Statistical Analysis Plan (SAP)

of

The TRACTION Trial

SAP version number:	1.0 (October 16, 2024)
Person writing the SAP:	[redacted]
Principal investigator:	
Protocol title:	Team-based Interventional Triage in Acute Coronary Syndrome Based on Non-Invasive Computed Tomography Coronary Angiography Versus Invasive Coronary Angiography – a Randomized, Non-Inferiority Trial: The TRACTION Trial
Protocol version:	5.0
ClinicalTrials.gov ID:	NCT06101862

Note: inspiration to write this SAP came from the SAP template provided by TransCelerate [39] as well as recommendations from Gamble et al. [16], Stevens et al. [37] and Evans and Ting [10].

Contents

1 Statistical analysis plan approval signature page

Person writing the SAP:	Name: [redacted] Position/Title: [redacted] [redacted] October 16, 2024 Signature & Date
Approved by:	Name: [redacted] Position/Title: [redacted] [redacted] October 16, 2024 Signature & Date

2 List of Abbreviations

- ASCVD: atherosclerotic cardiovascular disease meaning stroke, chronic coronary syndrome, acute coronary syndrome, peripheral artery disease
- BARC: Bleeding Academic Research Consortium
- CABG: Coronary Artery Bypass Grafting
- CCTA: Coronary Computed Tomography Angiography
- CI: Confidence Interval
- ECG: electrocardiogram
- eGFR: estimated Glomerular Filtration Rate
- HF: Heart Failure
- ICA: Invasive Coronary Angiography
- ITT: Intention-To-Treat
- MACE: Major Adverse Cardiac Event
- MI: Myocardial Infarction
- MRI: Magnetic Resonance Imaging
- NSTEACS: Non-ST-segment Elevation Acute Coronary Syndrome, where acute coronary syndrome is defined as non-ST-element myocardial infarction or unstable angina pectoris
- PCI: Percutaneous Coronary Intervention
- PP: Per Protocol
- Rb-PET: Positron Emission Tomography using Rubidium
- SAQ-7: Seattle Angina Questionnaire with seven items
- SPECT: Single-Photon Emission computed Tomography

3 Introduction

The TRACTION trial is a multicenter, investigator-initiated, parallel group, 1:1 randomized, assessor-blinded, non-inferiority trial. Patients admitted with NSTEACS and an indication for ICA are randomized to Coronary Computed Tomography Angiography (CCTA) and team-based interventional triage (intervention group) versus standard-of-care with conventional Invasive Coronary Angiography (ICA, control group). Team-based interventional triage is performed with conference between an interventional cardiologist and a CCTA cardiologist to determine the invasive treatment strategy. Block randomization was used without stratification. The primary outcome is major adverse cardiac event within one year, consisting of all-cause mortality, non-fatal myocardial infarction, hospitalization with refractory angina, or hospitalization with heart failure within 1 year.

The trial is conducted at the three invasive centers in Eastern Denmark (Gentofte Hospital, Roskilde Hospital, and Rigshospitalet) with participation from selected referring non-invasive hospitals. The inclusion of all patients is expected to last three years. Adding one year of follow-up to collect the necessary data, the database lock that will precede the statistical analyses detailed below is expected approximately four years after study start.

4 Objectives, Endpoints, and Estimands

4.1 Primary objective, endpoint and estimand

The primary objective is to show that CCTA is not inferior to ICA for the initial diagnostic examination in patients admitted with NSTEACS and a clinical indication for inpatient ICA. By non-inferior, we mean that the risks of major adverse cardiac events (MACE) within one year after admission is not larger in the CCTA arm than in the ICA arm, by more than the pre-specified non-inferiority margin of 5%. The rationale for choosing this margin is provided in Section 4.2. If non-inferiority is established, a subsequent objective will be to show that CCTA is superior to ICA. Specifically, the two interventions being investigated are:

> **CCTA:** Coronary computed tomography angiography (CCTA) and team-based interventional triage (intervention group). Participants will be examined as early as possible during admission. CCTA will be discussed at a coronary CT-team conference to determine the treatment strategy and guide the interventional procedure.
>
> **ICA:** standard-of-care with conventional ICA (control group). Participants will be examined with ICA during admission.

For both interventions, the treatment strategy after diagnostic testing (ICA or CCTA) are the same and corresponds to current standard of care in Denmark. Regardless of the allocation to ICA or CCTA, the treatment decision after ICA or CCTA will be either medical treatment, percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG). For patients where CABG is the preferred strategy, final decision of treatment strategy will be taken by the Heart Team multidisciplinary conference involving cardiac surgeons, as is routine. All participants will be referred to cardiac rehabilitation after discharge and receive guideline-recommended medical treatment, which may include antithrombotic medication (see Protocol Section 2.6.5), antianginal medication, statins, antihypertensives, among others.

The primary clinical question of interest is: *"Is the risk of MACE within one year from admission not more than 5% higher when using CCTA instead of ICA for initial diagnostic testing, among patients hospitalized with NSTEACS and an indication for ICA, regardless of any post-diagnostic management needed to ensure good clinical care?"*

To answer this question, we will use two co-primary estimands. The first corresponds to an **"intention-to-treat estimand"** and is described by the following attributes:

- Population: hospitalized patients with NSTEACS, an indication for inpatient ICA, and elevated troponin defined according to local hospital standards or ischemic ECG changes (see Protocol section 2.2.1), who do not meet the exclusion criteria (see protocol, Section 2.2.2).

- Endpoint: major adverse cardiac events (MACE) defined by the occurrence of the first of the following events:
  - death (all-cause mortality)
  - non-fatal myocardial infarction (MI)
  - hospitalization with refractory angina
  - hospitalization with heart failure (HF)

  Unblinded study nurses will extract the information needed to assess clinical outcomes by reviewing electronic hospital files and possibly also via telephone interviews. Data on all potential clinical outcome events will be extracted, anonymized, and blinded before review by an independent, blinded clinical endpoint adjudication board (see Protocol, Section 2.8).
- Treatment: The investigational interventions (referral for diagnostic testing, either CCTA or ICA) regardless of any subsequent treatment decision needed to ensure good clinical care and regardless of adherence ("treatment policy strategy", see [21]).
- The intercurrent events "any subsequent treatment decision" and "the patient does not adhere" to the randomized initial diagnostic testing are addressed by the treatment condition of interest attribute ("treatment policy strategy", see [21]). Note that a "subsequent treatment decision" after initial diagnostic testing via coronary computed tomography angiography (CCTA) can be to perform invasive coronary angiography (ICA). There are no other relevant intercurrent events anticipated.
- Population-level summary: Difference in 1-year risk between arms.

The second co-primary estimand corresponds to a **"per protocol estimand"** and only differs from the previous estimand for these attributes:

- Treatment: The investigational interventions (referral for diagnostic testing, either CCTA or ICA) regardless of any subsequent treatment decision needed to ensure good clinical care ("treatment policy strategy", see [21]). Here again, a "subsequent treatment decision" after initial diagnostic testing via coronary computed tomography angiography (CCTA) can be to perform invasive coronary angiography (ICA). Unlike with the previous estimand, here the patients need to adhere to the initial diagnostic testing referral ("hypothetical strategy", see [21]).
- The intercurrent event "any subsequent treatment decision" is addressed by the treatment condition of interest attribute ("treatment policy strategy", see [21]). Another possible intercurrent is "the patient does not adhere" to the randomized initial diagnostic testing, that is, either "the patient received ICA instead of CCTA as randomized" or "the patient received CCTA instead of ICA as randomized" or "the patient received neither ICA nor CCTA". A scenario is envisaged in which this does not occur ("hypothetical strategy", see [21]).

Rationale for these estimands. As we expect a very small proportion of non adherence ($\leq 3\%$), the two estimands are considered very similar. However, in general, it is often considered *"advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices*

of the set of subjects analysed" [20, Sec 5.2.3]. An important advantage of the "Intention-To-Treat" estimand over the "Per Protocol" estimand is that it can be estimated without bias under weaker assumptions, because of randomization [12]. However, its interpretation depends on the level of adherence and poor adherence will generally make the two arms more similar, hence facilitating demonstration of non-inferiority. The "per-protocol" estimand does not suffer from this, but inference for this estimand inevitably relies a few additional assumptions (e.g. unmeasured confounding), which might be questionable [18, 3]. Consequently some guidelines recommend that both should have equal importance [6].

4.2 Non inferiority margin choice

The non-inferiority margin was determined following discussions within the Steering Committee. An increase in absolute risk of up to 5% was deemed clinically acceptable, considering the potential substantial benefits associated with a CCTA-based strategy (reduced days spent in hospital, inter-hospital transports, and the risks of ICA-related adverse events). Historical data from previous trials, such as the RITA-3 trial [15], which established routine ICA as standard-of-care in patients with NSTEACS, showed a reduction in absolute risk of at least 5% for a composite outcome comparable to the MACE definition used in the current trial. Therefore, meeting a non-inferiority margin of 5% would indicate that the new CCTA-based strategy retains efficacy compared to the old strategy of selective ICA, against which the current standard of routine ICA was tested.

4.3 Secondary endpoints

There are ten secondary outcomes. The first four are the individual components of the primary outcome:

S1: death (all-cause mortality) within one year

S2: non-fatal MI within one year

S3: hospitalization with refractory angina within one year

S4: hospitalization with HF within one year

See above for how the data on these outcomes S1-S4 are collected. The five others are:

S5: cardiovascular death within one year. See Protocol, Section 2.3.6, for the exact definition.

S6: unplanned coronary revascularization within one year. See Protocol, Section 2.3.6, for the exact definition.

S7: Quality-of-life measured by EQ-5D-5L [9] at one year (accepted range 11.5 to 14 months; more specifically, 350 to 430 days) and summarized as a single number, termed as either the "index value" [9, Sec 4.1] or the "EQ-5D value" [8, Sec 4.1]. To compute this number, we will use the "value set" recently developed for Denmark by Jensen et al. [22]. This outcome has *"a scale anchored at 1 (meaning full health) and 0 (meaning a state as bad as being dead). The scale allows negative values to be assigned to health states that are considered worse than dead."* [8, p. 62].

S8: Symptom burden as measured by Seattle Angina Questionnaire with seven items (SAQ-7) at one year (accepted range 11.5 to 14 months; more specifically, 350 to 430 days), summarized by a single summary score (range is 0-100). The SAQ-7 captures angina frequency, physical limitation, and quality of life domains. The summary score is calculated as the mean of the three scores corresponding to the three domains [4, 38].

Answers to the questionnaires (for outcomes S7 and S8), are obtained electronically directly from the patient or via telephone interviews, if no answer is provided electronically.

S9: Total (aka cumulative) radiation dosage (converted to mSv) during index admission. See Protocol, Section 2.3.6, for the precise definition.

Data for outcome S9 are obtained by a dedicated research nurse, by reviewing the hospital files after the procedure. Clinicians record the relevant information in the hospital files as per routine.

S10: Length of index admission (in days). The period starts at time of admission of the hospitalization stay during which the patient is included in the trial and ends at discharge. Time of discharge is defined as time of discharge to home or care facility or time of death in patients who die before discharge.

Data for outcome S10 are also obtained from hospital files.

4.4 Secondary objectives and estimands

For all secondary outcomes, we consider that both the "intention-to-treat" (ITT) and the "Per protocol" (PP) estimands are relevant, as for the primary outcome and for similar reasons.

For each of secondary **binary** outcomes **S1**-**S6**, the secondary objectives are to estimate the risks of the outcome in the two arms ICA and CCTA; as well as the risk differences. The corresponding clinical questions of interest are:

"What are the risks of [this secondary outcome] when using CCTA instead of ICA for initial diagnostic testing, among hospitalized patients with NSTEACS and an indication for ICA, regardless of any post-diagnostic management needed to ensure good clinical care? What is the corresponding risk difference?"

Accordingly, the corresponding estimands are defined as the primary estimands of Section 4.1, except for the endpoint attributes, which are defined by the above items **S1** to **S6**.

For secondary **quantitative** outcomes **S7**-**S8**, which are quality of life scores, the corresponding secondary objectives are to estimate the means of the outcome in the two arms ICA and CCTA, among patients alive one year after index admission; as well as the differences in means. This corresponds to a "survivors analysis", using the terminology of Colantuoni et al. [7]. The corresponding clinical questions of interest are:

"What are the means of [this quality of life score] one year after index admission, when using CCTA instead of ICA for initial diagnostic testing, among patients who were hospitalized with NSTEACS and an indication for ICA, who are still alive one year after index admission, regardless of any post-diagnostic management needed to ensure good clinical care? What is the corresponding

difference in means?"

Accordingly, the corresponding estimands are defined similarly to the primary estimands of Sec. 4.1; the difference are only:

- Population: same as for the primary estimands except that the included patients are required to be **alive** one year after index admission.
- Endpoint: now defined by the above items **S7** or **S8**.
- Intercurrent events: same as for the primary PP and ITT estimands, except that "death within one year" is now handled by the population of interest attributes.
- Population-level summary: Difference in mean between arms.

For secondary **quantitaive** outcomes **S9** (total radiation dosage) and **S10** (length of index admission), the corresponding secondary objectives are to estimate the medians of the outcome in the two arms ICA and CCTA; as well as the differences in medians. The corresponding clinical questions of interest are:

"What are the medians of the [total radiation dosage during index admission / length of index admission], when using CCTA instead of ICA for initial diagnostic testing, among hospitalized patients with NSTEACS and an indication for ICA, regardless of any post-diagnostic management needed to ensure good clinical? What is the corresponding difference in medians?"

Accordingly, the corresponding estimand is defined as for the primary estimand, except for the Endpoint, which is now defined by item **S9** or **S10** above and the population-level summary, which is now defined as a difference in median between arms.

4.5 Exploratory endpoints

The main exploratory objectives are to evaluate CCTA and ICA with respect to the following nine exploratory endpoints:

(a) Stroke within one year

(b) Coronary revascularization within one year

(c) Serious adverse event before 1^{st} hospital discharge

(d) Procedure-related complication during ICA, CCTA, PCI or CABG (e.g. allergy, heart attack, cardiac arrest, death, kidney failure) before 1^{st} hospital discharge and within one year (2 endpoints)

(e) BARC 3 or 5 bleedings before 1^{st} hospital discharge and within one year (2 endpoints)

(f) Hospitalization for any cardiac reason or related to cardiovascular treatment within one year

(g) Time to revascularization or to decision not to perform revascularization

See Protocol, Section 2.3.6, for their precise definitions.

5 General Considerations

5.1 Statistical Hypotheses

The following two (confirmatory) 1-sided hypotheses are planned to be tested, in relation to the primary objectives (detailed in Sec. 4.1). The primary aim is to show that CCTA is not unacceptably worse than ICA, using a non-inferiority margin of 5% for the difference of one-year risk of MACE. The aim is to show this using the two estimands: the "intention-to-treat" estimand and "Per Protocol" estimand. Importantly, non-inferiority will only be concluded if both are significant. If this is achieved, a test for superiority will subsequently be made, using the "Intention-to-treat" estimand.

Non-inferiority hypothesis test:

- Null hypothesis: the risk of MACE within one-year after admission is **at least** 5% higher when using CCTA than when using ICA. Formally, $\mathcal{H}_0 : \pi_1 - \pi_0 \geq 5\%$, where π_1 and π_0 are the one-year risk after CCTA or ICA, respectively.

versus

- Alternative hypothesis: the risk of MACE within one-year after admission is **at most** 5% higher when using CCTA than when using ICA. Formally, $\mathcal{H}_1 : \pi_1 - \pi_0 < 5\%$.

Superiority hypothesis test:

$$\mathcal{H}_0 : \pi_1 - \pi_0 \geq 0 \quad \text{versus} \quad \mathcal{H}_1 : \pi_1 - \pi_0 < 0.$$

Operationally the hypotheses will be evaluated by two-sided tests at 5% and matching 95% two-sided CI rather than by a one-sided test at 2.5% and matching 97.5% one-sided CI. This is the standard case considered in [6, 13].

5.2 Multiplicity Adjustment

No multiple testing correction will be used, as formal hypothesis testing will be performed only for the primary estimands described in Section 4.1. Although there are two estimands and hence two hypotheses test, no multiple correction is needed as we require that both are significant to conclude non-inferiority, which corresponds to an "intersection union test" [2, Sec. 2.2.2].

Reporting for other endpoints/estimands will be limited to point estimates of effects with 95% (two-sided) confidence intervals. The widths of the intervals will not be adjusted for multiplicity and therefore it will not be possible to use them in place of formal hypothesis testing. This is in line with common recommendations [31].

Although, a test for superiority might be performed in addition to the test for non-inferiority for the primary endpoint (see Section 5.1), we will not adjust for multiple testing as there is no need here, since the hypothesis test will be performed only if non-inferiority is established; see e.g. [6, Sec. IV.1] or [13, Sec. G].

5.3 Handling of Missing Data

Missing data are not expected for the primary outcome or secondary outcomes S1-S6 and S10. Indeed, these outcomes should always be available from hospital files. Some missing data are

expected for outcomes S7-S8, as some patients might not answer the questionnaires electronically and also decline telephone interviews offered to them as an alternative. For outcome S9 (Total radiation dosage), some missing data are expected as a few clinicians will likely forget to record the information into the hospital files.

The main analyses of outcomes S7, S8 and S9 will be performed using a complete case analysis, which relies on the missing at random assumption. Sensitivity analyses relaxing this assumption will complement the main analyses of outcome S7 and S8, for completeness, although we expect very few missing data and non-informative missing data. The sensitivity analyses will be conducted as outlined in [25, 30] (more details are provided below). Similar post hoc analyses will be performed for exploratory outcomes, if relevant.

To handle missing baseline covariate data (if any), we will use a simple but valid imputation approach [29].

5.4 Other general considerations

5.4.1 Time zero in estimands versus in data analysis

The time will start at randomization when estimating the one-year risks of the primary outcome, the secondary outcomes S1-S6, as well as for similar risks (e.g. those shown by cumulative incidence curves), for all data analyses. However, for defining the corresponding estimands and thus also for the interpretation of the results, we will pretend that the time zero is the "time of admission". This is because the difference between these two times is negligible for interpretation purpose and to facilitate the interpretation and communication of the results. Hence, we write "one year after admission" to define the estimands but we will use "one year after randomization" in the corresponding statistical analyses.

5.4.2 Covariate adjustment and marginal treatment effects

The main target of inference are marginal estimands and marginal "treatment effects" [29, 36], also often termed as "unconditional" [14], similar to that of simple unadjusted analyses. Although the main analyses will use covariate adjustment, the primary goal of covariate adjustment will be to improve precision in estimating the marginal treatment effect [29, 36, 40]. Hence, we will use multiple regression together with standardization to leverage the information contained in the baseline covariates, while targeting marginal treatment effects. This method is promoted in a recent FDA guidance document [14] and it is known to have important robustness properties to model misspecification, see e.g. [41] and references therein.

5.4.3 Reporting of redundant results

For all secondary outcomes, we consider that both the "intention-to-treat" (ITT) and the "Per protocol" (PP) estimands are relevant, as for the primary outcome and for similar reasons. However, to facilitate a concise presentation of the results, we might focus on the ITT estimands in the main document presenting the results and provide the results for the PP estimands in supplementary material. A summary of the main differences between the two sets of results will be presented in the main text (if any). This should not be interpreted as us considering that the PP estimands are less important than the ITT estimands. This is just a pragmatic choice as we expect the results and their interpretation to be nearly identical (as we expect a very small proportion of non adherence). A similar reporting strategy will likely be applied to analyses related to the exploratory outcomes.

6 Analysis Sets

- The **"All participants analysis set"** consists of all randomized participants.
- The **"Per protocol analysis set"** consists of all randomized participants that have received the intervention (ICA or CCTA) to which they were randomized. That is, the dataset obtained after removing the patients who did not receive the randomized intervention from the "All participants analysis set". In this document, these patients are referred to as "non adherers".
- The **"As treated analysis set"** consists of all randomized participants that have received either ICA or CCTA. That is, the dataset obtained after removing the patients who did receive neither ICA nor CCTA from the "All participants analysis set". This analysis set will be similar to the "Per protocol analysis set" if no patient received CCTA instead of ICA as randomized and no patient received ICA instead of CCTA as randomized.
- The **"One-year survivors analysis set"** consists of all randomized participants alive one-year after randomization. That is, the dataset obtained after removing the patients who died within one year from the "All participants analysis set".
- The **"One-year survivors PP analysis set"** consists of all randomized participants alive one-year after randomization that have received the intervention (ICA or CCTA) to which they were randomized. That is, the dataset obtained after removing the patients who died within one year from the "Per protocol analysis set".
- The **"One-year complete case survivors analysis set"** consists of all randomized participants alive one-year after randomization without missing questionnaire data. That is, the dataset obtained after removing the patients with missing questionnaire data from the "One-year survivors analysis set". Patients with missing questionnaire data will be patients who did not respond to the questionnaire either electronically or via phone interviews. We will consider that the patients did not respond to the questionnaire electronically if they did not successfully submit their responses to the questionnaire. Preliminary responses saved by the patients but not submitted yet will be considered as missing.
- The **"One-year complete case survivors PP analysis set"** consist of all patients of the "One-year complete case survivors analysis set" that have received the intervention (ICA or CCTA) to which they were randomized.
- The **"Complete case radiation dosage analysis set"** consists of all randomized participants without missing radiation dosage data.
- The **"Complete case radiation dosage PP analysis set"** consist of all patients of the "Complete case radiation dosage analysis set" that have received the intervention (ICA or CCTA) to which they were randomized.

Table 1 page 19 provides an overview of the analysis sets that will be use to analyze the main outcomes.

7 Analyses supporting the primary objective

7.1 Main Analytical Approach

The main analysis for the ITT estimand will use the "All participants analysis set" and corresponds to an "intention-to-treat" analysis. The main analysis for the PP estimand will use the "Per protocol analysis set" and corresponds to a "per protocol analysis" analysis. Apart from the difference in analysis set being used, the two analyses will be performed identically, as described below in Sections 7.1.1 and 7.1.2.

7.1.1 Computation of p-values, confidence intervals and significance

We will perform "Wald-type" inference for the risk difference as follows. The 95% CIs will be computed as "Est. $\pm$ 1.96·SE", where "Est." denotes the point estimate and SE the corresponding standard error. For the test of non-inferiority, we will use the test statistic $Z_{NI}=$ (Est.$-$NI)/SE to compute the p-value, where NI denotes the non-inferiority margin and assuming that Z_{NI} follows a standard normal distribution. For the test of superiority, we will use $Z_{Sup} =$ (Est.)/SE instead. We will consider the results statistically significant for the non-inferiority test if the p-value is below 5%, or equivalently, if the upper limit of the 95%-CI is below the non-inferiority margin, which is set to 5%. We will consider the results statistically significant for the superiority test if the p-value is below 5%, or equivalently, if the upper limit of the 95%-CI is negative.

7.1.2 Computation of estimates and standard errors

We will use multiple logistic regression together with standardization to estimate the one-year risks in each arm with two-sided 95%-CIs, their difference with two-sided 95%-CI and the p-value for the non-inferiority null hypothesis (and the p-value for the superiority null hypothesis, if relevant). This approach corresponds to using the standardized estimator advocated in [36]. Robust standard errors will be computed by Bootstrapping. This approach is promoted and outlined in a recent FDA guidance document [14]. Specifically, we will adjust on the following baseline covariates:

- sex: male/female
- age group: ≤ 55, $(55, 65]$, $(65, 75]$ or > 75.
- diabetes: yes/no (data obtained by a dedicated research nurse, by reviewing the hospital files, especially the patient's history and his or her current medication list)
- heart failure: yes/no (defined as history of heart failure with preserved or reduced ejection fraction or history of left ventricular ejection fraction <40%; data obtained by a dedicated research nurse, by reviewing the hospital files)
- presence of elevated troponin according to local hospital standard: yes/no

These variables will be included in the multiple logistic model in addition to the binary variable that indicates the randomized assignment to CCTA or ICA. No interaction will be used in the multiple logistic regression model. Computation of the standardized estimator and its robust standard error via bootstrap will be performed as described in appendix A.5 in [36] (with 1000 bootstrap samples). To handle missing baseline covariate data (if any), we will use the mode imputation approach (across arms, not within), as it is simple and appropriate with the standardized estimator [29].

Note that we expect appropriately 15% of the patients with diabetes, 10% to 15% with heart failure and approximately 80% without elevated troponin, at baseline. These expected prevalences were defined from, among other things, the report from a similar trial [24]. We also expect that between 20% and 30% of patients are contained within each of the four pre-specified age groups.

7.2 Sensitivity Analyses for the PP estimand

As pointed out in [28], the PP estimand analysis relies on a few assumptions to be unbiased, mainly:

- conditionally on baseline covariates adjusted for, patients who adhere with their assigned treatment are exchangeable between treatment arms; this is sometimes referred to as "no unmeasured confounding"
- there is no treatment effect heterogeneity accross subgroups defined by the baseline covariates adjusted for
- the multiple logistic regression model is correctly specified

Similar remarks can be found in e.g. [18, 19]. This motivates the following two sensitivity analyses, for the PP estimand. These sensitivity analyses will be relevant only if some patients received ICA instead of CCTA as randomized or some patients received CCTA instead of ICA as randomized. Otherwise these analyses would be identical to the PP estimand main analysis. We expect to observe such patients, referred to as "non adherers" below, although not many ($\leq 3\%$).

First sensitivity analysis. We will use the "As treated analysis set" and performed an "as treated analysis". Specifically, we will proceed as for the main analysis described in Section 7.1 except that we will use the variable that indicates the initial diagnostic testing received (CCTA or ICA) instead of that assigned by the randomization, when they differ. As pointed out in [18], this "as treated" analysis relies on similar assumptions to the PP estimand analysis, but not exactly the same, as the analysis set includes more patients. Because of these assumptions, we will performed a second sensitivity analysis which relaxes these assumptions.

Second sensitivity analysis. In this analysis, we proceed as in the first sensitivity analysis except in the "averaging step", when computing the standardized estimator of the risk difference. Specifically, we will set the estimated individual risk difference to be equal to a specific value Δ for patients who did not receive the intervention to which they were randomized. We will repeat the analysis for $\Delta = 5\%, 6\%, \ldots, 10\%$ and the case $\Delta = 5\%$, which corresponds to the non-inferiority margin, will be of special interest. Indeed, it correpsonds to the sensitivity analysis adovated by Koch [23, Sec. 6] and mentioned in a FDA guidance document [13, Sec. F]. Interestingly, this approach prevents non adherers to make the results of the two arms more comparable than if they were no non adherers, but still keeps the non adherers into the analysis set. Therefore, this approach controls the type-I error for the non-inferiority hypothesis test, when assuming $\pi_1 - \pi_0 = \Delta$, for any $\Delta \geq 5\%$.

8 Analyses supporting secondary objectives

8.1 Analyses of binary secondary endpoints S1 to S6

For the ITT estimands, we will use the "All participants analysis set". For the PP estimands, we will use the "Per protocol analysis set".

For outcome **S3** (hospitalization with refractory angina) and **S5** (cardiovascular death), we expect very small risks, 2% and 2.5%, respectively [24]. Hence we will not adjust for baseline covariates in the computation of the CIs. Specifically, we will compute the two-sided 95%-CI for the risk difference using the Miettinen-Nurminen asymptotic score interval method [27], as it has been shown to perform well and to be *"safe to use"* by Fagerland et al. [11], even when few events are observed. For computation, we will use the `diffscoreci()` function from the `PropCIs` package of `R`, as suggested by Fagerland et al. [11]. This method is more appropriate than common alternatives when a small to moderate number of events is expected, but it is not overly conservative [11].

For outcomes **S1** (death), **S2** (non-fatal MI) and **S6** (unplanned coronary revascularization), we expect a risk of approximately 4%. For outcome **S4** (hospitalization with HF), we expect a risk of approximately 7%. These expected risk were derived from, among other things, the report from a similar trial [24]. The risks are not as small as for outcomes **S3** and **S5** and we expect to observe more than a few events for each of these outcomes, even among patients with diabetes or within any other level of each variable adjusted for in the primary analysis detailed in Section 7.1.2. Therefore, we plan to proceed as for the analysis of the primary outcome , using multiple logistic regression with standardization, to leverage the information contained in the baseline covariates (as detailed in Section 7.1). However, if the observed risks happen to be lower than expected, the adjustment set might need to be modified. Hence we pre-specify the following pragmatic backup strategy. For the analysis of each outcome, we will systematically remove from the adjustment set any covariate for which less than 5 events are observed for at least one level of the covariate.

8.2 Analyses of quantitative secondary endpoints S7 and S8

We expect a few missing data for outcomes **S7** and **S8**, as some patients might not answer to the questionnaires sent one year after randomization (and not reply to telephone interviews either). But, we expect very few missing data and non-informative missing data. Hence we pre-specify the main analysis to be a complete case analysis. That is, we will use the "One-year complete case survivors analysis set" for the analysis related to the ITT estimands and the "One-year complete case survivors PP analysis set" for those related to the PP estimands. The analytical approach is detailed in Section 8.2.1 below.

Because the complete case analysis rely on the missing at random assumption, sensitivity analyses that do not rely on this assumptions could provide valuable additional insights. For these sensitivity analyses, we will use the "One-year survivors analysis set" for the analysis related to the ITT estimands and the "One-year survivors PP analysis set" for those related to the PP estimands. The analytical approach is detailed in Section 8.2.2 below.

8.2.1 Main analysis of quantitative secondary endpoints S7 and S8

We will fit a multiple linear model (aka ANCOVA) to estimate and compare means. This method leverages the information contained in the baseline covariates and it is often used and recom-

mended [34, 14]. Specifically, we will adjust on the following baseline variables:

- all those used for the primary analysis, listed in Section 7.1.2

and additionally:

- admitting hospital

These variables will be included in the multiple linear model in addition to the binary variable that indicates the randomized assignment to CCTA or ICA. No interaction will be used in the linear model. We will compute standard errors by non-parametric bootstrap resampling, as for the analysis of the primary outcome (with 1000 bootstrap samples). To handle missing baseline covariate data (if any), we will use the mode imputation approach (across arms, not within), as for the analysis of the primary outcome. Means in each arm will be estimated by standardization, for consistency with the adjusted analysis of the mean difference. The 95% CIs will be computed as "Est. $\pm$ 1.96·SE", where "Est." denotes the point estimate and SE the corresponding standard error.

8.2.2 Sensitivity analysis of quantitative secondary endpoints S7 and S8

Sensitivity analyses will be conducted as outlined in [25, 30]. Especially, we will follow the *"Pattern Mixture Model Approach"* described in the Section *"Example: Single outcome with auxiliary data"* and subsection *"Example: Continuous Values of Y"*, pages 91-92 in [30]. The estimated mean of the missing outcome given baseline covariates will be imputed as described in equation (15) page 91 in [30], that is, as the estimated mean based on the linear model fitted on the complete case data (as described above in Section 8.2.1) plus a shift parameter, denoted by Δ_z, $z = 0, 1$, which is specific to the arm CCTA or ICA in which the patient with missing outcome belongs. Inference will be performed following the 5 steps described in page 93 of [30], in the Section "Inference"[1]. The results of the sensitivity analysis will include heatmaps (or contour plots) of the point estimates and of the lower and upper limits of the 95%-CI, for a grid of values of the two sensitivity paramters (Δ_0, Δ_1). Of note, the parameter value $\Delta_0 = \Delta_1 = 0$ corresponds to the main analysis (assuming "missing at random"). This approach is similar to a simple "tipping point sensitivity analyses" rountinely performed in the context of drug development, see e.g., [17]. This sensitivity analysis will provide insights about the extent to which the missing outcomes should differ from those observed to substantially change the results.

8.3 Analyses of quantitative secondary endpoints S9 and S10

Estimated medians within each treatment arm and their difference will be reported, together with two-sided 95%-CIs. Point estimates will be computed as sample medians (via the `median()` function in `R`) and standard errors by non-parametric bootstrap resampling, as for the analysis of the primary outcome (with 1000 bootstrap samples). The 95% CIs will be computed as "Est. $\pm$ 1.96·SE", where "Est." denotes the point estimate and SE the corresponding standard error.

For secondary outcome **S9** (radiation dosage), the "Complete case radiation dosage analysis set" will be used for the ITT estimands and the "Complete case radiation dosage PP analysis set" for the PP estimands. We expect very few missing values (if any) and non-informative missing data, hence no sensitivity analysis is pre-specificy to complement the above analysis, which assumes that

[1] Specifically, we will use the simple pre-specified funtions $g(x) = x$ and $d(Y_0, \Delta) = \Delta$, where Y_0 denote the baseline covariates, for the computation described page 93 in [30].

the data are missing completely at random, within each arm. This missing completely at random assumption is thought very plausible here. Indeed, clinicians record radiation dosage in the hospital files as per routine and the only cause of missing data that we can think of is a completely random omission.

For secondary outcome **S10** (length of index admission), the "All participants analysis set" will be used for the ITT estimands and the "Per protocol analysis set" for the PP estimands.

9 Analyses supporting exploratory objectives

For all exploratory endpoints described in items (a)–(f) in Section 4.5, we will report the empirical proportions in each arm and their difference, with a 95%-CI for the difference. Specifically, we will use the same (unadjusted) statistical method as that used for secondary endpoints **S3** and **S5** described in Section 8.1.

For composite exploratory endpoints, e.g., serious adverse event or procedure-related complication, we will report the proportions for "any", i.e., at least one. But, we will additionally report counts and empirical proportions for each component of the composite endpoints separately, to provide more granular descriptive statistics: e.g., counts and proportions of each type of serious adverse event.

For the analysis of the time to revascularization or to decision not to perform revascularization (item (g) in Section 4.5), we will report the medians in each arm and their difference, with a 95%-CI for the difference. Specifically, we will use the same statistical method as that used for secondary endpoints **S9** and **S10** described in Section 8.3.

We will use thee "All participants analysis set" for the ITT estimands and the "Per protocol analysis set" for the PP estimands.

	Analysis set								
Outcome(s)	"All"	"PP"	"AT"	"1-Y"	"1-Y pp"	"1-Y cc"	"1-Y cc pp"	"cc rad"	"cc rad pp"
Primary	X	X	X						
S1–S6	X	X							
S7–S8				X	X	X	X		
S9								X	X
S10	X	X							
(a)–(g)	X	X							

Note: "All", "PP", "AT", "1-Y", "1-Y pp", "1-Y cc", "1-Y cc pp", "cc rad" & "cc rad pp", stand for "All participants analysis set", "Per protocol analysis set", "As treated analysis set", "One-year survivors analysis set", "One-year survivors PP analysis set", "One-year complete case survivors analysis set", "One-year complete case survivors PP analysis set", "Complete case radiation dosage analysis set", and "Complete case radiation dosage PP analysis set", respectively.

Table 1: Analysis sets used to analyze the main outcomes.

10 Other analyses

10.1 Additional outcome analyses

10.1.1 Invasive and non-invasive examinations

We will report proportions and counts of patients undergoing at least one invasive or non-invasive examinations for ischemic heart disease within one year, within each arm as well as the between arm difference. The examinations can be CCTA, ICA, cardiac MRI, Rb-PET or SPECT. Specifically,

we will use the same (unadjusted) statistical method as that used for secondary endpoints **S3** and **S5** described in Section 8.1. We will perform the analysis twice, once using the "All participants analysis set" and once using the "Per protocol analysis set", for completeness.

10.1.2 Contrast dose during admission

For the analysis of the total (aka cumulative) contrast dose used during index admission, we will report the medians in each arm and their difference, with a 95%-CI for the difference. Specifically, we will use the same statistical method as that used for secondary endpoints **S9** and **S10** described in Section 8.3. See protocol Section 2.3.6 for the definition of total (aka cumulative) contrast dose used during index admission.

10.1.3 Questionnaire data

Outcomes **S7** and **S8** are derived from replies to the EQ-5D-5L and the SAQ-7 questionnaires, which contain 5 and 7 questions, with 5 and 6 possible answers, respectively [9, 4].

Following common recommendations, we will report the proportions of responses to each of the five questions of the EQ-5D-5L for each arm and the corresponding barplots as in Figure 2 in [9]. For the SAQ-7, it has been argued that the *"The SAQ Angina Frequency Score has a more clinically intuitive interpretation. Scores of 0 to 30 points translate to having daily angina, 31 to 60 points to having weekly angina, 61 to 99 points to having monthly angina, and 100 points to having no angina; this alignment has been confirmed through comparisons with daily angina diaries."* [38]. Hence we will also report the proportions of patients having **S8** within 0 to 30 points, 31 to 60 points, 61 to 99 points and 100 points in each arm (and the corresponding barplot). We will perform the analysis twice, once using the "All participants analysis set" and once using the "Per protocol analysis set", for completeness.

Additionally, descriptive statistics will also be presented for the time from randomization to the 'reply to the questionnaire at "one-year", per arm. Median, minimum, maximum, first and third quartiles will be reported.

10.1.4 Cumulative incidence curves

Cumulative incidence curves will be reported per arm to describe when each of the following outcomes occurred during the follow-up.

- MACE
- death (all-cause mortality)
- non-fatal myocardial infarction (MI)
- hospitalization with refractory angina
- hospitalization with heart failure (HF)
- cardiovascular death
- unplanned coronary revascularization

We will report such curves not only within one year from randomization but also afterwards. Indeed, many participants will have more than one year of follow-up. We expect that the patients can have up to 4 years of follow-up, as we plan to include patients for approximately 3 years and complete the trial 1 year after the last inclusion. Because the length of follow-up will varies from patient to patient due to staggered entries, we will observe usual right censored data. For the two outcomes MACE and death, we will therefore use the Kaplan-Meier estimator. Instead, the Aalen-Johansen estimator will be used for the others, to account for the competing risk of death in top of censoring, as commonly recommended [35]. We will use the `prodlim()` function of the `prodlim` package of `R` to compute point estimates as well as 95%-CI.

We will perform the analysis using the "All participants analysis set". We might reproduce the analysis using the "Per protocol analysis set", if relevant, although we do not expect it will be the case, as we expect very few non adherers.

10.1.5 Subgroup analyses

We plan to perform subgroup analyses for these subgroups:

- male versus female
- age $\geq$ 64 versus age $<$ 64
- Elevated troponin versus not elevated troponin
- GRACE score $\leq$ 140 versus $>$ 140
- Prior ASCVD versus no prior ASCVD
- History of heart failure versus no history of heart failure
- Diabetes mellitus versus no diabetes mellitus
- Current or former smoker versus never-smoker
- Anemia versus no anemia
- First 50% of included patients versus last 50% of included patients
- Patient randomized at an invasive center versus randomized at a non-invasive center (i.e, Rigshospitalet, Gentofte and Roskilde versus the other hospitals)

For each subgroup, we would like to replicate the main analyses describe in Section 7.1. However, the smaller sample sizes (as well as the definition of some subgroups) implies that it might be inappropriate to use the same adjustment set, for some of the subgroups. Hence, for each subgroup analysis, we will systematically remove from the adjustment set any covariate for which less than 5 events (MACE) are observed for at least one level of the covariate. For instance, we will remove heart failure from the adjustment set in the subgroup analysis of patients with diabetes if less than 5 events are observed among patients with heart failure (which could happen as both heart failure and diabetes are not expected to be very common at inclusion). Also, for the subgroup analysis with age $\geq$ 64, the categorical age variable in the adjustment set will be updated as ≤ 75 versus > 75. For the subgroup analysis with age $<$ 64, it will be updated as ≤ 55 versus > 55. Note that some of these subgroups were analyzed in the VERDICT trial [24]. We refer to the protocol, Section 2.3.4, for the definition of the variables that define the subgroups.

10.1.6 Length of index admission

To complement the primary analysis of outcome **S10** (see Section 8.3), we will also report the medians number of days alive after discharge of index admission within 28 days within each arm and their difference (with 95% CI). The rationale for this additional analysis is to not make a short length of admission due to death look "good". This will facilitate the interpretation of the results for outcome **S10** from a clinical point of view. By contrast, directly comparing the median length of admission without distinguishing end of admission due to death or discharge alive (as done for the primary analysis of this outcome **S10**) should facilitate the interpretation of the results for outcome **S10** from an "hospitalization cost" point of view. However, because we expect very few deaths during index admission, these two approaches to analyze outcome **S10** are expected to end up being very similar. Choosing a time horizon of 28 days was thought as relevant here and it is a common choice [26, 1]. As for the primary analysis of outcome **S10**, we will compute standard errors by Bootstrapping.

10.1.7 Costs analyses

Costs analyses described in the protocol will be specified and performed later.

10.2 Descriptive analyses

10.2.1 Procedure details

We will descriptively summarize the:

- procedure (PCI) duration during index admission
- total fluoroscopy time during index admission

among subjects who received a PCI. Minimun, maximum, median, first and third quartiles will be reported, within each of the two subgroups: those who received CCTA and those who receive ICA ("As Treated Analysis").

10.2.2 Radiation dosage

Total (aka cumulative) radiation dosage (converted to mSv) of all cardiac diagnostic examinations and interventions (CCTA, PCI, ICA, SPECT, Rb-PET) within 1 year will be descriptively summarized within each of the two subgroups: those who received CCTA and those who receive ICA ("As Treated Analysis"). Minimun, maximum, median, first and third quartiles will be computed, using a complete case analysis. The proportion of missing data will also be reported.

10.2.3 Examination and randomization timing

We will descriptively summarize the:

- time from index admission to randomization
- time from randomization to examination (i.e., initial diagnostic testing, ICA or CCTA)
- time from index admission to examination

among subjects who received CCTA and among those who receive ICA ("As Treated Analysis"). Minimun, maximum, median, first and third quartiles will be computed, using a complete case analysis.

10.2.4 Non-adherence

We will report the proportions of non adherers. Specifically, within each of the two arms, "randomized to ICA" and "randomized to CCTA", the counts and proportions of patients who received the other examination (ICA instead of CCTA or CCTA instead of ICA) and the counts and proportions of patients who did not received any examination (neither ICA nor CCTA). Additional descriptive statistics will summarize the data recorded by the clinicians to document the reason for such deviation to the randomized assignment to CCTA or ICA.

10.2.5 Recruitment

Recruitment of the patients will be summarized via descriptive statistics. Especially, start and end dates of recruitment will be presented as well as a flowchart, inspired by the CONSORT guidelines and template [33].

10.2.6 Screening data

Screening data, about assessment for eligibility, are collected. They will be presented in a in a flow diagram, following the CONSORT guidelines [33].

10.2.7 Baseline characteristics

Baseline characteristics will be descriptively summarized per arm, using the "All participants analysis set". The list of baseline variables to be summarized includes:

- sex (male/female)
- Age (year)
- Admitting hospital
- Body mass index (kg/cm^2)
- Smoking status (current, former or never)
- eGFR (ml/min/1.73m^2)
- Hemoglobin (mmol/L)
- Left ventricular ejection fraction (%)
- Elevated troponin (yes/no)
- Ischemic ECG changes (yes/no)
- GRACE score

- ASCVD (yes/no)
- Maximal troponin prior to randomization (ng/L)
- History of obstructive lung disease (yes/no)
- History of stroke (yes/no)
- History of myocardial infarction (yes/no)
- History of Hypercholesteorlemia (yes/no)
- History of Hypertension (yes/no)
- History of heart failure (yes/no)
- Diabetes mellitus (Type 1, Type 2 or none)
- Acetylsalicylic acid (yes/no)
- Antithrombotics (yes/no)
- Anticoagulants (yes/no)

For quantitative variables, we will present median, first and third quartiles and also minimum and maximum. Categorical variables will be summarized by counts and percentages. The number and proportions of missing values (if any) will be reported for each variable, per arm. Hypothesis tests will not be performed to compare baseline characteristics, but clinical importance of any imbalance will be noted. This is in line with usual recommendations [32].
Supplementary tables will present similar descriptive statistics within three subgroups of patients: those who did no receive any examinations, those randomized to ICA who received CCTA and those randomized to CCTA who received ICA. These supplementary tables might be used to discuss the plausibility of the missing at random or missing completely at random assumptions assumed to hold for some analyses.

10.2.8 Outcomes

Outcomes (primary, secondary and exploratory) will also be descriptively summarized, per arm (using relevant analysis sets described in Section 6). Similar descriptive statistics as for baseline characteristics will be used. Mean and standard deviation will be presented instead of median, first and third quartiles (or in addition to those) whenever appropriate. Importantly, the number and proportions of missing data (if any) will be reported.

11 Sample size determination and power calculations

11.1 Sample size determination

A sample size of $n = 2,300$ patients (1,150 in each group) was calculated as follows. The desired power was 90% and the type one error was set to 5%. The expected proportion of primary endpoint in both groups was 15%. This assumption was made based on data from a previous trial [24].

Using a non inferiority margin of 5% for the difference in proportions and the usual asymptotic normal approximation gives

$$\text{Power} \approx \Phi\left(\frac{(0.15-0.15)+0.05}{\sqrt{0.15(1-0.15)/1073+0.15(1-0.15)/1073}}-1.96\right)\approx 90\%\ ,$$

where Φ is the cumulative standard normal distribution function; see e.g. Chow et al. [5, Sec. 4.2.2]. This suggests a sample size of $n/2$ =1073 per group, but the sample size was further rounded up to 1150 per group, so $n = 2300$ in total, to account for the possibility that a few patient might withdraw their consent. Further loss of follow-up is not accounted for in the calculation as it is not expected. Note that in the above calculation the quantile 1.96 is used to reflect a two-sided test with type-I error control at 5% (or equivalently a one-sided test at 2.5%). We expect very few non adherers, hence this sample size calculation is expected to be relevant for both the ITT and the PP estimand analyses.

11.2 Additional power calculations and remarks

1. Using $n/2 = 1150$ in the above formula instead of 1073, the power is computed as 91.9%. This can be interpreted as the power with sample $n = 2300$ assuming that no patient will withdraw consent.

2. With a sample size of $n/2 = 1073$ (and no missing data or withdraw of consent), assuming a proportion of 13%, 14%, 16% or 17% instead of 15% for the CCTA group leads to a power of 99.7%, 97.7%, 72.6% or 47.5%, repectively. This shows that the power to show non-inferiority is very large if CCTA is superior to ICA, but also that the power may be very low if CCTA is non-inferior at the 5% margin level but still "inferior" in the sense that the risk is 2% higher in the CCTA group than in the ICA group. In that case, the power is indeed only 47.5%.

3. Assuming that CCTA is superior to ICA because it leads to a risk of 12% or 13% instead of 15% in the ICA group, the power to show superiority is computed as of 53.0% and 26.7%, respectively (again using $n/2 = 1073$).

4. Note that the width of the 95% confidence interval of the primary outcome computed upon completion of the trial, for the risk difference between the groups, is expected to be between $2 \times 2.9\% = 5.8\%$ and $2 \times 3.1\% = 6.2\%$, assuming a risk of 15% in the CCTA group and a risk between 12% and 17% in the CCTA group. That is, we expect a margin of error between $\pm 2.9\%$ and $\pm 3.1\%$.

5. Note that the above sample size and power calculations do not account for the planned baseline covariates adjustment (detailed in section 7.1). This approach is often thought as reasonable and recommended, although it should lead to slightly conservative sample size and power calculations [14, 29].

References

[1] Beyersmann, J., Friede, T., and Schmoor, C. (2021). Design aspects of COVID-19 treatment trials: Improving probability and time of favorable events. *Biometrical Journal*.

[2] Bretz, F., Hothorn, T., and Westfall, P. (2010). *Multiple comparisons using R*. Chapman & Hall/CRC.

[3] Brittain, E. and Lin, D. (2005). A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Statistics in Medicine*, 24(1):1–10.

[4] Chan, P. S., Jones, P. G., Arnold, S. A., and Spertus, J. A. (2014). Development and validation of a short version of the seattle angina questionnaire. *Circulation: Cardiovascular Quality and Outcomes*, 7(5):640–647.

[5] Chow, S.-C., Shao, J., Wang, H., and Lokhnygina, Y. (2008). *Sample size calculations in clinical research, Second Edition*. CRC press.

[6] CMP (2000). Points to consider on switching between superiority and non-inferiority. Technical report, EMA, `https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-and-non-inferiority_en.pdf`.

[7] Colantuoni, E., Scharfstein, D. O., Wang, C., Hashem, M. D., Leroux, A., Needham, D. M., and Girard, T. D. (2018). Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ*, 360.

[8] Devlin, N., Parkin, D., and Janssen, B. (2020). *Methods for analysing and reporting EQ-5D data*. Springer Nature.

[9] EuroQol Research Foundation (2019). *EQ-5D-5L User Guide*. Available from: https://euroqol.org/publications/userguides.

[10] Evans, S. and Ting, N. (2015). *Fundamental concepts for new clinical trialists*. CRC Press.

[11] Fagerland, M. W., Lydersen, S., and Laake, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical Methods in Medical Research*, 24(2):224–254.

[12] Fleming, T. R. (2008). Current issues in non-inferiority trials. *Statistics in Medicine*, 27(3):317–332.

[13] Food and Drug Administration (2016). Guidance for industry non-inferiority clinical trials. `https://www.fda.gov/media/78504/download`.

[14] Food and Drug Administration (2023). Adjusting for covariates in randomized clinical trials for drugs and biological products guidance for industry. `https://www.fda.gov/media/148910/download`.

[15] Fox, K., Poole-Wilson, P., Henderson, R., Clayton, T., Chamberlain, D., Shaw, T., Wheatley, D., and Pocock, S. (2002). Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. *The Lancet*, 360(9335):743–751.

[16] Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P. R., Altman, D. G., Montgomery, A., Lim, P., et al. (2017). Guidelines for the content of statistical analysis plans in clinical trials. *JAMA*, 318(23):2337–2343.

[17] Gorst-Rasmussen, A. and Tarp-Johansen, M. J. (2022). Fast tipping point sensitivity analyses in clinical trials with missing continuous outcomes under multiple imputation. *Journal of Biopharmaceutical Statistics*, 32(6):942–953.

[18] Hernán, M. A. and Hernández-Díaz, S. (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical trials*, 9(1):48–55.

[19] Hernán, M. A., Robins, J. M., et al. (2017). Per-protocol analyses of pragmatic trials. *New England Journal of Medicine*, 377(14):1391–1398.

[20] ICH E9 (1998). Statistical principles for clinical trials. Technical report, EMA/CHMP/ICH, `www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline`.

[21] ICH E9 (R1) (2017). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Technical report, EMA/CHMP/ICH, `www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline`.

[22] Jensen, C. E., Sørensen, S. S., Gudex, C., Jensen, M. B., Pedersen, K. M., and Ehlers, L. H. (2021). The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. *Applied Health Economics and Health Policy*, 19:579–591.

[23] Koch, G. G. (2008). Comments on 'Current issues in non-inferiority trials' by Thomas R. Fleming, Statistics in Medicine, DOI: 10.1002/sim. 2855. *Statistics in Medicine*, 27(3):333–342.

[24] Kofoed, K. F., Kelbæk, H., Hansen, P. R., Torp-Pedersen, C., Høfsten, D., Kløvgaard, L., Holmvang, L., Helqvist, S., Jørgensen, E., Galatius, S., et al. (2018). Early versus standard care invasive examination and treatment of patients with non-st-segment elevation acute coronary syndrome: Verdict randomized controlled trial. *Circulation*, 138(24):2741–2750.

[25] Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H. (2012). The Prevention and Treatment of Missing Data in Clinical Trials. *New England Journal of Medicine*, 367(14):1355–1360.

[26] McCaw, Z. R., Tian, L., Vassy, J. L., Ritchie, C. S., Lee, C. C., Kim, D. H., and Wei, L. J. (2020). How to quantify and interpret treatment effects in comparative clinical studies of covid-19. *Annals of Internal Medicine*, 173(8):632–637.

[27] Miettinen, O. and Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine*, 4(2):213–226.

[28] Morgan, K. E., White, I. R., Leyrat, C., Stanworth, S., and Kahan, B. C. (2023). Applying the estimands framework to non-inferiority trials: guidance on choice of hypothetical estimands for non-adherence and comparison of estimation methods. *arXiv preprint arXiv:2312.00494*.

[29] Morris, T. P., Walker, A. S., Williamson, E. J., and White, I. R. (2022). Planning a method for covariate adjustment in individually randomised trials: a practical guide. *Trials*, 23(1):328.

[30] National Research Council (2010). The prevention and treatment of missing data in clinical trials. Washington, DC: National Academies Press, 2010.

[31] NEJM (2023). Statistical Reporting Guidelines of the New England Journal of Medicine, section Multiplicity considerations. https://www.nejm.org/author-center/new-manuscripts. Accessed: 2023-10-24.

[32] Schulz, Altman, and Moher, for the CONSORT Group (2023). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. https://www.goodreports.org/reporting-checklists/consort/. Accessed: 2023-11-17.

[33] Schulz, K. F., Altman, D. G., and Moher, D. (2010). Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and pharmacotherapeutics*, 1(2):100–107.

[34] Senn, S. (2005). Baseline balance and valid statistical analyses: common misunderstandings. *Applied Clinical Trials*, 14(3):24–27.

[35] Stegherr, R., Schmoor, C., Lübbert, M., Friede, T., and Beyersmann, J. (2021). Estimating and comparing adverse event probabilities in the presence of varying follow-up times and competing events. *Pharmaceutical Statistics*, 20(6):1125–1146.

[36] Steingrimsson, J. A., Hanley, D. F., and Rosenblum, M. (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary clinical trials*, 54:18–24.

[37] Stevens, G., Dolley, S., Mogg, R., and Connor, J. T. (2023). A template for the authoring of statistical analysis plans. *Contemporary Clinical Trials Communications*, page 101100.

[38] Thomas, M., Jones, P. G., Arnold, S. V., and Spertus, J. A. (2021). Interpretation of the Seattle Angina Questionnaire as an outcome measure in clinical trials and clinical care: a review. *JAMA cardiology*, 6(5):593–599.

[39] TransCelerate (2024). Common Statistical Analysis Plan Template v5. `https://www.transceleratebiopharmainc.com/assets/clinical-content-reuse-solutions/`. Accessed: 2024-05-01.

[40] Tsiatis, A. A., Davidian, M., Zhang, M., and Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in Medicine*, 27(23):4658–4677.

[41] Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M., and Rosenblum, M. (2023). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, 118(542):1152–1163.