Statistical Analysis Plan

(SAP)

Trial ID: TORNADO

Title of Trial

Treatment of rhinosinusitis with nasal polyposis with dupilumab and mepolizumab: A randomized, multi-centre, head-to-head comparison in real-world Danish patients

Trial ID: 2022-50-22-50-14-00

Investigational Medicinal Products:	dupilumab and mepolizumab
Trial phase:	Phase IV
Sponsor:	Christian von Buchwald, Rigshospitalet, Copenhagen, Denmark
GCP-Monitor:	The GCP Unit, Frederiksberg Hospital
Clinicaltrials.gov number:	NCT05942222
Data protection approval number:	Pactius P-2023-108
Date:	25.06.2025
SAP Version:	1.0
Protocol version:	1.8
SAP revision history and justification:	(None)

Roles and responsibility

Sponsor:

Name: Christian von Buchwald

Title: MD, Professor, Consultant, DMSc

Department: Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Copenhagen University, Rigshospitalet

Address: Inge Lehmansvej 7, 2100 Copenhagen

Telephone: +45 35452370

E-mail: Christian.von.buchwald@regionh.dk

Signature: _____ Date 25.06.2025

Chief Investigator and SAP author:

(Authorised to sign SAP and revisions)

Name: Christian Korsgaard Pedersen

Title: MD, Ph D.-student

Department: Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Copenhagen University, Rigshospitalet

Address: Inge Lehmansvej 7, 2100 Copenhagen

Telephone: +45 35457873

E-mail: Christian.korsgaard.pedersen.01@regionh.dk

Signature: _____ Date 25.06.2025

Senior statistician

Name: Janne Petersen

Title: M.Sc., Associate Professor, Ph.D.

Department: Copenhagen Phase IV Unit and Section for Data, Biostatistics and Pharmacoepidemiology

Address: Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital

SAP co-author

Name: Tonny Studsgaard Petersen

Title: MD, Associate Professor, Consultant, Ph.D.

Department: Department of Clinical Pharmacology Bispebjerg Hospital

Address: Bispebjerg Bakke 23, 2400 København NV

SAP co-author

Name: Vibeke Backer

Title: Professor, Consultant

Department: Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Copenhagen University, Rigshospitalet

Address: Inge Lehmansvej 7, 2100 Copenhagen

Table of contents

0A Summary table - **TORNADO-24**

Table 0A Statistical analysis plan summary – **TORNADO-24**

Objective	Outcome		Hypothesis	Method of analysis*
	Name	Type		
Primary objective				
To test for non-inferiority of DUPI to MEPO on the co-primary outcomes: improvement of symptom burden and reduction of polyp size after 24 weeks of treatment.	Sino-nasal Outcome Test 22 **(SNOT-22)**	Continuous	DUPI is non-inferior to MEPO	Linear mixed model (LMM) analysis on the ITT population
	Nasal polyp score **(NPS)**	Continuous	DUPI is non-inferior to MEPO	LMM analysis on the ITT population
Secondary objectives (will be tested hierarchically and only if the primary objective is significant)				
Secondary objective 1				
To compare the **smell improvement** after 24 weeks of treatment	Smell identification test 16 **(SSIT-16)**	Continuous	DUPI is superior to MEPO	LMM analysis on the ITT population
Secondary objective 2				
To compare the improvement of **symptom burden** after 24 weeks of treatment	**SNOT-22**	Continuous	DUPI is superior to MEPO	LMM analysis on the ITT population
Secondary objective 3				
To compare the **reduction of nasal polyp size** after 24 weeks of treatment	**NPS**	Continuous	DUPI is superior to MEPO	LMM analysis on the ITT population
Secondary objective 4				
To compare **the improvement in nasal congestion** after 24 weeks of treatment	**Nasal Congestion Score (NCS)**	Continuous	DUPI is superior to MEPO	LMM analysis on the ITT population
Secondary objective 5				
To compare the **improvement of comorbid asthma** after 24 weeks of treatment	**Asthma Control Test (ACT)**	Continuous	DUPI is superior to MEPO	LMM analysis on the ITT population
Explorative objectives				
To compare other disease-relevant outcomes.	**ACQ-7, VAS-CRS, VAS-smell, VAS-allergic rhinitis, VAS-asthma, VAS-NERD, Tympanometry,**	Depending on data	No formal hypothesis: exploratory comparisons between DUPI and MEPO to identify	Appropriate model depending on outcome type: Linear regression for continuous outcomes,

	responders according to Danish and EPOS 24-week criteria, FeNO, patient-reported treatment satisfaction		potential differences in treatment effects.	logistic regression for categorical outcomes.
Subgroup analyses: Will be performed for each of the co-primary outcomes: SNOT-22, NPS, and the key secondary outcome: SSIT-16. These subgroup analyses are exploratory and hypothesis-generating only due to power limitations.				
Age at inclusion	Age groups <40,40-65,>65	Categorical	Older age is linked to slower mucosal healing and poorer symptom improvement	LMM analysis on the ITT population
Body mass index (BMI)	15-25, >25	Categorical	Higher BMI is possibly linked to slower mucosal healing and poorer symptom improvement	LMM analysis on the ITT population
Gender (at birth)	Male, female	Categorical	Gender may affect the treatment effect	LMM analysis on the ITT population
Region	Capital, Zealand, Southern, Central, Northern	Categorical	Effect may be different in different regions of the country	LMM analysis on the ITT population
Number of previous ESS surgeries	1, 2, 3 or more	Categorical	Patients with a higher number of previous surgeries are less likely to achieve an SSIT-16 score >8.	LMM analysis on the ITT population
Comorbid asthma	Yes/no	Categorical	Marker of higher type-2 inflammatory burden; associated with larger SNOT-22/NPS reductions after biologics	LMM analysis on the ITT population
NSAID-exacerbated respiratory disease (N-ERD)	Yes/no	Categorical	Marker of higher type-2 inflammatory burden; associated with larger SNOT-22/NPS	LMM analysis on the ITT population

			reductions after biologics	
Blood eosinophilic count (*10^9/L)	Low: <150 Moderate: 150-500 High: >500	Categorical	Patients with higher levels of eosinophils in the blood will improve more than patients with lower levels, regardless of treatment.	LMM analysis on the ITT population
Total IgE-levels	<100, 100–500, >500 IU/mL	Categorical	Patients with higher IgE levels will improve more than patients with lower levels, regardless of treatment.	LMM analysis on the ITT population
Sensitivity analyses				
Type	**Outcome Name**	**Outcome type**	**Hypotheses**	**Method of analysis**
Per-protocol analysis	SNOT-22, NPS	Continuous	Results are robust to protocol violations.	LMM analysis of the per-protocol population (i.e., no major deviations)
Complete-case analysis	SNOT-22, NPS	Continuous	Results are robust to missing data.	LMM analysis of participants with complete data for primary outcomes; no imputation applied

*All analyses will be controlled for smell score (0-8 or >8), gender (M/F), and region (Capital/Zealand/South/Central/North) as they are stratification variables at randomization.
EOS: eosinophils; ESS: endoscopic sinus surgery; DUPI: dupilumab; ITT: Intention to treat; LMM: linear Mixed Model; MEPO: mepolizumab; NPS: Nasal Polyp Score; SNOT-22: Sino-Nasal Outcome Test; VAS: Visual Analogue Scale.

0B Summary table – TORNADO-48

Table 0B Statistical analysis plan summary – **TORNADO-48**

Objective	**Outcome**		**Hypothesis**	**Method of analysis***
	Name	**Type**		
Primary objective				
To test for non-inferiority of a treatment strategy with dupilumab as the initial treatment (DUP-	Sino-nasal Outcome Test 22 **(SNOT-22)**	Continuous	DUP-FIRST is non-inferior to MEP-FIRST	LMM analysis on the ITT population

FIRST) and response-guided crossover at 24 weeks to a strategy with mepolizumab as the initial drug (MEP-FIRST) on improvement of symptom burden and polyp size after 48 weeks.	Nasal polyp score **(NPS)**	Continuous	DUP-FIRST is non-inferior to MEP-FIRST	LMM analysis on the ITT population
Secondary objectives (will be tested hierarchically and only if the primary objective is statistically significant).				
Secondary objective 1				
To compare **smell improvement** after 48 weeks of treatment	Smell identification test 16 **(SSIT-16)**	Continuous	DUP-FIRST is superior to MEP-FIRST	LMM analysis on the ITT population
Secondary objective 2				
To compare improvement of **symptom burden** after 48 weeks of treatment	**SNOT-22**	Continuous	DUP-FIRST is superior to MEP-FIRST	LMM analysis on the ITT population
Secondary objective 3				
To compare **reduction of nasal polyp size** after 48 weeks of treatment	**NPS**	Continuous	DUP-FIRST is superior to MEP-FIRST	LMM analysis on the ITT population
Secondary objective 4				
To compare **improvement in nasal congestion** after 48 weeks of treatment	**Nasal Congestion Score (NCS)**	Continuous	DUP-FIRST is superior to MEP-FIRST	LMM analysis on the ITT population
Secondary objective 5				
To compare **improvement of comorbid asthma** after 24 weeks of treatment	**Asthma control Test (ACT)**	Continuous	DUP-FIRST is superior to MEP-FIRST	LMM analysis on the ITT population
Explorative objectives				

To compare other disease relevant outcomes.	**Lund-Mackay score, ACQ-7, VAS-CRS, VAS-smell, VAS-allergic rhinitis, VAS-asthma, VAS-NERD, Tympanometry, EPOS 48-week "excellent response", FeNO, patient reported treatment satisfaction**	Depending on data	No formal hypothesis: exploratory comparisons between DUPI and MEPO to identify potential treatment effects.	Appropriate model depending on outcome type: Linear regression for continuous outcomes, logistic regression for categorical outcomes.
Subgroup analyses: Will be performed for each of the primary outcomes: SNOT-22 and NPS. The subgroup analyses are exploratory and hypothesis-generating due to power limitations.				
Age at inclusion	Age groups <40,40-65,>65	Categorical	Older age is linked to slower mucosal healing and poorer symptom improvement	LMM analysis on the ITT population
Body mass index (BMI)	15-25, >25	Categorical	Higher BMI is possibly linked to slower mucosal healing and poorer symptom improvement	LMM analysis on the ITT population
Gender (at birth)	Male, female	Categorical	Gender may affect the treatment effect	LMM analysis on the ITT population
Region	Capital, Zealand, Southern, Central, Northern	Categorical	Effect may be different in different regions of the country	LMM analysis on the ITT population
Number of previous ESS surgeries	1, 2, 3 or more	Categorical	Patients with a higher number of previous surgeries are less likely to achieve an SSIT-16 score >8.	LMM analysis on the ITT population
Comorbid asthma	Yes/no	Categorical	Marker of higher type-2 inflammatory burden; associated with larger SNOT-22/NPS	LMM analysis on the ITT population

			reductions after biologics	
NSAID-exacerbated respiratory disease (N-ERD)	Yes/no	Categorical	Marker of higher type-2 inflammatory burden; associated with larger SNOT-22/NPS reductions after biologics	LMM analysis on the ITT population
Blood eosinophilic count (*10^9/L)	Low: <150 Moderate: 150-500 High: >500	Categorical	Patients with higher levels of eosinophils in the blood will improve more than patients with lower levels, regardless of treatment.	LMM analysis on the ITT population
Total IgE-levels	<100, 100–500, >500 IU/mL	Categorical	Patients with higher IgE levels will improve more than patients with lower levels, regardless of treatment.	LMM analysis on the ITT population
Sensitivity analyses				
Type	**Outcome**		**Hypotheses**	**Method of analysis**
	Name	**Type**		
Per-protocol analysis	SNOT-22, NPS	Continuous	Results are robust to protocol violations.	LMM analysis of the per protocol population (i.e. no major deviations and possibly including crossover at week 24)
Complete-case analysis	SNOT-22, NPS	Continuous	Results are robust to missing data.	LMM analysis of participants with complete data for primary outcomes; no imputation applied

*All analyses will be controlled for smell score (0-8 or >8), gender (M/F) and region (Capital/Zealand/South/Central/North) as stratification variables at randomization.

EOS: eosinophils; DUP-FIRST: A 48-week treatment strategy where mepolizumab is the initial treatment, but there is a response-guided crossover after 24 weeks of treatment; ITT: Intention to treat; LMM: linear

Mixed Model analysis; MEP-FIRST: As with DUP-FIRST, but with mepolizumab as the initial treatment; NPS: Nasal Polyp Score; SNOT-22: Sino-Nasal Outcome Test; VAS: Visual Analogue Scale.

1 Introduction

This document outlines the proposed presentation and analysis of the findings from the ongoing open label, "head-to-head", multicenter RCT **"TORNADO"**, which directly compares two monoclonal antibodies, mepolizumab (MEPO) and dupilumab (DUPI), on key outcomes in Danish patients with uncontrolled Chronic Rhinosinusitis with Nasal Polyps (CRSwNP). To ease interpretation and appropriately present the full longitudinal design, including the response-guided crossover after 24 weeks, the authors have chosen to deviate from the original protocol and report the results in two separate papers: TORNADO-24 (week 0 to 24) and TORNADO-48 (week 0 to 48). This approach reflects the substantial changes in treatment allocation after week 24 and the volume of data collected. Both papers will follow the strategy outlined in the present document and adhere to the EQUATOR Network's guidelines for statistical analysis plans (SAP)(1). Subsequent exploratory analyses will not be constrained by this strategy but will be carefully considered if proposed by oversight committees, journal editors, or referees, in line with the principles of this plan. Any deviations from the SAP will be documented and justified in the final papers. The analysis will be conducted by the primary investigator in close collaboration with qualified and experienced statisticians. The trial is funded by the five Danish Regions and by a salary grant from a not-for-profit foundation. The trial registration numbers are EU CT 2022-50-22-50-14-00 and ClinicalTrials.gov ID: NCT05942222.

This statistical analysis plan was written after the inclusion period ended, but while follow-up data was still collected. Hence, at the time of writing this SAP, no data review or statistical analysis has been conducted.

2 Background and Rationale

In 2022, the Danish Medicines Council approved two monoclonal antibodies for treating severe CRSwNP, mepolizumab (MEPO) (Nucala, Glaxo-Smith-Kline)) and dupilumab (DUPI) (Dupixent, Sanofi). In Denmark, the treatment is fully subsidized by the state, and consequently, medication costs are a priority when making treatment recommendations. After lengthy discussions and a systematic comparison process, the Danish Medicines Council decided to recommend mepolizumab as the first-line drug (2). Since there were no existing head-to-head trials, they reached this consensus based on an indirect comparison between the two phase 3 trials (SYNAPSE and LIBERTY NP SINUS-52) in which patients with severe CRSwNP were treated for 52 weeks (2–4). The indirect comparison concluded that DUPI caused both a greater reduction in polyp size than MEPO (difference: 1.6 points on the Nasal Polyp Score (NPS)) and a larger improvement in symptom score (7.6 points on the Sino-Nasal Outcome Test 22 (SNOT22) score). However, the difference was only larger than the minimal clinically important difference (MCID) for NPS (the MCID of NPS is 1.0), and not for SNOT-22 (MCID: 12 points). Moreover, since the phase 3 trials differed significantly in several key areas such as statistical methods and baseline characteristics of the study populations (disease severity, co-morbidity, etc.), the Medicine Council found no compelling reasons to recommend one drug over the other. Hence, the least expensive of the two, MEPO, was recommended as the first-line therapy in minimum 80% of patients, although with a response-guided cross-over after 24 weeks if the treatment fails to achieve a sufficient clinical response (1–4). Based on experience from the SINUS LIBERTY-52, it was recommended to increase the inter-dose interval in DUPI patients for whom a successful clinical response is achieved after 24 weeks of treatment.

Lastly, due to a suspected (but not confirmed) superiority of DUPI to MEPO, the Danish Medicines Council suggested that a randomized comparison between the drugs be performed to enable more solid treatment recommendations. Hence, we perform this randomized trial, which will be the first "head-to-head" comparison of DUPI and MEPO for uncontrolled CRSwNP.

Various reviews and network meta-analyses published since the approval process in 2022 generally find that DUPI ranks above MEPO on improvement of central CRSwNP disease outcomes. And although some of these reviews include real-world studies (that generally find even better efficacy than the phase-3 trials), these reviews, too, rely on indirect comparisons of heterogeneous studies (5–12). Several of the real-life studies have confirmed – and even exceeded – the promising results of the phase 3 trials (13–19) but to our knowledge, a head-to-head comparison between DUPI and MEPO is still not available. A recent real-life, observational multicenter study from France likewise found superior treatment effects of DUPI vs MEPO, i.e., significantly more DUPI patients achieved an "excellent" treatment response based on the EUFOREA-EPOS criteria from 2023 (OR 4.03; p=0.01). But this result, too, should be interpreted with caution, especially due to skewed treatment allocation towards DUPI (20).

We will present the TORNADO trial results in two papers with different main objectives. The first paper (in this document referred to as "TORNADO-24") will reflect a "pure" comparison of DUPI vs. MEPO after 24 weeks of treatment. The second paper ("TORNADO-48") aims to investigate which intention-to-treat treatment strategy is superior after 48 weeks; that is, we seek to answer the question: "Which is the best first-line drug, irrespective of whether cross-over took place after 24 weeks."

The trial was proposed by the Danish Medicines Council but was designed and planned - and will be carried out – by the author group, without involvement from either the pharmaceutical industry or legislative organisations. Salary funding for the primary investigator is provided by a neutral, not-for-profit foundation. The remaining part of the trial (medication, tests, clinical staff salary etc.) is funded by the five public "Danish Regions": *The Capital Region of Denmark, Region Zealand, The Region of Southern Denmark, The Central Denmark Region,* and *The North Denmark Region.*

Since MEPO is the first-line treatment in Denmark, it will be treated as the "comparator drug", while DUPI is considered the experimental drug.

2.2 Objectives

Below are the objectives for the separate reports TORNADO-24 and TORNADO-48.

2.2.1 TORNADO-24 objectives:

TORNADO-24 – Primary objective:

To establish if dupilumab 300 mg. every 2 weeks is **non-inferior** to mepolizumab 100 mg. every 4 weeks in patients with severe chronic rhinosinusitis with nasal polyps who are uncontrolled despite previous sinus surgery within the last three years (or contra-indication to surgery) and standard of care treatment, on the co-primary endpoints Sino-Nasal Outcome Test (SNOT-22) and Nasal Polyp Score (NPS) measured after 24 weeks of treatment.

TORNADO-24 – Secondary objectives:

The secondary objectives aim to determine **whether dupilumab is superior to mepolizumab** on the following disease-relevant outcomes after 24 weeks:

1. **Smell improvement**: Change from baseline in Sniffin' Sticks Identification Test 16 (SSIT-16)
2. **Reduction in symptom burden**: change from baseline in SNOT-22
3. **Reduction in polyp size**: Change from baseline in total Nasal Polyp Score (NPS)
4. **Improvement in nasal congestion**: Change from baseline in nasal congestion score (NCS)

5. **Improved asthma control**: Change from baseline in Asthma Control Test (ACT) score

Additional exploratory analyses will include:

- Proportion of patients fulfilling the Danish week-24 response criteria
- Patient-reported improvement (via VAS) of over CRS symptoms, smell, allergic rhinitis, and NERD symptoms
- Patient-reported treatment satisfaction (via VAS)
- Asthma control: Change from baseline in Asthma Control Questionnaire (ACQ-7)
- Proportion of "good to excellent" responders on the EPOS/EUFOREA week 24 criteria (21)
- Normalisation of middle ear pressure (via tympanometry)
- Reduction in fractional exhaled nitric oxide (FeNO)

2.2.2 TORNADO-48 objectives:

TORNADO-48 – Primary objective:

To determine whether a treatment strategy with dupilumab as the initial drug (**"DUP-FIRST"**) is non-inferior to a strategy with mepolizumab as the initial drug (**"MEP-FIRST"**) in improving symptom burden and polyp size at 48 weeks.

The **DUP-FIRST strategy** involves dupilumab 300 mg every two weeks followed by:

- Either a response-guided taper to 300 mg every four weeks at week 24 (if clinical response is achieved),
- Or crossover to mepolizumab (100 mg every four weeks) in case of insufficient response (figure 1).

The **MEP-FIRST strategy** involves mepolizumab 100 mg every four weeks with the same response-guided options at week 24 (but no tapering).

Non-inferiority will be assessed using the co-primary outcomes:

- Sino-Nasal Outcome Test-22 (SNOT-22)
- Nasal Polyp Score (NPS),

both measured at week 48.

TORNADO-48 – Secondary objectives:

To determine if the **DUP-FIRST strategy is superior to the MEP-FIRST strategy** after 48 weeks of treatment on the following disease-relevant outcomes:

1. **Smell improvement**: Change from baseline in SSIT-16
2. **Reduction in symptom burden**: change from baseline in SNOT-22
3. **Reduction in polyp size**: Change from baseline in NPS
4. **Nasal congestion improvement**: Change from baseline in NCS
5. **Asthma control**: Change from baseline in Asthma Control Test (ACT) score

Furthermore, we will conduct exploratory analyses on several outcomes, including:

- Proportion of patients fulfilling the Danish week-48 response criteria
- Improvement in Lund-Mackay CT score
- Patient-reported improvement of smell, allergic rhinitis and NERD symptoms (via VAS).
- Patient-reported treatment satisfaction (via VAS)

- Asthma control: Change from baseline in ACQ-7
- Proportion of "good to excellent" responders on the EPOS/EUFOREA week 48 criteria (21)
- Normalisation of middle ear pressure (via tympanometry)
- Reduction in fractional exhaled nitric oxide (FeNO)

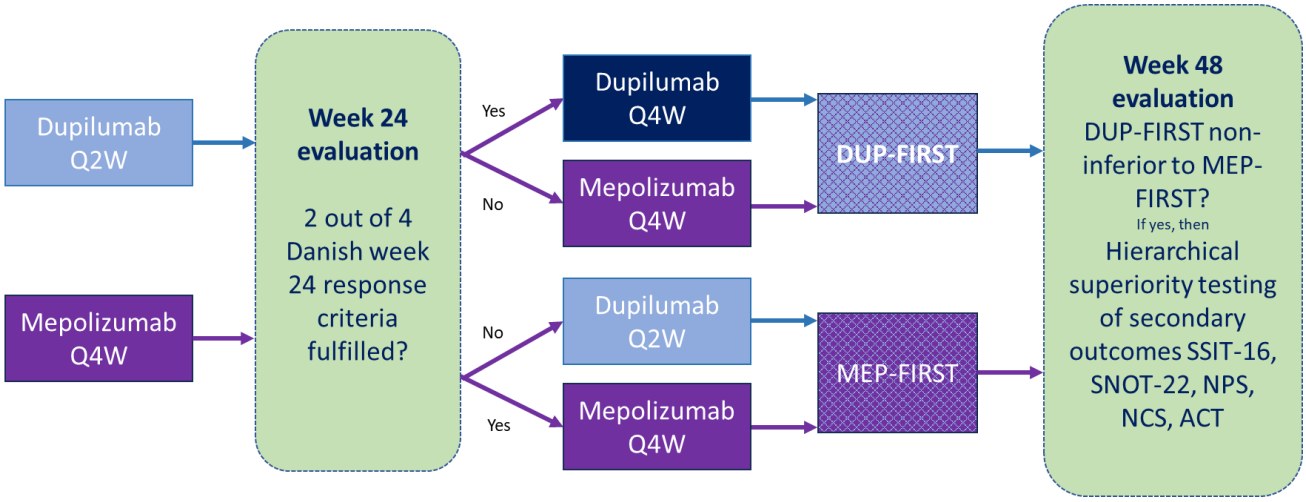


Figure 1 – TORNADO-48 design: DUP-FIRST/MEP-FIRST: dupilumab/mepolizumab as the initial treatment with possible response-guided crossover after 24 weeks; **Q2W/Q4W:** Inter-dose-interval 2 or 4 weeks; **SSIT-16:** Sniffin' Sticks Identification Test 16; **SNOT-22:** Sino-nasal Outcome Test 22 items, **NPS:** Nasal polyp score, **NCS:** Nasal Congestion Score, **ACT:** Asthma Control test

3. Study Methods

3.1 Overall design

This is a pragmatic, randomised, multicenter, open label, "head-to-head" trial comparing dupilumab (DUPI) to mepolizumab (MEPO) with adaptive crossover and tapering of DUPI responders at week 24 in patients with severe chronic rhinosinusitis with nasal polyps (CRSwNP) who are uncontrolled despite standard of care treatment and endoscopic sinus surgery (ESS) within three years (or contraindication to surgery).

The study population consists of approximately 212 participants from 9 tertiary centres in Denmark. Participants were randomised 1:1 to receive either DUPI or MEPO in standard doses, i.e., dupilumab 300 mg subcutaneously every 2 weeks (Q2W) or mepolizumab 100 mg every 4 weeks (Q4W). Participants were stratified by region, gender, and smell function. This pragmatic trial follows the strategy proposed by the Danish Medicines Council, i.e., participants treated with DUPI who are responsive at week 24 will be tapered to injections every 4 weeks (Q4W), and participants in either arm who are unresponsive at week 24 will crossover(2).

All patients screened for possible treatment with a biologic in Denmark agree to have their data recorded in a nationwide electronic registry (REDCap). At the screening visit, the following data are collected: basic demographic data, previous treatments for CRSwNP and asthma (including surgery, topical treatment, and systemic steroids), patient-reported outcomes via a QR-code link (SNOT-22, ACT, NCS, VAS-scores, etc., see protocol), adherence to topical treatment, etc. Furthermore, all results of clinical tests performed by nurses or doctors are recorded (NPS, SSIT-16, FEV1, CT scans, blood samples, etc.). In the end, a decision is made whether the patient fulfills the criteria for treatment with a biologic (see section 5.2.1). Patients who

qualified for treatment in the enrolment period (and did not have any exclusion criteria, section 5.2.3) were offered to enrol and, hence, be randomized 1:1 to DUPI or MEPO. If they chose not to enrol, they would be treated outside of protocol. They were also told that the follow-up programme and their option of crossing over after 24 weeks in case of no response were the same regardless of whether they enrolled in the trial.

The trial has five planned visits (Figure 2): A baseline visit (w.0) where treatment is initiated, a blood sample and self-administration training visit (week 4 +/- 7 days), a safety visit (week 16 +/- 7 days), a response evaluation visit (week 24 +/- 7 days) and a final response evaluation visit (week 48 visit +/- 7 days). At the visits in w.0, w.24 and w. 48., subjects will complete patient-reported outcome measures (PROMs) and undergo clinical examinations of lower- and upper-airway disease (lung function test, smell test, nasal endoscopy, etc.). At the w. 16 visit only PROM data and safety data will be collected. If the subject crosses over at the week 24 evaluation (see later), an additional safety visit is added 16 weeks after starting the new treatment.

The **visit in week 4** has two objectives: to check for possible drug-induced hyper eosinophilia (a rare side-effect of DUPI treatment (protocol section 9.2.1)), and to train the subject in self-administering the injections (up to that point, the medication has been administered by clinical staff).

The **visit in week 16** is primarily a safety visit, i.e., possible AEs/SAEs are recorded and assessed, but it also contains certain PROMS (SNOT-22, ACT, VAS scores, ACQ-6 (i.e., excluding FEV1 measurement).

The focus of the major **halfway visit in week 24** is treatment response. If the treatment response criteria are met (protocol table 2) the subject continues with the same treatment. If that treatment is DUPI, the inter-dose interval is prolonged from injections every second week (Q2W) to every four weeks (Q4W). This procedure mimics the recommendations written in the Danish treatment criteria by the Danish Medicine Council (22). If the response criteria are not fulfilled, the subjects cross over to the opposite drug (see Figure 1).

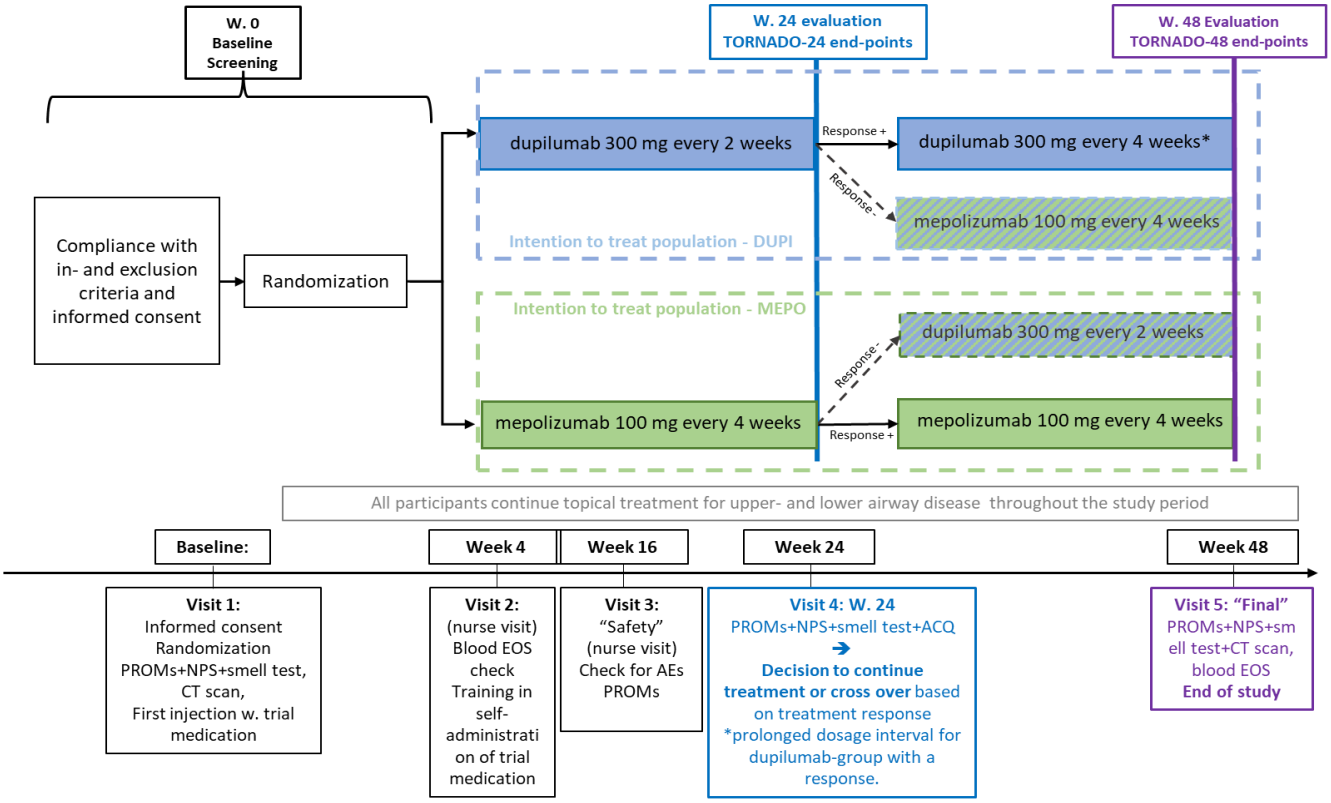


Figure 2: Overall trial design, TORNADO.

3.2 Randomization

Eligible patients were randomized using a secure, central, web-based randomization system developed and hosted within REDCap (Yale University). An allocation table was uploaded by the authors. The allocation table contains more than four times the number of expected participants to ensure equal distribution. We applied variable block randomisation in blocks of 4-8 subjects. Users of the system had no insight into the next allocation (access to the randomization list is blocked by the software). Following allocation, the specific participant ID is locked against further randomization attempts, and all randomization outcomes are saved in a digital log. The implementation of the randomization procedure followed a validation procedure under monitoring from the Good Clinical Practice (GCP) Unit in Copenhagen, Denmark. Randomization occurred in a 1:1 proportion at inclusion and subjects were stratified according to gender, smell olfactory test score, and geographical location.

3.3 Sample Size

This study was designed to be conducted within a time frame of 2-3 years, yielding a realistic sample size of approximately 220 patients. This estimate was based on a Danish registry study, which projected that 120 new patients per year would qualify for treatment with monoclonal antibodies (23). Based on these numbers, we expected to enrol 176 patients without dropouts and 220 patients, accounting for potential dropouts within the study timeframe.

3.3.1 Power for the co-primary objective: Non-inferiority

The abovementioned participant number provides the study a power of >95% to show non-inferiority of dupilumab to mepolizumab for both co-primary endpoints with the following criteria:

- About 88 participants in each arm
- Level of significance for both endpoints of a one-sided test, $p<0.025$ and
- including previously found standard deviation (SD) values 1.9 for NPS and 22 for SNOT-22
- an expected superior effect of 0.7 for NPS, and 7 for SNOT-22 and
- a minimal clinically relevant difference (MCID) of 1 for NPS and 12 for SNOT-22, and
- assuming no negative correlation between the two co-primary endpoints, respectively

3.3.2 Power for proving superiority equaling the *minimal clinically important difference* values of the key secondary objectives

Smell: The study is powered >95% for proving MCID-equivalent superiority on smell improvement of DUPI versus MEPO with the following criteria:

- A hybrid-estimated SD value of 1.85 for Sniffin' Sticks 16 identification test (based on a hybrid of SD's for different age groups in a large normative value study (Oleszkiewicz et al (24)
- An expected superior improvement of 2 points on the Sniffin Sticks 16 Identification test (this is a estimate calculated from the difference of 5 points on the University of Pennsylvania Smell Identification Test that was found in a Network Meta-analysis by Oykhman et al (10). A reduction of 5 points out of 40 equals 1/8 of the total interval, which again equals 2 points out of 16 on the SSIT-16).
- An MCID of 3 for SIT16 (25)

SNOT-22: The study has about 56% power to show superiority of DUPI on SNOT-22 based on the criteria in 3.3.1.

NPS: The study has about 69% power to show superiority of DUPI on NPS based on the criteria in 3.3.1.

3.4 Framework

The analyses in both papers—TORNADO-24 and TORNADO-48—will be conducted in a sequential manner. First, a **non-inferiority** test assesses whether DUPI provides a statistically non-inferior improvement on the co-primary endpoints NPS and SNOT-22 compared to MEPO on the predefined non-inferiority margins (see section 6.3.1)

If non-inferiority is confirmed, we will perform **superiority analysis**. The sequence of the hierarchical tests in TORNADO-24 and TORNADO-48 is described in section 6.3.2.

If statistical superiority is not achieved at any step in the sequence, subsequent superiority tests will be considered exploratory and not used for formal inferential conclusions. Each test in the hierarchy must achieve statistical significance at the 5% level before moving to the next. If statistical superiority is achieved, we will discuss in the publication whether the improvement is clinically relevant or not.

The rationale is to establish that DUPI is at least as effective as MEPO (i.e., non-inferior) for the co-primary endpoints, given the clinical importance of ensuring no loss of efficacy. If non-inferiority is established, we then test superiority to determine whether DUPI may offer additional benefits for smell improvement, symptom burden, polyp size, and comorbid asthma. This pre-specified hierarchical testing controls the overall Type I error rate that is otherwise associated with multiple testing.

3.5 Interim analysis and stopping guidance.

An interim analysis will not be performed. TORNADO-48 data will be analysed following the strategy in this document, regardless of the findings in TORNADO-24.

The sponsor reserves the right to terminate the trial if safety concerns emerge.

If the trial is prematurely terminated or suspended, the investigator will promptly inform enrolled subjects and ensure appropriate therapy and follow-up. Furthermore, the investigator and/or sponsor will promptly inform the pertinent ethics committee and regulatory authorities.

3.6 Timing of final analysis

We will analyse data in two stages: after 24 weeks of follow-up (TORNADO-24) and after 48 weeks of follow-up (TORNADO-48). Statistical analyses will be undertaken only after recruitment has ended and all follow-up data have been collected. TORNADO-24 data will be analysed after the last patient's last 24-week visit (expected in early July 2025) and after all patients have had their 24-week data monitored by the Good Practice Unit. TORNADO-48 analysis will be performed after the last patient's last 48-week visit and finalisation by the GCP unit.

3.7 Timing of the outcome assessments

See section 10.1 of the protocol for a detailed description of outcome timings and visit windows.

4. Statistical Principles

4.1 Confidence intervals and P values

Non-inferiority analyses TORNADO-24 and TORNADO-48: non-inferiority will be claimed if the null hypothesis is rejected with an alpha level of 0.025 (one-sided) for each of the two co-primary endpoints in the non-inferiority analysis.

Superiority analyses, TORNADO-24 and TORNADO-48: The significance level for the superiority analysis is p=0.05. Results will be reported with a p-value and 95% confidence intervals where applicable. Statistical

significance will be claimed for p-values <0.05 or when not including 0 or 1 on a relative or absolute scale, respectively.

If a test fails to confirm a given hypothesis all subsequent tests will be regarded as explorative. Explorative tests will still be presented with 95%CI and p-values, but it will be made clear when the tests switch from superiority to explorative.

4.2 Protocol deviations

Only major protocol deviations will be presented in the final reports. The definition of major protocol deviation is whether they impact completeness, accuracy, and/or reliability of the study data, or if the rights, safety, or well-being of a participant is threatened. **Major protocol** deviations in this trial will be placed in one of the following categories:

- Eligibility violations:
  - Participants who did not fulfil key eligibility criteria but were included in the study.
- Treatment deviations:
  - <80% dosage of the assigned IMP: either due to poor compliance, violated dosage interval windows, wrong injection technique, incorrect dose or other problems that lead to underdosing of the IMP.
  - Need for major rescue treatment (surgery and/or systemic corticosteroids)
  - use of disallowed medications that could affect primary outcomes (other monoclonal antibodies/type-2 specific medication or systemic corticosteroids)
- Discontinuation deviations:
  - Participants who met discontinuation criteria but were not withdrawn from the study (examples are pregnancy, rescue treatment)
- Study visits deviations:
  - Week 24 visit window exceeded more than the maximum allowed (i.e., +/- 6 weeks)
  - Week 48 visit window exceeded more than the maximum allowed (i.e., +/- 8 weeks)

4.3 Analysis-populations

Screening population: will consist of patients who were assessed at one of Denmark's nine "Global airway" clinics, i.e., public hospital ENT clinics that treat patients with severe CRSwNP with biologics and enter data in the national "Global Airways" REDCap database. The total number in the screening population will be shown on the CONSORT diagram in the final manuscripts (Figure 2a and Figure 2b) but will not otherwise be described.

Intention-to-Treat (ITT) Population: The ITT set will consist of all patients who fulfilled all inclusion criteria and no exclusion criteria and were randomised to a study treatment (see figures 2a and 2b). For easier reference, the ITT population in TORNADO-24 and TORNADO-48 will be named ITT-24 and ITT-48, respectively. The group initially treated with dupilumab from baseline will be referred to as **DUP-FIRST,** and similarly for the **MEP-FIRST** group.

Per-Protocol (PP) Population: The Per-Protocol population (PP) will consist of all randomised participants who 1) receive only the assigned intervention, i.e. no external interventions that distort efficacy signals (such as OCS or sinus surgery, regardless if this was done as rescue treatment), 2) remain compliant with the assigned treatment (i.e. ≥80% of planned doses) 3) attend the scheduled visits within the allowable windows (see Table 2)) and 4) avoid major protocol deviations (see Section 4.2). Participants will be analysed according to the treatment they were originally assigned (with/without response-guided crossover) for PP analyses in TORNADO-24 and TORNADO-48. For shorthand, these populations will be referred to as PP-24 and PP-48.

As-treated (AT) populations: The AT population will describe the four possible treatment trajectories at week 48, i.e., containing response-guided cross-over. The groups are:

- DUPDUP: Received dupilumab throughout
- DUPMEP: Started on dupilumab, switched to mepolizumab
- MEPMEP: Received mepolizumab throughout
- MEPDUP: Started on mepolizumab, switched to dupilumab

Complete-case analysis (CCA): includes only participants with complete data on all variables required for a given analysis.

5.0 Trial population

5.1 Screening data

Screening data will not be presented separately. In Denmark, all potentially eligible patients for biologic treatment of CRSwNP are referred to one of the nine participating ENT clinics. Those who meet the inclusion criteria and none of the exclusion criteria are offered participation in TORNADO. Since TORNADO follows the same treatment schedule as standard care, most patients choose to participate.

5.2 Eligibility

TORNADO-24: Eligible participants are patients who have been referred to one of the study sites to be evaluated for possible treatment with biologics for CRSwNP. The inclusion criteria are equal to the Danish treatment criteria for treating CRSwNP with biologics and are shown below along with the exclusion criteria. Of special note is "eosinophilic blood cell counts of ≥1.5x109cells/L (at baseline, i.e., before first injection)" which was added as an amendment due to increased attention of the risk of secondary hyper-eosinophilia and possible hyper-eosinophilic syndrome caused by DUPI treatment (see section 9.2.1 of the protocol).

TORNADO-48: The ITT population from TORNADO-24 that has not withdrawn their consent or been discontinued.

5.2.1 TORNADO-24 *Inclusion criteria:*

Patients must have:

- Bilateral polyps in nose and sinuses
- ESS within the previous three years (unless unfit for surgery[1])
- Adherence to local treatment with saline irrigation and topical nasal steroids for at least three months (unless contraindicated)
- Evidence of type 2 inflammation (as per EPOS2020(26))

Furthermore, patients must fulfil at least three out of the following five criteria:

[1] in this study defined as either a severe somatic disease, for which other specialist advise against surgery, e.g., cardiac disease, pulmonary disease, coagulation disorder or similar OR/AND severe anxiety which can either be due to previous traumatic experiences with surgery or the postoperative period, post-traumatic stress disorder, severe anxiety disorder or similar. In cases of doubt, investigators can ask for a written statement from the general practitioner or a psychiatrist/psychologist).

- Need for systemic corticosteroids (at least two courses in previous year OR long-term treatment >3 months) or contraindication to systemic steroids
- Significantly impaired quality of life (SNOT-22 score≥50)
- Significant loss of smell (SSIT-16 score 0-8)
- Large nasal polyps (NPS ≥5, with at least a score of 2 on either side)
- Asthma diagnosis (requiring inhaled corticosteroids)

5.2.2 TORNADO-24 Exclusion criteria:

- Systemic corticosteroid treatment within the last three months
- Endoscopic sinus surgery (ESS) within the last six months
- Hypersensitivity to the active substance or any of the excipients in the two IMPs
- Not able to understand spoken and/or written Danish
- Treatment with any biologic drug aimed at type II disease within the previous six months
- Prior treatment failure with one of the two IMPs for any indication (treatment failure is defined as failure to achieve the desired therapeutic outcome or effectively manage a condition within an expected timeframe)
- Eosinophilic blood cell count of ≥1.5x109cells/L (at baseline, i.e., before first injection)
- Pronounced fear of needles
- Pregnant or breastfeeding patients

5.2.3 TORNADO-48: Inclusion criteria:

- ITT population from TORNADO-24

5.2.4 TORNADO-48: Exclusion criteria:

- Withdrawal of consent after the week 24 evaluation

5.3 Recruitment

The flow of participants through the study is displayed in the CONSORT diagrams below.

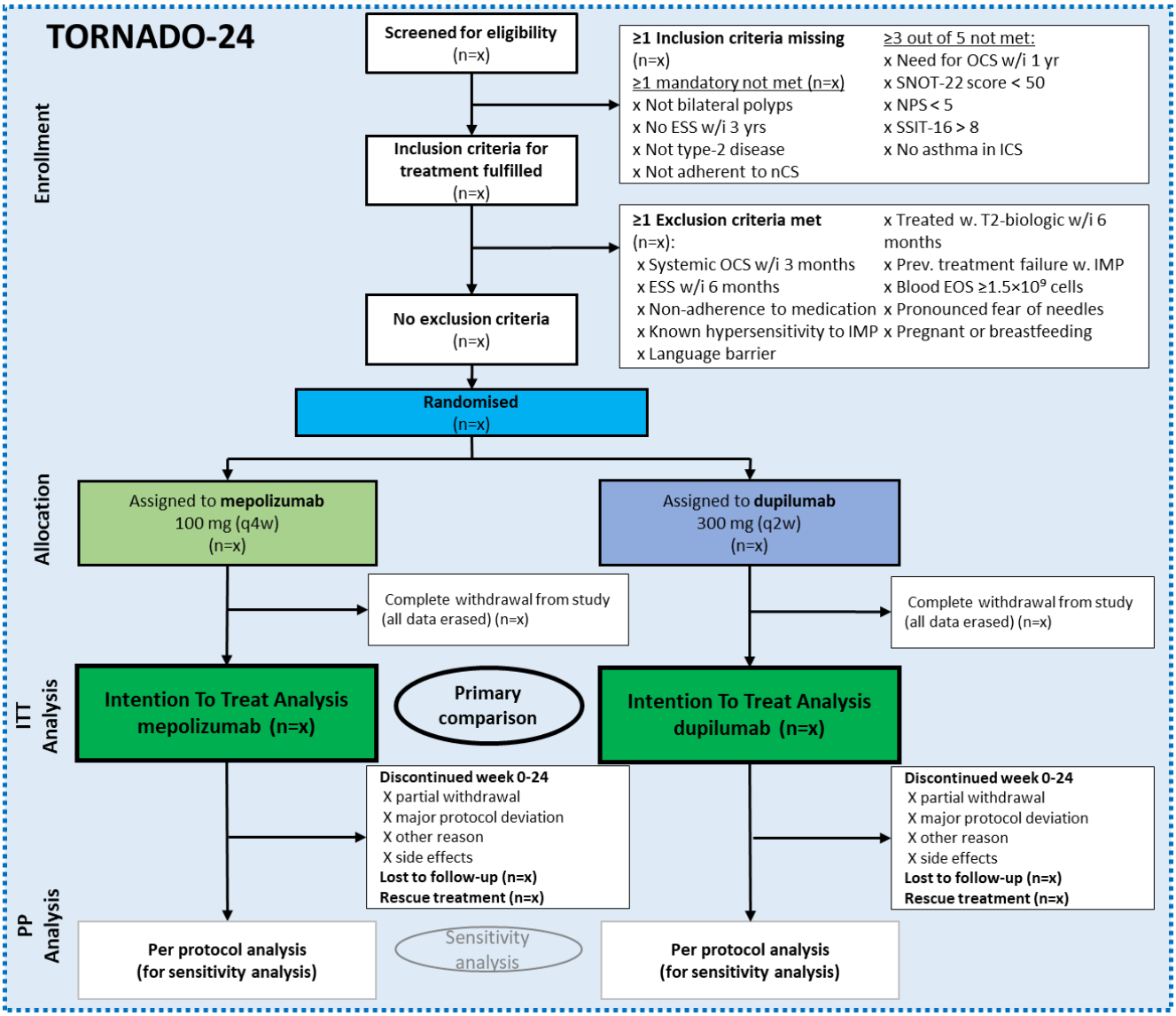


Figure 2a – CONSORT diagram 1 TORNADO-24

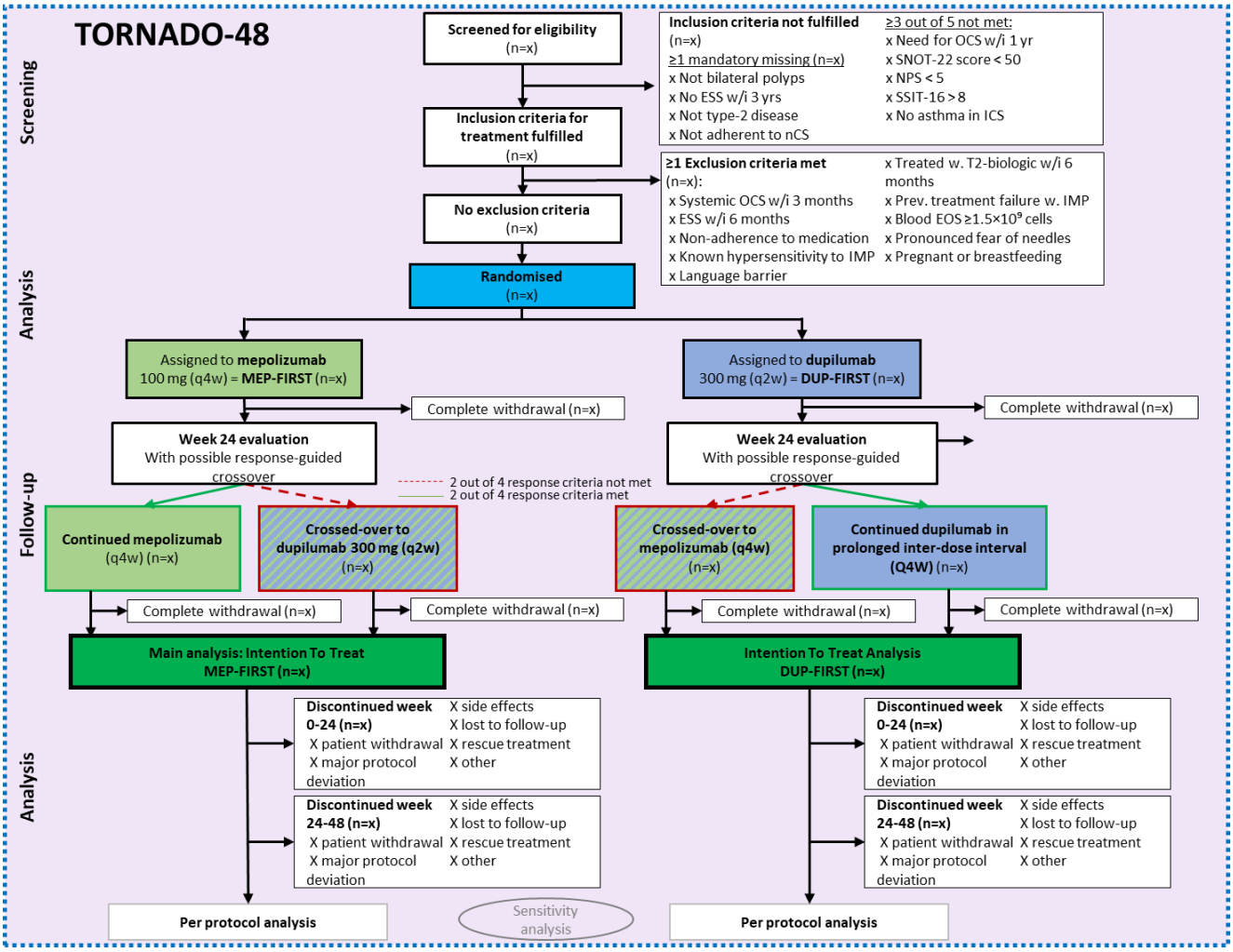


Figure 2b – CONSORT diagram TORNADO-48

5.4 Withdrawal/follow-up

Level of withdrawal:

- Partial Withdrawal: Participants discontinue the study drug but continue to show up for scheduled follow-up assessments.
- Withdrawal from the study: the participant withdraws from taking the study drug and does not come to future follow-up visits, but any available data will still be used in the analysis.
- Complete Withdrawal: Participants may withdraw entirely from the study, meaning that no further data collection or follow-up is conducted, and all their previously collected data is erased.

Timing of Withdrawal/Lost to Follow-up Data:

The date on which a participant discontinues intervention or withdraws from the study will be noted in the case report forms (CRFs) and/or in a note to file send to the sponsor. A participant will be classified as "lost to follow-up" if they cannot be contacted or do not return for scheduled visits despite documented attempts. The date of last contact will be recorded.

Summaries of withdrawals and loss to follow-up will be generated for each scheduled visit (e.g., Baseline, Week 24, final visit).

Reasons and details of how withdrawal/lost to follow-up data will be presented:

All instances of participant withdrawal or loss to follow-up will be presented in a CONSORT flow diagram to illustrate the progression of participants through the study. If no reason is provided by the participant, the withdrawal will be documented as "Unknown" in the dataset.

5.5 Baseline Patient Characteristics

The descriptive summaries of baseline characteristics will be presented in a baseline table that will differ between TORNADO-24 and TORNADO-48 (see MOCK-UP Tables 1a and 1b). Since the study involves two treatment arms, a column will be provided for each treatment group (i.e., the drug they were first assigned to), alongside the total summary for all participants. We will perform simple descriptive statistics to check for equal distribution between treatment groups, but comparative analyses will not be done.

Continuous Variables

Normally distributed variables (based on visual inspection of histograms or descriptive statistics) will be summarized by mean and standard deviation (SD). If variables appear skewed, they may be summarized by median and interquartile range (IQR) instead.

Categorical Variables

Summarized by frequency counts (n) and percentages. Categories will be clearly defined (e.g., gender: m/f, comorbidity status as present vs. absent etc.).

For ordinal measures (e.g., certain symptom severity scales), if used as categorical data, they will be displayed as counts per category.

6.0 Outcomes and timings

This section describes outcomes and timings.

6.1 Outcomes

The various outcomes are explained in detail in the protocol section 10.11 and 10.12.

Table 1 contains information about each endpoint that will be included in the analyses along with their timing.

Table 1: Outcomes used in TORNADO-24 and TORNADO-48: Explanations and timing

Outcome	Description	Scale	Outcome definition
KEY OUTCOMES			
Sino-Nasal Outcome Test 22 (SNOT-22) Visits: w0+w16+w24+w48 (Co-primary endpoint in non-inferiority analysis) (Also #2 in superiority test hierarchy)	A 22-item questionnaire covering five symptom domains: nasal, ear/face, sleep, function, and emotion. Likert responses, ranging from no problem (=0) to problem as bad as it can be (=5). The recall period is 4 weeks.	The total score ranges from 0 (no symptoms) to 110 (maximum symptom burden).	Change scores from baseline The MCID for medically managed CRSwNP is 12.0 points(27).
Nasal polyp size (NPS) Visits: w0+w24+w48 (Co-primary endpoint in non-inferiority analysis) (Also #3 in superiority test hierarchy)	Scored with the Nasal Polyp Score (NPS)	Ranges from 0 = no nasal polyps to 4 = polyps reaching the floor of the nasal cavity. Each side is scored 0-4, total score range: 0-8. In this study we will grade polyps protruding from the middle meatus or medial to the middle turbinate as NPS=3.	Sum change scores (left + right nostril) from baseline. The MCID in this study was set to ≥1 with reference to the Danish Medicines Council (22)
Sniffin' Sticks Identification Test 16 (SSIT-16) Visits: w0+w24+w48 (#1 in superiority test hierarchy)	Measures olfactory identification ability. It is a forced choice test including 16 everyday-life odors with four alternatives for each odor.	Range: 0-16, higher score indicating better olfactory ability. Normosmia: 15-16 Hyposmia: 9-14 Anosmia: 0-8	Change scores from baseline. The MCID is 3.0 (25)

Nasal Congestion Score (NCS) Visits: w0+w16+w24+w48 (#4 in superiority test hierarchy)	Self-rated score of nasal congestion	Ranges from 0 (no congestion) to 3 (completely congested)	Change scores from baseline. The MCID is 0.5 (28)
Asthma Control Test (ACT) Visits: w0+w16+w24+w48 (#5 in superiority test hierarchy)	A patient self-administered tool for quantification of asthma control.	5 items on symptoms and daily functioning, with 4-week recall period. Item scores ranging from 1-5 (1 being worst) and sum score range from 5 (poor control) to 25 (complete control of asthma).	Change scores from baseline. The MCID is 3 points between groups and over time (29)
OTHER OUTCOMES			
Response Criteria w. 24 (TORNADO 24) Visit: w24	The response criteria set by the Danish Medicines Council. To continue treatment at week 24, 2 out of the 4 needs to be fulfilled.	Improvement of: ≥1 points on NPS, ≥12 points on SNOT-22, ≥0.5 points on ACQ-7 and/or improvement of smell from anosmia to at least hyposmia (SSIT-16 improved from 0-8, to 9 or more)	Proportion of participants fulfilling at least 2 criteria (regardless of which) at week 24
Response Criteria w. 48 (TORNADO 48) Visit: w48	The response criteria set by the Danish Medicines Council. To continue treatment at week 48, 2 out of the 4 needs to be fulfilled.	Improvement of: ≥2 points on NPS, ≥12 points on SNOT-22, ≥0.5 points on ACQ-7 and/or improvement of smell from anosmia to at least hyposmia (SSIT-16 improved from 0-8, to 9 or more)	Proportion of participants fulfilling at least 2 criteria (regardless of which) at week 48.
EPOS/EUFOREA response criteria	The response criteria set by the EPOS/EUFOREA group(21)	Based on 5 domains—polyp size, need for rescue treatment, SNOT-22, smell, and comorbidities. Patients are classified as no to poor, moderate, or good to excellent responders	Proportion of participants achieving an “good to excellent” response (4-5 points out of 5 possible) at week 24/48

		depending on how many domains show clinically meaningful improvement	
Asthma Control Questionnaire (ACQ-7) Visits: w0+w16+w24+w48	A measure of asthma control status via five subjective, patient-reported items and two objective items (medical use and lung function (FEV1))	Scores ranging from 0 (no impairment) to 6 (extreme impairment), using the past seven days as a recall period. A total score of ≤1.2 points indicates well-controlled asthma.	Change scores will be calculated as follow-up minus baseline score. The MCID is 0.5 point(22)
CRS-VAS Visits: w0+w16+w24+w48	Self-rated CRS severity	VAS-score of 0 to 100, 100 indicating the worst possible disease severity.	Change scores will be calculated as follow-up minus baseline score. In this study, the MCID was set to 15 mm
Smell-VAS Visits: w0+w16+w24+w48	A self-rated score of olfactory ability	VAS-score of 0 to 100, 100 indicating worst self-rated olfactory ability.	Change scores will be calculated as follow-up minus baseline scores. In this study, the MCID was set to 30 mm(30)
Asthma-VAS Visits: w0+w16+w24+w48	Self-rated asthma severity	VAS-score of 0 to 100, 100 indicating the worst possible disease severity.	Change scores will be calculated as follow-up minus baseline scores. In this study, the MCID was set to 15 mm
Allergic rhinitis-VAS Visits: w0+w16+w24+w48	Self-rated allergic rhinitis severity	VAS-score of 0 to 100, 100 indicating the worst possible disease severity.	Change scores will be calculated as follow-up minus baseline scores. In this study, the MCID was set to 15 mm
NERD-VAS Visits: w0+w16+w24+w48	Self-rated severity of sensitivity to NSAIDs	VAS-score of 0 to 100, 100 indicating the worst possible disease severity.	Change scores will be calculated as follow-up minus baseline scores. In this study, the MCID was set to 15 mm
Tympanometry Visits: w0+w24+w48	Middle ear pressure measured by tympanometry	A-curve = normal tympanic membrane mobility, B-curve = flat curve indicating either middle ear fluid or perforation of the TM, C-curve = negative middle ear pressure indicating eustachian tube dysfunction	Proportion with Eustachian tube dysfunction (i.e., B- or C-curves) at baseline, week 24, and week 48.

Need for rescue treatment Visits: continuously	Proportion requiring rescue treatment, i.e., systemic corticosteroids/antibiotics or sinus surgery.	Percentage of cohort	Proportion calculated at 24 and 48 weeks.
AEs/SAEs leading to treatment cessation. Continuously	Proportion of subjects with AEs/SAEs leading to treatment cessation	Percentage of cohort	Proportion calculated at 24 and 48 weeks.
Lund-Mackay Score Visits: w0+w48	A CT-based measure of paranasal sinus opacification.	Each of 10 sinuses are scored from 0-2 (2 meaning complete opacification). The ostio-meatal complex is scored as either "open" (score=0) or "obstructed" (score=2)	Change scores as follow-up minus baseline scores.
Lung fractional exhaled nitrous oxide (FeNO) Visits: w0+w24+w48	FeNO is a marker associated with eosinophilic airway inflammation. It is measured in parts per billion (ppb)	Levels above 25 ppb is indicative of eosinophilic airway inflammation	Change scores will be calculated as follow-up minus baseline scores.

6.2 Trial visit windows

Table 2 shows visit names, their target day, and the maximum allowed window for the visit.

Table 2: Trial visits

Time point	Target day	Visit window[a]
Baseline/screening:	0	0
First injection (max. 4 weeks from baseline)	0	0-28
Week 16 visit (+/-2 weeks)	112	84-140
Week 24 visit (+/-6 weeks)	168	126-210
Week 48 visit (+/-8 weeks)	336	280-392

a: Visits that exceed this interval will be regarded as "missing" – see section 6.4

b: See table 1 for explanations of each endpoint

7.0 TORNADO-24 analyses

A hierarchical (sequential) testing process will be used for the analyses in TORNADO-24. For both co-primary outcomes, non-inferiority will be explicitly tested first. Only if non-inferiority is proven for both outcomes will superiority analyses be carried out. Any superiority analyses that are conducted after non-inferiority has been established will be regarded as exploratory. 95%-confidence intervals and p-values will still be presented.

7.1 Primary analysis (non-inferiority tests) – TORNADO-24

For the primary analysis of the co-primary outcome SNOT-22, we will apply a linear mixed-effects model (LMM) to the ITT population from TORNADO-24. The model will include fixed effects for treatment group, visit number (t=0, t=16 weeks, t=24 weeks), their interaction, and the stratification variables gender, region, and baseline smell score. A random intercept for participant ID will account for within-subject correlation. Residual plots and model diagnostics will be assessed for normality of residuals, homoscedasticity, and influential points.

The same method will be applied to the co-primary outcome **NPS**, except this model will not contain data from week 16 (as it is not measured at this time point).

Non-inferiority will be concluded only if the lower bound of the 95% confidence intervals for both outcomes SNOT-22 and NPS lies above the non-inferiority margins of -12 points and -1 points (=MCID), using a one-sided $\alpha = 0.025$.

If the conditions are not met for **both** outcomes, non-inferiority will not be declared, and any subsequent superiority analyses will be considered exploratory.

7.2 Secondary analyses (superiority analyses) – TORNADO-24

A linear mixed-effects model will be used for superiority analyses. This model will include fixed effects for treatment group, visit, and their interaction, as well as the stratification variables gender, region, and baseline smell score. To take into consideration within-subject correlation across time, a random intercept for subjects will be incorporated. To control the family-wise Type I error rate at an overall $\alpha = 0.05$, we will employ hierarchical (sequence) testing. The following predetermined order will be used to determine superiority:

1. Smell Identification Test (SSIT-16)
2. Sino-Nasal Outcome Test 22 (SNOT-22)
3. Nasal Polyp Score (NPS)
4. Nasal Congestion Score (NCS)
5. Asthma Control Test (ACT)

We have placed smell improvement highest in the hierarchy since loss of smell is repeatedly reported as the most bothersome symptom of CRSwNP(26,31,32).

Statistical significance will be claimed if the null hypothesis is rejected at the alpha level of 0.05 (two-sided).

7.3 Additional analyses in TORNADO-24

7.3.1 Explorative analyses - TORNADO-24

The planned exploratory analyses in TORNADO-24 are:

- Proportion not meeting Danish response criteria
- Smell-VAS
- Allergic rhinitis-VAS
- Patient reported treatment satisfaction (on a VAS score)
- ACQ-7
- NSAID-exacerbated respiratory disease (NERD-VAS)
- Tympanometry improvement (proportion with normalisation, i.e., from B-curve to A-curve or C-curve to A-curve)
- EPOS/EUFOREA week 24 criteria – proportion "good to excellent" responders (4-5 of the following: reduced polyp size ≥1; reduced need for OCS/salvage surgery; improved SNOT-22 score <40 + >MCID; improved sense of smell (hyposmic); reduced impact of comorbidities (≥MCID – in this setting we chose MCID for ACQ=≥0.5)
- Reduction in fractional exhaled nitric oxide (FeNO)

Exploratory results will be interpreted with caution due to the potential for inflated type I error from multiple testing. No formal adjustments for multiple comparisons will be applied unless specifically indicated.

7.3.2 Safety analysis – TORNADO-24

We will present the rates of AEs, SAEs, and the need for major rescue treatment (surgery or OCS) in descriptive tables for both drugs individually. Only AEs occurring in more than 5% will be presented.

7.3.3 Sensitivity analyses, TORNADO-24

To assess the robustness of our findings, we will conduct the following sensitivity analyses for the co-primary and secondary outcome:

- Per-protocol analysis, including only patients without major protocol violations, to assess the effect of adherence and compliance.
- Complete-case analysis and multiple imputation to evaluate the impact of missing data assumptions on the co-primary outcomes (will only be performed in case of >5% missing data)

The results of these analyses will be compared to the primary (intention-to-treat) analysis to gauge consistency and identify potential bias due to protocol deviations or missing data mechanisms.

7.3.4 Subgroup analyses – TORNADO-24

Subgroup analyses will explore whether the treatment effect on various outcomes differs according to various baseline characteristics. The outcomes tested will be SNOT-22, NPS, and SSIT-16. The baseline subgroups of interest will be age group (<40, 40–65, >65), sex assigned at birth, region, number of previous ESS surgeries (1, 2, ≥3), asthma diagnosis, blood eosinophil count (<150, 150–500, >500 cells/µL), aspirin-exacerbated respiratory disease (yes/no), and total IgE levels (<100, 100–500, >500 IU/mL). Each subgroup will be evaluated separately by including an interaction term between treatment and subgroup level in the linear mixed model. These analyses are exploratory and will not be used to support formal claims of heterogeneity. We expect to present the results in a Forest plot (see example in the Mock-up Figure 3). We will, however, refrain from this if no consistent pattern of subgroup modification is observed.

7.3.5 Dropout characteristics - TORNADO-24

In the case of five or more dropouts in an arm (=consent withdrawn) we will produce supplementary tables with baseline characteristics of dropouts in an identical way as Mock-up table 1A (Section 8, "supplementary").

7.4 Missing data and imputation – TORNADO-24

We expect low proportions of missing values, i.e., less than 10% for exploratory and tertiary outcomes and less than 5% for primary and major secondary outcomes. Rather than treatment-related adverse events or dropouts, missing data will mostly result from logistical reasons like incomplete registrations or missed appointments/violated visit windows (see section 6.2). As a result, we assume that the data are **missing at random (MAR)**, and the principal analyses will be based on this assumption using **LMMs** and the implicit maximum likelihood estimation**.**

If **more than 5% is missing** for a given outcome in either treatment arm, **multiple imputation** will be performed to assess the robustness of the results.

Multiple imputation will apply suitable regression models and be repeated at least 20 times. The results will be analysed separately and averaged. The model will incorporate baseline variables (e.g., age, gender, treatment group, region), the most recent follow-up data (e.g., from week 16 or week 24), and key clinical covariates such as eosinophil count, number of ESS surgeries, etc.
Unless otherwise indicated, continuous outcomes will be presumed normally distributed; binary and ordinal outcomes will be imputed using logistic or ordinal regression, as applicable.

Imputation will be used to replace data for a follow-up visit if a participant completes it outside of the permitted visit window (see Section 6.2). If a participant misses the window for the week 24 visit but attends the week 48 visit on time, the data from week 48 may be used in the TORNADO-48 analyses.

Regarding treatment failures
In cases of treatment failure, i.e. where participants receive major rescue therapy (i.e., sinus surgery or systemic steroids), they will be considered as treatment failures, and thereby considered to have no further improvement after the last pre-rescue measurement. Post-rescue data will not be used for analysis of the primary endpoint, and the participant's post-rescue values will be imputed as described previously. Participants will remain in the ITT population to have AE/SAE registered, but all clinical outcomes will be registered with imputed values.

8. TORNADO-48 analyses

This section describes the planned analyses in TORNADO-48

As was the case in TORNADO-24, a hierarchical (sequential) testing process will be used for the analyses in TORNADO-48. See section 7.0 for further explanation.

TORNADO-48 includes a response-dependent, response-guided crossover at week 24. As a result, a direct treatment-to-treatment comparison is not feasible. Instead, TORNADO-48 is designed as an **estimand-focused, treatment policy analysis** aimed at assessing which first-line treatment strategy yields superior outcomes. Given the expected low crossover rate (approximately 10–20%), the number of participants in crossover subgroups will likely be insufficient for conclusive as-treated or predictive analyses. However, we will conduct descriptive analyses of crossover patterns and report outcomes separately for the four treatment pathways: DUP–DUP, DUP–MEP, MEP–MEP, and MEP–DUP.

8.1 Primary analyses (non-inferiority) – TORNADO-48

As in TORNADO-24 the primary analysis of the co-primary outcome **SNOT-22** will apply a LMM to the **ITT population from TORNADO-48**. The model will include fixed effects for treatment group (DUPFIRST/MEPFIRST), visit (t=0, t=16, t=24, t=48), their interaction, and the stratification variables gender, region, and baseline smell score. A random intercept for participant ID will account for within-subject correlation. Residual plots and model diagnostics will be assessed for normality of residuals, homoscedasticity, and influential points.

The same method will be applied to the co-primary outcome **NPS**, except this model will not contain data from week 16 (as it is not measured).

Non-inferiority will be concluded only if the lower bound of the 95% confidence intervals for both outcomes SNOT-22 and NPS lies above the non-inferiority margins of -12 points and -1 points (=MCID), using a one-sided $\alpha = 0.025$.

If the conditions are not met for **both** outcomes, non-inferiority will not be declared, and any subsequent superiority analyses will be considered exploratory.

8.2 Secondary analyses (superiority analyses) – TORNADO-48

Superiority analysis will be conducted on the **ITT population**, i.e. determined by which treatment was started first regardless of cross-over (DUPFIRST or MEPFIRST).

We will use a linear mixed-effects model including fixed effects for treatment group, visit, and their interaction, as well as the stratification variables gender, region, and baseline smell score. To take into consideration within-subject correlation across time, a random intercept for subjects will be incorporated. To control the family-wise Type I error rate at an overall $\alpha = 0.05$, we will employ hierarchical (sequence) testing. The following predetermined order will be used to determine superiority:

1. Smell Identification Test (SSIT-16)
2. Sino-Nasal Outcome Test 22 (SNOT-22)
3. Nasal Polyp Score (NPS)
4. Nasal Congestion Score (NCS)
5. Asthma Control Test (ACT)

Statistical significance will be claimed if the null hypothesis is rejected at the alpha level of 0.05 (two-sided) for each of the two co-primarily tested outcomes.

8.3 Additional Analyses – TORNADO-48

8.3.1 Explorative analyses, TORNADO-48

In the case the primary or secondary analyses reach a statistical non-significant result, the remaining analyses from section 6.3.1 and/or 6.3.2 will be exploratory from that point.

Other planned exploratory analyses are:

- Improvement in Lund-Mackay CT score
- Smell-VAS
- Allergic rhinitis-VAS
- Asthma-VAS
- NSAID-exacerbated respiratory disease (NERD-VAS)
- Tympanometry improvement (proportion with normalisation, i.e., from B-curve to A-curve or C-curve to A-curve)
- EPOS/EUFOREA week 24 criteria – proportion “good to excellent” responders (4-5 of the following: reduced polyp size ≥1, reduced need for OCS/salvage surgery, improved SNOT-22 score <40 + >MCID)
- Reduction in fractional exhaled nitric oxide (FeNO)

Exploratory results will be reported descriptively (e.g., means, standard deviations, confidence intervals) without formal hypothesis testing due to expected small sample sizes in the cross-over groups. Interpretation will be with caution due to the potential for inflated type I error from multiple testing. No formal adjustments for multiple comparisons will be applied unless specifically indicated.

8.3.2 As-treated analysis – TORNADO-48

We will present summary outcomes (e.g., change in SNOT-22 and NPS) across the four observed treatment trajectories:

- DUPDUP: Received dupilumab throughout
- DUPMEP: Started on dupilumab, switched to mepolizumab
- MEPMEP: Received mepolizumab throughout
- MEPDUP: Started on mepolizumab, switched to dupilumab

These comparisons will be reported descriptively (e.g., means,/medians, standard deviations, confidence intervals) without formal hypothesis testing due to expected small sample sizes in the cross-over groups.

8.3.3 Safety analysis – TORNADO-48

We will present the rates of AEs, SAEs and need for major rescue treatment (ESS surgery or OCS) in descriptive tables for both drugs individually. Only AEs occurring in more than 5% will be presented.

8.3.4 Sensitivity analyses - TORNADO-48

To assess the robustness of our findings, we will conduct the following sensitivity analyses for the co-primary and secondary outcomes:

- Per-protocol analysis, including only patients without major protocol violations, to assess the effect of adherence and compliance.
- Complete-case analysis and multiple imputation to evaluate the impact of missing data assumptions on the primary outcomes (will only be performed in case of >5% missing data)
- .

The results of these analyses will be compared to the primary (intention-to-treat) analysis to gauge consistency and identify potential bias due to protocol deviations or missing data mechanisms.

8.3.5 Subgroup analyses - TORNADO-48

As in TORNADO-24 we will perform subgroup analyses on age group (<40, 40–65, >65), BMI category (15–25, >25), sex assigned at birth, region, number of previous ESS surgeries (0, 1, 2, ≥3), asthma diagnosis, blood eosinophil count (<150, 150-500, >500 cells/µL), aspirin-exacerbated respiratory disease (yes/no), and total IgE levels (<100, 100–500, >500 IU/mL). Each subgroup will be evaluated separately by including an interaction term between treatment and subgroup level in the linear mixed model. These analyses are exploratory and will not be used to support formal claims of heterogeneity. We expect to present the results in a Forest plot (see example in the Mock-up Figure 3). We will, however, refrain from this if no consistent pattern of subgroup modification is observed.

8.3.6 Dropout characteristics - TORNADO-48

In the case of five or more dropouts in an arm (=consent withdrawn) we will produce supplementary tables with baseline characteristics of dropouts in an identical way as Mock-up table 1B (Section 8, "supplementary").

8.3.7 Analysis of crossovers – TORNADO-48

Crossover analyses will focus on patients who—due to failed response criteria and per the protocol—switch treatment at the week 24 evaluation. These analyses are considered exploratory and hypothesis-generating only, given the expected low number of crossover participants and limited statistical power. The results will be interpreted with caution and will not inform formal efficacy conclusions.

If ≥10 participants in an arm cross over at week 24, we will:

- Produce supplementary baseline descriptive tables identical to Table 1, comparing baseline characteristics of crossover and non-crossover patients.
- Compare changes in outcome values from week 24 to week 48 with changes from baseline to week 24. This will be performed using a paired t-test if test score values are normally distributed or a non-parametric test if normality assumptions are not met.

If fewer than 10 participants cross over, no formal statistical comparisons will be performed. In such cases, only descriptive summaries of baseline characteristics and outcome changes will be presented.

8.4 Missing data – TORNADO-48

We expect low proportions of missing values, i.e., less than 10% for exploratory and tertiary outcomes and less than 5% for primary and major secondary outcomes. Rather than treatment-related adverse events or dropouts, missing data will mostly result from logistical reasons like incomplete registrations or missed appointments/violated visit windows (see section 6.2). As a result, we assume that the data are **missing at random (MAR)**, and the principal analyses will be based on this assumption using **LMMs** and the implicit maximum likelihood estimation**.**

If **more than 5% is missing** for a given outcome in either treatment arm, **multiple imputation** will be performed to assess the robustness of the results.

Multiple imputation will apply suitable regression models and be repeated at least 20 times. The results will be analysed separately and averaged. The model will incorporate baseline variables (e.g., age, gender, treatment group, region), the most recent follow-up data (e.g., from week 16 or week 24) and key clinical covariates such as eosinophil count, number of ESS surgeries etc.
Unless otherwise indicated, continuous outcomes will be presumed to be normally distributed; binary and ordinal outcomes will be imputed using logistic or ordinal regression, as applicable.

Imputation will be used to replace data for a follow-up visit if a participant completes it outside of the permitted visit window (see Section 6.2). If a participant misses the window for the week 24 visit but attends the week 48 visit on time, the data from week 48 may be used in the TORNADO-48 analyses.

Regarding treatment failures

In cases of treatment failure, i.e., where participants receive major rescue therapy (i.e., sinus surgery or systemic steroids), they will be considered as treatment failures, and thereby considered to have no further improvement after the last pre-rescue measurement. Post-rescue data will not be used for analysis of the primary endpoint, and the participant's post-rescue values will be imputed as described previously. Participants will remain in the ITT population to have AE/SAE registered, but all clinical outcomes will be registered with imputed values.

9.0 Harms

Clinical staff will report any non-serious adverse events (AEs) and serious AEs (SAEs) during follow-up using specific reporting tools in the eCRF and following ICH-GCP guidelines. AEs/SAEs will be classified by severity (mild, moderate, severe), expectedness, and causality (related or unrelated to the investigational medicinal product), based on the investigator's clinical judgment and predefined criteria in the protocol. We will report the number and proportion of participants experiencing each AE and SAE. Only AEs occurring in ≥5% of participants in either arm will be presented in the main text tables. Comparisons of AE rates between treatment groups will be descriptive only. No formal hypothesis tests will be conducted on safety. For each AE, results will include the percentage of participants affected and 95% confidence intervals.

10.0 Statistical Software

All statistical analyses will be performed using R version 4.3.3 (R Core Team, 2048, available at https://cran.r-project.org) or a later version if an update is released during the study analysis period. Packages in R such as lme4, LmerTest, or LMMstar, or a validated equivalent will be used for linear mixed model analyses.

11.0 References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA - Journal of the American Medical Association. 2017;318(23):2337–43.

2. Danish Medicines Council. Medicinrådets anbefaling vedr. mepolizumab til behandling af svaer kronisk rhinosinuitis med naesepolypper (CRSwNP). 2022 Aug.

3. Bachert C, Han JK, Desrosiers M, Hellings PW, Amin N, Lee SE, et al. Efficacy and safety of dupilumab in patients with severe chronic rhinosinusitis with nasal polyps (LIBERTY NP SINUS-24 and LIBERTY NP SINUS-52): results from two multicentre, randomised, double-blind, placebo-controlled, parallel-group phase 3 trials. The Lancet. 2019 Nov 2;394(10209):1638–50.

4. Han JK, Bachert C, Fokkens W, Desrosiers M, Wagenmann M, Lee SE, et al. Mepolizumab for chronic rhinosinusitis with nasal polyps (SYNAPSE): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med. 2021 Oct 1;9(10):1141–53.

5. Hopkins C, Han JK, Fokkens W, Wagenmann M, Guyot P, Khan AH, et al. Dupilumab versus mepolizumab for chronic rhinosinusitis with nasal polyposis: an indirect treatment comparison. J Allergy Clin Immunol Pract [Internet]. 2024; Available from: http://www.ncbi.nlm.nih.gov/pubmed/39326524

6. Peters AT, Han JK, Hellings P, Heffler E, Gevaert P, Bachert C, et al. Indirect Treatment Comparison of Biologics in Chronic Rhinosinusitis with Nasal Polyps. Journal of Allergy and Clinical Immunology: In Practice. 2021 Jun 1;9(6):2461-2471.e5.

7. Cai, Shiru; Xu, Shenglong; Zhao, Yan; Zhang L. Efficacy and Safety of Biologics for Chronic Rhinosinusitis With Nasal Polyps. Allergy [Internet]. 2025; Available from: https://doi.org/10.1111/all.16499

8. Cai S, Xu S, Lou H, Zhang L. Comparison of Different Biologics for Treating Chronic Rhinosinusitis With Nasal Polyps: A Network Analysis. Journal of Allergy and Clinical Immunology: In Practice [Internet]. 2022;10(7):1876-1886.e7. Available from: https://doi.org/10.1016/j.jaip.2022.02.034

9. Papacharalampous GX, Constantinidis J, Fotiadis G, Zhang N, Bachert C, Katotomichelakis M. Chronic rhinosinusitis with nasal polyps (CRSwNP) treated with omalizumab, dupilumab, or mepolizumab: A systematic review of the current knowledge towards an attempt to compare agents' efficacy. Int Forum Allergy Rhinol. 2023;(March 2023):96–109.

10. Oykhman P, Paramo FA, Bousquet J, Kennedy DW, Brignardello-Petersen R, Chu DK. Comparative efficacy and safety of monoclonal antibodies and aspirin desensitization for chronic rhinosinusitis with nasal polyposis: A systematic review and network meta-analysis. Journal of Allergy and Clinical Immunology. 2022 Apr 1;149(4):1286–95.

11. Tănase MI, Tanase M, Cosgarea M, Radeanu GD, Pop SS. Biologic Treatments for Chronic Rhinosinusitis With Nasal Polyps (CRSwNP): A Comparative Review of Efficiency and Risks. 2025;17(II):1–7.

12. Wu Q, Zhang Y, Kong W, Wang X, Yuan L, Zheng R, et al. Which Is the Best Biologic for Nasal Polyps: Dupilumab, Omalizumab, or Mepolizumab? A Network Meta-Analysis. Int Arch Allergy Immunol. 2021;

13. Cantone E, Cassiano B, Pezzella P, Russo MB, Detoraki A. The Effects of Mepolizumab on CRSwNP: Real-Life Evidence. J Pers Med. 2024;14(11).

14. Domínguez-Sosa MS, Cabrera-Ramírez MS, Marrero-Ramos M del C, Dávila-Quintana D, Cabrera-López C, Carrillo-Díaz T, et al. Real-Life Effectiveness of Mepolizumab in Refractory Chronic Rhinosinusitis with Nasal Polyps. Biomedicines. 2023;11(2):1–15.

15. Albrecht T, Sailer MM, Capitani F, van Schaik C, Löwenheim H, Becker S. Real-world evidence for the effectiveness and safety of dupilumab in patients with CRSwNP after 1 year of therapy. World Allergy Organization Journal. 2023 May 1;16(5).

16. Huber P, Förster-Ruhrmann U, Olze H, Becker S, Bärhold F, Cuevas M, et al. Real-world data show sustained therapeutic effects of dupilumab in chronic rhinosinusitis with nasal polyps (CRSwNP) over 3 years. Allergy: European Journal of Allergy and Clinical Immunology. 2024;(May):3108–17.

17. Cavaliere C, Loperfido A, Ciofalo A, Di Michele L, Begvarfaj E, Bellocchi G, et al. Real-Life Evidence of Mepolizumab Treatment in Chronic Rhinosinusitis with Nasal Polyps: A Multicentric Study. J Clin Med. 2024;13(12):1–12.

18. De Corso E, Pasquini E, Trimarchi M, La Mantia I, Pagella F, Ottaviano G, et al. Dupilumab in the treatment of severe uncontrolled chronic rhinosinusitis with nasal polyps (CRSwNP): A multicentric observational Phase IV real-life study (DUPIREAL). Allergy: European Journal of Allergy and Clinical Immunology. 2023;(January):1–15.

19. van der Lans RJL, Otten JJ, Adriaensen GFJPM, Hoven DR, Benoist LB, Fokkens WJ, et al. Two-year results of tapered dupilumab for CRSwNP demonstrates enduring efficacy established in the first 6 months. Allergy: European Journal of Allergy and Clinical Immunology. 2023;(February):1–14.

20. Mortuaire G, Béquignon E, Daveau C, Papon J françois, Lecanu J baptiste, Favier V. Responders to biologics in severe uncontrolled chronic rhinosinusitis with nasal polyps : a multicentric observational real-life study • Patient data. 2025;

21. Viskens A, Backer V. EPOS/EUFOREA update on indication and evaluation of biologics in chronic rhinosinusitis with nasal polyps 2023. Rhinology. 2023;(In press):194–202.

22. Danish Medicines Council. Medicinrådets samling af vurderinger af laegemidler til svaer kronisk rhinosinuitis med naesepolypper. 2022 Oct.

23. Eriksen PRG, Jakobsen KK, Aanæs K, Backer V, von Buchwald C. The potential role of biological treatment of chronic rhinosinusitis with nasal polyps: A nationwide cohort study*. Rhinology. 2021;59(4):374–9.

24. Oleszkiewicz A, Schriever VA, Croy I, Hähner A, Hummel T. Updated Sniffin' Sticks normative data based on an extended sample of 9139 subjects. European Archives of Oto-Rhino-Laryngology [Internet]. 2019;276(3):719–28. Available from: http://dx.doi.org/10.1007/s00405-018-5248-1

25. Gudziol V, Lötsch J, Hähner A, Zahnert T, Hummel T. Clinical significance of results from olfactory testing. Laryngoscope. 2006;116(10):1858–63.

26. Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S, et al. European Position Paper on Rhinosinusitis and Nasal Polyps 2020. Vol. 103, Dieudonné Nyenbue Tshipukane. 2020.

27. Phillips KM, Hoehle LP, Caradonna DS, Gray ST, Sedaghat AR. Minimal clinically important difference for the 22-item Sinonasal Outcome Test in medically managed patients with chronic rhinosinusitis. Clinical Otolaryngology. 2018;43(5):1328–34.

28. Braid J, Islam L, Gugiu C, Omachi TA, Doll H. Meaningful changes for efficacy outcomes in patients with chronic rhinosinusitis with nasal polyps. World Allergy Organization Journal [Internet]. 2023;16(5):100776. Available from: https://doi.org/10.1016/j.waojou.2023.100776

29. Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: Reliability, validity, and responsiveness in patients not previously followed by asthma specialists. Journal of Allergy and Clinical Immunology. 2006 Mar;117(3):549–56.

30. Mullol J, Lund VJ, Wagenmann M, Han JK, Sousa AN, Smith SG, et al. Mepolizumab improves sense of smell in severe chronic rhinosinusitis with nasal polyps: SYNAPSE. Rhinology Journal. 2023;0(0):0–0.

31. Luke L, Lee L, Gokani SA, Boak D, Boardman J, Philpott C. Understanding the Impact of Chronic Rhinosinusitis with Nasal Polyposis on Smell and Taste: An International Patient Experience Survey. J Clin Med. 2023;12(16).

32. Bachert C, Hellings PW, Mullol J, Hamilos DL, Gevaert P, Naclerio RM, et al. Dupilumab improves health-related quality of life in patients with chronic rhinosinusitis with nasal polyposis. Allergy: European Journal of Allergy and Clinical Immunology. 2020 Jan 1;75(1):148–57.

12. Supplementary: suggested tables and figures

12.1 TORNADO-24:

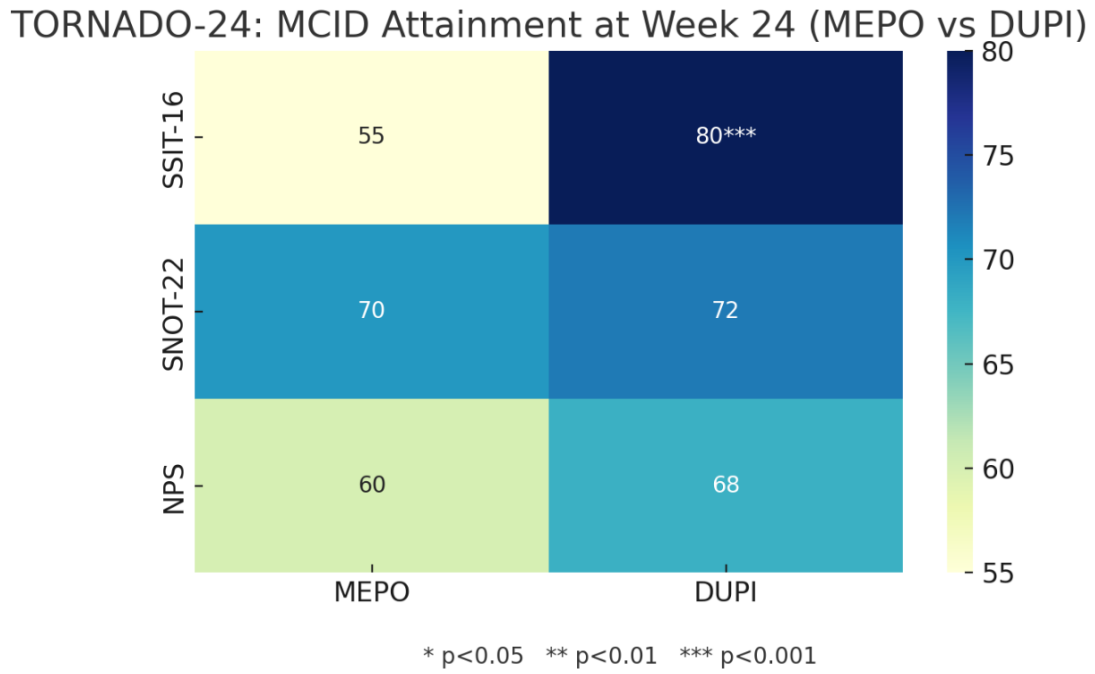
Mock-up Table 1a – Baseline Characteristics TORNADO-24	Mepolizumab	Dupilumab	Total
Characteristic	**(N = xxx)**	**(N = xxx)**	**(N = xxx)**
Age at treatment start, year (IQR)			
Gender (at birth) – no. (%) Female Male			
Smoking history (%) Never Current smoker Former smoker			
Body Mass Index 15-25 >25			
Region Capital Region Region Zealand Region of Southern Denmark Central Denmark Region North Denmark Region			
Nasal Polyp Surgery – no. (%) ≥1 2 ≥3			
Time since last surgery – years			
Blood eosinophils (EOS) Total blood EOS – (x10^9/L)			
Blood IgE Total IgE (U/mL)			
Tissue eosinophils Low (<10 EOS/HPF) Intermediate (10-100 EOS/HPF) High (>100 EOS/HPF)			
SCS for nasal polyps Total amount previous five years (mg)			
Type-2 comorbidity: Asthma diagnosis – no. (%) FeNO (ppb) N-ERD[a] – no. (%) Atopy [b] – no. (%			

Danish Treatment Criteria ≥2 courses of SCS within previous year Nasal Polyp Score ≥5 SNOT-22 ≥50 Smell Test Score ≤8 Asthma in ICS treatment			

a: Based on patient's reply to the question: "Do you have hypersensitivity against acetoacetic acid?".
B: One or more positive results from skin prick testing and/or specific IgE blood tests
Abbreviations: ACQ-7: Asthma Control Questionnaire 7 items, EOS: eosinophils, FeNO: Fraction of exhaled nitrous oxide, HPF: High power field, ICS: Inhaled corticosteroids, IgE: Immunoglobulin E, IQR: inter quartile range, N-ERD: NSAID-exacerbated respiratory disease, NPS: Nasal Polyp Score, NSAID: non-steroidal anti-inflammatory drugs, PPB: parts per billion, SCS: systemic corticosteroids, SNOT-22: Sino-Nasal Outcome Test.

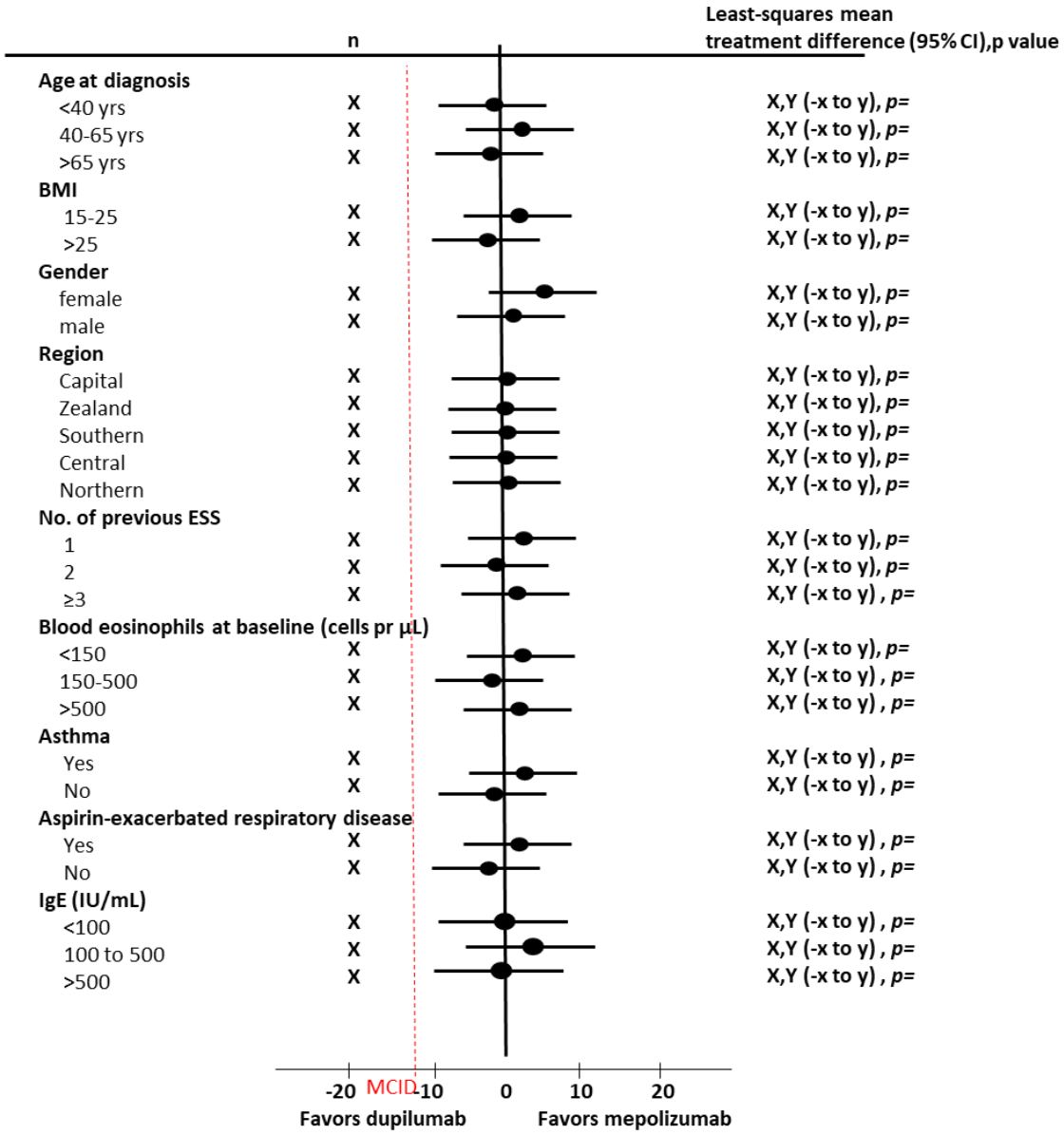
Mockup Figure 1A: TORNADO-24: Graphs showing changes in outcome over time.

Mock-up Figure 2A



Mock-up table 2A - TORNADO-24: Results of superiority analysis				
Outcome	**No. in test hierarchy[¥]**	**Dupilumab (n = …)** Baseline → Week 24 Change since baseline	**Mepolizumab (n = …)** Baseline → Week 24 Change since baseline	**Between treatment difference, week 24 Dupilumab vs. Mepolizumab (95%CI; p value)**
Sniffin' Sticks 16 Identification test (SIT16) (scale: 0-16, lower worse)	2	X_1 (SD) → x_1 (SD) **ΔX_1 (SD)**	Y_1 (SD) → y_1 (SD) **ΔY_1 (SD)**	**ΔX_1- ΔY_1** (-X.x to -Y.y), p=0.XX $p_{superiority}$=0.xx
SNOT-22 (scale:0-110, higher worse)	1 and 3	X_2 (SD) → x_2 (SD) **ΔX_2 (SD)**	Y_1 (SD) → y_1 (SD) **ΔY_2 (SD)**	**ΔX_2- ΔY_2** (-X.x to -Y.y), $p_{superiorit}$=0.xx
Nasal congestion score (NCS) (scale: 0-3, higher worse)	6	X_3 (SD) → x_3 (SD) **ΔX_3 (SD)**	Y_1 (SD) → y_1 (SD) **ΔY_3 (SD)**	**ΔX_3- ΔY_3** (-X.x to -Y.y), p=0.XX $p_{superiority}$=0.xx
Nasal polyp size (NPS) (scale: 0-8, higher worse)	1 and 4	X_4 (SD) → x_4 (SD) **ΔX_4 (SD)**	Y_1 (SD) → y_1 (SD) **ΔY_4 (SD)**	**ΔX_4- ΔY_4** (-X.x to -Y.y), $p_{superiority}$=0.xx
Asthma Control Test (scale: 0-25, lower worse)		X_5 (SD) → x_5 (SD) **ΔX_5 (SD)**	Y_1 (SD) → y_1 (SD) **ΔY_5 (SD)**	**ΔX_5- ΔY_5** (-X.x to -Y.y), p=0.XX $p_{superiority}$=0.xx
Color explanation				
Improvement better than MCID		Significant or statistically superior improvement ($p < 0.05$)	Improvement is not statistically different ($p < 0.05$)	Uncertain improvement

Mock-up Figure 3A: Forest plot of sub-group analyses (Example with SNOT-22)



Mock-up table 3A, TORNADO-24: Summary of adverse events week 0-24			
Adverse events		**dupilumab**	**mepolizumab**
		Number of patients (%)	
Any adverse event			
Any serious adverse event[a]			
Any adverse event resulting in discontinuation of the drug			
Any serious adverse reaction[b]			
Adverse event occurring in at least 5% of the randomisation group			
	Examples listed below: Headache Injection site reactions		

	Muscle or joint pain Upper respiratory tract infection Skin rashes/swelling/redness Nasopharyngitis		
a: Serious adverse events are events occurring within the study period. They include: hospitalization, death, life-threatening injury or disease, permanent damage to organs or bodily functions, hospital admission or prolonged hospitalization, events where medicinal or surgical treatment is necessary to avoid the previously mentioned, death of a fetus/inborn anomalies/other negative effects of a fetus. b: (Serious adverse reactions are serious adverse events that are judged by an investigator and the sponsor to likely be related to the study drug.			

12.2 TORNADO-48

Mock-up table 1B – Baseline Characteristics TORNADO-48	**Assigned to Mepolizumab (MEP-FIRST) (n = xxx)**		**Assigned to Dupilumab (DUP-FIRST) (n = xxx)**		**Total (N = xxx)**
	Stayed on MEPO	Crossed over to DUPI	Stayed on DUPI	Crossed over to MEPO	
Characteristic					
Age at treatment start, year (IQR)					
Gender (at birth) – no. (%) Female Male					
Body Mass Index 15-25 >25					
Region Capital Region Region Zealand Region of Southern Denmark Central Denmark Region North Denmark Region					
Crossover after 24 weeks due to insufficient response - no (%)					
Previous endoscopic sinus surgery – no. (%) ≥1 ≥3 ≥5					
EOS Total blood EOS – (x10^9/L)					
IgE Total IgE (U/mL)					
Tissue eosinophils Low (<10 EOS/HPF)					

Intermediate (10-100 EOS/HPF) High (>100 EOS/HPF)					
SCS for nasal polyps Total amount previous five years (mg)					
Type-2 comorbidity at inclusion: Asthma diagnosis – no. (%) FeNO (ppb) N-ERD[a] – no. (%) Atopy [b] – no. (%)					
Danish Treatment Criteria ≥2 courses of SCS within previous year Nasal Polyp Score ≥5 SNOT-22 ≥50 Smell Test Score ≤8 Asthma in ICS treatment					
a: Based on patient's reply to the question: "Do you have hypersensitivity against acetoacetic acid?". B: One or more positive results from skin prick testing and/or specific IgE blood tests. *Abbreviations:* ACQ-7: Asthma Control Questionnaire 7 items, EOS: eosinophils, FeNO: Fraction of exhaled nitrous oxide, HPF: High power field, ICS: Inhaled corticosteroids, IgE: Immunoglobulin E, IQR: inter quartile range, N-ERD: NSAID-exacerbated respiratory disease, NPS: Nasal Polyp Score, NSAID: non-steroidal anti-inflammatory drugs, PPB: parts per billion, SCS: systemic corticosteroids, SNOT-22: Sino-Nasal Outcome Test.					

MOCK-UP Table 2B: Response criteria fulfillment per treatment (will alternatively be shown as histograms)						
	% achieved in **DUPI-FIRST** at **week 24**	% achieved in **MEPO-FIRST** at **week 24**	% achieved in **MEPO-MEPO** at **week 48**	% achieved in **DUPI-DUPI** at **week 48**	% achieved in **MEPO-DUPI** at **week 48**	% achieved in **DUPI-MEPO** at **week 48**
NPS reduction week 24: ≥1 week 48: ≥1						
SNOT-22 reduction ≥12						
Smell improved from anosmia to hyposmia/normosmia						
ACQ-7 improvement ≥0.5						

Mock-up Table 3B: Numerical changes according to treatment groups

Outcome	DUP-FIRST (n = …)		MEP-FIRST (n = …)		Between-treatment path difference, week 48
	Change from baseline to Week 24	Change from week 24 to week 48	Change from baseline to Week 24	Change from week 24 to week 48	DUP-FIRST vs. MEP-FIRST (95%CI; p value)
SSIT-16 (scale: 0-16, lower worse)	X(SD)→x(SD) **ΔX (SD)**	X(SD)→x(SD) **ΔX (SD)**	Y (SD)→y (SD) **ΔY (SD)**	Y (SD)→y (SD) **ΔY (SD)**	**$\Delta SSIT\text{-}16_{w.0\text{-}48}$** (95%CI), p=0.XX $p_{superiority}$=0.xx
SNOT-22 (scale:0-110, higher worse)	X(SD)→x(SD) **ΔX (SD)**	X(SD)→x(SD) **ΔX (SD)**	Y (SD)→y (SD) **ΔY (SD)**	Y (SD)→y (SD) **ΔY (SD)**	**$\Delta SNOT\text{-}22_{w.0\text{-}48}$** (95%CI), p=0.XX $p_{superiority}$=0.xx
NCS (scale: 0-3, higher worse)	X(SD)→x(SD) **ΔX (SD)**	X(SD)→x(SD) **ΔX (SD)**	Y (SD)→y (SD) **ΔY (SD)**	Y (SsD)→y (SD) **ΔY (SD)**	**$\Delta NCS_{w.0\text{-}48}$** (95%CI), p=0.XX $p_{superiority}$=0.xx
NPS (scale: 0-8, higher worse)	X(SD)→x(SD) **ΔX (SD)**	X(SD)→x(SD) **ΔX (SD)**	Y (SD)→y (SD) **ΔY (SD)**	Y (SD)→y (SD) **ΔY (SD)**	**$\Delta NPS_{w.0\text{-}48}$** (95%CI), p=0.XX $p_{superiority}$=0.xx
ACT (scale: 0-25, lower worse)	X(SD)→x(SD) **ΔX (SD)**	X(SD)→x(SD) **ΔX (SD)**	Y (SD)→y (SD) **ΔY (SD)**	Y (SD)→y (SD) **ΔY (SD)**	**$\Delta ACT_{w.0\text{-}48}$** (95%CI), p=0.XX $p_{superiority}$=0.xx

Color explanation			
Improvement better than MCID	Significant or statistically superior improvement ($p < 0.05$)	Improvement is not statistically different ($p < 0.05$)	Uncertain improvement

Mock-up table 4B, TORNADO-48: Summary of adverse events week 24-48			
Adverse events		**Week 0-24**	**Week 24-48**
		Number of patients (%)	
DUPDUP[a]			
	Any adverse event		
	Any serious adverse event		
	Adverse event occurring in at least 5% of the randomization group		
MEPMEP[b]			
	Any adverse event		
	Any serious adverse event		
	Adverse event occurring in at least 5% of the randomization group		
DUPMEP[c]			
	Any adverse event		
	Any serious adverse event		
	Adverse event occurring in at least 5% of the randomization group		
MEPDUP[d]			
	Any adverse event		
	Any serious adverse event		
	Adverse event occurring in at least 5% of the randomization group		

a: DUPDUP: Received dupilumab throughout all 48 weeks

b: DUPMEP: Started on dupilumab, switched to mepolizumab after 24 weeks

c: MEPMEP: Received mepolizumab throughout all 48 weeks

d: MEPDUP: Started on mepolizumab, switched to dupilumab after 24 weeks

Mock-up Figure 1B – Changes in outcome over time and number of participants in each arm including after cross-over

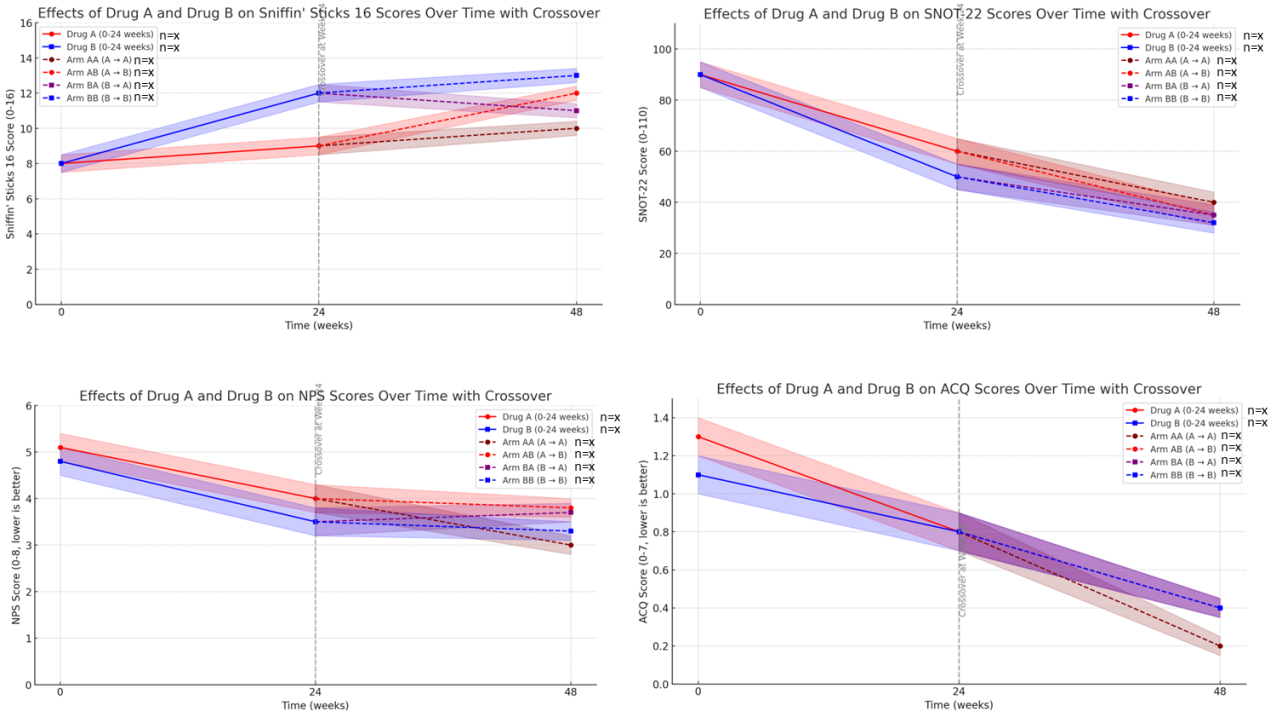


Figure 2B: EPOS/EUFOREA response groups per treatment allocation group

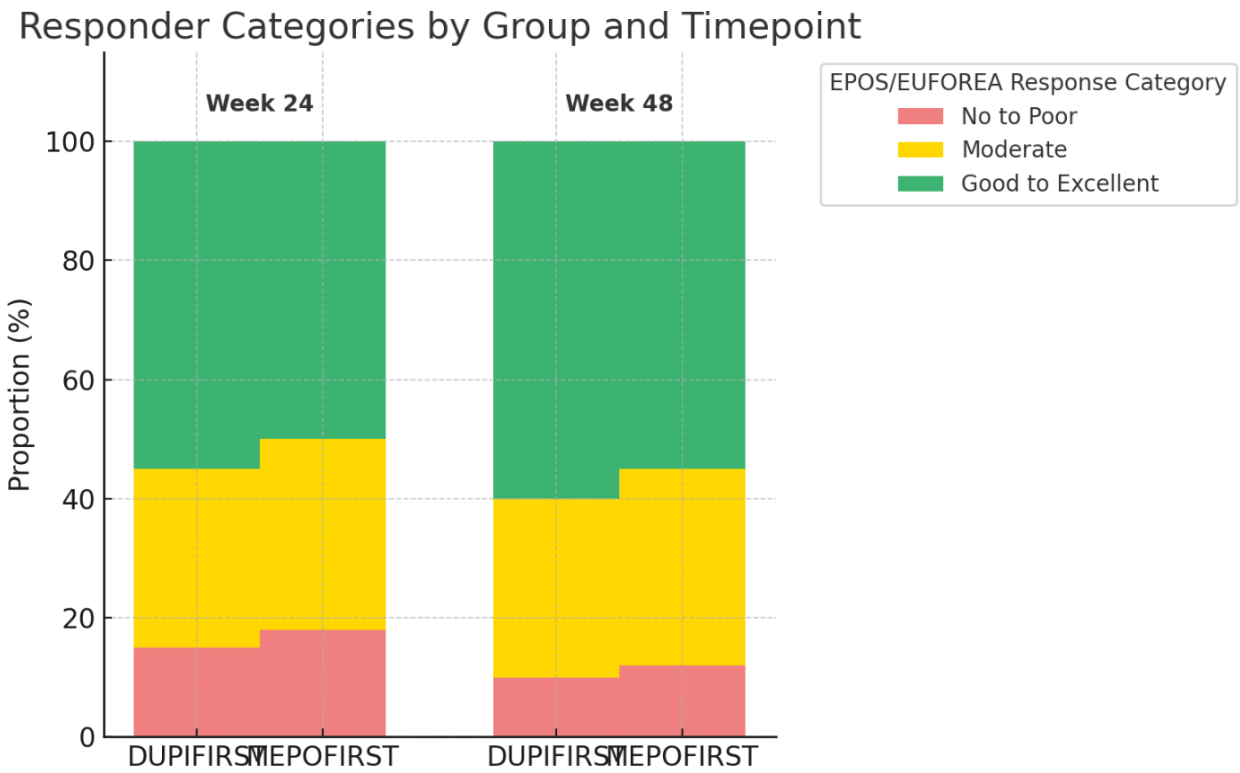


Figure 3B – MCID response proportion by treatment path

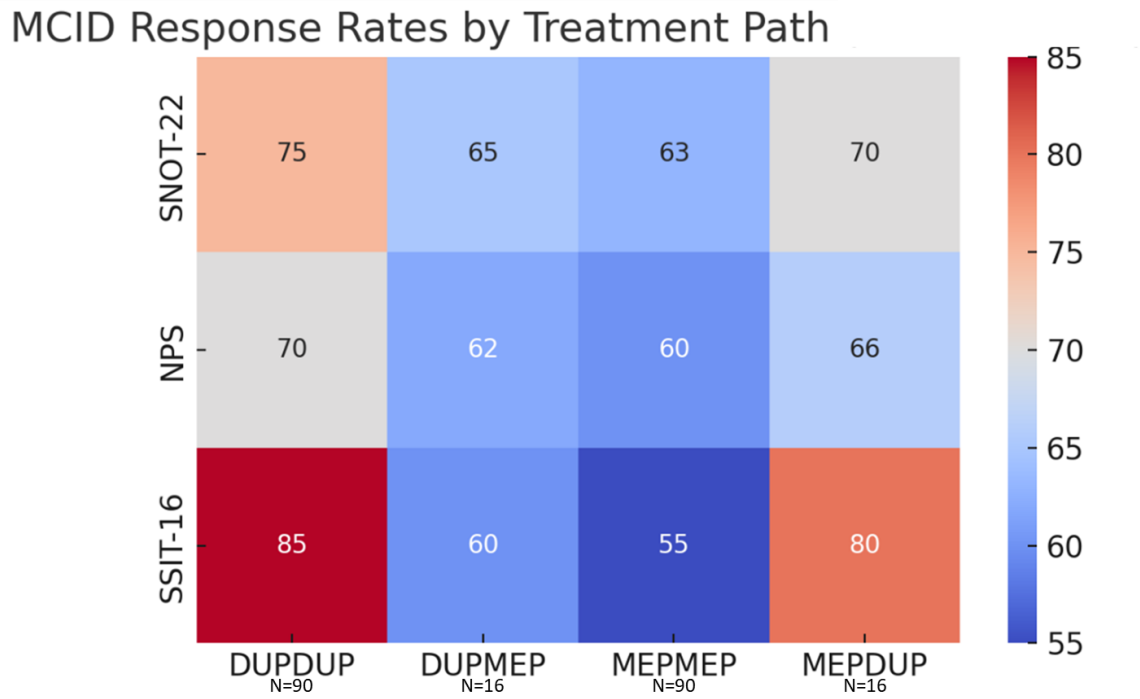


Figure 3: Heatmap showing the percentage reaching the MCID at week 48.
The respective MCID's are: SNOT-22=12, NPS=1, SSIT-16=3

Figure 4B – Sub-group analysis for SNOT-22