free survival, excluding patients from the analysis who have received additional adjuvant treatment after surgery. Number of patients who have received adjuvant treatment will be reported. The sensitivity analysis will be analyzed according to the survival outcome analyses as described at §6.1.1 and §6.1.6.

7.4. Harms

Potential harms and safety reporting have been described in the general study protocol (§9).

7.5. Statistical software

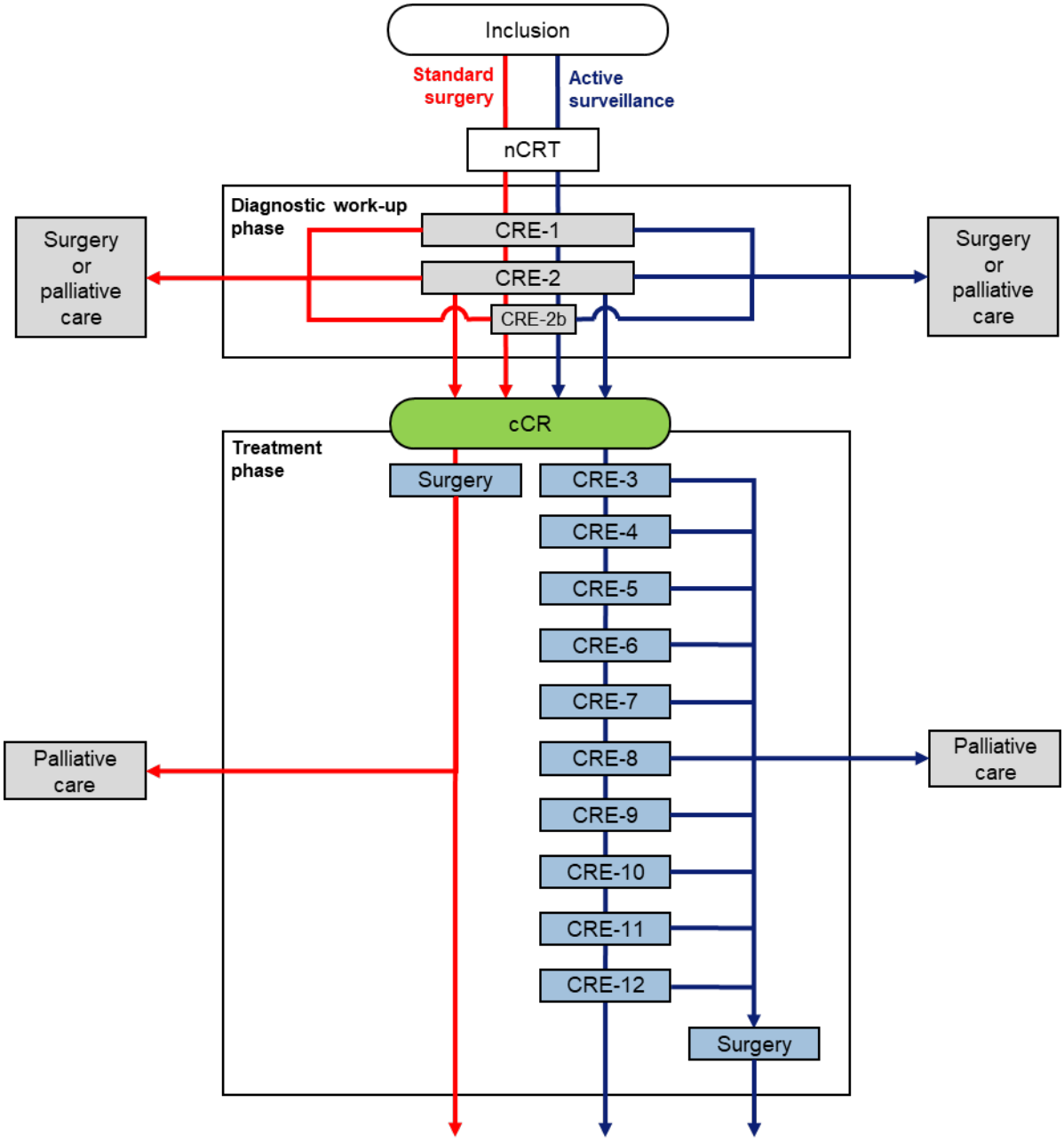
All analyses will be performed in R (latest available version at time of analysis, R Foundation for Statistical Computing, Vienna, Austria) with reporting in RMarkDown.

7. References

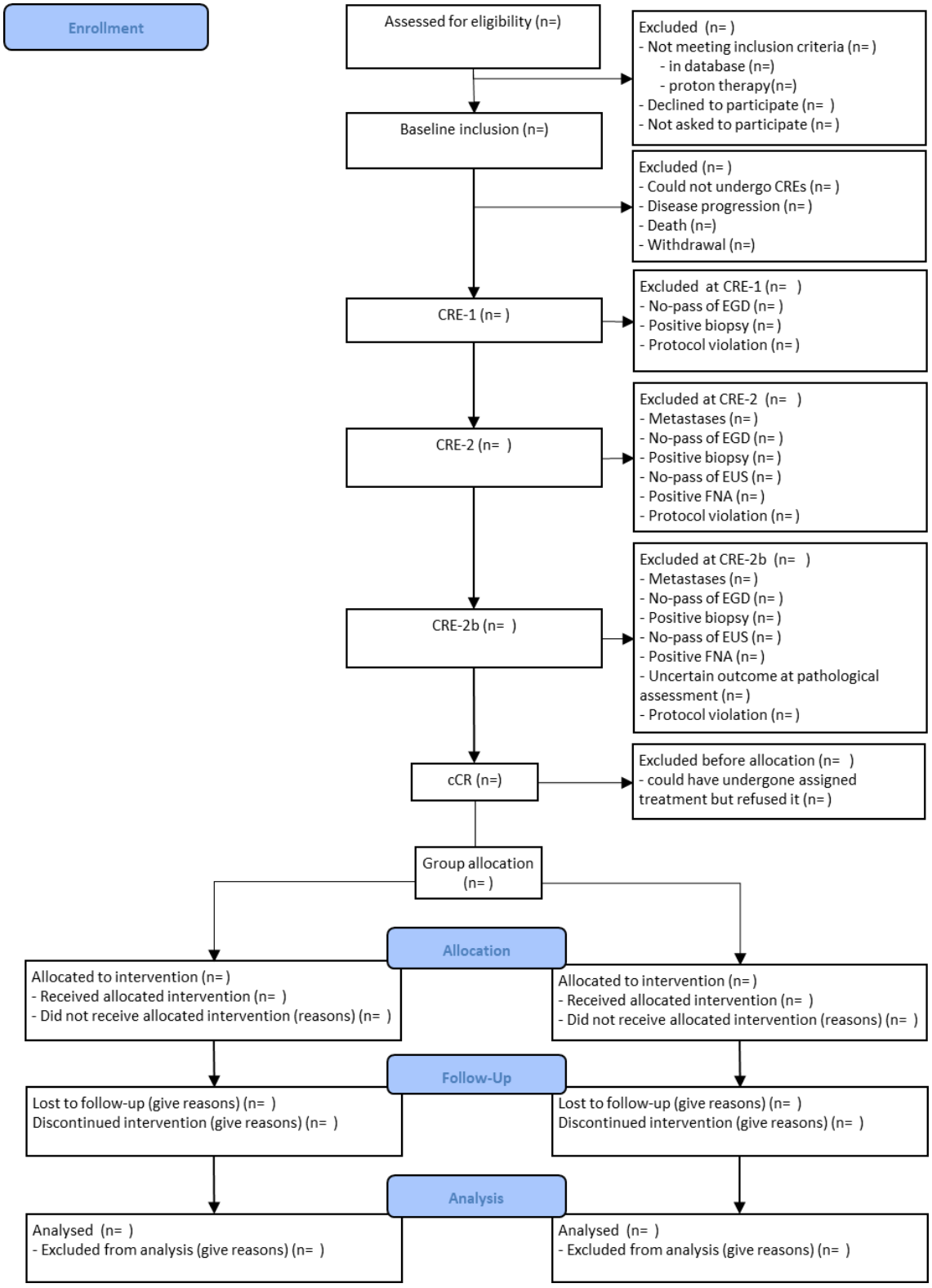
1. Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations. Bmj. 2020 Jul 1;370:m2215.
2. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. Bmj. 2015 Feb 6;350:h391.
3. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. Bmj. 2018 Nov 9;363:k1614.
4. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383.
5. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. NY: Springer New York; 2000.
6. Custers JA, van den Berg SW, van Laarhoven HW, Bleiker EM, Gielissen MF, Prins JB. The Cancer Worry Scale: detecting fear of recurrence in breast cancer survivors. Cancer Nurs. 2014 Jan-Feb;37(1):E44-50.
7. Lagergren P, Fayers P, Conroy T, Stein HJ, Sezer O, Hardwick R, et al. Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-OG25, to assess health-related quality of life in patients with cancer of the oesophagus, the oesophago-gastric junction and the stomach. Eur J Cancer. 2007 Sep;43(14):2066-2073.
8. McKernan M, McMillan DC, Anderson JR, Angerson WJ, Stuart RC. The relationship between quality of life (EORTC QLQ-C30) and survival in patients with gastro-oesophageal cancer. Br J Cancer. 2008 Mar 11;98(5):888-893.
9. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol. 2011 Jan 1;29(1):89-96.
10. EuroQol Research Foundation. EQ-5D-5L User Guide, 2019. Available from: https://euroqol.org/publications/user-guides.
11. Noordman BJ, Verdam MGE, Lagarde SM, Hulshof M, van Hagen P, van Berge Henegouwen MI, et al. Effect of neoadjuvant chemoradiotherapy on health-related quality of life in esophageal or junctional cancer: results from the randomized CROSS trial. J Clin Oncol. 2018 Jan 20;36(3):268-275.
12. Twisk J, de Vente W. Attrition in longitudinal studies. How to deal with missing data. J Clin Epidemiol. 2002 Apr;55(4):329-337.
13. Cohen J. Statistical power analysis for the behavioral sciences: Routledge; 2013.
14. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003 May;41(5):582-592.
15. Rice TW, Blackstone EH, Rusch VW. 7th edition of the AJCC Cancer Staging Manual: esophagus and esophagogastric junction. Ann Surg Oncol. 2010 Jul;17(7):1721-1724.
16. Chirieac LR, Swisher SG, Ajani JA, Komaki RR, Correa AM, Morris JS, et al. Posttherapy pathologic stage predicts survival in patients with esophageal carcinoma receiving preoperative chemoradiation. Cancer. 2005 Apr 1;103(7):1347-1355.

17. Low DE, Alderson D, Cecconello I, Chang AC, Darling GE, D'Journo XB, et al. International consensus on standardization of data collection for complications associated with esophagectomy: Esophagectomy Complications Consensus Group (ECCG). Ann Surg. 2015 Aug;262(2):286-294.
18. Rice TW, Kelsen DP, Blackstone EH. Esophagus and esophagogastric junction. In: Amin MB, Edge SB, Greene FL, David RB, D.R., Brookland RK, Washington M, K., et al., editors. AJCC Cancer Staging Manual, 8th ed. New York: Springer; 2017. p. 185-202.
19. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? Stat Med. 2012 May 20;31(11-12):1089-1097.
20. Hakkaart-van Roijen L. Handleiding voor kostenonderzoek, methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg: CvZ; 2010.
21. Gold M, Siegel J, Russell L, Weinstein M. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
22. van Vliet EP, Steyerberg EW, Eijkemans MJ, Kuipers EJ, Siersema PD. Detection of distant metastases in patients with oesophageal or gastric cardia cancer: a diagnostic decision analysis. Br J Cancer. 2007 Oct 8;97(7):868-876.
23. Glick HA, Briggs AH, Polsky D. Quantifying stochastic uncertainty and presenting results of cost-effectiveness analyses. Expert Rev Pharmacoecon Outcomes Res. 2001 Oct;1(1):25-36.
24. Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M, et al. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices--budget impact analysis. Value Health. 2007 Sep-Oct;10(5):336-347.

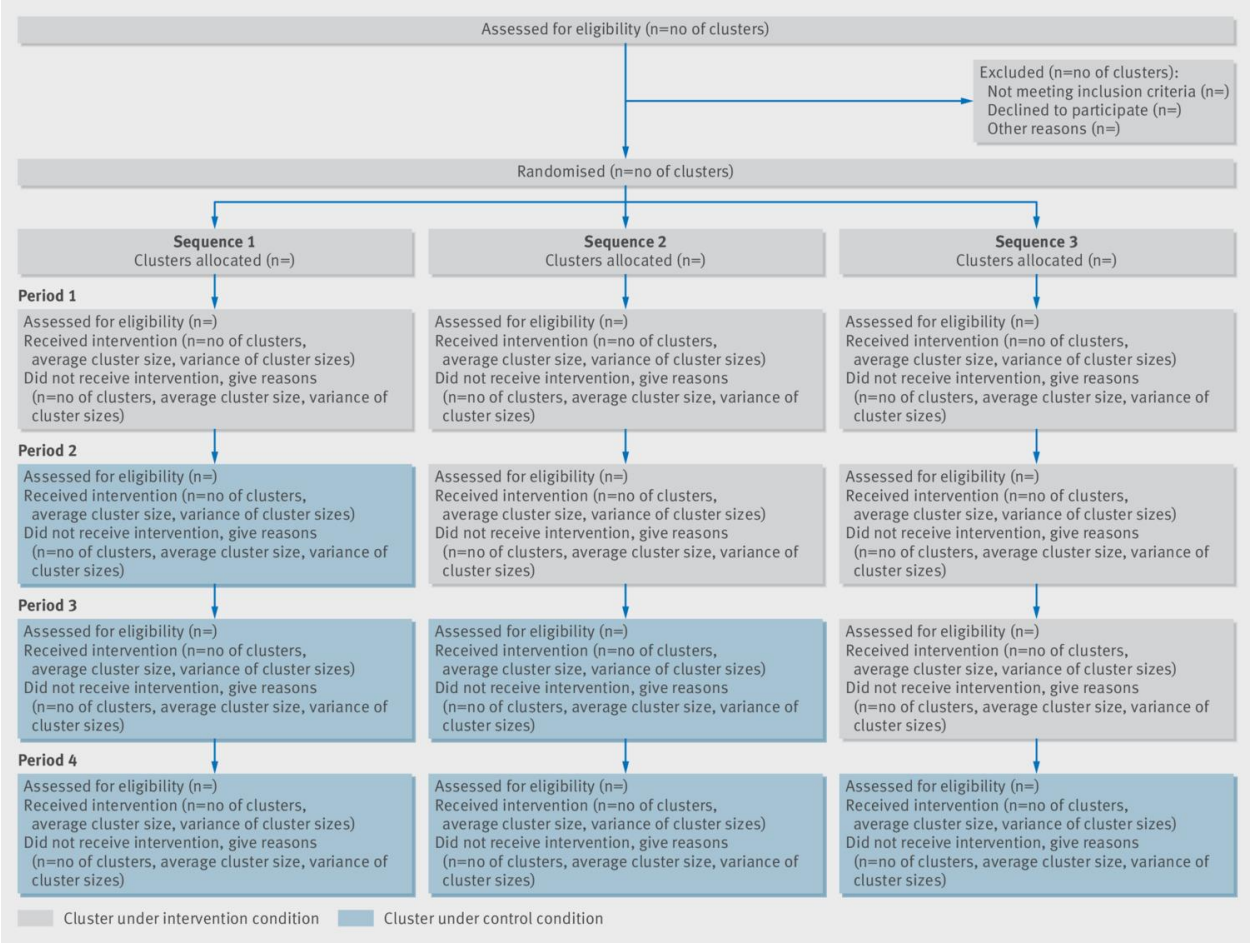
Appendix A: Study overview of the SANO trial.



Appendix B: CONSORT flow chart



Appendix C: CONSORT flow chart for stepped-wedge randomized trials.



APPENDIX D