Appendix

Table of Contents

S1. Additional methods

Eligibility criteria

Inclusion criteria

In order to be eligible to participate in this study, a subject must meet all the following criteria:

1. Adults (Age 18 years or above)
2. Diagnosis and history of chronic, moderate-to-severe AD for at least 3 years before the screening visit.
3. Diagnosis and history of chronic AD (by the UK diagnostic criteria):
    a. Subjects must have an itchy skin condition in the last 12 months
    b. Plus, three or more or the following criteria:
        i. Onset below the age of two
        ii. History of flexural involvement
        iii. History of a generally dry skin
        iv. Personal history of other atopic diseases
        v. Visible flexural dermatitis
4. Currently treated with one of the JAK-1 inhibitors (upadacitinib/abrocitinib), and controlled disease for at least 12 weeks.
5. Signed and dated informed consent has been obtained prior to any protocol related procedures.
6. Willing and able to comply with the clinical study protocol.
7. Subjects able to read and understand, and willing to sign the informed consent form (ICF).

Exclusion criteria

A potential subject who meets any of the following criteria will be excluded from participation in this study:

1. Treatment with allergen immunotherapy within 6 months before the baseline visit.
2. Treatment with leukotriene inhibitors within 4 weeks before the baseline visit.
3. Treatment with systemic glucocorticosteroids within 4 weeks before the baseline visit.
4. Subjects who have received treatment with any non-marketed drug substance (that is, an agent which has not yet been made available for clinical use following registration) within 4 weeks prior to the baseline visit.
5. Chronic or acute infection requiring treatment with oral or intravenous antibiotics, anti-virals, anti-parasitics, anti-protozoals, or anti-fungals within 4 weeks before the baseline visit or superficial skin infections within 1 week before the baseline visit.
6. Current participation in any other interventional clinical trial.

S2. Additional methods

Table S1. Study design and overview of outcome measures

	Observation period	**Intervention period**			
		T0	**T1**	**T2**	**T3**
Outcome measures	Screening	Baseline	+4 weeks	+8 weeks	+12 weeks
Physician-reported					
Patient characteristics	X				
EASI	X	X	X		X
IGA	X	X	X		X
Patient-reported					
POEM	X	X	X	X	X
NRS Peak Pruritus	X	X	X	X	X
DLQI	X	X	X	X	X
ADCT	X	X	X	X	X
Other					
AEs	X	X	X	X	X
TCS use		X			X
Patient-reported healthcare utilization		X			X

Screening, T0, T1 and T3 represent visits to the outpatient clinic. T2 refers to a phone consultation. There is a four-week interval between the screening and baseline assessments. If patients maintained controlled disease at baseline, the observation period is followed by a reduction in the JAK1 inhibitor dose (intervention period). Patients may return to their original dose if disease is not controlled.

Abbreviations: ADCT, Atopic Dermatitis Control Tool; AEs, adverse events; EASI, Eczema Area and Severity Index; DLQI, Dermatology Life Quality Index; IGA, Investigator Global Assessment; NRS Peak Pruritus, Numeric Rating Scale Peak Pruritus of the last 24 hours; POEM, Patient Oriented Eczema Measure; T0, start of JAK1 dose reduction; T1, visit week 4; T2, visit week 8, T3, visit week 12; TCS, topical corticosteroids;

S3. Additional methods

Sample size calculation

The single proportion test was used to calculate the sample size. Based on expert opinion, an expected proportion of patients with sustained JAK1 inhibitor reduction at the end of follow-up of 20% was considered relevant. With an alpha of 0.05 and a margin of error of 10%, the study required a sample size of 62.

Sample size calculation was performed using a two-sided single proportion test with the following formula: $n = \left(\frac{Z^2 . \ p \ . \ (1-p)}{E^2}\right)$. In this formula, n represents the required sample size, M represents the margin of error, and p represents the estimated value of the proportion. Z represents the confidence level, which is set at 95%.

S4. Additional methods

Statistical analyses: Linear mixed-effects model

In this study, the independence assumption for running a linear regression model was not met as multiple observations were nested within the same patients. Therefore, a linear mixed-effects model was used to analyze mean changes in continuous outcome measures over time. Outcomes of interest were included as dependent variables, and scores at screening and time (Time_days, as continuous variable in days) as independent variables.

To determine the optimal model structure, we first examined whether natural cubic splines could better capture potential nonlinear trends in the data. First, we created a scatterplot over time to visually inspect the trajectory. If the scatterplot suggested a non-linear pattern, such as a noticeable change in trend at a specific time point, we included a spline with a single knot (to prevent overfitting). automatically determined using ns(Time_days, df=2). Nonlinear trends were observed in all outcomes, and cubic splines were included in all models.

Next we included a time-varying covariate (T_falen_timevarying_2) to indicate whether a patient was on the reduced dose or on the previous dose regimen. This time-varying covariate indicates the time point at which an individual transitions from the reduced dose to the previous dose. By using this covariate, we can capture dynamic changes over time. For patients who returned to their previous regimen, observations up to the time of return were included in the dose reduction analysis and shown in our manuscript. Observations after returning to the original dose were included in the analyses evaluating the effects of returning to the previous regimen. This figures are shown in the Supplementary Information (S11-S12).

Next, we include an interaction term between time and the time-varying covariate to explore whether the effect of time on scores varied.. However, the interaction was found to be non-significant in all models (likelihood ratio test), indicating that it did not contribute meaningfully to the models' explanatory power. As a result, we excluded this interaction from the final models to maintain simplicity, allowing for easier interpretation of the results.

A model with a random intercept and fixed slope was preferred to ensure model stability and interpretability. The final model for all outcomes was specified as follows (example of the ADCT model. Note that baseline_ADCT are the scores at 'screening'):

*model_ADCT<- lmer(ADCT ~ baseline_ADCT + ns(Time_days, df = 2) * T_falen_timevarying_2 + (1|participant_id),*

data=DOJAK_ADCT, REML=TRUE)

After selecting the right model, several assumptions of the linear mixed model were tested. First, the response variables were plotted against the fitted values to determine whether the response variable was a reasonable linear function. Second, QQ plots and histograms of the residuals were plotted to test for normality. If the residuals deviated from normality, transformations of the response variables were used to improve the model fit. Some observations contained values of zero. In contrast to square root transformations, logarithmic transformations require adding a small amount to numbers that are exactly zero. Therefore, square root transformations of dependent variables were used. Only the POEM score was not transformed because it met the assumption of normality.

Finally, the Estimated Marginal Means (EMMs) were calculated for different timepoints (day 0, 28, 56 and 84) and were transformed back to the original scale for interpretation. We performed pairwise

comparisons of the EMMs between specific timepoints, adjusting for multiple testing. Plots were created to visualize the predicted scores over time. Plots included a dashed horizontal line indicating the threshold value for mild disease and treat-to-target goals for AD.

S5. Additional methods

Statistical analyses: Continuation ratio model

To analyze the IGA scores, we used a continuation ratio model (CRM) to handle the ordinal nature of the data. The model estimates the odds of being in a higher category of the IGA score compared to the lower categories as a function of the covariates (cohort status, IGA scores at screening, time (continuous variable), time-varying covariate). A time-varying covariate was included in the model to indicate whether a patient was on the reduced dose or on their previous dose regimen. This time-varying covariate captures the time point at which an individual transitions from the reduced dose regimen to their previous dose regimen. For patients who returned to their previous regimen, observations up to the time of return were included in the dose reduction analysis and shown in our manuscript. Observations after returning to the original dose were included in the analyses evaluating the effects of returning to the previous regimen, and shown in the Supplementary Information (S11-S12). To fit the forward formulation of the CRM, the data were reconstructed by creating a dichotomous outcome[1]. In addition, a cohort variable was created that indicated the current condition of a cohort of subjects. For visualization, we generated effect plots to illustrate the marginal probabilities associated with IGA scores, while marginalizing over the random effects. The predicted values of each IGA score were transformed to obtain probabilities, and upper and lower confidence intervals were also calculated to illustrate the uncertainty around the predictions.

We constructed the model using the formula (note that baseline_IGA are the scores at 'screening'):

```
IGA <- mixed_model(IGA_NEW ~ cohort + Time_days + baseline_IGA + T_falen_timevarying_2,
                   random = ~1 | participant_id,
                   data = cr_data_IGA,
                   max_coef_value = 15,
                   family = binomial())
```

S6. Additional methods

Overview of R packages used

- lme4: Fitting mixed-effects models for both categorical (IGA) and continuous outcomes.
- haven: Importing SPSS data for analysis in R.
- dplyr: Data manipulation and transformation to prepare datasets
- devtools: Tools for developing R packages and installing development versions.
- purrr: Functional programming for working with lists and data frames.
- sjPlot: Generating tables and plots for model output visualization for all PROMs.
- GLMMadaptive: Fitting generalized linear mixed models, specifically for ordinal data.
- emmeans: Calculating estimated marginal means.
- ggplot2: Data visualization, for creating plots.
- ggeffects: Visualization of effects from model variables.
- rms: Tools for building and validating regression models. Useful for model selection and validation.
- Matrix: Advanced matrix operations and data structures.
- glmmTMB: Fitting mixed models with additional capabilities for various data types.
- tab_model: Displaying model output in table form. Helpful for presenting results.
- modelr: Making model and data analysis workflows easier.
- nlme: Fitting linear and nonlinear mixed models for continuous outcomes.
- effects: Visualizing effects of regression models.
- lattice: Creating graphs focused on multivariate data, for IGA scores.
- AICcmodavg: Calculating AICc and model averages for model selection. Helps in selecting the best models for both ordinal and continuous outcomes.
- lmerTest: Extension of lme4 for significance testing of model parameters in LMM.
- pbkrtest: Conducting custom hypothesis tests in mixed models.

- splines: Working with (cubic) splines in the LMM models. Applicable for modeling non-linear relationships in continuous outcomes.

S7. Additional results

Table S2. Outcomes at screening and baseline, in all patients (n= 60)

	Scores at screening	**Scores at baseline**
Outcomes		
EASI, median (IQR)	1.6 (0.5-2.9)	1.6 (0.8-2.6)
IGA, *n* (%)		
Clear	4 (7)	2 (3)
Almost clear	41 (68)	42 (70)
Mild	14 (23)	15 (25)
Moderate	1 (2)	1 (2)
Severe	0 (0)	0 (0)
POEM, median (IQR)	4.0 (2.0-7.8)	4.0 (1.3-8.0)
NRS Peak Pruritus, median (IQR)	2.0 (1.0-3.0)	2.0 (1.0-2.0)
DLQI, median (IQR)	2.0 (1.0-4.0)	2.0 (1.0-4.0)
ADCT, median (IQR)	3.0 (1.0-4.0)	2.0 (1.0-5.0)

Disease remained controlled in 60/60 patients after the four-week observation period (screening-baseline): Eczema Area and Severity Index (EASI; 0-72) ≤7, a Numeric Rating Scale Peak Pruritus of the last 24 hours (NRS Peak Pruritus; 0-10) ≤4, or an Atopic Dermatitis Control Tool (ADCT; 0-24) ≤6.

Abbreviations: ADCT, Atopic Dermatitis Control Tool; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, investigator Global Assessment; IQR, Interquartile Range; n, Number of patients; NRS Peak Pruritus, Numeric Rating Scale Peak Pruritus of the last 24 hours; POEM, Patient Oriented Eczema Measure;

S8. Additional results

Table S3. IGA probabilities

	Time (weeks)			
	Baseline (n=60)	Week 4 (n=56)	Week 8 (n=52)	Week 12 (n=48)
IGA score				
0	0.011255543	0.008467292	0.006365306	0.004782617
1	0.711784921	0.652884487	0.587447395	0.517876593
2	0.266896124	0.322441369	0.380680639	0.438201613
3	0.010063412	0.016206852	0.025506660	0.039139177

This table represents the estimated probabilities for each IGA score at different time points based on the continuation ratio model. A time-varying covariate was included in the model to indicate whether a patient was on the reduced dose or on their previous dose regimen. This time-varying covariate captures the time point at which an individual transitions from the reduced dose regimen to their previous dose regimen. For patients who returned to their previous regimen, observations up to the time of return were included in the dose reduction analysis. Abbreviations: IGA, Investigator Global Assessment.

S9. Additional results

Table S4. Adverse events before and after JAK1 inhibitor dose reduction, for all patients who maintained their reduced dose until the end of follow-up (n=48)

	Change in AEs over time					**New AEs**
	Baseline	**End of follow up**				
	Total (n=48)	Stable	Worsened	Decreased	Remission	
Adverse events, *n* (%)						
Acne*	16 (33)	6 (38)	0 (0)	5 (31)	5 (31)	2
Active infections						
Folliculitis	1 (2)	0 (0)	0 (0)	0 (0)	1 (100)	0
Herpes simplex infection**	6 (13)	1 (17)	1 (17)	4 (67)	0 (0)	1
Herpes zoster infection	1 (2)	0 (0)	0 (0)	0 (0)	1 (100)	1
Upper respiratory tract infection	1 (2)	1 (100)	0 (0)	0 (0)	0 (0)	0
Fatigue	4 (8)	2 (50)	0 (0)	2 (50)	0 (0)	0
Gastrointestinal disorders	1 (2)	0 (0)	0 (0)	0 (0)	1 (100)	0
Hair loss	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
Headache	3 (6)	1 (33)	0 (0)	1 (33)	1 (33)	1
Laboratory abnormalities						
Anemia	5 (10)	2 (40)	0 (0)	1 (20)	2 (40)	1
Hypercholesterolemia	4 (8)	2 (50)	1 (25)	0 (0)	1 (25)	0
Liver dysfunction	10	4 (40)	0 (0)	3 (30)	3 (30)	0
Kidney dysfunction	6 (13)	1 (17)	0 (0)	1 (17)	4 (67)	0
Thrombocytosis	8 (17)	4 (50)	0 (0)	2 (25)	2 (25)	1
Nausea	1 (2)	1 (100)	0 (0)	0 (0)	0 (0)	0
Weight gain	1 (2)	1 (100)	0 (0)	0 (0)	0 (0)	0
Weight loss	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0

New adverse events are defined as adverse events that emerged during the 12-week intervention period. Weight gain is defined as a gain of ≥5kg; Weight loss is defined as a loss of ≤5kg; Anemia is defined as Hemoglobin <8.6 mmol/L for men and <7.4 mmol/L for woman; Hypercholesterolemia is defined as cholesterol >6.7 mmol/L; Kidney dysfunction is defined as an increase of creatinine of 10%; Thrombocytosis is defined as thrombocytes >370x10^9/L; Liver dysfunction is defined as ALAT >35 U/L or GGT >35 U/L. Abbreviations: AEs, adverse events; n, number of patients.

*Acne was treated with metronidazole crème in three patients.
**Herpes simplex was treated with valaciclovir in four patients.

Table S5. Adverse events at baseline and week 12, for all patients who returned to their previous dose during follow-up (n=12)

	Change in AEs over time					**New AEs**
	Baseline	**End of follow up**				
	Total (n=12)	Stable	Worsened	Decreased	Remission	
Adverse events, *n* (%)						
Acne	5 (42)	3 (60)	1 (20)	1 (20)	0 (0)	1
Active infections						
Folliculitis	1 (8)	1 (100)	0 (0)	0 (0)	0 (0)	0
Herpes simplex infection*	2 (17)	1 (50)	1 (50)	0 (0)	0 (0)	0
Upper respiratory tract infection	2 (17)	2 (100)	0 (0)	0 (0)	0 (0)	0
Fatigue	1 (8)	1 (100)	0 (0)	0 (0)	0 (0)	0
Laboratory abnormalities						
Anemia	2 (17)	1 (50)	0 (0)	0 (0)	1 (50)	0
Liver dysfunction	2 (17)	2 (100)	0 (0)	0 (0)	0 (0)	1
Thrombocytosis	2 (17)	1 (50)	1 (50)	0 (0)	0 (0)	0
Nausea	1 (8)	0 (0)	0 (0)	1 (100)	0 (0)	0

New adverse events are defined as adverse events that emerged during the 12-week intervention period. Anemia is defined as Hemoglobin <8.6 mmol/L for men and <7.4 mmol/L for woman; Thrombocytosis is defined as thrombocytes >370x10^9/L; Liver dysfunction is defined as ALAT >35 U/L or GGT >35 U/L. Abbreviations: AEs, adverse events; n, number of patients.

*Herpes simplex was treated with valaciclovir in one patient

S10. Additional results

Table S6. Topical treatment use at baseline and week 12 for all patients who maintained their reduced dose until the end of follow-up (n=48) and for all patients who returned to their previous dose during follow-up (n=12)

	Baseline			**Week 12**		
	Total (n=60)	Reduced dose (n=48)	Returned to previous dose (n=12)	Total n=60)	Reduced dose (n=48)	Returned to previous dose (n=12)
No TCS	22 (37)	16 (33)	6 (50)	15 (25)	13 (27)	2 (17)
Class I	1 (2)	0 (0)	1 (8)	0 (0)	0 (0)	0 (0)
Class II	3 (5)	3 (6)	0 (0)	4 (7)	3 (6)	1 (8)
Class III	22 (37)	18 (38)	4 (33)	26 (43)	21 (44)	5 (42)
Class IV	8 (13)	8 (17)	0 (0)	10 (17)	8 (17)	2 (17)
Calcineurin inhibitor	4 (7)	3 (6)	1 (8)	5 (8)	3 (6)	2 (17)

Abbreviations: n, number.

Table S7. Shift in topical treatment use at week 12 for all patients who maintained their reduced dose until the end of follow-up (n=48) and for all patients who returned to their previous dose during follow-up (n=12)

	Week 12		
	Total (n=60)	Reduced dose (n=48)	Returned to previous dose (n=12)
Same class TCS	37 (62)	32 (67)	5 (42)
Higher class TCS	16 (27)	10 (21)	6 (50)
Lower class TCS	7 (12)	6 (13)	1 (8.3)

Abbreviations: n, number; TCS, topical steroid

S11. Additional results

Table S8. Patient reported healthcare utilization before and after JAK1 inhibitor dose reduction, for all patients who maintained their reduced dose until the end of follow-up (n=48)

	Time	
	Baseline	**End of follow up**
Consultations (study visits excluded), n (%)		
Dermatologist consultation	3 (6)	6 (13)
Other doctor consultation	1 (2)	0 (0)
Emergency department	0 (0)	0 (0)
AD-related out-of-hospital consultations, n (%)		
GP visit	1 (2)	1 (2)
Social worker visit	0 (0)	0 (0)
Medication		
Prescribed medication[a]	4 (8)	5 (10)
Non-prescribed medication[b] for AD complication	4 (8)	2 (4)
Work absenteeism	1 (2)	2 (4)
Out of the pocket costs	4 (8)	2 (4)

[a]Prescribed medication includes antihistamines, oral antibiotics and anti-virals.

[b]Non-prescribed medication includes non-prescribed emollients, antihistamines and lubricants.

[c]Out of the pocket costs include shower gels, scrubs, neutral detergent, shampoos.

Abbreviations: AD, atopic dermatitis; GP, general practice; Absenteeism in half days as reported by patients. The definition was based on an 8-hour workday.

Table S9. Patient reported healthcare utilization before and after JAK1 inhibitor dose reduction, for all patients who returned to their previous dose the end of follow-up (n=12)

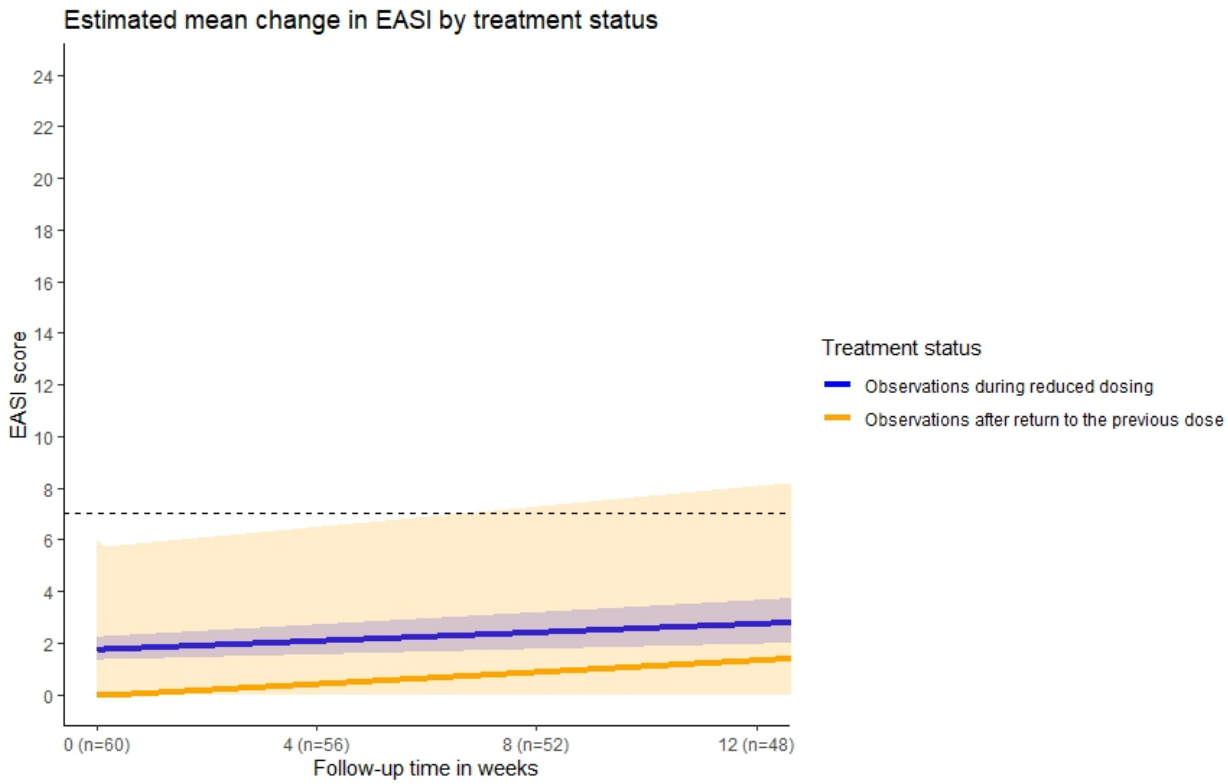
	Time	
	Baseline	**End of follow up**
Consultations (study visits excluded), n (%)		
Dermatologist consultation	0 (0)	2 (17)
Other doctor consultation	0 (0)	1 (8)
Emergency department	0 (0)	0 (0)
AD-related out-of-hospital consultations, n (%)		
GP visit	0 (0)	(0)
Social worker visit	0 (0)	(0)
Medication		
Prescribed medication[a]	0 (0)	2 (17)
Non-prescribed medication for AD complication	0 (0)	0 (0)
Work absenteeism	0 (0)	0 (0)
Out of the pocket costs	0 (0)	(0)

[a]Prescribed medication includes antihistamines and anti-virals

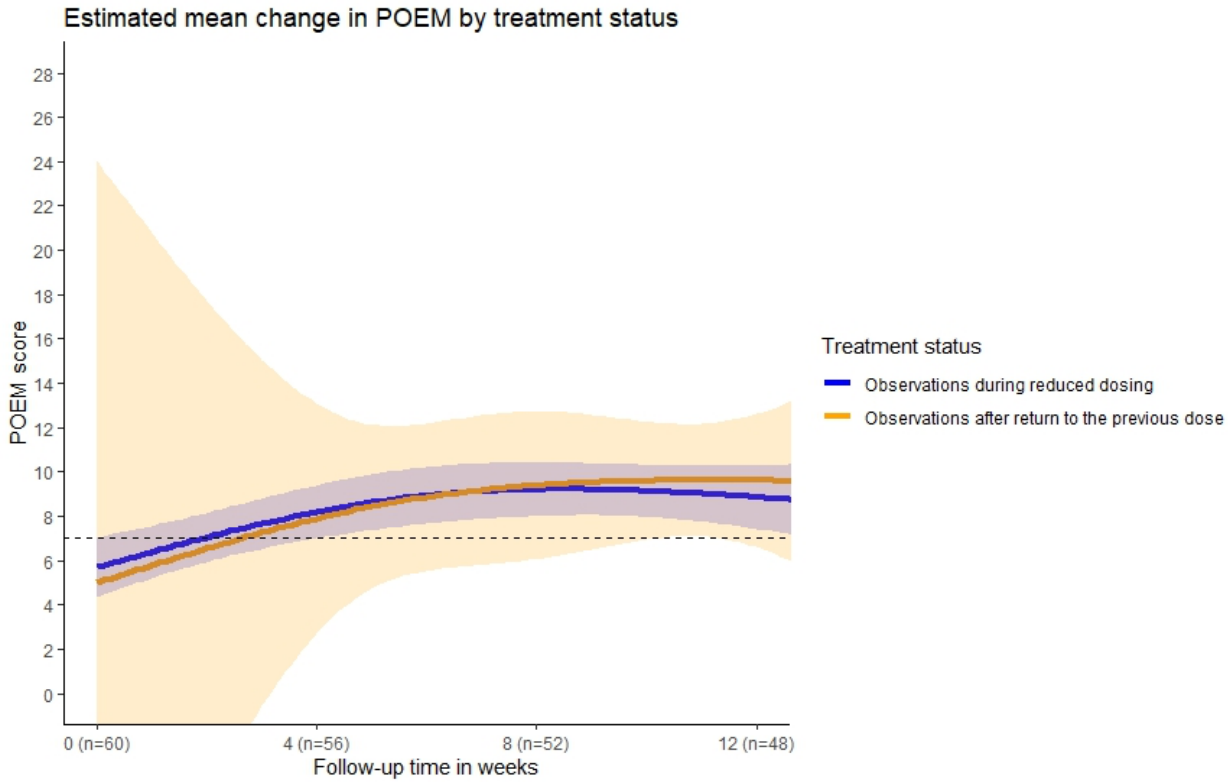
S12. Additional results

Figure S1

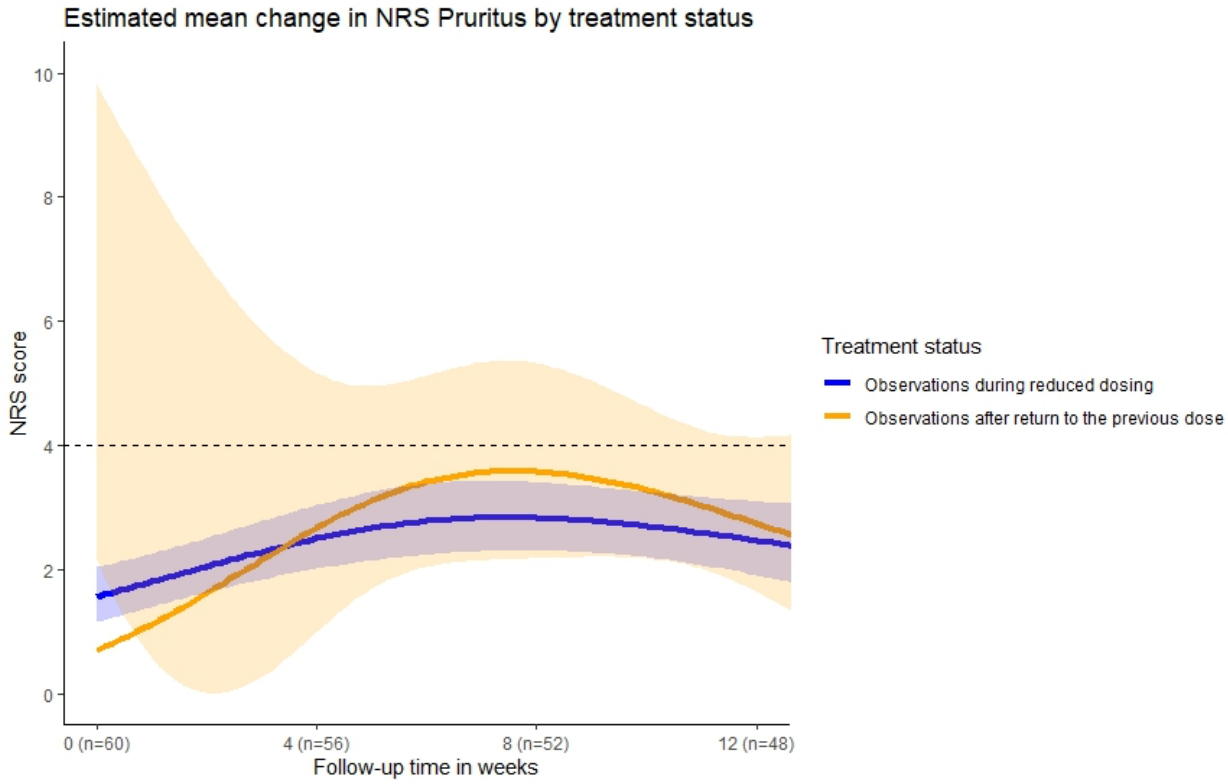
a.



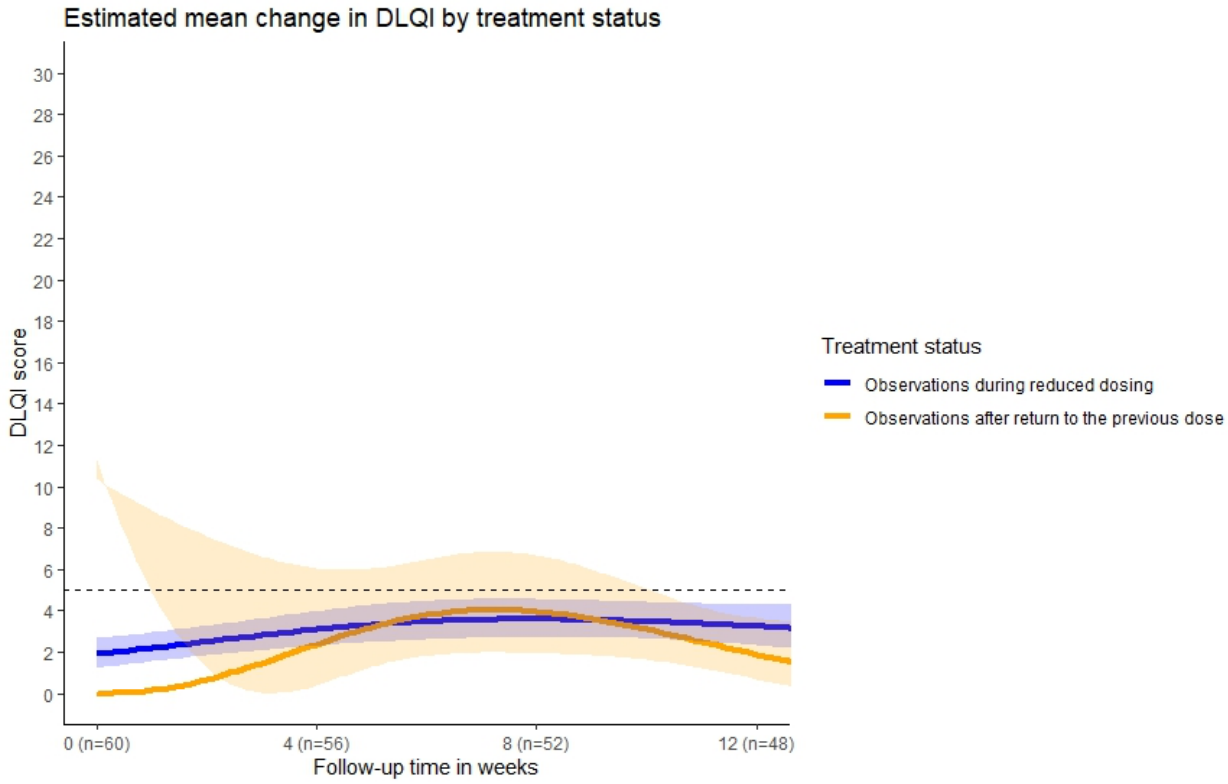
b.


Estimated mean change in POEM by treatment status
POEM score
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
0 (n=60)
4 (n=56)
8 (n=52)
12 (n=48)
Follow-up time in weeks
Treatment status
Observations during reduced dosing
Observations after return to the previous dose

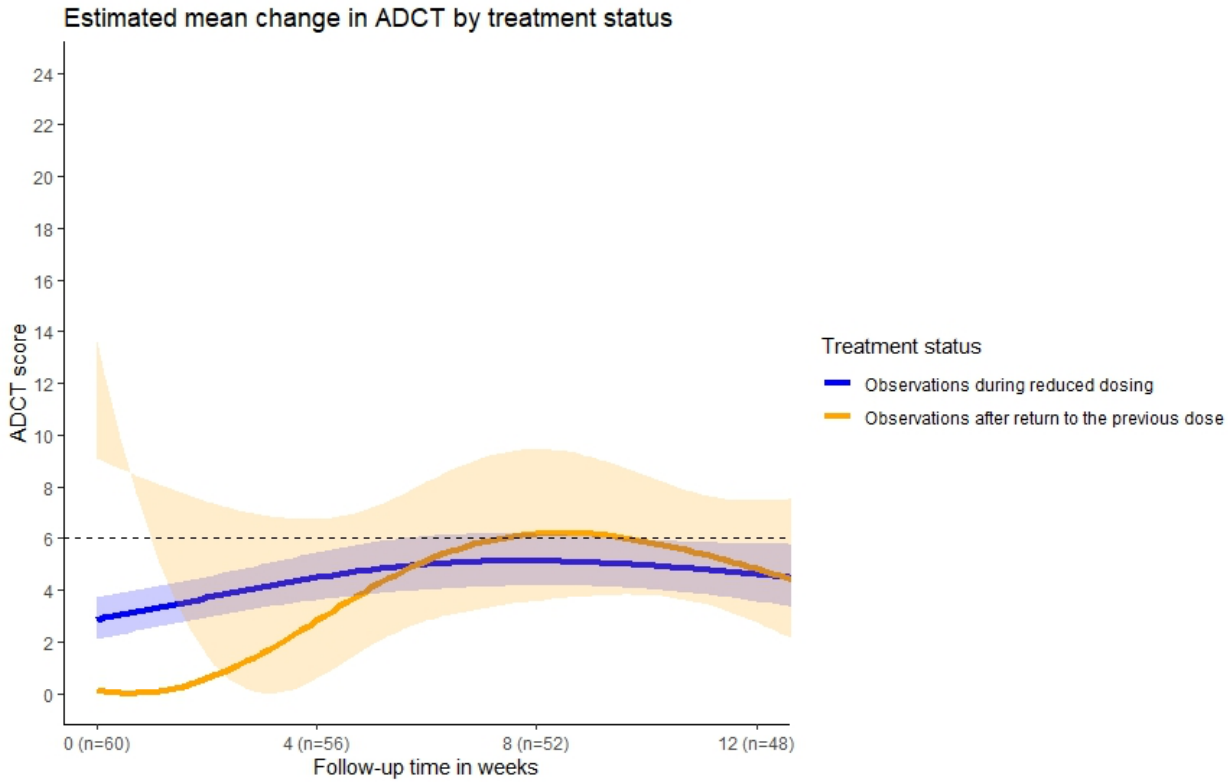
c.


Estimated mean change in NRS Pruritus by treatment status
NRS score
10
8
6
4
2
0
0 (n=60)
4 (n=56)
8 (n=52)
12 (n=48)
Follow-up time in weeks
Treatment status
Observations during reduced dosing
Observations after return to the previous dose

d.


Estimated mean change in DLQI by treatment status
DLQI score
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
0 (n=60)
4 (n=56)
8 (n=52)
12 (n=48)
Follow-up time in weeks
Treatment status
Observations during reduced dosing
Observations after return to the previous dose

e.

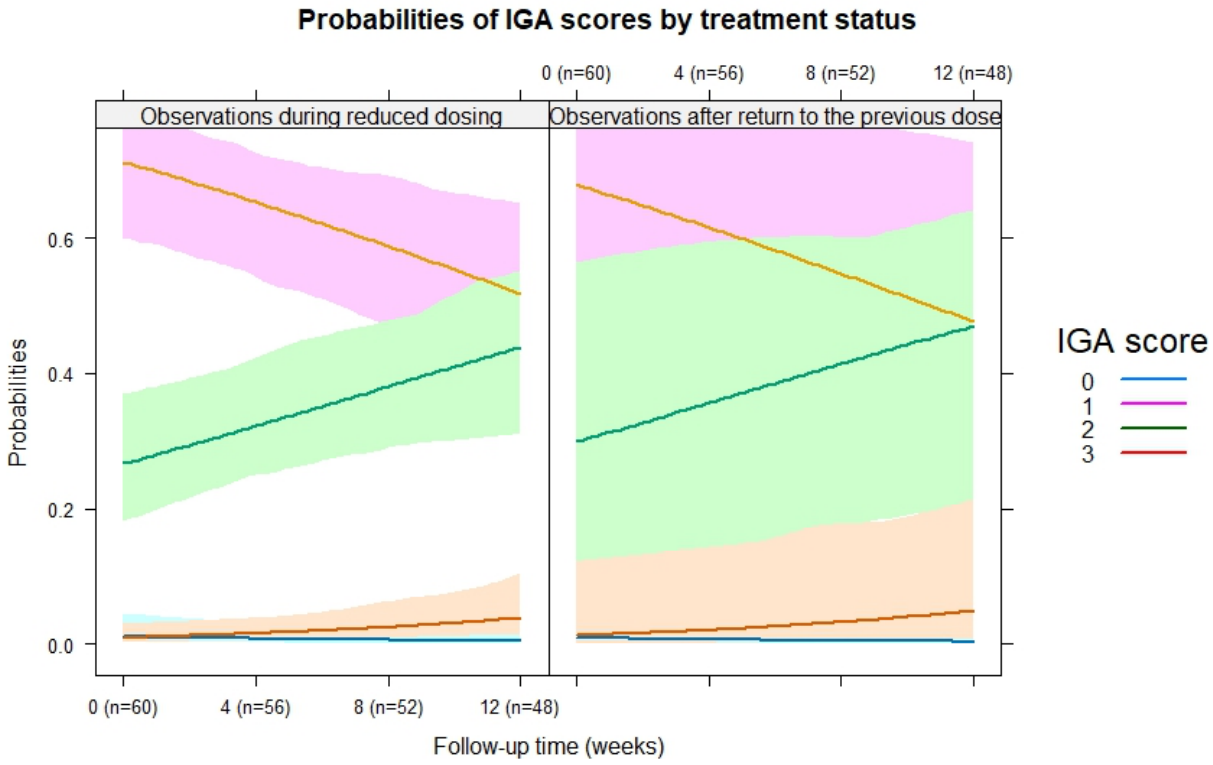


Estimated mean change in Eczema Area and Severity Index (EASI, 0-72) (Fig. 2a), Patient Oriented Eczema Measure (POEM, 0-28) (Fig. 2b), Numeric Rating Scale Peak Pruritus of the last 24 hours (NRS Peak Pruritus, 0-10) (Fig. 2c), Dermatology Life Quality Index (DLQI, 0-30) (Fig. 2d) and Atopic Dermatitis Control Tool (ADCT, 0-24) (Fig 2e). Linear-mixed effects models with cubic splines were used to model nonlinear changes over time. A time-varying covariate was included in the model to indicate whether a patient was on the reduced dose or on their previous dose regimen. This time-varying covariate captures the time point at which an individual transitions from the reduced dose regimen to their previous dose regimen. For patients who returned to their previous regimen, observations up to the time of return were included in the dose reduction analysis. Observations after returning to the original dose were included in the analyses evaluating the effects of returning to the previous regimen. The blue line represents all observations during the reduced dose over time. The orange line represents observations after returning to the previous regimen. The light blue and light orange shaded area indicate the 95% Confidence Interval (CI). Lower scores indicate mild disease activity. Horizontal lines are shown to represent thresholds for mild disease and treat-to-target goals for AD: EASI ≤7, POEM ≤7, NRS Peak

Pruritus ≤4, DLQI ≤5, ADCT ≤6. Disease activity decreased after returning to the previous dose, indicating that returning to the original dose had a positive effect on disease stability. Missing data in the dose reduction analyses: POEM, DLQI and ADCT, DLQI scores were missing at week 4 and week 8 (n=2). NRS Peak Pruritus was missing at week 8 (n=1). Missing data in the analysis after return to the previous dose: EASI and IGA scores were missing at week 12 (n=1). Abbreviations: n = number of patients on a reduced dose at each timepoint; CI, confidence interval.

S13. Additional results

Figure S2



IGA probabilities in the dose reduction analysis and analysis with observations after return to the previous dose over time. The figure illustrates the predicted marginal probabilities for each IGA score (0,1,2,3) at specified follow-up time points (baseline, week 4, week 8, week 12). The probabilities reflect the likelihood of patients being in each IGA category based on the continuation ratio model. A time-varying covariate was included in the model to indicate whether a patient was on the reduced dose or on their previous dose regimen. This time-varying covariate captures the time point at which an individual transitions from the reduced dose regimen to their previous dose regimen. For patients who returned to their previous regimen, observations up to the time of return were included in the dose reduction analysis. Observations after returning to the original dose were included in the analyses evaluating the effects of returning to the previous regimen. The data indicate that the probability of having an IGA 1 decreased over time, while the probability of an IGA 2 increased. IGA score 4 is not represented, as it was not observed in the study population. Missing data in the analysis after return to the previous dose: IGA scores were missing at week 12 (n=1). Abbreviations: IGA, Investigator Global Assessment.

References

1 Rizopoulos D. Mixed Models for Ordinal Data 2023. In. 2023.