MC2 Therapeutics

Statistical Analysis Plan
for
Clinical Study Reporting

of the Phase II Clinical Study MC2-25-C3

A parallel group (2-arm), randomised, double-blind, 12-week trial to explore the efficacy and safety of MC2-25 cream and MC2-25 vehicle in women diagnosed with vulvar lichen sclerosus (VLS)

FINAL Version 1.0 as of 30 October 2024

Author: David Carr, Senior Principal Statistical Writer, Estimondo GmbH

Contributor: Frank Freischläger, Senior Principal Statistical Writer, Estimondo GmbH

For review and approval, see the MC2-25-C3 SAP for CSR Version 1.0 Review and Approval Sheet on file.

Confidential

Document History

Version	Effective Date	Description of Changes
1.0	30 Oct 2024	New

Table of Contents

List of Tables

List of Abbreviations and Acronyms

Abbreviation	Term
ADaM	Analysis Data Model
AE	Adverse event
Ala-Gln	Alanyl-Glutamine
ANCOVA	Analysis of covariance
ATC	Anatomic therapeutic chemical
CDISC	Clinical Data Interchange Standards Consortium
CFR	Code of Federal Regulation
CGA	Clinician's Global Assessment
CGIC	Clinician's Global Impression of Change
CKD	Chronic Kidney Disease
CKD-aP	Chronic Kidney Disease-associated Pruritus
ClinRO	Clinician-Reported Outcome
COA	Clinical Outcome Assessment
CRF	Case report form
CRO	Contract Research Organisation
CSR	Clinical study report
CTP	Clinical trial protocol
eCRF	Electronic Case Report Form
EQ-5D	EuroQOL five dimensions
EMA	European Medicines Agency
EoT	End of Treatment
EU	European Union
FAS	Full analysis set
FTU	Fingertip Unit
FU	Follow-Up
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
hCG	Human Chorionic Gonadotropin
HRQoL	Health-Related Quality of Life
IB	Investigator's Brochure
ICH	International Council on Harmonization
ICIQ-UI-SF	International Consultation on Incontinence Questionnaire Urinary Incontinence Short Form
IEC	Independent Ethics Committee
IMP	Investigational Medicinal Product
IRB	Institutional Review Board
LOCF	Last observation carried forward
LS	Lichen Sclerosus
LSR	Local skin reaction
MCMC	Markov Chain Monte Carlo
MedDRA	Medical dictionary for regulatory activities
MFAS	Modified Full Analysis Set
MI	Multiple imputation
MMRM	Mixed Model of Repeated Measures
PGIC	Patient's Global Impression of Change
PPS	Per-protocol set
PRO	Patient Reported Outcome
PT	Preferred term
RSI	Reference Safety Information
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Safety Analysis Set
SAS®	Statistical Analysis System®
SAS/STAT®	SAS® Statistics software module
SD	Standard deviation
SDTM	Standard Data Tabulation Model
Skindex-29	Skin specific multidimensional quality-of-life instrument

Abbreviation	Term
SmPC	Summary of Product Characteristics
SoA	Schedule of Assessments
SOC	System Organ Class
SS	Screened Set
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TFL	Tables, figures, listings
TOMS	Technical Operational Measures
UTI	Urinary Tract Infection
UV-A	Ultraviolet A
UV-B	Ultraviolet B
VAS	Visual analogue scale
VASSI	Vulvar lichen sclerosus Area and Sign Severity Index
VLS	Vulvar Lichen Sclerosus
VQLI	Vulvar disease Quality of Life Index
w/w	Weight by weight
WI	Worst Itch
WI-NRS	Worst Itch – Numeric Rating Score
WPP	Worst Penetrative sex related Pain
WPP-NRS	Worst Penetrative sex related Pain – Numeric Rating Score
WP	Worst Pain
WP-NRS	Worst Pain – Numeric Rating Score

1 Introduction

The Phase II trial MC2-25-C3 will be analyzed and reported in accordance with the clinical trial protocol (CTP) [1] and the International Conference on Harmonization (ICH) guidelines ICH E3 [2], ICH E6 [3] and ICH E9 [4]. MC2-25-C3 is a trial to compare the MC2-25 cream with MC2-25 vehicle in women with vulvar lichen sclerosus (VLS).

This statistical analysis plan (SAP) should be read in conjunction with the CTP and the electronic case report form (CRF) [5]. This version of the plan has been developed using the CTP version 4.0 dated 31 July 2024 and the SDTM-annotated CRF. Any further changes to the protocol or CRF may necessitate updates to the SAP.

Based on this SAP, deliverables are derived analysis datasets following the CDISC ADaM standard, tables, figures and listings (TFL).

2 Study Design and Objectives

2.1 Study Design

This is a multicentre, phase 2, randomised, double-blind, 2-arm, parallel-group and vehicle-controlled trial in VLS conducted in Denmark. Random allocation to one of two parallel treatment groups is in a 1:1 ratio to MC2-25 cream or MC2-25 vehicle.

A total of 40 patients are planned to be enrolled, randomised and observed until final evaluation and completion. Patients prematurely terminating the trial will not be replaced. Following protocol amendment #3, and the associated protocol v4.0, recruitment was terminated at 33 patients.

The maximum trial duration for each patient will be at most 18 weeks and includes a screening period of up to 4 weeks (allowing for wash out of other treatments and emollients before baseline as/if applicable), a 12-week treatment period, and a 2-week post-treatment follow-up (FU) period.

A total of eleven visits are planned (six of these will be in-clinic, 4 by phone, and the follow-up visit either in-clinic or by phone). The in-clinic visits will be held at Screening 1 (Visit 1, Day -28 to -8), Baseline (Visit 3, Week 0, Day 0), Visits 4, 6, and 8 at Weeks 1, 4 and 8, respectively, planned for Days 7±1, 28±4, and 56±4, respectively, Visit 10 as end of treatment Visit (Week 12, Day 84±4, or at early termination). The follow-up visit at Visit 11 (Week 14, Day 98±4) will be either an in-clinic or a phone visit, depending on the follow up assessments considered necessary by the investigator. The phone visits will be held at Screening 2 (Visit 2, Day -11 to -8), Visits 5, 7, and 9 at Weeks 3, 7 and 11, respectively, planned for Days 18±2, 46±2, and 74±2, respectively. An unscheduled visit can be conducted to perform any of the post-baseline assessments or procedures included in the schedule of assessments (SoA) if deemed necessary by the Investigator. If Screening 1 visit is performed on Day -8, then Screening 2 visit will be omitted. For a visit schedule, see CTP Section 1.3 [1].

2.2 Treatments, Study Groups and Treatment Assignment

Patients will be randomly allocated to either receive *MC2-25 cream*, white oil-in-water, topical emulsion containing the active ingredient, Alanyl-Glutamine (Ala-Gln) in the aqueous phase at a concentration of 30 mg/g or *MC2-25 vehicle*, which is identical to *MC2-25 cream* without the active ingredient.

MC2-25 cream, or MC2-25 vehicle, must be applied by the patient after each toilet visit following urination (during day and night) starting on the day of the Baseline visit. Assuming, a maximum of 12 toilet visits are needed over a period of 24 hours and a maximum of 0.5 g cream per application, the maximum daily dose is expected to be around 6.0 g.

The sponsor, sponsor representatives, and all patients, monitors, and site staff will be blinded throughout the trial. Instructions for emergency unblinding are provided in CTP Section 10.2.1. The staff at the investigational medicinal product (IMP) manufacturing site and the external statistician responsible for preparing the randomisation list are not blinded.

Each patient will be assigned a unique randomisation number determining which study treatment the patient will receive. These randomisation numbers will be based on a supply list accompanying the IMP provided to the sites and the same randomisation number named as kit no. will also be listed on the applicable IMP. At each dispense, this kit no. will be documented in the eCRF linking the IMP to the correct patient. No patient ID, other than the kit no. will be written on the IMP.

2.3 Study Objectives

2.3.1 Primary Objectives

To explore the efficacy of MC2-25 cream compared to MC2-25 vehicle in vulvar lichen sclerosus (VLS).

2.3.2 Secondary Objectives

- To explore the safety of MC2-25 cream compared to MC2-25 vehicle in VLS.
- To explore the burden of VLS on women's lives.

2.4 Study Endpoints

2.4.1 Efficacy

The primary endpoint is the mean change in weekly mean Worst Itch Numeric Rating Score (WI-NRS) recorded in the patient's diary from Baseline to Week 12 for MC2-25 cream compared to MC2-25 vehicle.

(Weekly mean WI-NRS is calculated as the average of WI-NRS values recorded in the patient's diary for 7 days prior to the in-clinic visits).

The secondary endpoints (all defined from Baseline to Week 12 for MC2-25 cream compared to MC2-25 vehicle) are:

- Mean change in weekly mean Worst Pain Numeric Rating Score (WP-NRS) recorded in the patient's diary (Weekly mean WP-NRS is calculated as the average of WP-NRS values recorded in the patient's diary for 7 days prior to the in-clinic visits.)
- Percentage of patients obtaining a ≥4-point improvement in weekly mean WI-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥4-point improvement in weekly mean WP-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥3-point improvement in weekly mean WI-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥3-point improvement in weekly mean WP-NRS recorded in the patient's diary
- Change in Skindex-29 domains

Primary and secondary endpoints will be confirmatorily analyzed as described in Sections 5.2 to 5.5. Other endpoints will be exploratorily analyzed.

Other endpoints (all defined from Baseline to Week 12) include:

- Percentage of patients obtaining a complete response (score ≤1) in weekly mean WI-NRS recorded in the patient's diary
- Percentage of patients obtaining a complete response (score ≤1) in weekly mean WP-NRS recorded in the patient's diary
- Mean change in Worst Itch Numeric Rating Score (WI-NRS) recorded at in-clinic visits
- Mean change in Worst Pain Numeric Rating Score (WP-NRS) recorded at in-clinic visits
- Mean change in Worst Penetrative sex related Pain Numeric Rating Score (WPP-NRS) recorded at in-clinic visits
- Percentage of patients obtaining a ≥3-point improvement in WI-NRS recorded at in-clinic visits
- Percentage of patients obtaining a ≥3-point improvement in WP-NRS recorded at in-clinic visits
- Percentage of patients obtaining a ≥3-point improvement in WPP-NRS recorded at in-clinic visits
- Percentage of patients who reported an improvement* in Patient's Global Impression of Change (PGIC) for Worst Itch (WI), Worst Pain (WP) or Worst Penetrative sex related Pain (WPP)
- Percentage of patients who reported an Important improvement in Patient's Global Impression of Change (PGIC) for Worst Itch (WI), Worst Pain (WP) or Worst Penetrative sex related Pain (WPP)
- Mean change in VQLI total score
- Mean change in Vulvar lichen sclerosus Area and Sign Severity Index (VASSI)
- Mean change in VASSI Area scores
- Mean change in VASSI Sign scores
- Mean change in Clinician's Global Assessment (CGA) of VLS
- Percentage of patients with a ≥2-point improvement in Clinician's Global Assessment (CGA) of VLS
- Percentage of patients with an improvement* in the Clinician's Global Impression of Change (CGIC) assessment
- Burden of VLS on women's lives as reported on the "VLS patient journey sheet"

Note: * Improvement = 'Very much better', 'Much better' or 'A little better'

Patient reported outcomes (PRO) assessments

Worst Itch Numeric Rating Score (WI-NRS) of VLS

The patient's WI-NRS score must be recorded by the patient in a diary once daily (i.e., every day around the same timepoint, e.g. morning, afternoon or evening) for a total of 7 days (i.e. starting 7 days before the next in-clinic visit and ending on the day before the next in-clinic visit) leading up to each in-clinic visit day at Baseline, Week 1, Week 4, Week 8, and Week 12 or End of Treatment (EoT) using an 11-point numeric rating scale ranging from 0 to 10:
Please score the intensity of the Worst Itch associated with your lichen sclerosus, that you have experienced in the past 24 hours: 0 (no itch), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (worst imaginable itch).
In addition, the patient's WI-NRS score must be reported to the clinical staff at in-clinic visits at Screening 1, Baseline, Week 1, Week 4, Week 8, and Week 12 (or EoT) using the 11-point numeric rating scale above.

Worst Pain Numeric Rating Score (WP-NRS) of VLS

The patient's WP-NRS score must be recorded by the patient in a diary once daily (i.e., every day around the same timepoint, e.g. morning, afternoon or evening) for a total of 7 days (i.e. starting 7 days before the next in-clinic visit and ending on the day before the next in-clinic visit) leading up to each in-clinic visit day at Baseline, Week 1, Week 4, Week 8, and Week 12 (or EoT) using an 11-point numeric rating scale ranging from 0 to 10:
Please score the intensity of the Worst Pain associated with your lichen sclerosus, that you have experienced in the past 24 hours: 0 (no pain), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (worst imaginable pain).*
*Pain includes sensations such as Burning, Stinging, Soreness and Irritation.
In addition, the patient's WP-NRS score must be reported to the clinical staff at in-clinic visits at Screening 1, Baseline, Week 1, Week 4, Week 8, and Week 12 (or EoT) using the 11-point numeric rating scale above.

Worst Penetrative sex related Pain Numeric Rating Score (WPP-NRS) of VLS

At Screening 1, Baseline, Week 4, Week 8, and Week 12 (or EoT) visits all patients must indicate whether they are sexually active or not by answering Question 1 below. Patients who answer No to Question 1, must subsequently answer only Question 2. Patients who answer Yes to Question 1 must subsequently answer Questions 2, 3, and 4, and finally in Question 5 report their Worst Penetrative sex related Pain on an 11-point numeric rating scale ranging from 0 to 10 as also described below:

1. Have you had penetrative sex during the past 4 weeks? (Yes/No) If No, please answer Question 2 only. If Yes, please answer Questions 2, 3, 4 and 5.
2. Have you abstained from penetrative sex during the past 4 weeks due to your vulvar lichen sclerosus? (Yes/No)
3. How many times have you had penetrative sex during the past 4 weeks? (X times)
4. Have you used lubricants during penetrative sex during the past 4 weeks? (Yes/No)
5. Please score the intensity of the Worst Penetrative sex related Pain* you have experienced due to your lichen sclerosus in the past 4 weeks: 0 (no penetrative sex related pain), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (worst imaginable penetrative sex related pain).

*Pain includes sensations such as Burning, Stinging, Soreness and Irritation.

Patient's Global Impression of Change (PGIC) of VLS

PGIC in Worst Itch (WI), Worst Pain (WP) and Worst Penetrative sex related Pain (WPP) must be reported to the clinical staff at Week 4, Week 8, and Week 12 (or EoT) visits using a 7-item verbal rating scale:

Please choose the response below that best describes the overall change in your [Worst Itch*] associated with your lichen sclerosus since you started applying the trial medication:

- Very much better
- Much better
- A little better
- No change
- A little worse
- Much worse
- Very much worse

*Repeat with Worst Pain and Worst Penetrative sex related Pain.

Patient's rating of importance of improvements of VLS

Patients reporting an improvement in PGIC (i.e., 'Very much better', 'Much better' or 'A little better') must report the importance of the improvement at Week 4, Week 8, and Week 12 (or EoT) visits using a 2-item Importance Questionnaire:

Question 1: You have reported an improvement in your [Worst Itch*] associated with your lichen sclerosus since you started applying the trial medication. Was the improvement important to you? (Yes/No). If yes, please proceed to question 2.

Question 2: In which way(s) was the improvement in your [Worst Itch*] associated with your lichen sclerosus important to you? (*Free text*).

*Repeat with Worst Pain and Worst Penetrative sex related Pain.

Menopausal status and Menstrual cycle status

Menopausal status (Pre-menopausal or Menopausal) will be recorded for all women, while menstrual cycle status (i.e., whether the patient is menstruating or not, reason in case of no menstruation, number of days since the first day of the latest menstruation period, average duration of menstruation period in days, regularity of menstrual cycle) must be evaluated by all pre-menopausal women; both statuses will be reported at Baseline, and Week 12 (or EoT) visits.

International Consultation on Incontinence Questionnaire Urinary Incontinence Short Form (ICIQ-UI-SF)

Incontinence status must be evaluated by the patient at Baseline and Week 12 (or EoT) visits using the ICIQ-UI-SF questionnaire.

The ICIQ-UI-SF is a validated questionnaire with a recall period of 4 weeks comprising a total of 4 questions about when, how often and how much urine is leaked and how much urine leaking interferes with everyday life. The ICIQ-UI-SF score ranges from a maximum score of 21 (indicating maximum burden) and a minimum score of 0 (indicating no burden). A score ≥ 1 indicates urinary incontinence [6].

The ICIQ-UI-SF sum score is the sum of items 3-5 as follows:

Item 3. *How often do you leak urine?* (0=never; 1=about once a week or less; 2=two or three times a week; 3= about once a day; 4=several times a day; 5=all the time)
Item 4. *We would like to know how much urine you think leaks. How much urine do you usually leak?* (0=none; 2=a small amount; 4=a moderate amount; 6=a large amount)

Item 5. *Overall, how much does leaking urine interfere with your everyday life? Please ring a number between 0 (not at all) and 10 (a great deal)* (0-10).

Item 6 is the self-diagnostic question for type of UI and is not included in the final score:
When does urine leak? (please tick all that apply to you) (never, urine does not leak; leaks when you cough or sneeze; leaks when you are asleep; leaks when you are physically active/exercising; leaks when you are finished urinating and are dressed; leaks for no obvious reason).

Skindex-29

Skindex-29 is a skin specific multidimensional quality-of-life instrument. It must be evaluated by the patient, concentrating on their vulvar area, at Baseline, Week 4, Week 8, and Week 12 (or EoT) visits using the Skindex-29 Quality of Life Index questionnaire. It inquires how often (1. Never, 2. Rarely, 3. Sometimes, 4. Often, 5. All the time) during the previous 4 weeks the patient experienced the effect described in each item. Seven items address the Symptoms domain, ten items the Emotions domain, and twelve items the Functioning domain. All responses are transformed to a linear scale of 100, varying from 0 (no effect) to 100 (effect experienced all the time) as 25 times (score-1). Skindex-29 scores are reported as three scale scores, corresponding to the three domains; a scale score is the average of a patient's non-missing responses to items in a given domain [7]. The questionnaire comprises 30 items; however, Item 18 is excluded from the Skindex-29 score as per the guidelines of the originators of the test. If any scale has more than 25% of the responses missing, the scale is missing.

Table 1 Skindex-29 domains and their items

Functioning	**Emotions**	**Symptoms**
2. MSC affects how well I sleep	3. I worry that MSC may be serious	1. My skin hurts
4. MSC makes it hard to work or do hobbies	6. MSC makes me feel depressed	7. MSC burns or stings
5. MSC affects my social life	9. I worry about getting scars from MSC	10. My skin itches
8. I tend to stay at home because of MSC	12. I am ashamed of MSC	16. Water bothers MSC (bathing, washing hands)
11. MSC affects how close I can be with those I love	13. I worry that MSC may get worse	19. My skin is irritated
14. I tend to do things by myself because of MSC	15. I am angry about MSC	24. My skin is sensitive
17. MSC makes showing affection difficult	21. I am embarrassed by MSC	27. MSC bleeds
20. MSC affects my interactions with others	23. I am frustrated by MSC	
22. MSC is a problem for the people I love	26. I am humiliated by MSC	
25. MSC affects my desire to be with people	28. I am annoyed by MSC	
29. MSC interferes with my sex life		
30. MSC makes me tired		

MSC = My skin condition

VQLI

The Vulvar Quality of Life Index (VQLI) questionnaire is a vulvar specific quality-of-life instrument. It must be evaluated by the patient at Baseline, Week 4, Week 8, and Week 12 (or EoT) visits.

The questionnaire comprises a total of 15 questions with a recall period of 1 month under the following domains:

- Symptoms (Questions 1–2)
- Feelings and Emotions (Questions 3–5)
- Activities of daily living (Questions 6–10)
- Relationships (Question 11)

- Sexual function (Questions 12–13);
- Future health concerns (Question 14)
- Treatment (Question 15).

Each of the questions are graded zero to three on a Likert scale, with 0 representing no symptom or complaint ('Not at all' or 'Not applicable'), 1 ('A little'), 2 ('A lot') and 3 ('Very much'). The individual domain scores are the sum of the respective item scores (coded 0-3), while the total score is the sum of all 15 items, ranging from 0 to 45 [8].

VLS patient journey sheet

The Burden of living with VLS before entering the trial must be described by the patient by filling in the 'VLS patient journey sheet' at the Screening 1 visit. The patient journey sheet is a questionnaire which collects information about:

- which issues the VLS patient has experienced in relation to VLS within four areas: symptoms, psychological impacts, social impacts, and functional impacts (the issues are entered as free text)
- the patient can name up to four of the issues experienced in relation to VLS they have found most burdensome.

Clinician reported outcomes (ClinRO) assessments

Vulvar lichen sclerosus Area and Sign Severity Index (VASSI)

The Vulvar lichen sclerosus Area and Sign Severity Index (VASSI) evaluates the severity of 3 non-permanent signs of VLS in 5 different vulvar areas. It will be performed at Screening 1, Baseline, Week 1, Week 4, Week 8 and Week 12 (or EoT) visits for all patients.

Areas: The 5 different vulvar areas are Clitoris and periclitoral skin (**C**), Right interlabial sulcus and labium minus (**R**), Left interlabial sulcus and labium minus (**L**), Posterior fourchette and perineum (**P**), Anus and perianal skin (**A**))

Signs: The 3 non-permanent signs of VLS are **Fissures/erosions** (longitudinal/patchy ruptures of the skin surface), **Ecchymoses** (bleedings within the skin) and **Hyperkeratoses** (patches/plaques of bright white skin with a 'powdery' texture).

Severity: The severity must be scored on a 4-point scale from 0-3 (i.e., 0 (no lesion), 1 (length of lesion ≤5 mm), 2 (length of lesion > 5mm and ≤ 10 mm) or 3 (length of lesion > 10 mm)) for each non-permanent sign in each vulvar area (i.e., the light grey cells in Table 2). For endpoint analyses, the VASSI Area scores, VASSI Sign scores and Total VASSI score (i.e., the white cells in Table 2) will be calculated as follows:

1. A VASSI Area score (**$VASSI_{C, -R, -L, -P, -A}$**) is calculated by summing up the 3 severity scores for Fissures/erosions, Ecchymoses and Hyperkeratoses in the particular vulvar area [range 0-9].
2. A VASSI Sign score (**$VASSI_{F, -E, -H}$**) is calculated by summing up the severity scores for a particular sign across all 5 vulvar areas [range 0-15]
3. The Total VASSI (**VASSI**) score is calculated by summing the 5 VASSI Area scores or 3 VASSI Sign scores [range 0-45]

Table 2 Vulvar lichen sclerosus Area and Sign Severity Index (VASSI)

	Vulvar areas					
Signs	Clitoris and periclitoral skin (C)	Right interlabial sulcus and labium minus (R)	Left interlabial sulcus and labium minus (L)	Posterior fourchette and perineum (P)	Anus and perianal skin (A)	**VASSI Sign scores**
Fissures erosions (F)						$VASSI_F$
Ecchymoses (E)						$VASSI_E$
Hyperkeratoses (H)*						$VASSI_H$
VASSI Area scores	$VASSI_C$	$VASSI_R$	$VASSI_L$	$VASSI_P$	$VASSI_A$	**Total VASSI score**

Light grey cells must be filled in by the investigator with severity scores. The severity must be scored on a 4-point scale (i.e. 0 (no lesion), 1 (length of lesion ≤ 5 mm), 2 (length of lesion > 5 mm and ≤ 10 mm) or 3 (length of lesion > 10 mm) for each non-permanent sign in each vulvar area; White cells will be calculated.

Clinician's Global Assessment (CGA) of VLS

The Clinician's Global Assessment (CGA) of VLS evaluates the global (overall) severity of VLS in the entire vulvar region (i.e., including the clitoris, periclitoral skin, interlabial sulci, labia minora, posterior fourchette and perineum, anal and perianal areas). The CGA will be evaluated on a 5-point scale ranging from Clear to Severe at Screening 1, Baseline, Week 1, Week 4, Week 8, and Week 12 (or EoT) visits, using the following Table 3 below:

Table 3 Clinician's Global Assessment (CGA) of VLS

CGA Score	**Description**
Clear	No visible Non-permanent signs* of VLS (Fissures/Erosions, Ecchymoses or Hyperkeratoses) in the vulvar region. Note: Permanent signs# of VLS may be present
Almost clear	Barely visible Non-permanent signs* of VLS (Fissures/Erosions, Ecchymoses or Hyperkeratoses) in the vulvar region.
Mild	Slightly visible Non-permanent signs* of VLS (Fissures/Erosions, Ecchymoses or Hyperkeratoses) in the vulvar region
Moderate	Easily visible Non-permanent signs* of VLS (Fissures/Erosions, Ecchymoses or Hyperkeratoses) in the vulvar region.
Severe	Markedly visible Non-permanent signs* of VLS (Fissures/Erosions, Ecchymoses or Hyperkeratoses) in the vulvar region

*Non-permanent signs of VLS include fissures/erosions (longitudinal/ respectively patchy ruptures of the skin surface), ecchymoses (bleedings inside the skin) and hyperkeratoses (patches/plaques of bright white skin with a 'powdery' texture). Hyperkeratoses must be distinguished from Sclerosis (areas of yellowish/ivory white skin with a smooth/waxy/firm texture) and pallor (areas of pale whitish skin that differ from hyperkeratosis in that they are not "powdery")
Permanent signs of VLS include e.g., atrophy, sclerosis, pallor, scarring or fusion.

Clinician's Global Impression of Change (CGIC) of VLS

Clinician's Global Impression of Change (CGIC) in non-permanent signs of VLS including fissures/ erosions (longitudinal/respectively patchy ruptures of the skin surface), ecchymoses (bleedings inside the skin) and hyperkeratoses (patches/plaques of bright white skin with a 'powdery' texture) will be assessed at Week 1, Week 4, Week 8, and Week 12 (or EoT) visits using the below 7-item verbal rating scale.

The possible responses are as follows:

- Very much better
- Much better

- A little better
- No change
- A little worse
- Much worse
- Very much worse

2.4.2 Safety and Tolerability

Adverse Events

All adverse events (AE) are recorded with the following attributes:

- Start and End dates
- Ongoing status
- Date site aware of AE
- Severity (mild, moderate, severe)
- Relationship to study drug (not related, unlikely related, possibly related, probably related)
- Action taken with study drug (not changed, temporarily interrupted, withdrawn, not applicable – no study drug, unknown)
- Outcome (fatal, not recovered/not resolved, recovered/resolved, recovered/resolved with sequelae, recovering/resolving, unknown)
- Recovered/resolved with sequelae specified (free text)
- Was medication given to treat this AE?
- Serious AE (SAE)
- Criteria for SAE (death, life-threatening, inpatient hospitalisation or prolongation of existing hospitalisation, congenital anomaly or birth defect, important medical event)

Any AE is treatment-emergent, if fulfilling the definition in Section 4.4.5.

Any AE is a drug-related AE, if relationship to IMP is not assessed as *not related*.

Any AE is leading to drug withdrawal, if action taken with IMP is *drug withdrawn*.

Verbatim terms of the description of AEs will be coded according to MedDRA version 27.0 and classified according to primary system organ classes (SOC) and preferred term (PT).

Laboratory Parameters

There will be no scheduled sampling for clinical laboratory analyses in this trial, apart from the urine pregnancy tests mentioned below.

If there is a suspicion of a medical condition (considered related or not to the IMP), relevant laboratory samples may be taken and analysed at the discretion of the investigator; none of these samples will be stored for future analysis.

During the course of the trial, urine pregnancy tests will be performed at Screening 1, Baseline, Week 1, Week 4, Week 8, Week 12 (or EoT) visits, and at the follow-up visit (Week 14), if it is an in-clinic visit.

Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure, heart rate and temperature) will be recorded at Baseline (Week 0) and Week 12 (or Early termination (EoT)) visits. Vital signs are also recorded at Week 14 (Follow-up) if judged necessary by the investigator, and if it is an in-clinic visit. Any abnormalities must

be assessed as “clinically significant” or “not clinically significant” by the investigator. If a vital sign is both abnormal and clinically significant, an AE event must be reported.

Physical Examination

A complete (i.e., including all body systems as well as the vulvar region) physical examination is required at Screening 1 and Week 12 (or EoT) visits. An abbreviated physical examination (including the vulvar region as minimum; other body systems may be included depending on present signs, symptoms, or investigation outcomes) is required at the Baseline visit (and if judged necessary by the investigator) at the FU visit.
Any abnormalities must be assessed as either “clinically significant” or “not clinically significant” by the investigator. If a physical examination is both abnormal and clinically significant, an AE event must be reported.

2.4.3 Drug Accountability, Extent of Exposure and Study Treatment Compliance

Patients will at visit 6, visit 8 and visit 10 be asked to return the dispensed carton box with the used and unused tubes. All returned boxes with used and unused tubes will be weighed together to determine the amount of IMP used. The calculations for drug used will use: a) a carton box packed with 8 full labelled tubes weighs on average 658.112 g; b) a full labelled tube weighs on average 73.928 g.

The *amount of drug used* in grams is derived as the sum of the difference in drug weight dispensed and returned by kit; when there are tubes missing, assuming all missing tubes were fully used. This principle will be used to calculate the *total amount of drug used*.

The *duration of study treatment* is derived as the time in days from the date of first study drug application to the date of last study drug application. Per day, the number of toilet visits and the number of study drug applications are reported in a subject diary. If diary data is coming from separate diaries on the same day, it will be pooled to obtain the total number of toilet visits for that given day. *Compliant days* are days, where the number of toilet visits does not exceed the number of study drug applications by more than one. *Non-compliant days* are days of the treatment period, where the number of study drug applications is lower than the number of toilet visits minus 1, where the number or study drug applications or the number of toilet visits is missing, and where no diary entry is given. *Overall compliance (%)* is 100 times the number of compliant days divided by the duration of study treatment in days. Treatment compliance will be categorized to less than 80% versus 80% or more and to less than 70% versus 70% or more.

Duration of study treatment, number of compliant and non-compliant days, overall compliance, total amount of drug used, and weekly amount of drug used will be presented by summary tables by treatment group.

2.4.4 Concomitant Medication and Non-Drug Therapies

All medications and treatments (i.e., prescription drugs [including COVID-19 vaccination programs], over the counter [OTC] drugs and vitamins, herbal and dietary supplements, and lubricants used with sex), hereafter referred to as therapies, taken within 30 days prior to screening or during the trial will be recorded. Information about the total daily dose, route of administration, start and discontinuation dates, and indication are to be recorded in the eCRF.

All verbatim terms of drug and non-drug therapies will be encoded according to the WHO Drug Dictionary in its latest annual version and classified by primary Anatomic Therapeutic Chemical (ATC) class level 2 and PT.

A therapy or procedure is *prior* if starting and ending before or at the Baseline visit (Day 0). A therapy or procedure is *prior ongoing* if starting before or at the Baseline visit (Day 0) and ending after the Baseline

visit (Day 0) or ongoing at study completion. A therapy or procedure is *concomitant*, if starting at or after the Baseline visit and ending after the Baseline visit (Day 0) or ongoing at study completion.

Frequency counts and percentages of patients with at least one prior, prior ongoing, or concomitant therapy will be presented in separate summary tables by ATC level 2 and 4, PT, treatment group and overall, respectively.

2.5 Baseline Characteristics

2.5.1 Disease Characteristics

All patients are suffering from vulvar lichen sclerosus. At screening and/or at randomisation, the trial population is described by

- VLS History (diagnosis and symptoms (year, who made diagnosis, biopsy (y/n), family members with VLS, regular VLS treating physician visits, type of treating physician) at Screening 1 visit.
- In-clinic values for Worst Itch Numeric Rating Score (WI-NRS), Worst Pain Numeric Rating Score (WP-NRS), and Worst Penetrative sex related Pain Numeric Rating Score (WPP-NRS) at Screening 1 and Baseline visits; weekly diary means for Worst Itch Numeric Rating Score (WI-NRS), Worst Pain Numeric Rating Score (WP-NRS) at Baseline visit)
- Vulvar lichen sclerosus Area and Sign Severity Index (VASSI) at Screening 1 and Baseline visits
- Clinician's Global Assessment of VLS (CGA) at Screening 1 and Baseline visits
- Menopausal status at Baseline visit
- International Consultation on Incontinence Questionnaire Urinary Incontinence Short Form (ICIQ-UI-SF) at Baseline visit
- The Vulvar Quality of Life Index (VQLI) at Baseline visit
- Skin specific multidimensional quality-of-life (Skindex-29) at Baseline visit

Baseline disease characteristics will be presented by summary tables by treatment group and overall.

2.5.2 Demographics

At Screening, age, ethnic origin (Hispanic or Latino, or not), and race (White, Black or African America, Native Hawaiian or Other Pacific Islander, Asian, American Indian or Alaska Native, or Other) will be recorded.

Demographics will be presented by summary tables by treatment group, and overall, for the following populations: Full Analysis Set, Modified Full Analysis Set, Safety Analysis Set, and Per-Protocol Set.

2.5.3 Medical and Surgical History

Entries of medical and surgical history, including medical diagnoses and surgical procedures, are recorded with start and end times, as well as ongoing status, at the Screening 1 visit. Verbatim terms of diagnoses and surgeries will be coded according to MedDRA version 27.0 classified according to primary SOCs and PTs. Frequency counts and percentages of patients with at least one entry in the medical and surgical history and/or at least one concurrent diagnosis will be presented in summary tables by primary SOC, PT, treatment group, and overall. An entry in the medical history is a concurrent diagnosis if ongoing at the Screening 1 visit.

2.5.4 Prior, Prior Ongoing, and Concomitant Medication

A therapy or procedure is *prior* if starting and ending before or at the Baseline visit (Day 0). A therapy or procedure is *prior ongoing* if starting before or at the Baseline visit (Day 0) and ending after the Baseline

visit (Day 0) or ongoing at study completion. A therapy or procedure is *concomitant*, if starting at or after the Baseline visit and ending after the Baseline visit (Day 0) or ongoing at study completion.

Frequency counts and percentages of patients with at least one prior, prior ongoing, or concomitant therapy will be presented by summary tables by ATC level 2 and 4, PT, treatment group, and overall.

2.5.5 Vital Signs at Baseline

Systolic and diastolic blood pressure, heart rate, and body temperature at Week 0, as well as height and weight from demographics at the Screening 1 visit, along with body mass index, derived as weight in kg divided by the square of height in meters, will be presented in summary tables by treatment group, and overall.

2.5.6 Physical Examination at Baseline

A complete physical examination (including all body systems as well as the vulvar region) is required at Screening 1; the body systems are as follows:

- vulva region
- general appearance
- dermatological,
- thyroid,
- head, neck, throat
- heart / lungs
- abdomen
- lymph nodes
- musculoskeletal
- cardiovascular
- neurological

In each case, both abnormality and clinically significant status will be reported; these will be combined to create three categories: normal, abnormal/non-clinically significant, and abnormal/clinically significant. These will be presented in summary tables (frequency counts and percentages) by treatment group, and overall.

3 General Specifications for Analysis

3.1 Analysis Software

All analyzes will be carried out on the Statistical Analysis System (SAS®) Analytics Pro, version 9.4 including statistical procedures in the SAS/STAT® version 14.1 module [9].

3.2 Descriptive Statistics

All endpoints will be summarized using summary statistics for each treatment group, by visit, where appropriate.

3.2.1 Summarizing Categorical Data

Categorical including binary data will be summarized by frequency counts and percentages. The denominator of the percentages will be indicated if ambiguous. All categories will be displayed in summary tables including those with zero counts. A category of missing counts will be added. Per post-baseline visit,

summaries of secondary endpoints of improvement will also include p-values of exact binomial tests on improvement rates within each treatment group.

3.2.2 Summarizing Continuous Data

Continuous data including scores will be summarized by arithmetic mean, standard deviation (SD), median, quartiles, and ranges. Quartiles are given as *(lower quartile, upper quartile)* and ranges as *(minimum, maximum)*. The number of non-missing values will also be displayed. Per post-baseline visit, summaries of changes from baseline in primary and secondary endpoints, on WI-NRS and WP-NRS, will also include p-values of t-tests on the change within each treatment group.

3.3 Disposition of Patients

A patient is regarded as *screened* if they have signed informed consent and carried out any of the assessments at the Screening visit 1. A patient who fulfils the trial eligibility requirements (Visit 1 and the Baseline Visit) will be *randomised*, i.e., randomly assigned to treatment with IMP and given a randomisation number. A screened patient who either did not fulfill the inclusion and exclusion criteria, or withdrew from the trial before randomisation, is defined as a *screen failure*.

Patient disposition will be summarized by the number (frequency counts and percentages) of patients

- screened
- failed screening
- randomised
- started IMP treatment
- prematurely discontinued IMP treatment
- completed IMP treatment
- prematurely withdrawn from trial
- completed trial

Patient disposition will be summarized by treatment group and overall.

3.4 Premature Discontinuation of Treatment and/or Study Participation

On patient disposition (see Section 3.3) counts and percentages will also be given by reason for screen failure and premature discontinuation of IMP treatment and/or withdrawal from trial.

3.5 Protocol Deviations

Important protocol deviations may include, but are not limited to, IMP non-compliance, missing consecutive application days, and intake of prohibited concomitant therapies, which may have an impact on the primary endpoint. Important deviations related to study inclusion or exclusion criteria, conduct of the trial, or patient assessment, will also be reported.

All observed protocol deviations will be listed. Protocol deviations will be classified as major or minor in a minuted meeting for blind review of the database conducted prior to final database lock and unblinding. All protocol deviations and major protocol deviations will be presented in frequency tables.

4 Study Specifications for Analysis

4.1 Interim Analysis

The current trial is a phase 2 trial and is powered to detect statistical significance on the primary endpoint and to inform further clinical development. The expected recruitment period is relatively short, and the double-blind treatment period is only 12 weeks. Thus, no interim analysis is planned.

4.2 Analysis Populations

4.2.1 Screened Set (SS)

All patients who sign the informed consent form.

4.2.2 Full Analysis Set (FAS)

All patients who are randomised. Patients will be analysed according to the randomised treatment.

4.2.3 Modified Full Analysis Set (MFAS)

All patients who are randomised and returned for at least one post-baseline scheduled visit with data on weekly mean WI-NRS. Patients will be analysed according to the randomised treatment. Analyses with MFAS will only be carried out if the number of patients in the MFAS differs from the number of patients in the FAS.

4.2.4 Safety Analysis Set (SAS)

All patients who are randomised and dispensed the trial medication at Randomisation/Day 0, excluding patients to whom trial medication is not dispensed or who return all the trial medication unused. Patients will be analysed according to the actual treatment.

4.2.5 Per-Protocol Set (PPS)

A subset of the MFAS patients who completed the trial with WI-NRS data for Week 12 and are deemed to have no important protocol deviations that could interfere with the objectives of this trial. Important deviations of eligibility criteria and other deviations from the protocol will be assessed. Important deviations from the protocol may lead to exclusion of a patient or data points from the PPS, which could have interfered with the administration of the treatment or the evaluation of treatment effect. All such decisions will be identified and documented before the final trial database is unblinded. Important protocol deviations may include, but are not limited to, randomized treatment not taken, medication non-compliance (<80%), missing 3 or more consecutive application days, and intake of prohibited concomitant therapies, which may have an impact on the primary endpoint.

4.2.6 Usage of Analysis Populations

The SAS will be used for all safety analyses. FAS, MFAS, and PPS will be used for efficacy analyses. The SS will be used for analyses of disposition, screened failures, and general comments.

4.3 Subgroups and Confounding Factors for Analysis

The mixed model for repeated measures (MMRM) for continuous endpoints (primary and secondary) will be adjusted by menopausal status at baseline, ICIQ-UI-SF score at baseline, and the respective baseline

value of the endpoint being modelled. Similarly for the two binary secondary endpoints, which will be analyzed using a logic regression at Week 12, the models will be adjusted by menopausal status at baseline, ICIQ-UI-SF score at baseline, and the baseline value for each scale the binary endpoint is derived from (WI-NRS or WP-NRS).

4.4 Time Points and Time Periods

4.4.1 Periods, Visits and Study Days

The screening period will last from Screening 1 Visit to Baseline visit (Week 0, Day 0). The treatment period will last from randomisation (Week 0) to end of treatment (Week 12 or EoT). The follow-up period will last from end of treatment (Week 12 or EoT) to end of follow-up (Week 14).

A total of eleven visits are planned for each patient (six of these will be in-clinic, 4 phone calls, and the follow-up visit either in-clinic or a phone call). The in-clinic visits will be held at Screening 1, Baseline visit, Weeks 1, 4, 8, and 12, with a follow-up visit at Week 14 (either in-clinic or phone visit). The phone visits will be held at Screening 2, Weeks 3, 7 and 11. The reference date is the date of first treatment; from this, for each visit, actual study days will be derived as the time passed in number of days since the reference date plus 1. Further unscheduled visits may be done as deemed necessary.

4.4.2 Dates

Start and end dates, and the dates of last study drug application and last contact will be related to Day 1, given as *study day*. Negative study days indicate dates before the Baseline visit (Week 0)).

It is assumed that treatment in the treatment period occurs from the first dispense of IMP till the last IMP returned, unless otherwise reported.

4.4.3 Identification of Baseline Values

In general, baseline values for each variable are the values reported at baseline (Baseline visit, Day 0). If values at baseline are missing, values at the Screening 1 Visit will be used as baseline values, if available.

4.4.4 Definitions of Change from Baseline

Changes from baseline are defined as post-baseline value minus baseline value. Reductions are negative changes. Percentage changes from baseline are defined as 100 times changes from baseline divided by the baseline value.

4.4.5 Treatment Emergency of Adverse Events

In general, treatment emergent AEs are defined as events starting on or after the day of start of IMP that were absent pre-treatment, or events present prior to the first dose but increased in severity after the first dose. Otherwise, an AE is not treatment-emergent.

4.4.6 Treatment Relation of Concomitant Medication

A medication or therapy is related to treatment by timing. It is prior, prior ongoing, or concomitant as defined in Section 2.4.4 above.

4.5 Units and Unit Conversion

4.5.1 Date and Time Formats and Conversions

Dates are recorded by day, month and year. For imputation of partial dates, see Section 4.7.1 below. Days are calculated following rules in Sections 4.4.1, 4.4.2 and 4.6.1.

It is assumed that no unit conversion is required.

4.6 Data Derivations

For the derivation of total and sum scores in efficacy endpoints see Section 2.4.1 above.

4.6.1 Durations

Time durations, i.e. lengths of time periods, will be derived as end date minus start date plus 1 day in days.

4.6.2 Visit Windowing

Post-baseline in-clinic visits are to be held at Weeks 1, 4, 8, and 12 within the following days from target day: study day 7±1, 28±4, 28±4, 56±4, 84±4. In addition, unscheduled EoT visits will also be taken into account, if need be.

Non-overlapping visit windows are defined as follows:

Table 4 Analysis visit time windows

Assessments at following Weeks	Time window	Start of time window (study days)	End of time window (study days)
1, 4, 8, 12	Week 1 visit	1	17
	Week 4 visit	18	42
	Week 8 visit	43	70
	Week 12 visit	71	91
4, 8, 12	Week 4 visit	1	42
	Week 8 visit	43	70
	Week 12 visit	71	91

Primary and secondary efficacy endpoints will be summarized and analysed by *analysis visits*. For the visits shown in the table above, analysis visits are defined as the visits held closest to the target day within the analysis visit window as defined above. Any other visits within the same visit window will not be considered for analysis of primary or secondary efficacy endpoints. If the time distances of two visits from the target day are identical, the later visit is selected for analysis.

Other efficacy endpoints and safety endpoints will be summarized and analysed by nominal visits as appropriate.

4.6.3 Classifications

There are no data-specific classifications in this study (i.e. laboratory).

4.7 Handling of Missing Data and Data Imputation

4.7.1 Partial Dates

If a date relevant for analysis is only partially available, it will be imputed by the central day of the time period of uncertainty, if this rule will be in line with potential dependencies of other complete dates in chronology. The time period of uncertainty is, for example, the calendar month, if the day is missing, or the calendar year, if day and month are missing. If the imputed date must be before or after a reference date, or between two reference dates, and the central day of the time period of uncertainty would not be, the date will be imputed by the day of the time period of uncertainty closest to the central day but fulfilling the dependencies.

The approach of partial date imputation will be applied to visit dates, start and end dates of adverse events, start and end dates of concomitant medication, and the dates of last study drug application and of last contact, if partial.

4.7.2 Missing Data

Data is missing if unavailable. Data is invalid if not considered for analysis. Invalidity of data is defined below, where applicable. On some analyses as indicated below, missing or invalid data will be replaced by multiple imputation (MI).

For analyses on MFAS, PPS, and SAS, no imputations will be made for missing data. Of the three different analyses on FAS, in one analysis MI will be implemented for missing and invalid data.

Multiple Imputation

Post-baseline missing or invalid data will be multiply imputed within treatment groups in two steps [10]. At first, any potential intermediate missing data will be imputed 20 times using a Markov Chain Monte Carlo (MCMC) method to obtain a monotone missing data pattern. In a second step, imputation is done using a monotone regression method, fitting an analysis of covariance (ANCOVA) model stepwise visit by visit. The ANCOVA model adjusts for the effects of menopausal status at baseline, ICIQ-UI-SF score at baseline as factors, and the baseline value of the respective endpoint as a covariate. For each treatment group and each copy of the dataset, the estimated parameters and their variances are used to impute the endpoint. Starting with Week 1, the procedure is repeated on Week 4, Week 8 and Week 12, adding the possibly imputed values of the previous visits as covariates. On MI, missing score data is handled like missing continuous data.

Worst-case Imputation

For the secondary endpoints with 3- and 4 -point improvements in weekly mean WI-NRS and WP-NRS from the patient's diary, a *worst-case imputation* will be employed, where no improvement will be assumed for any missing endpoints.

5 Statistical Methods

5.1 Conventions for Inference

The two-sided level of significance is set to $\alpha = 0.05$, the one-sided level of significance to $\alpha = 0.025$.

5.2 Statistical Hypotheses

Let μ_M, and μ_V be estimators of treatment effect for each endpoint for MC2-25 cream, and MC2-25 vehicle. The following statistical hypothesis will be tested.

$$H_0: \mu_M = \mu_V \text{ versus } H_1: \mu_M \neq \mu_V$$

Statistical significance will be tested at the two-sided 5% significance level unless otherwise specified. Superiority is concluded if a p-value of statistical hypothesis testing is below the significance level and the estimate of treatment difference is in a favourable direction for MC2-25 cream. Estimates of treatment effects and differences from baseline at time points other than Week 12 will be reported exploratively.

5.3 Multiplicity Adjustment and Estimands

No multiple testing adjustment will be employed in this study.

The primary endpoint will be analyzed by different estimands, following a treatment policy strategy with imputation (primary estimand) (Section 4.7.2), a treatment policy strategy without imputation (secondary estimand), and a while on treatment strategy without imputation (tertiary estimand).

For the primary estimand, all available data will be used, regardless of treatment and treatment adherence, with missing data imputed as described in Section 4.7.2. For the secondary estimand, all available data will be used, regardless of treatment and treatment adherence, and missing data will not be imputed.

For the tertiary estimand, all available data will be included up to 7 days after the last treatment intake, regardless of treatment and treatment adherence, and missing data will not be imputed. Data more than 7 days after the last treatment intake is invalid and will not be used.

5.4 Approaches for Analysis

All collected, recorded, and derived data for analysis will be listed in individual patient data listings for all randomised patients. For both WI-NRS and WP-NRS, 4- and 3-point improvements, as well as complete response (score ≤1) will be flagged in listings; for WPP-NRS, a 3-point improvement will be flagged in a listing. For PGIC in Worst Itch (WI), Worst Pain (WP) and Worst Penetrative sex related Pain (WPP), as well as CGIC, an improvement (‘Very much better’, ‘Much better’ or ‘A little better’) will be flagged in listings. For CGA of VLS, a ≥2-point improvement will be flagged in a listing.

5.4.1 Binary Data

The binary secondary endpoints (Section 2.4.1) will be analysed using a logistic model with effects of treatment, menopausal status at baseline, ICIQ-UI-SF score at baseline, and the baseline value for each scale the binary endpoint is derived from (WI-NRS or WP-NRS) as factors. The estimated log odds ratios between the treatments at Week 12 will be derived together with the associated standard error and 95% confidence intervals. These binary endpoints will be summarized in descriptive tables and visualized in bar charts, in both cases stratified by treatment group and visit.

5.4.2 Categorical Data

Categorical data will be described by summary tables of counts and percentages. This includes zero counts of all possible but unobserved categories, if any.

5.4.3 Continuous Longitudinal Data

Continuous data will be described by summary tables of sample characteristics by visit and for changes from baseline for visits as defined in Section 3.2.2 above. Primary and secondary continuous endpoints will be visualized by box plots by treatment group and visit.

Mixed Model of Repeated Measures (MMRM)

Comparisons will be done using an MMRM model including treatment, visit, treatment-by-visit interaction, as factors, and menopausal status at baseline, ICIQ-UI-SF score at baseline, and the baseline value of the respective endpoint as covariates.

The statistical model is

Endpoint = β_0 + β_1○treatment + β_2○Visit + β_3○Treatment-by-visit + β_4○menopausal status + β_5○ICIQ-UI-SF score + β_6○baseline value + error,

where β_1 - β_6 are parameters of fixed effects and β_0 an intercept. The estimated treatment differences between MC2-25 cream and MC2-25 vehicle, will be reported together with the associated standard error, the 95% confidence interval, and a p-value of a test on the respective contrast corresponding to the hypothesis of no difference between treatments (superiority) (Section 5.2).

5.5 Application of Methods

The following subsections describe the application of methods per endpoint, summarized in Section 5.5.5.

Multiple Imputation Procedure

Applicable to all primary and secondary endpoints with MI, the procedural steps are as follows. On 20 multiple complete imputed datasets as produced according to the instructions in Section 4.7.2, each dataset will be analyzed the same way. After pooling the estimates and standard errors, results will be aggregated following Rubin's rule to yield one estimate and associated standard error, and to calculate the 95% confidence interval for the treatment difference or odds ratio between treatments, respectively [11]. In data listings, multiply imputed values are described by mean values.

5.5.1 Primary Endpoint

The primary endpoint is the mean change in weekly mean Worst Itch Numeric Rating Score (WI-NRS) recorded in the patient's diary from Baseline to Week 12.

Primary Analysis

The analysis is for superiority of MC2-25 cream versus MC2-25 vehicle using the treatment policy strategy on the FAS using MI to impute missing data. The estimated treatment difference will be reported together with the associated standard error, the 95% confidence interval, and the respective p-value of testing hypotheses H_0.

Sensitivity Analyses

Four sensitivity analyses are planned on the primary endpoint:

a) following the treatment policy strategy on the FAS without imputation
b) following the treatment policy strategy on the MFAS without imputation
c) following the treatment policy strategy on the PPS without imputation
d) following the while on treatment strategy on the FAS without imputation

Details are given in Sections 5.2, 5.3, and 4.7.2.

5.5.2 Secondary Endpoints

The secondary endpoints follow two quite different analysis strategies.

The two continuous endpoints (all defined from Baseline to Week 12):

- Mean change in weekly mean Worst Pain Numeric Rating Score (WP-NRS) recorded in the patient's diary (Weekly mean WP-NRS is calculated as the average of WP-NRS values recorded in the patient's diary for 7 days prior to the in-clinic visits.)
- Change in Skindex-29 domains

will be analysed in an identical fashion to the primary endpoint described in Section 5.5.1, which includes one analysis on the FAS with imputation, and a further four sensitivity analyses, all without imputation.

The four binary endpoints (all defined from Baseline to Week 12):

- Percentage of patients obtaining a ≥4-point improvement in weekly mean WI-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥4-point improvement in weekly mean WP-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥3-point improvement in weekly mean WI-NRS recorded in the patient's diary
- Percentage of patients obtaining a ≥3-point improvement in weekly mean WP-NRS recorded in the patient's diary

will be analysed using the treatment policy strategy on the FAS in a logistic regression model with a factor for treatment, and covariates menopausal status at baseline, ICIQ-UI-SF score at baseline, as well as the baseline value of the scale the binary endpoint is derived from (WI-NRS or WP-NRS). The estimated log odds ratios between treatments at Week 12 will be reported, along with the 95% confidence interval. The logistic regression models will be carried out on worst case imputation data, where missing endpoint data is assumed to be no response.

5.5.3 Other Endpoints

All other endpoints in Section 2.4.1 will be described descriptively in appropriate summary tables.

5.5.4 Safety Endpoints

Safety endpoints will be described by summary tables on the Safety Analysis Set (SAS).

Treatment-emergent AEs, treatment-emergent drug-related AEs, treatment-emergent serious AEs (SAE), treatment-emergent drug-related SAEs, treatment-emergent AEs leading to drug withdrawal and treatment-emergent drug-related SAEs leading to drug withdrawal will be presented in summary tables of the number and percentage of patients with at least one event, and the number of events, by primary SOC and by PT within SOC. For treatment-emergent AEs and treatment-emergent drug-related AEs, these summaries will also be provided by maximum severity. Maximum severity is defined by the most severe AE per patient, SOC and PT. An overview will summarize incidences in all listed AE categories including deaths.

Vital signs will be summarized by visit including changes from baseline, as well as a shift table with categories: normal, abnormal/non-clinically significant, and abnormal/clinically significant. For physical examination, frequency counts of the complete examination (including all body systems and the vulvar region) at Screening 1 and Week 12 (or EoT) visits as well as the results of the abbreviated physical examination at the baseline and FU visits (if judged necessary by the investigator), will be reported in both

a table with percentages as well as a shift table with categories: normal, abnormal/non-clinically significant, and abnormal/clinically significant. Frequency counts and percentage of urine pregnancy tests will also be presented.

5.5.5 Overview of Analyses

Analyses of other and safety endpoints are descriptive and exploratory. Planned analyses of primary, secondary and other endpoints for efficacy, including patient related outcomes, are summarized in the following table.

Table 5 Overview of primary and sensitivity analyses of endpoints for efficacy (with strategy, analysis set and imputation)

Sequence	Endpoint	Primary Analysis	Sensitivity Analyses
Primary	Change in weekly mean WI-NRS from patient's diary from Baseline to W12 from patient's diary	treatment policy, FAS, MI	1: treatment policy, FAS, 2: treatment policy, MFAS 3: treatment policy, PPS 4. while on treatment, FAS
Secondary	Change in weekly mean WP-NRS from patient's diary from Baseline to W12	treatment policy, FAS, MI	1: treatment policy, FAS, 2: treatment policy, MFAS 3: treatment policy, PPS 4. while on treatment, FAS
	Change in Skindex-29 domains from Baseline to W12	treatment policy, FAS, MI	1: treatment policy, FAS, 2: treatment policy, MFAS 3: treatment policy, PPS 4. while on treatment, FAS
	A ≥4-point improvement in weekly mean WI-NRS from patient's diary	treatment policy, FAS, worst case imputation	
	A ≥4-point improvement in weekly mean WP-NRS from patient's diary	treatment policy, FAS, worst case imputation	
	A ≥3-point improvement in weekly mean WI-NRS from patient's diary	treatment policy, FAS, worst case imputation	
	A ≥3-point improvement in weekly mean WP-NRS from patient's diary	treatment policy, FAS, worst case imputation	

W = Week. Worse case imputation = missing endpoint data is assumed to be no response

5.6 Changes and Clarifications to Analyses Planned in the Study Protocol

There are two changes to the sub-group analyses specified in section 9.4.6 in the protocol:

- The stratification by country will not be carried out as the study is carried in only one country
- Similarly, the stratification by site is deemed superfluous due to both the smaller sample size than expected, and possibly sites with very small sample sizes

In section 9.4.8 the exploratory endpoint burden living with VLS is only assessed at Screening and will only be listed, not described in summary tables.

Statistical within-treatment group pre-post testing was added to descriptive summary tables of selected endpoint as specified above.

5.7 Implementation

The SAP will be implemented by means of statistical programming specifications with technical details of implementation and the level of validation of programming. Derived datasets will be produced following raw-to-derived data specifications, tables, figures and listings (TFL) according to approved TFL shell versions.

6 References

[1] MC2 Therapeutics: Clinical Trial Protocol, Version 3.0, dated 23 August 2023.

[2] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (1995). Topic E3. Structure and Content of Clinical Study Reports.

[3] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2016). Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice. Topic E6(R2). Guideline for Good Clinical Practice.

[4] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (1998). Topic E9. Statistical Principles for Clinical Trials.

[5] MC2 Therapeutics: Case Report Form, Version 1.0, dated 9 October 2023.

[6] Avery, K., Donovan, J., Peters, T. J., Shaw, C., Gotoh, M., & Abrams, P. (2004). ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodyn*, *23*(4), 322-330. https://doi.org/10.1002/nau.20041.

[7] Chren, M. M., Lasek, R. J., Flocke, S. A., & Zyzanski, S. J. (1997). Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. *Arch Dermatol*, 133(11), 1433-1440, https://www.ncbi.nlm.nih.gov/pubmed/9371029.

[8] Saunderson, R. B., Harris, V., Yeh, R., Mallitt, K. A., & Fischer, G. (2020). Vulvar quality of life index (VQLI) - A simple tool to measure quality of life in patients with vulvar disease. *Australas J Dermatol*, *61*(2), 152-157. https://doi.org/10.1111/ajd.13235.

[9] SAS Institute Inc. SAS/STAT® 14.1 User's Guide. SAS Institute Inc. 2016.

[10] O'Kelly M, Ratitch B (2014): Clinical Trials with Missing Data: A Guide for Practitioners. Wiley, Chichester, UK.

[11] Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY, USA: Wiley & Sons; 1987.

7 Deliverables

Based on a TFL shells document, the following tables, figures and listings will be produced.

7.1 In-text Tables

In-text tables will be produced identical to selected post-text tables by content and in the layout of the report.

7.2 In-text Figures

In-text figures will be copied from selected post-text figures.

7.3 Post-text Tables

The following post-text tables will be produced. All tables will summarize results by treatment group.

Table 6 Table of Content of Post-Text Tables

14.1	**Disposition, Baseline Characteristics, Exposure and Concomitant Therapies**
Table 14.1.1	Patient Disposition - Summary (Screened Set)
Table 14.1.2	Analysis Sets - Summary (Full Analysis Set)
Table 14.1.3.1	Inclusion and Exclusion Criteria - Summary (Full Analysis Set)
Table 14.1.3.2	Protocol Deviations - Summary (Full Analysis Set)
Table 14.1.3.3	Major Protocol Deviations - Summary (Full Analysis Set)
Table 14.1.4.1	Demographics - Summary (Full Analysis Set)
Table 14.1.4.2	Demographics - Summary (Modified Full Analysis Set)
Table 14.1.4.3	Demographics - Summary (Safety Analysis Set)
Table 14.1.4.4	Demographics - Summary (Per-Protocol Set)
Table 14.1.5.1	Baseline Disease Characteristics - Summary (Full Analysis Set)
Table 14.1.5.2	Baseline Disease Characteristics - Summary (Modified Full Analysis Set)
Table 14.1.5.3	Baseline Disease Characteristics - Summary (Safety Analysis Set)
Table 14.1.5.4	Baseline Disease Characteristics - Summary (Per-Protocol Set)
Table 14.1.6	Vital Signs, Height, and Weight at Baseline - Summary (Full Analysis Set)
Table 14.1.7.1	Medical and Surgical History - Summary by SOC and PT (Full Analysis Set)
Table 14.1.7.2	Concurrent Diagnoses at Screening - Summary by SOC and PT (Full Analysis Set)
Table 14.1.8.1	Prior Drug and Non-Drug Therapies - Summary by ATC Class and PT (Full Analysis Set)
Table 14.1.8.2	Prior Ongoing Drug and Non-Drug Therapies - Summary by ATC Class and PT (Full Analysis Set)
Table 14.1.8.3	Concomitant Drug and Non-Drug Therapies - Summary by ATC Class and PT (Full Analysis Set)
Table 14.1.9	Exposure and Treatment Compliance - Summary (Full Analysis Set)
14.2	**Efficacy Data**
14.2.1	**Primary Efficacy Endpoint**
Table 14.2.1.1	Weekly mean WI-NRS from patient's diary - Summary by Visit, Observed Cases (Full Analysis Set)
Table 14.2.1.2	Weekly mean WI-NRS from patient's diary - Summary by Visit, Averaged Multiple Imputations, Treatment policy Strategy (Full Analysis Set)
Table 14.2.1.3	Weekly mean WI-NRS from patient's diary - Summary by Visit, Observed Cases (Modified Full Analysis Set)
Table 14.2.1.4	Weekly mean WI-NRS from patient's diary - Summary by Visit, Observed Cases (Per-Protocol Set)
Table 14.2.1.5	Weekly mean WI-NRS from patient's diary - Summary by Visit, Observed Cases, While on treatment Strategy (Full Analysis Set)
Table 14.2.1.6	Change in weekly mean WI-NRS from patient's diary at Week 12 - Primary Analysis using MI, Treatment policy Strategy (Full Analysis Set)
Table 14.2.1.7	Change in weekly mean WI-NRS from patient's diary at Week 12 - Sensitivity Analysis, Treatment policy Strategy (Full Analysis Set)
Table 14.2.1.8	Change in weekly mean WI-NRS from patient's diary at Week 12 - Sensitivity Analysis, Treatment policy Strategy (Modified Full Analysis Set)

7.4 Post-text Figures

The following post-text figures will be produced.

Table 7 Table of Content of Post-Text Figures

Figure 14.2.1.1	Weekly mean WI-NRS from patient's diary – Box Plot by Treatment and Visit, Treatment policy Strategy (Full Analysis Set)
Figure 14.2.2.1	Weekly mean WP-NRS from patient's diary – Box Plot by Treatment and Visit, Treatment policy Strategy (Full Analysis Set)
Figure 14.2.2.2	A ≥ 4-point improvement in weekly mean WI-NRS from patient's diary – Bar chart by Treatment and Visit (Full Analysis Set)
Figure 14.2.2.3	A ≥ 4-point improvement in weekly mean WP-NRS from patient's diary – Bar chart by Treatment and Visit (Full Analysis Set)
Figure 14.2.2.4	A ≥ 3-point improvement in weekly mean WI-NRS from patient's diary – Bar chart by Treatment and Visit (Full Analysis Set)
Figure 14.2.2.5	A ≥ 3-point improvement in weekly mean WP-NRS from patient's diary – Bar chart by Treatment and Visit (Full Analysis Set)

7.5 Statistical Output Documentation

Original SAS® outputs of inferential statistical procedures will be available upon request.

7.6 Data Listings

Table 8 Table of Content of Data Listings

16.1	
16.1.7	**Study Initiation and Randomization Scheme and Codes**
Listing 16.1.7.1	Informed Consent (Full Analysis Set)
Listing 16.1.7.2	Subject Enrollment and Randomization (Full Analysis Set)
16.2	**Patient Data Listings**
16.2.1	**End of Study**
Listing 16.2.1.1	Screen Failure (Screened Set)
Listing 16.2.1.2	Subject Disposition (Full Analysis Set)
16.2.2	**Protocol Deviations**
Listing 16.2.2	Protocol Deviations (Full Analysis Set)
16.2.3	**Analysis Populations**
Listing 16.2.3	Analyses Populations (Full Analysis Set)
16.2.4	**Baseline Characteristics**
16.2.4.1	**Demographics**
Listing 16.2.4.1	Demographic Data, Height, and Weight (Full Analysis Set)
16.2.4.2	**Diagnosis**
Listing 16.2.4.2	Details of VLS History (Full Analysis Set)
16.2.4.3	**Prior and Concomitant Therapy**
Listing 16.2.4.3	Prior and Concomitant Drug and Non-Drug Therapy – Details and Coding (Full Analysis Set)
16.2.4.4	**Medical History**
Listing 16.2.4.4	Medical and Surgical History – Details and Coding (Full Analysis Set)
16.2.5	**Other Collected Data**
Listing 16.2.5.1	Inclusion Criteria (Full Analysis Set)
Listing 16.2.5.2	Exclusion Criteria (Full Analysis Set)
Listing 16.2.5.3	Dates of Visits and Times Between Visits (Full Analysis Set)
Listing 16.2.5.4	IMP Assignment Handed Out by Visit (Full Analysis Set)
Listing 16.2.5.5	IMP Assignment Returned by Visit (Full Analysis Set)
Listing 16.2.5.6	Application Diary Week by Visit (Full Analysis Set)
Listing 16.2.5.7	Overall Study Drug Exposure and Compliance (Full Analysis Set)
Listing 16.2.5.8	General Comments (Screened Set)
16.2.6	**Individual Efficacy Data**
Listing 16.2.6.1.1	Weekly WI-NRS and WP-NRS from patient's diary by Visit (Full Analysis Set)
Listing 16.2.6.2.1	Weekly mean WI-NRS and WP-NRS calculated score from patient's diary by Visit (Full Analysis Set)
Listing 16.2.6.2.2	Skindex-29 individual answers by Visit (Full Analysis Set)
Listing 16.2.6.2.3	Skindex-29 domains by Visit (Full Analysis Set)
Listing 16.2.6.3.1	WI-NRS from in-clinic visits by Visit (Full Analysis Set)
Listing 16.2.6.3.2	WP-NRS from in-clinic visits by Visit (Full Analysis Set)
Listing 16.2.6.3.3	WPP-NRS from in-clinic visits by Visit (Full Analysis Set)
Listing 16.2.6.3.4	PGIC for WI, WP, and WPP by Visit (Full Analysis Set)
Listing 16.2.6.3.5	VQLI by Visit (Full Analysis Set)
Listing 16.2.6.3.6	Imaging of VLS by Visit (Full Analysis Set)

Listing 16.2.6.3.7	VASSI Total, Area, and Sign Scores by Visit (Full Analysis Set)
Listing 16.2.6.3.8	ICIQ-UI-SF by Visit (Full Analysis Set)
Listing 16.2.6.3.9	Menopausal status and Menstrual cycle status by Visit (Full Analysis Set)
Listing 16.2.6.3.10	CGA of VLS by Visit (Full Analysis Set)
Listing 16.2.6.3.11	CGIC by Visit (Full Analysis Set)
Listing 16.2.6.3.12	Burden of VLS on women's lives by Visit (Full Analysis Set)
D 16.2.7	**Safety & Tolerability Listings**
16.2.7.1	**Adverse Events Listings**
Listing 16.2.7.1.1	Adverse Events - Timing, Details, and Coding (Safety Analysis Set)
Listing 16.2.7.1.2	Serious Adverse Events - Timing, Details, and Coding (Safety Analysis Set)
Listing 16.2.7.1.3	Adverse Events Leading to Drug Withdrawal - Timing, Details, and Coding (Safety Analysis Set)
16.2.7.2	**Laboratory Parameters**
Listing 16.2.7.2	Urine Pregnancy Test by Visit (Safety Analysis Set)
16.2.7.3	**Vital Signs**
Listing 16.2.7.3	Vital Signs by Visit (Safety Analysis Set)
16.2.7.4	**Physical Examination**
Listing 16.2.7.4	Physical Examination by Visit (Safety Analysis Set)

Estimondo GmbH

Statistical Analysis Plan
for
Clinical Study Reporting

of the Phase II Clinical Study MC2-25-C3

A parallel group (2-arm), randomised, double-blind, 12-week trial to explore the efficacy and safety of MC2-25 cream and MC2-25 vehicle in women diagnosed with vulvar lichen sclerosus (VLS)

FINAL Version 1.0 as of 30 October 2024

Review and Approval Sheet

Reviewer and Approver:	Frank Freischläger Head Operations, Estimondo GmbH	05/11/2024 Frank Freischläger Frank Freischläger (Nov 5, 2024 13:53 GMT+1) Date, Signature
Client Approval:	Irene Sandholdt, Senior Project Manager, CroxxMed ApS	05/11/2024 Irene Sandholdt Date, Signature
Sponsor Approval:	Morten Præstegaard, Chief Operating Officer, MC2 Therapeutics	12/11/2024 Morten Præstegaard Date, Signature

MC2_25_C3_Final_SAP_v1_RAS

Final Audit Report 2024-11-12

Created:	2024-11-05
By:	Frank Freischläger (frank@freischlaeger.com)
Status:	Signed
Transaction ID:	CBJCHBCAABAAiQ4TjZpdUIAlC1Z-uTajlkW7NXzaFY_J

"MC2_25_C3_Final_SAP_v1_RAS" History

Document created by Frank Freischläger (frank@freischlaeger.com)
2024-11-05 - 12:52:14 PM GMT- IP address: 31.150.224.133

Document emailed to Frank Freischläger (frank.freischlaeger@estimondo.com) for signature
2024-11-05 - 12:52:17 PM GMT

Email viewed by Frank Freischläger (frank.freischlaeger@estimondo.com)
2024-11-05 - 12:53:04 PM GMT- IP address: 31.150.224.133

Document e-signed by Frank Freischläger (frank.freischlaeger@estimondo.com)
Signature Date: 2024-11-05 - 12:53:44 PM GMT - Time Source: server- IP address: 31.150.224.133

Document emailed to Irene Sandholdt (isa@croxxmed.com) for signature
2024-11-05 - 12:53:45 PM GMT

Email viewed by Irene Sandholdt (isa@croxxmed.com)
2024-11-05 - 2:09:13 PM GMT- IP address: 104.47.51.190

Document e-signed by Irene Sandholdt (isa@croxxmed.com)
Signature Date: 2024-11-05 - 2:19:21 PM GMT - Time Source: server- IP address: 77.72.48.114

Document emailed to Morten Præstegaard (mpr@mc2therapeutics.com) for signature
2024-11-05 - 2:19:22 PM GMT

Email viewed by Morten Præstegaard (mpr@mc2therapeutics.com)
2024-11-05 - 11:26:59 PM GMT- IP address: 172.225.208.14

Signer Morten Præstegaard (mpr@mc2therapeutics.com) entered name at signing as Morten Praestegaard
2024-11-12 - 10:54:08 AM GMT- IP address: 77.72.48.114

Document e-signed by Morten Praestegaard (mpr@mc2therapeutics.com)
Signature Date: 2024-11-12 - 10:54:10 AM GMT - Time Source: server- IP address: 77.72.48.114

Agreement completed.

2024-11-12 - 10:54:10 AM GMT