BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058799
Article Type:	Original research
Date Submitted by the Author:	19-Nov-2021
Complete List of Authors:	Thomas, Caitlin; Evidera Raibouaa, Afaf; Eli Lilly and Company Wollenberg, Andreas ; University Hospital of Ludwig-Maximilian University of Munich, Department of Dermatology and Allergy Capron, Jean-Philippe; Eli Lilly and Company Krucien, Nicolas; Evidera Karn, Hayley; Evidera Tervonen, T; Evidera; University of Groningen, University Medical Center Groningen, Department of Epidemiology
Keywords:	DERMATOLOGY, Eczema < DERMATOLOGY, THERAPEUTICS, Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Caitlin Thomas [1], Afaf Raibouaa [2], Andreas Wollenberg [3], Jean-Philippe Capron [2], Nicolas Krucien [1], Hayley Karn [1], Tommi Tervonen [1, 4]

[1] Patient-Centered Research, Evidera, London, UK

[2] Eli Lilly & Co., Indianapolis, IN, USA

[3] Department of Dermatology and Allergy, University Hospital, Ludwig-Maximilian University of Munich, Germany

[4] Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Corresponding author: Tommi Tervonen, Evidera, The Ark, 201 Talgarth Rd, London W6 8DL, UK. Phone: +44 (0)7788 237 412. Email: tommi.tervonen@evidera.com

3332 words

ABSTRACT

Objectives Understanding patients' preferences for atopic dermatitis (AD) therapies, including new targeted therapies, can aid shared decision-making between clinicians and patients and support health technology assessments. We aimed to quantify patient preferences for efficacy, safety, and convenience features of AD treatments.

Design and setting Online discrete choice experiment (DCE) survey.

Participants Adults in the UK, France, and Spain who had used AD treatments during the past 2 years.

Primary and secondary outcome measures Preferences for attributes were analysed using a multinomial logit model. Willingness to make trade-offs was expressed as the maximum acceptable decrease (MAD) in the probability of achieving clear/almost clear skin at week 16.

Results The survey was completed by 404 patients (44.1±12.0 years; 65% female; 64% moderate/severe eczema; 68% naïve to self-injecting). Participants most valued increasing the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%, followed by reducing the risks of serious infections from 6% to 0% and of eye inflammation from 20% to 0%. Participants were willing to accept a decrease in the possibility of achieving clear/almost clear skin to obtain a treatment that can be paused (MAD = 24.1%), a once- or twice-daily oral pill over subcutaneous injection every 2 weeks (MAD = 16.6%), a 2-day over 2-week onset of action (MAD = 11.3%), and the ability to use the treatment for flare management (MAD = 5.8%).

Conclusions Although patients with AD most valued treatment benefits and risks, they were willing to tolerate reduced efficacy to obtain a rapid onset, oral administration, and a treatment that can be paused.

Keywords: Dermatology, Eczema, Health Economics, Therapeutics

Strengths and limitations of this study:

- This is the first study to elicit the preferences of patients from France and Spain for attributes of atopic dermatitis treatments
- Stated preferences were analysed using hypothetical scenarios with a fixed set of attributes, and patients may consider factors beyond the attributes included in this study when choosing a treatment
- Patients self-reported their diagnosis, and the patient sample included patients with and without experience of systemic treatments

INTRODUCTION

Atopic dermatitis (AD) is mostly treated using emollients and moisturizers, topical corticosteroids and calcineurin inhibitors, and, for severe cases, systemic immunosuppressants.[1, 2] However, emollients and moisturisers may not be sufficiently effective, and conventional systemic immunosuppressants have many potential side effects and are not generally recommended for long-term maintenance of AD.[3, 4] New targeted therapies for treating AD are now available. Dupilumab, a subcutaneously administered human monoclonal antibody inhibiting interleukin-4 and interleukin-13 signalling, was licensed in the US and the European Union in 2017 for the treatment of AD.[5] Baricitinib and upadacitinib, oral small-molecule inhibitors of Janus kinases, were recently licensed in the European Union for the treatment of moderate-to-severe AD in patients who are candidates for systemic therapy.[6, 7] Several additional targeted therapies are in development, including a variety of monoclonal antibodies inhibiting interleukin signalling.[1, 2, 8]

These new targeted therapies have different efficacy, risks, and non-clinical attributes, especially the mode of administration. Studies in other chronic diseases have shown that patients may prefer oral over parenteral treatment because they perceive some barriers to parenteral administration, which may lead to reduced adherence.[9-11] Because non-health benefits cannot be captured in traditional cost-effectiveness analysis, understanding to what extent they are valued by patients can help guide health technology assessment discussions[12-16] and inform shared decision-making at the point of care.[17]

Preferences for different treatment attributes, such as their benefits, risks, mode of administration, and convenience features, can be elicited from patients using discrete choice experiments (DCEs).[18] In DCEs, participants are presented with a series of tasks where they have to select between different hypothetical treatment options, each of which is composed of one level from each attribute in such a way that they are forced to make trade-offs, such as a higher risk of an adverse event but improved efficacy. DCEs have the advantage that the results can be used to quantify to what extent participants value each of the different attributes and estimate the trade-offs they would be willing to make. We hypothesized that patients with AD would not value all attributes relevant for their treatment choices equally. In the current study, we used a DCE to elicit the preferences of patients for key efficacy, safety, and convenience attributes of targeted AD therapies and examine the trade-offs they are willing to make between them.

MATERIALS AND METHODS

An online DCE survey was conducted between October and December 2019 in adults with AD living in the UK, France, or Spain. In the DCE survey, participants completed a series of choice tasks in which they selected between hypothetical treatment options described by a set of attributes with different levels. Treatment attributes and levels included in the DCE were identified through a targeted literature review of Embase and MEDLINE for quantitative and qualitative preference studies and a review of product labels for AD treatments (see **Online Supplemental Methods** for details). Attributes included the following: chance of achieving clear or almost clear skin at week 16, chance of achieving a meaningful reduction in itch at week 16, risk of eye inflammation, risk of serious infections, administration, flare management, long-term disease management, monitoring, and speed of onset (**Table 1**).

Table 1. Treatment attributes and levels included in the main discrete choice experiment

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Itch reduction	Eczema (Atopic Dermatitis) causes your skin to itch. Treatments for Eczema (Atopic Dermatitis) increase the probability of achieving a meaningful reduction in itch severity.	**2 out of 10 (20%):** There is a 20% chance of achieving a meaningful reduction in itch severity (reference level) **4 out of 10 (40%):** There is a 40% chance of achieving a meaningful reduction in itch severity **5 out of 10 (50%):** There is a 50% chance of achieving a meaningful reduction in itch severity
Skin appearance	Eczema (Atopic Dermatitis) affects the way your skin looks due to flaking, redness, swelling, oozing, crusting, bleeding. Treatment for Eczema (Atopic Dermatitis) may improve your skin condition, but different treatments have different impacts. In this survey, we will ask you to consider the chance of achieving clear skin after 16 weeks starting the treatment.	**1 out of 10 (10%):** After taking treatment for 16 weeks, there is a 10% chance you will have clear/almost-clear skin (reference level) **2 out of 10 (20%):** After taking treatment for 16 weeks, there is a 20% chance you will have clear/almost-clear skin **4 out of 10 (40%):** After taking treatment for 16 weeks, there is a 40% chance you will have clear/almost-clear skin

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Eye inflammation	All treatments have some risk of negative side effects. Some treatments can cause minor eye infections. You may have swollen eyelids, feel sensitivity to light, feel itching or burning in your eyes, or have pink discoloration of the white in your eyes. This can be treated but may require interruption to treatment. Other treatments do not increase your risk of getting an eye inflammation.	**0 out of 100 (0%):** Your treatment does not increase the chance of an eye inflammation **10 out of 100 (10%):** There is a 10% chance of experiencing an eye inflammation **20 out of 100 (20%):** There is a 20% chance of experiencing an eye inflammation (reference level)
Serious infections	All treatments have some risk of negative side effects. Some treatments reduce your immune system's effectiveness at fighting off illness and can result in serious infections, such as pneumonia or blood poisoning, that may require treatment and hospitalisation; you may be hospitalised for around one week. There is always a very low risk of serious infection and this low risk may be increased.	**0 out of 100 (0%):** Your treatment does not increase the risk of serious infection **3 out of 100 (3%):** 3 out of 100 people will experience a serious infection **6 out of 100 (6%):** 6 out of 100 people will experience a serious infection (reference level)
Speed of onset	All medications for Eczema (Atopic Dermatitis) take some time to start working. Some medications will start to work in 2 days, but others can take 1 or 2 weeks.	**2 days:** Your medication will begin to work 2 days after starting the treatment **1 week:** Your medication will begin to work one week after starting the treatment **2 weeks:** Your medication will begin to work two weeks after starting the treatment (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Flare management	For some treatments, your doctor can increase your dose if your symptoms get worse (flare-ups). After the flare is controlled, reducing the dose again may also be an option. However, other treatments cannot be adjusted in this way and you will remain on a fixed dose, even if your symptoms change.	**Yes:** Your doctor can increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves **No:** Your doctor cannot increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves (reference level)
Long-term disease management	Some treatments for Eczema (Atopic Dermatitis) need to be used continuously, without the option to stop and restart therapy when you want. Interruption of treatment, also known as a treatment holiday, can lead to a loss of efficacy over time. This means the therapy may not work as well when you restart treatment. These treatments must be used continuously and cannot be paused. Other treatments can be stopped and restarted (treatment holiday), with no impact on how effective the treatment is. Some treatments should not be used for the long-term, as they can have life threatening side effects, if used for a long period of time.	**Yes, with the possibility for pauses:** Treatment can be taken long term, and can be paused with no impact on how effective the treatment is **Yes, without the possibility for pauses:** Treatment can be taken long term, but must be taken continuously for there to be no impact on how effective the treatment is **Should not be used long-term:** You can pause the treatment, but using for the long-term may result in life threatening side effects (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Administration	Treatments are not all given/taken in the same way; for instance, some are pills, others are injections or topical creams. In this study we will only be considering pills and injections.	**Oral pill, once or twice daily** **Injection under the skin, every 2 weeks:** This is a subcutaneous injection, below the skin, but above muscle, usually injected into the thigh/stomach area. You can administer the injection yourself or a health care professional can administer it. If you choose to administer it yourself, you may need to be trained by a nurse on the injection technique. Treatment is once every two weeks. (reference level)
Check-ups	Some treatments require periodic blood tests taken by your doctor, because although you may not feel any symptoms, some Eczema (Atopic Dermatitis) medications can have a negative impact on your body.	**Frequent check-ups required:** Blood tests every 2 weeks during the initial 3 months of therapy and then monthly if the patient is stable (reference level). **Occasional check-ups required:** Blood tests at beginning of treatment, after 12 weeks, and then routinely, as determined by your doctor, while on treatment. **No check-ups required**

In each choice task, participants were asked to choose between different treatment options, each composed of one level from each of the attributes. Sensitivity of participants to changes in levels for each attribute were measured relative to the reference level, which is the level that patients least prefer. For example the reference level for risks is the highest level and for efficacy the reference level is the lowest level.

To ensure the feasibility and robustness of the DCE, cognitive pilot interviews were conducted in the UK, France, and Spain (n=5 per country). The cognitive pilot interviews examined whether the chosen attributes and levels were relevant, tradeable, and understandable to participants.[19] In addition, the cognitive pilot interviews assessed the complexity and clarity of the overall questionnaire. Each interview lasted approximately 60 min. Participants were provided a description of the study and completed the initial version of the study survey instrument online while sharing their screen with an interviewer. While participants completed the DCE, interviewers probed them using a semi-structured discussion guide. Patients were asked if they thought any attributes were missing that they would want to know about when selecting a treatment. No missing attributes were identified.

The online DCE survey was initially tested in 29 to 30 participants per country. Minor updates were made to the visual presentation of the survey. Recruitment targets were to include an additional 115 participants in the UK, 115 in Spain, and 85 in France.

Ethics approval

The study was conducted according to good practice for stated preference research[16] and was approved by Ethical & Independent Review Services (Independence, MO, USA; study number 19100-01). In addition, the study was conducted in accordance with International Council on Harmonisation Guidelines for Good Clinical Practice, the ethical principles of the Declaration of Helsinki, the European Union General Data Protection Regulation, and all local laws and regulations.

Participants

Participants were recruited via recruiter databases, social media, patient associations, and online patient panels. Adults (≥18 years) living in the UK, France, or Spain with a self-reported diagnosis of AD for ≥ 12 months were eligible if they had received a topical or systemic therapy for AD in the past 2 years. Participants also had to be able to speak, read, and write the official language of the respective country. Potential participants were excluded if they had a diagnosis of psoriasis, acne, lupus erythematosus, skin cancer, or any other condition that could interfere with participation in and completion of the interview. To account for the possibility that preferences differ between participants with and without self-injectable experience, the study was initially designed to include a target of 40% of participants with prior self-injectable experience, although this was reduced to 30% during the study to allow enough participants to be recruited.

All participants provided online informed consent before participating. Participants in the cognitive pilot consented to being audio-recorded. Participants were remunerated for completing the study.

DCE survey

The DCE was generated using Ngene software v1.2.1 (ChoiceMetrics, Sydney, Australia) using a D-efficient design that was assessed against good experimental design properties. The design was optimized for the estimation of a multinomial logit (MNL) model, and, where appropriate, directional priors. The experimental design of the DCE included 36 experimental choice tasks split into three blocks, such that each participant would complete only 12 experimental choice tasks. Participants in the pilot interviews did not struggle with the number of attributes in the choice tasks. Full profiles (where no attributes were fixed to a set level to simplify the design) were therefore used. In each choice task, participants were asked to choose between two hypothetical treatment options (A and B) and an opt-out of staying with their “old treatment”, wherein each treatment option was composed of one level from each of the attributes (**Figure 1**). If a participant selected the “old treatment” option, they answered a follow-up question asking them to choose between treatment options A and B. The order of the 12 experimental choice tasks and of the attribute groups (benefits, risks, other) within the choice options was randomised across participants to minimise the influence of ordering effects.[20, 21] In addition to the 12 experimental choice tasks, participants answered two choice tasks to assess internal validity.[22] Task 13 was a repeat of the third experimental choice task seen by the participant and was intended to check the stability of their choices. Task 14 was a dominated-choice test in which one treatment option was as good as or better than the other option for all attributes and was intended to test attendance to the tasks.

In addition to the DCE, participants completed a sociodemographic/clinical questionnaire and, the Set of Brief Screening Questions to assess health literacy[23] and five of the seven items from the Numeracy Scale to assess numeracy[24] to assess their ability to understand the attributes and levels presented and their engagement in the survey.

Validity assessments

For the dominance test, a respondent was considered to have failed the test if they chose the inferior (dominated) option as their preferred treatment. A respondent was considered to have failed the stability test if they made different choices in the initial and repeated tasks. A respondent was classified as a serial non-participant if they chose the same treatment option for all 12 experimental choice tasks. Decision-making was considered dominated when the respondent chose their preferred treatment option based on a single attribute in all 12 experimental choice tasks. For each choice task, response times in the lower 10% of the corresponding distribution were classified as fast and those in the upper 10% as slow. Attendance to the DCE survey was classified as inadequate if ≥80% of a participant's responses for the 12 experimental choice tasks were classified as too fast or too slow.

Statistical analysis

Statistical analysis was performed using R version 3.6.1 (R Foundation, Vienna, Austria). DCE preference data were analysed using a MNL model within the random utility maximization framework[25] (see **Online Supplemental Methods** for details). This model assumed that respondents chose the alternative that resulted in the highest utility (a measure of desirability) based on the included attributes and up to a random error.[26] The main results from this model were part-worth utility estimates, which reflect participants' sensitivities to changes in the treatment attributes. A dummy coding strategy was implemented to estimate preferences for discrete changes in the treatment attributes. In addition, the MNL model included two alternative-specific constants, one that captured left-right bias (tendency to select the option presented on the left of the choice tasks) and one that captured a preference for the old therapy option.

A second MNL model with linearly coded attributes for the skin appearance attribute was also estimated to support the computation of the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. The acceptability of the underlying hypothesis of linearity in preferences for changes in the skin appearance attribute was first verified (see **Online Supplemental Methods** for details). The MAD analysis measured the percentage decrease in the chance of achieving clear or almost clear skin at week 16 a respondent was willing to accept for changes in other attributes. The 95% confidence intervals for the MAD in achieving clear or almost clear skin at week 16 were obtained using the Delta method.[27]

Subgroup analyses were performed according to country (France, Spain, UK), age (<40, 40–50, and >50 years), gender (female, male), Patient Oriented Eczema

Measure (POEM) overall score (0–7 [clear or almost clear/mild], 8–16 [moderate], severe/very severe [17–28]),[28] and self-reported eczema severity (very mild/mild, moderate/severe/very severe).

Patient and public involvement

Cognitive pilot interviews were held with 15 patients to test understandability of the DCE survey. Other than participating in the DCE survey as respondents, patients were not involved in recruitment or study conduct. Investigators were blinded to the identities of the study participants, so the results of the study were not directly disseminated to them.

RESULTS

Participants

The DCE survey included 404 participants (114 in France, 145 in Spain, and 145 in the UK) who were recruited between October and December 2019. Most participants were female (65%) with an average age of 44.1 years (**Table 2**). Most participants were employed full time (56%) and had completed university education or higher (58%). The majority of participants had moderate-to-very severe AD according to POEM scores (62%) and self-reported eczema severity (67%) but good-to-excellent self-reported overall health (69%). Topical corticosteroids (66%) were the most frequently used class of medications at the time of the survey, followed by systemic immunosuppressant therapies (27%) and biologics (18%). Topical betamethasone (29%) and hydrocortisone (24%) were the most frequent currently used individual medications.

Table 2. Participant characteristics

Characteristic	N=404
Sex, n (%)	
Male	142 (35)
Female	262 (65)
Age, mean (SD)	44.1 (12.0)
Employment status	
Full time	227 (56)
Part time	75 (19)
Homemaker/housewife	21 (5)
Student	10 (2)
Unemployed	30 (7)
Retired	35 (9)
Disabled	12 (3)
Other	2 (0)
Education, n (%)	
No formal qualifications	1 (0)
Primary school or secondary education	38 (9)
College or some university	43 (11)
Completed vocational or professional certification	83 (21)
Completed university degree	148 (37)
Completed doctorate, post-doctorate, or equivalent	88 (22)
Other	3 (1)
Overall health, n (%)	
Excellent	20 (5)
Very good	96 (24)
Good	161 (40)
Fair	98 (24)
Poor	29 (7)
Prior experience with self-injectables (any)[*]	
Yes	129 (32)
No	275 (68)
Self-rated eczema severity, n (%)	
Very mild	19 (5)
Mild	116 (29)
Moderate	212 (52)
Severe	45 (11)
Very severe	12 (3)
POEM overall score, n (%)	
Clear or almost clear (0–2)	32 (8)
Mild eczema (3–7)	121 (30)
Moderate eczema (8–16)	192 (48)
Severe eczema (17–24)	47 (12)
Very severe eczema (25–28)	12 (3)
Class of AD medication currently used, n (%)†	
Topical corticosteroids	265 (66)
Topical calcineurin inhibitors	32 (8)
Phototherapy/UV treatment	20 (5)
Systemic immunosuppressant therapies	109 (27)

Characteristic	N=404
Biologics	72 (18)
Most frequently used current AD medications, n (%)†	
Betamethasone	119 (29)
Hydrocortisone	97 (24)
Prednisone	61 (15)
Clobetasol propionate	46 (11)

*Participants were not asked whether their prior use of self-injectables was for AD.
†Not mutually exclusive.
Abbreviations: AD, atopic dermatitis; POEM, Patient Oriented Eczema Measure; SD, standard deviation

Validity assessments

Overall, participants appeared to have paid adequate attention to the DCE choice tasks: 89% passed the dominance test, 64% chose the same answers in the repeated choice task, and 97% spent an adequate amount of time on the choice tasks (**Online Supplemental Table 1**). Also, for 63% of participants, decisions were not dominated by a single attribute, and only 5% always chose the opt-out old therapy option.

Overall preferences for treatment attributes

Of the treatment attributes included in the DCE survey, participants most valued improving symptoms and reducing the risk of side effects (**Figure 2** and **Online Supplemental Table 2**). The most valued change was an improvement from 20% to 50% in the chance of achieving a meaningful reduction in itch at week 16, although preferences did not significantly differ between an improvement to a 40% or 50% chance of achieving a meaningful reduction in itch. The next-most valued changes, in descending order, were a decrease in the risk of serious infections from 6% to 0%, a decrease in the risk of eye inflammation from 20% to 0%, and an improvement in the chance of achieving clear or almost clear skin from 10% to 40%.

Participants also valued changes in the non-clinical attributes. The most valued change was switching from a treatment that can be used long-term but cannot be paused without affecting efficacy to one that can be used long-term with the possibility for pauses, without affecting efficacy.

An oral pill once or twice daily was preferred over a subcutaneous treatment every 2 weeks, and a 2-day onset of action was preferred over a 2-week onset of action, although participants did not have a significant preference for a 1-week over a 2-week onset of action. Participants also preferred a treatment that can manage flares by modifying the dose according to symptoms over one that cannot be used to manage flares, although this was less important than changes in other non-clinical attributes.

Subgroup analyses

Results were similar for the three included countries (UK, Spain, and France) (**Online Supplemental Figure 1**), by age (**Online Supplemental Figure 2**), by gender (**Online Supplemental Figure 3**), by POEM overall score (**Online Supplemental Figure 4**), and by self-reported eczema severity (**Online Supplemental Figure 5**). However, participants who had experience of self-injecting were more willing to accept self-injection and placed less importance on reducing the risk of serious infections than those who did not have experience self-injecting (**Online Supplemental Figure 6**).

Willingness to make trade-offs between treatment attributes

Participants would be willing to tolerate reduced efficacy to obtain changes in other treatment attributes. Specifically, they would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 50.1% (95% CI, 38.5%–61.8%)

to increase the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%; 48.6% (95% CI, 35.2%–62.0%) to reduce the risk of serious infections from 6% to 0%; and 42.3% (95% CI, 30.0%–54.5%) to reduce the risk of eye inflammation from 20% to 0% (**Table 3**). They would also be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 24.1% (95% CI, 16.5%–31.6%) to switch from a treatment that can be used long-term but cannot be paused without losing efficacy to one that can be paused without losing efficacy; 16.6% (95% CI, 9.2%–24.0%) to switch from a subcutaneous treatment every 2 weeks to an oral pill once or twice daily; and 5.8% (95% CI, 0.5%–11.1%) to obtain a treatment whose dosage can be modified to manage flares over one that cannot. Further, they would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 20.9% (95% CI, 12.3%–29.5%) to switch from a treatment that requires frequent check-ups to one that does not require check-ups; and 16.1% (95% CI, 8.7%–23.5%) to switch from a treatment that requires frequent check-ups to one that requires occasional check-ups.

Table 3. Maximum acceptable decrease in the probability of achieving clear or almost clear skin at week 16

Attribute/level	**Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)**
Itch reduction	
2 out of 10 (20%)	Reference
4 out of 10 (40%)	38.7 (28.8, 48.6)
5 out of 10 (50%)	50.1 (38.5, 61.8)
Eye inflammation	
20 out of 100 (20%)	Reference
10 out of 100 (10%)	17.9 (10.5, 25.4)
0 out of 100 (0%)	42.3 (30.0, 54.5)
Serious infections	
6 out of 100 (6%)	Reference
3 out of 100 (3%)	20.6 (12.7, 28.6)
0 out of 100 (0%)	48.6 (35.2, 62.0)
Speed of onset	

Attribute/level	Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)
2 weeks	Reference
1 week	0.2 (−6.5, 6.9)
2 days	11.3 (4.4, 18.2)
Flare management	
No	Reference
Yes	5.8 (0.5, 11.1)
Long-term disease management	
Yes, without the possibility for pauses	Reference
Should not be used long-term	4.3 (−2.7, 11.3)
Yes, with the possibility for pauses	24.1 (16.5, 31.6)
Administration	
Injection under the skin every 2 weeks	Reference
Oral pill once or twice daily	16.6 (9.2, 24.0)
Check-ups	
Frequent check-ups required	Reference
Occasional check-ups required	16.1 (8.7, 23.5)
No check-ups required	20.9 (12.3, 29.5)

Abbreviations: CI, confidence interval

DISCUSSION

The current study, which included 404 participants across the UK, France, and Spain, was the largest to date to examine the preferences of patients with AD and the first to include samples from France and Spain. It showed that adults with AD who had recently been treated with topical and/or systemic therapy most valued increasing the benefits and reducing the risks of their treatments, although attributes specific to new targeted therapies, such as mode of administration and long-term disease management, also had a significant effect on choices. Participants were willing to tolerate a significant decrease in the possibility of achieving clear or almost clear skin to obtain a treatment that is more convenient, including an oral pill once or twice daily in place of a subcutaneous injection every 2 weeks, the ability to pause the treatment without losing efficacy, the ability to modify the dosage to manage flares, and the possibility of requiring only occasional or no check-ups instead of frequent check-ups. Further, participants with self-injectable experience were more willing to accept self-injection than participants without self-injectable experience. Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age, sex, or disease severity.

Two other recent DCEs have examined the treatment preferences of patients with AD. A DCE in the US and UK including 320 adults with moderate-to-severe AD[29] found, as in the current study, that patients preferred an oral pill over subcutaneous injection and valued a rapid onset of action and increasing the chance of achieving clear or almost clear skin at week 16. A DCE including 323 patients in Japan ≥ 15 years of age with moderate to very severe AD and 121 dermatologists treating patients with AD[30] found that, as in the current study, both groups considered benefits and adverse effects the most important attributes of injectable treatments, although preferences for some treatment attributes differed between the groups. For example, patients placed more value on efficacy of improving rashes and treatment costs than dermatologists, while dermatologists valued time until response more than patients. Patients also preferred adding new treatments to current treatments as add-ons and receiving treatments at clinics, while physicians preferred reducing the number of current treatments and having patients self-administer at home. These differences in the preferences of patients and physicians emphasize the need for studies like the current one that are specifically designed to provide insight into patients' preferences.

Internal validity of the current DCE was examined using tests of choice stability and dominance, as well as by considering response times, health literacy, and numeracy.[22] The results were in line with existing research[31] and suggested that participants paid adequate attention to the survey. A potential limitation of this study is that the attributes and levels were not identified through a separate qualitative research phase but rather through a targeted review of previous quantitative and qualitative studies of patients with AD and product labels for AD treatments. We do not expect that this influenced the results because the same attributes (onset of itch relief, probability of skin clearance, frequency or ease of administration/convenience, and safety) were also identified through the qualitative phase of the US/UK study.[29]

Due to the need to limit the participants' cognitive burden, not all potentially relevant attributes could be included in the DCE survey. However, cognitive pilot interviews of 15 patients with AD indicated that the attributes and levels were relevant and that no attributes were missing. Overall, participants also found the length and complexity of the survey acceptable. A further limitation is the inclusion of patients with non-severe AD, who would possibly not receive systemic therapies.[2] However, there is value in including these patients, because patients' disease severity may vary over time and treatment recommendations may change. Finally, although few differences were found in preferences by age, sex, or country, care should be taken when generalizing to underrepresented AD populations, such as patients with very severe AD, children, or patients in lower income countries. Moreover, our sample included a high proportion of participants with university education and may therefore not be fully representative of the general AD population.

In conclusion, this study showed that patients with AD most valued treatment benefits and reducing risks but were willing to accept a decrease in efficacy, as measured by the possibility of obtaining clear or almost clear skin at week 16, to obtain an oral treatment with a rapid onset of action. This information may help clinicians make shared decisions with patients about the most suitable treatment for

AD. It can also support reimbursement applications, ensuring that health technology assessment decisions align with the preferences of individuals living with AD.

Author contributions: C.T. contributed to conception and planning of the study, acquisition and interpretation of the data, and drafting and critical revision of the manuscript. A.R. contributed to conception and design of the study, interpretation of the data, and critical revision of the manuscript. A.W. and J.-P.C. contributed to design of the study, interpretation of the data, and critical revision of the manuscript. N.K. contributed to analysis and interpretation of the data and critical revision of the manuscript. H.K. contributed to conception and design of the study and critical revision of the manuscript. T.T. contributed to conception and planning of the study, interpretation of the data, and critical revision the manuscript. All authors approved the final version of the manuscript.

Acknowledgements: Medical writing support was provided by Philip S. Leventhal, PhD, Holly Richendrfer, PhD, and Stephen Gilliver, PhD of Evidera and was paid for by Eli Lilly & Co.

Competing Interests: C.T., N.K., H.K., and T.T. are employees of Evidera, which was paid by Eli Lilly & Co. for work related to this study. A.R. and J.-P.C. are employees of Eli Lilly & Co. A.W. received personal fees from Eli Lilly & Co for work related to this study.

Funding: This work was supported by Eli Lilly & Company. Grant/award number is not applicable

Data availability statement: The datasets generated and/or analysed during the study are not publicly available, because consent was not sought from participants to allow sharing of data with third parties.

REFERENCES

1. Fishbein AB, Silverberg JI, Wilson EJ, et al. Update on Atopic Dermatitis: Diagnosis, Severity Assessment, and Treatment Selection. *J Allergy Clin Immunol Pract* 2020;8:91-101.
2. Wollenberg A, Christen-Zach S, Taieb A, et al. ETFAD/EADV Eczema task force 2020 position paper on diagnosis and treatment of atopic dermatitis in adults and children. *J Eur Acad Dermatol Venereol* 2020;34:2717-44.
3. Wollenberg A, Barbarot S, Bieber T, et al. Consensus-based European guidelines for treatment of atopic eczema (atopic dermatitis) in adults and children: part II. *J Eur Acad Dermatol Venereol* 2018;32:850-78.
4. Sidbury R, Davis DM, Cohen DE, et al. Guidelines of care for the management of atopic dermatitis: section 3. Management and treatment with phototherapy and systemic agents. *J Am Acad Dermatol* 2014;71:327-49.
5. Seegraber M, Srour J, Walter A, et al. Dupilumab for treatment of atopic dermatitis. *Expert Rev Clin Pharmacol* 2018;11:467-74.
6. OLUMIANT® Summary of Product Characteristics: Eli Lilly Nederland B.V.; 2020. Available from: https://ec.europa.eu/health/documents/community-register/2020/20201019149480/anx_149480_en.pdf. Accessed 02 December 2020.
7. Rinvoq: Summary of opinion (post authorisation): Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency; 2021. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-

authorisation-summary-positive-opinion-rinvoq-x-06-g_en.pdf. Accessed 09 November 2021.

8. Deleanu D, Nedelea I. Biological therapies for atopic dermatitis: An update (Review). *Exp Ther Med* 2018;17:1061-67.

9. Casciano R, Malangone E, Ramachandran A, et al. A quantitative assessment of patient barriers to insulin. *Int J Clin Pract* 2011;65:408-14.

10. Edel Y, Sagy I, Pokroy-Shapira E, et al. A Cross-sectional Survey on the Preference of Patients with Rheumatoid Arthritis for Route of Administration of Disease-Modifying Anti-Rheumatic Drugs: Oral Target-Specific Versus Parenteral Biologic. *Isr Med Assoc J* 2020;22:154-59.

11. Mohr DC, Boudewyn AC, Likosky W, et al. Injectable medication for the treatment of multiple sclerosis: The influence of self-efficacy expectations and infection anxiety on adherence and ability to self-inject. *Ann Behav Med* 2001;23:125-32.

12. Bouvy JC, Cowie L, Lovett R, et al. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. *Patient* 2020;13:145-49.

13. Huls SPI, Whichello CL, van Exel J, et al. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. *Value Health* 2019;22:1318-28.

14. Janus SI, Weernink MG, van Til JA, et al. A Systematic Review To Identify the Use of Preference Elicitation Methods in Health Care Decision Making. *Value Health* 2014;17:A515-6.

15. Mühlbacher AC, Juhnke C, Beyer AR, et al. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. *Value Health* 2016;19:734-40.

16. Johnson FR, Zhou M. Patient Preferences in Regulatory Benefit-Risk Assessments: A US Perspective. *Value Health* 2016;19:741-45.

17. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. *Patient* 2017;10:263-66.

18. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics* 2018;37:201-26.

19. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Econ* 2009;18:321-36.

20. Carlsson F, Mørkbak MR, Olsen SB. The first time is the hardest: A test of ordering effects in choice experiments. *J Choice Model* 2012;5:19-37.

21. Heidenreich S, Phillips-Beyer A, Flamion B, et al. Benefit–risk or risk–benefit trade-offs? Another look at attribute ordering effects in a pilot choice experiment. *Patient* 2021;14:65-74.

22. Johnson FR, Yang J-C, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. *Value Health* 2019;22:157-60.

23. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:588-94.

24. Lipkus IM, Samsa G, Rimer BK. General Performance on a Numeracy Scale among Highly Educated Samples. *Med Decis Making* 2016;21:37-44.

25. Manski CF. The structure of random utility models. *Theory Decis* 1977;8:229-54.

26. Thurstone LL. A law of comparative judgment. *Psychol Rev* 1927;34:273-86.

27. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ* 2007;16:827-40.

28. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. *Arch Dermatol* 2004;140:1513-9.

29. Boeri M, Sutphin J, Hauber B, et al. Quantifying patient preferences for systemic atopic dermatitis treatments using a discrete-choice experiment. *J Dermatolog Treat* 2020:1-10.

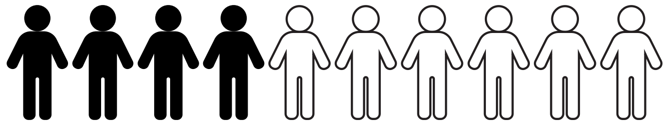
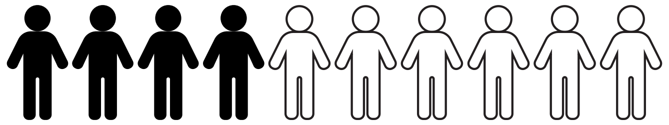
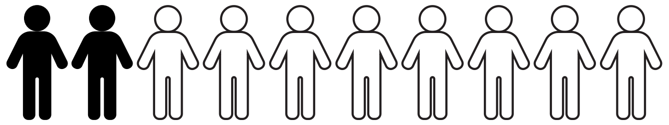
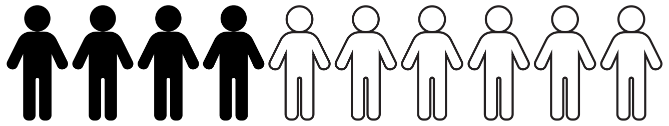
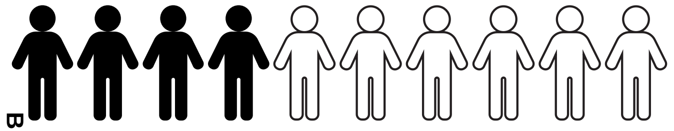
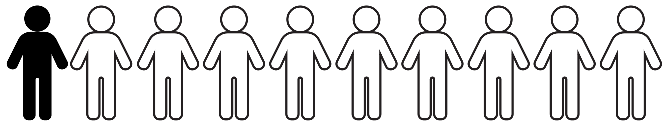
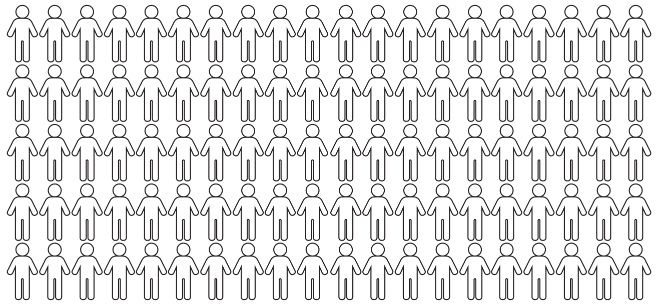
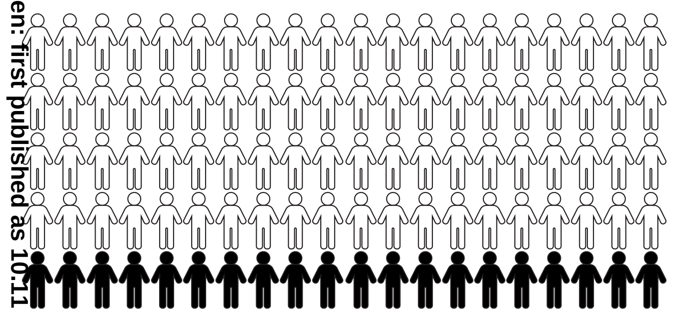
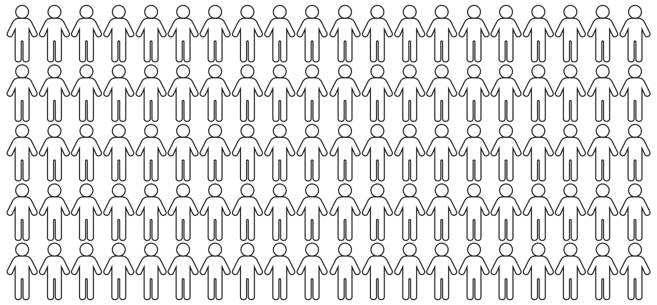
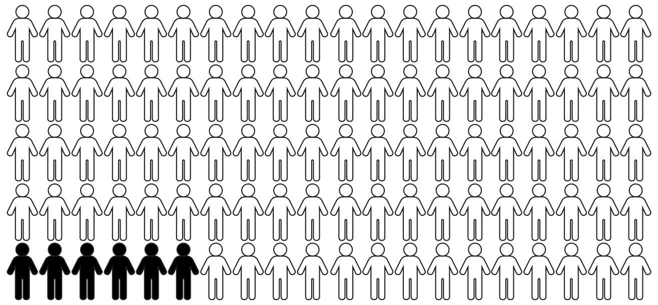
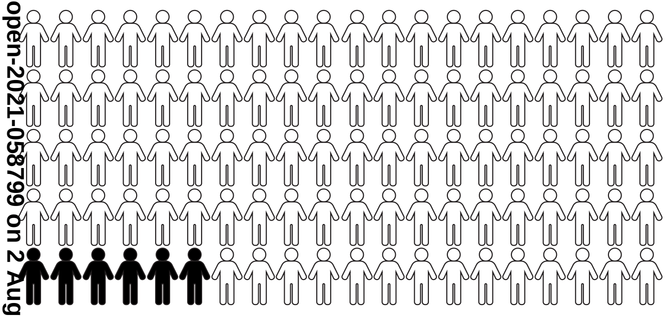
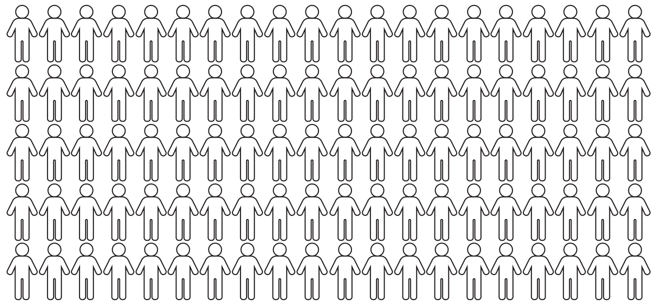
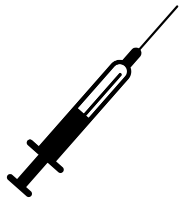
30. Okubo Y, Ho K-A, Fifer S, et al. Patient and physician preferences for atopic dermatitis injection treatments in Japan. *J Dermatolog Treat* 2019;31:821-30.

31. Sandorf ED. Did you miss something? Inattentive respondents in discrete choice experiments. *Environ Resour Econ* 2018;73:1197-235.

FIGURE LEGENDS

Figure 1. Example choice task

Figure 2. Multinomial logit results: part-worth utilities

		Treatment A	**Treatment B**	**Your old Treatment**
Benefits	**Itch Reduction**	4 out of 10 (40%) 	4 out of 10 (40%) 	2 out of 10 (20%) 
	Skin Appearance	4 out of 10 (40%) 	4 out of 10 (40%) 	1 out of 10 (10%) 
Side Effects	**Eye Inflammation**	0 out of 100 (0%) 	20 out of 100 (20%) 	0 out of 100 (0%) 
	Serious Infections	6 out of 100 (6%) 	6 out of 100 (6%) 	0 out of 100 (0%) 
Other	**Speed of Onset**	1 week	2 weeks	2 weeks
	Flare Management	✓	✗	✓
	Long-term Disease Management	Yes, with the possibility of pauses	Yes, without the possibility of pauses	Should not be used long-term
	Administration	Oral pill Once or twice daily	Injection under the skin  Every two weeks	Oral pill Once or twice daily
	Check-ups	Occasional check-ups required	No check-ups required	Frequent check-ups required
	Choice	☐	☐	☐
	[*if old treatment is chosen*] **Choice**	☐	☐	

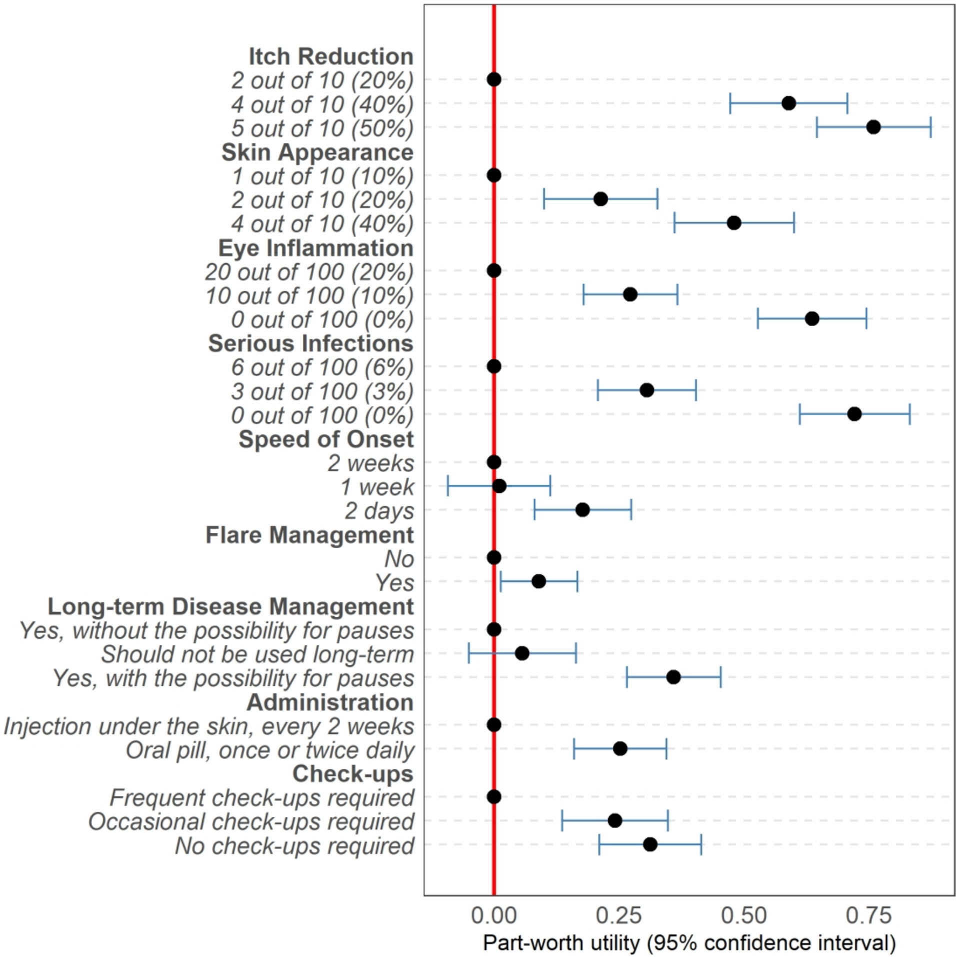


Figure 2. Multinomial logit results: part-worth utilities

177x177mm (300 x 300 DPI)

Supplemental Methods

Selection of attributes and levels

The aim of the literature review was to determine the key attributes and levels to be included in the DCE. This involved both a targeted literature review and a product label review.

The targeted literature review involved separate searches in major medical literature databases (Embase and MEDLINE); a search for qualitative studies that considered the patient perspective on AD treatments; and a search for patient preference-specific studies, which considered AD treatments. Once key themes within the literature review were identified, the attributes were classified into corresponding categories.

The search strategy for the qualitative studies focused on studies that conducted interviews or focus groups, which mentioned AD or eczema, and their available treatments, as well as quality of life or patient preferences. The search excluded any non-adult studies, animal studies, clinical trials, and editorial notes. The search strategy for patient preference-specific studies sought studies that were explicitly patient preference in design, such as those utilising DCEs. Additionally, the studies had to mention AD or eczema.

The targeted literature search identified 33 potential studies. No duplicates were found, and all 33 were screened for eligibility. The abstracts were screened sequentially by two reviewers, and a third reviewer compared the rationale for inclusion and exclusion of studies to obtain the final list of full texts to screen. Fifteen studies were excluded because they did not involve adults with AD, one because it did not mention outcomes of interest; and seven because they were other study types not focusing on patient preferences.

Of the seven remaining studies, four were excluded because a full text was not available. The final remaining three studies included one quantitative[30] and two qualitative studies.[31, 32] In the quantitative study, the most important treatment attribute was the appearance of eczema (dryness/flakiness). In the two qualitative studies, itch reduction (symptom control), monitoring of symptoms, flexibility of treatment regimens to control flares, appearance (dryness/flakiness), and skin pain were identified themes.

Additionally, a product label search was conducted. Ten product labels for medications indicated for use in AD were reviewed in detail, including baricitinib (Olumiant®), dupilumab (Dupixent®), clobetasol propionate (Clobex®), tacrolimus (Protopic®), prednisone (Rayos®), cyclosporin (Neoral®), methotrexate, azathioprine (Imuran®), mycophenolate mofetil (CellCept®), and phototherapy. Itch reduction was most commonly reported as the percentage of patients achieving a meaningful (≥4-point reduction in the itch numerical rating scale) reduction in itch at week 16. Skin appearance was most commonly measured by the proportion of patients achieving clear or almost clear skin at week 16 (Investigator's Global Assessment scores of 0 or 1). The review of product labels also identified conjunctivitis as a differentiating and common side-effect of dupilumab that is not associated with other systemic therapies. Risk of serious infections were associated with other treatments, such as baricitinib and cyclosporine. The product label review also highlighted different modes and frequency of administration for systemic treatments, which included daily oral medication or subcutaneous administration every 2 weeks. Monitoring was also required for baricitinib and cyclosporine, but not for dupilumab.

Model specification

The analysis of all DCE responses followed random utility theory.[24, 33, 34] The model assumes that each respondent (*n*) chooses the alternative (*j*) in every DCE question (*t*) that results in the highest utility (a measure of desirability) of all available alternatives. Utility in a random utility model is defined as:

$$u(\boldsymbol{x}_{jnt}) = v(\boldsymbol{x}_{jnt}) + \varepsilon_{jnt}$$

Here the systematic utility component $v(\boldsymbol{x}_{jnt})$ is a function of the DCE attributes and ε_{jnt} is a type 1 extreme value distributed random error. Two models are presented: a dummy-coded MNL model and an MNL model with skin appearance coded linearly, which is required to estimate the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. For the former, the utility function was defined as:

$$\begin{aligned} u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} + \beta_1 \text{40\%\_itch\_reduction}_{jnt} + \beta_2 \text{50\%\_itch\_reduction}_{jnt} \\ + \beta_3 \text{20\%\_skin\_appearance}_{jnt} + \beta_4 \text{40\%\_skin\_appearance}_{jnt} + \beta_5 \text{10\%\_eye\_inflammation}_{jnt} \\ + \beta_6 \text{0\%\_eye\_inflammation}_{jnt} + \beta_7 \text{3\%\_serious\_infections}_{jnt} + \beta_8 \text{0\%\_serious\_infections}_{jnt} \\ + \beta_9 \text{1\_week\_onset}_{jnt} + \beta_{10} \text{2\_days\_onset}_{jnt} + \beta_{11} \text{flare\_management}_{jnt} + \beta_{12} \text{long\_term\_no}_{jnt} \\ + \beta_{13} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{14} \text{oral\_admin}_{jnt} + \beta_{15} \text{no\_check\_ups}_{jnt} \\ + \beta_{16} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt} \end{aligned}$$

The constants $\alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}}$ controlled for potential bias to select the left option (Treatment A), and the Old Treatment, $\beta_1$to β_{16} were the estimated marginal utilities (i.e., estimated preference parameters), ε_{jnt} was an extreme value type I distributed error that allowed the function to be estimated in a logit model.[34] All attributes were dummy-coded. The reference level was the assumed worst-case option. Each of the estimated marginal utilities measured respondents' sensitivity to deviations from the reference level of the corresponding attribute. The sign (+ or –) of a marginal utility denotes whether patients valued this deviation positively or negatively. Only the initial choices (A vs. B vs. old therapy) were considered for the analysis of preferences. The initial and follow-up choices can be combined to allow for a more precise measurement of preferences. However, it is appropriate to combine these two types of choices only when they generate approximately the same information about participants' preferences. This condition was verified in two ways. Two MNL models were separately estimates for the initial (4,848 observations) and follow-up choices (1,126 observations), and then their preference estimates were compared. The Pearson correlation coefficient between the two sets of estimates was relatively low (0.32) as was the coefficient of determination for the linear regression (0.104), indicating poor agreement between the sets of estimates. A third MNL model was estimated on the combined initial and follow-up choices (5,974 observations), and its statistical performance was compared with the MNL model based on initial choices only. The adjusted McFadden pseudo-R^2 was lower for the model based on combined choices (7.3%) than for the initial model (8.3%), indicating that combining the initial and follow-up choices had a detrimental effect on the explanatory power of the model.

The linear coding of skin appearance was required to derive meaningful MAD measures. This measure was obtained by estimating the baseline utility function with skin appearance being coded as linear (i.e., one marginal utility is estimated instead of β_3 and β_4 for skin appearance). The utility function was defined as:

$$\begin{aligned} u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} &+ \beta_1 \text{40\%\_itch\_reduction}_{jnt} + \beta_2 \text{50\%\_itch\_reduction}_{jnt} \\ &+ \beta_3\text{\_skin\_appearance}_{jnt} + \beta_4 \text{10\%\_eye\_inflammation}_{jnt} + \beta_5 \text{0\%\_eye\_inflammation}_{jnt} \\ &+ \beta_6 \text{3\%\_serious\_infections}_{jnt} + \beta_7 \text{0\%\_serious\_infections}_{jnt} + \beta_8 \text{1\_week\_onset}_{jnt} \\ &+ \beta_9 \text{2\_days\_onset}_{jnt} + \beta_{10} \text{flare\_management}_{jnt} + \beta_{11} \text{long\_term\_no}_{jnt} \\ &+ \beta_{12} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{13} \text{oral\_admin}_{jnt} + \beta_{14} \text{no}_{\text{check}_{\text{ups}}}{}_{jnt} \\ &+ \beta_{15} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt} \end{aligned}$$

Each marginal utility was then divided by the marginal utility for skin appearance:

$$\mathrm{MAD}_k = \frac{\hat{\beta}_k}{\hat{\beta}_3}$$

Such linear encoding is based on the underlying assumption of linearity in preferences, wherein a one-unit change in the attribute has a constant effect on respondents' choices and does not depend on the absolute value of the attribute level (e.g., a one-unit change from 15% to 16% increase in the chance of achieving clear or almost clear skin at week 16 has the same effect as the change from 20% to 21%). The validity of the assumption of linearity in preferences was tested by analysing the trend in risk estimates from the dummy-coded MNL model. Estimates were obtained for every attribute level in the dummy-coded MNL model (i.e., 3 levels for skin appearance). The linearity of skin appearance was tested by fitting a linear regression and evaluating its coefficient of determination. The assumption of linearity in skin appearance was accepted with a coefficient of 0.81, which exceeds the threshold of 0.7 to verify linearity.

Supplemental Table 1. Validity assessments

Assessment	Full sample N=404	France N=114	Spain N=145	UK N=145
Choice stability, n (%)				
Passed the test	260 (64)	71 (62)	94 (65)	95 (66)
Failed the test	144 (36)	43 (38)	51 (35)	50 (34)
Choice dominance [a], n (%)				
Passed the test	359 (89)	109 (96)	130 (90)	120 (83)
Failed the test	45 (11)	5 (4)	15 (10)	25 (17)
Serial non-participation [b], n (%)				
Never select the same option	384 (95)	108 (95)	136 (94)	140 (97)
Always select treatment A	0 (0)	0 (0)	0 (0)	0 (0)
Always select treatment B	1 (0)	1 (1)	0 (0)	0 (0)
Always select old therapy	19 (5)	5 (4)	9 (6)	5 (3)
Dominated decision making [c], n (%)				
Itch reduction	6 (1)	1 (1)	2 (1)	3 (2)
Skin appearance	1 (<1)	0 (0)	1 (1)	0 (0)
Eye inflammation	3 (1)	1 (1)	1 (1)	1 (1)
Serious infections	8 (2)	3 (3)	3 (2)	2 (1)
Speed of onset	0 (0)	0 (0)	0 (0)	0 (0)
Flare management	1 (<1)	0 (0)	0 (0)	1 (1)
Long-term disease management	0 (0)	0 (0)	0 (0)	0 (0)
Administration	21 (5)	8 (7)	8 (6)	5 (3)
Check-ups	2 (<1)	1 (1)	0 (0)	1 (1)
None	362 (90)	100 (88)	130 (90)	132 (91)
Response time for DCE choice task section only [d], n (%)				
Adequate	391 (97)	111 (97)	143 (99)	137 (95)
Inadequate	13 (3)	3 (3)	2 (1)	8 (5)
Time to complete DCE choice task section only, n (%)				
<5 min	236 (58)	64 (56)	93 (64)	79 (54)
5-10 min	123 (30)	38 (33)	38 (26)	47 (32)
10-15 min	28 (7)	7 (6)	9 (6)	12 (8)
15-20 min	4 (1)	1 (1)	2 (1)	1 (1)
>20 min	13 (3)	4 (4)	3 (2)	6 (4)

Abbreviations: DCE, discrete choice experiment

[a] A respondent was considered to have failed the test if they chose the inferior (dominated) option as their preferred treatment.

[b] A respondent was classified as a serial non-participant if they choose the same option for all 12 experimental choice tasks.

[c] Decision making was considered dominated when the respondent choses the best option on one attribute in all 12 experimental tasks.

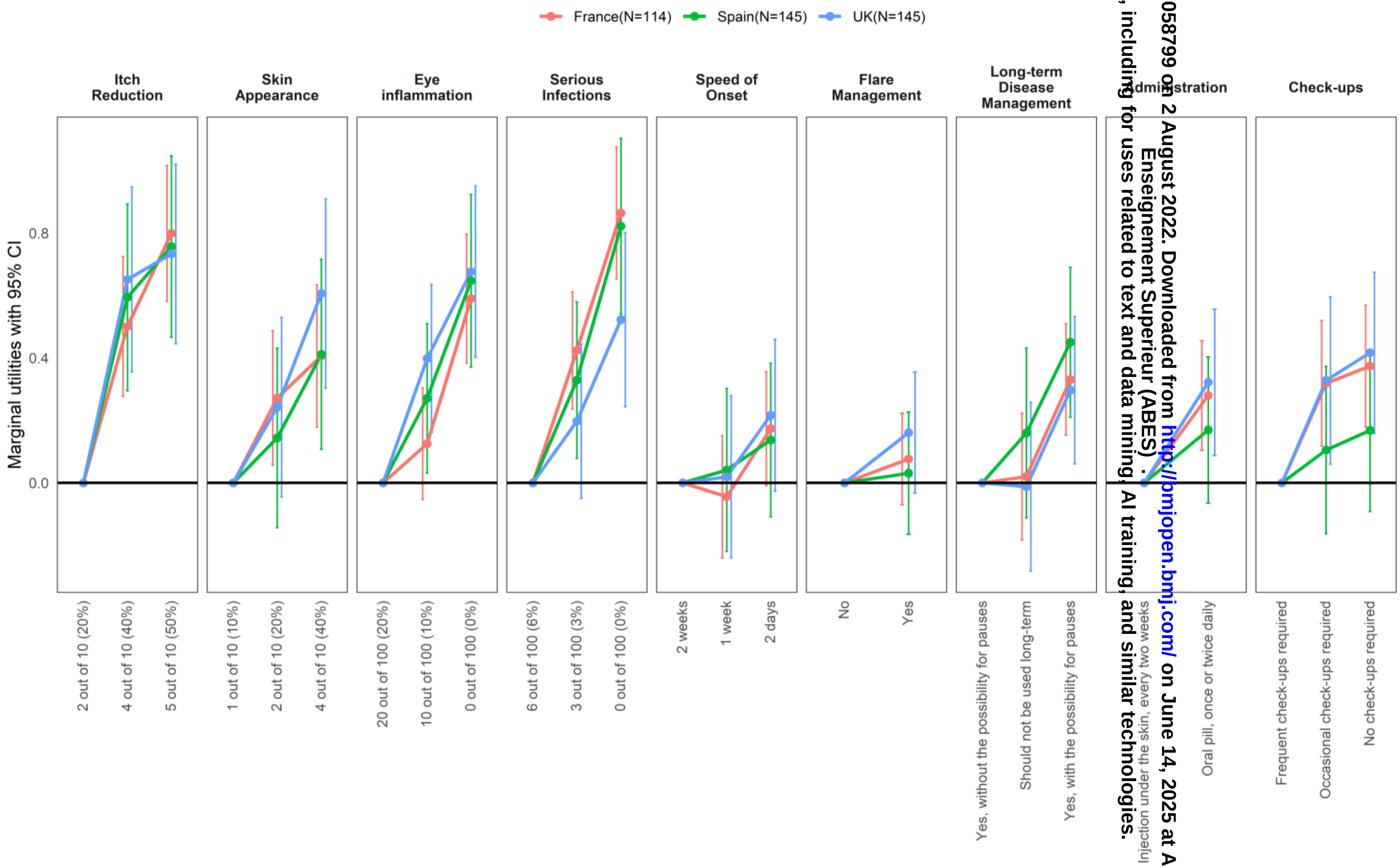
[d] Response times in the lower 10% of the distribution were classed as too fast, and those in the upper 10% of the distribution as too slow. A participant was considered to have had an adequate response time if <80% of choice tasks were answered too fast or too slow.

Supplemental Table 2. Multinomial logit results: maximum likelihood estimates

Attribute	Level	MLE (SE)	95% CI
Alternative specific constant	Old therapy	1.46 (0.12)***	[1.23; 1.69]
	Option A	-0.04 (0.04)	[-0.11; 0.03]
Itch reduction	2 out of 10 (20%)	Reference	-
	4 out of 10 (40%)	0.59 (0.06)***	[0.47; 0.71]
	5 out of 10 (50%)	0.76 (0.06)***	[0.65; 0.87]
Skin appearance	1 out of 10 (10%)	Reference	-
	2 out of 10 (20%)	0.21 (0.06)***	[0.10; 0.33]
	4 out of 10 (40%)	0.48 (0.06)***	[0.36; 0.60]
Eye inflammation	20 out of 100 (20%)	Reference	-
	10 out of 100 (10%)	0.27 (0.05)***	[0.18; 0.37]
	0 out of 100 (0%)	0.64 (0.06)***	[0.53; 0.75]
Serious infections	6 out of 100 (6%)	Reference	-
	3 out of 100 (3%)	0.31 (0.05)***	[0.21; 0.40]
	0 out of 100 (0%)	0.72 (0.06)***	[0.61; 0.83]
Speed of onset	2 weeks	Reference	-
	1 week	0.01 (0.05)	[-0.09; 0.11]
	2 days	0.18 (0.05)***	[0.08; 0.27]
Flare management	No	Reference	-
	Yes	0.09 (0.04)*	[0.01; 0.17]
Long-term disease management	Yes, without the possibility for pauses	Reference	-
	Should not be used long-term	0.06 (0.05)	[-0.05; 0.16]
	Yes, with the possibility for pauses	0.36 (0.05)***	[0.27; 0.45]
Administration	Injection under the skin, every 2 weeks	Reference	-
	Oral pill, once or twice daily	0.25 (0.05)***	[0.16; 0.35]
Check-ups	Frequent check-ups required	Reference	-
	Occasional check-ups required	0.24 (0.05)***	[0.14; 0.35]
	No check-ups required	0.31 (0.05)***	[0.21; 0.41]

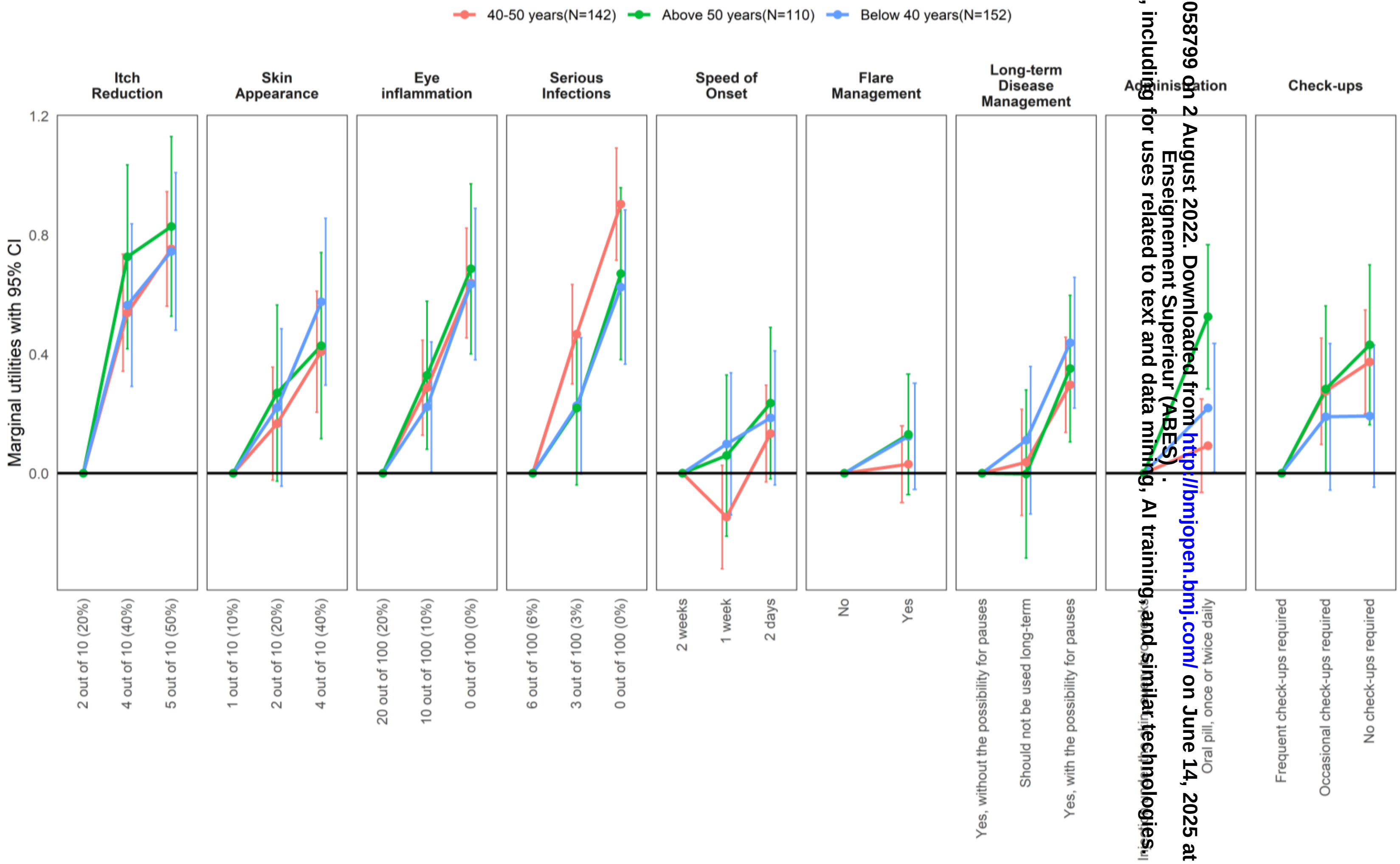
Number of observations	4848
Model log-likelihood at convergence	-4867
Adjusted pseudo R^2	0.08
Bayesian information criterion	9887

Abbreviations: CI, confidence interval; MLE, maximum likelihood estimate; SE, standard error

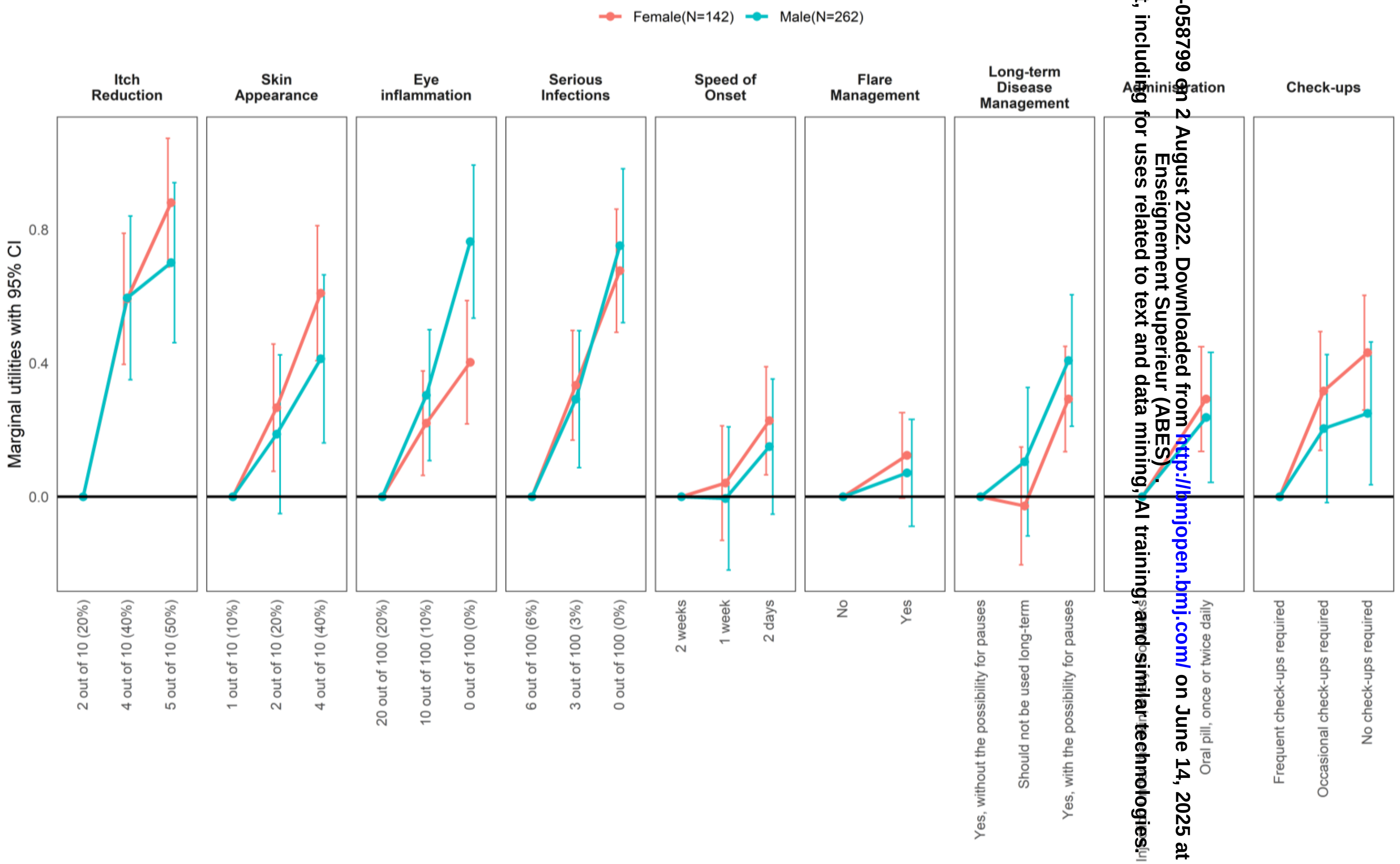


Supplemental Figure 1. MNL results by country

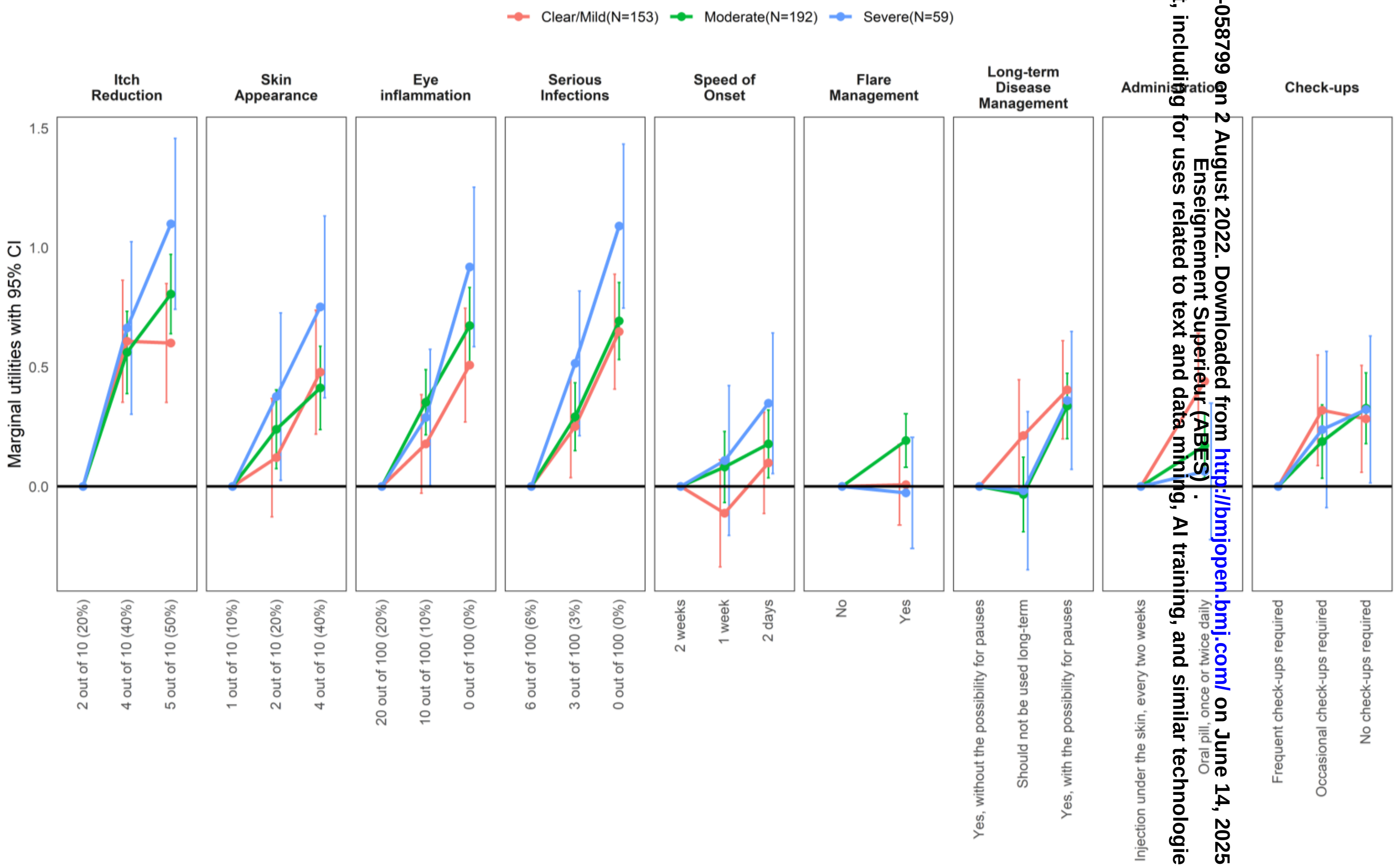
Abbreviation: CI, confidence interval



Supplemental Figure 2. MNL results by age
Abbreviation: CI, confidence interval

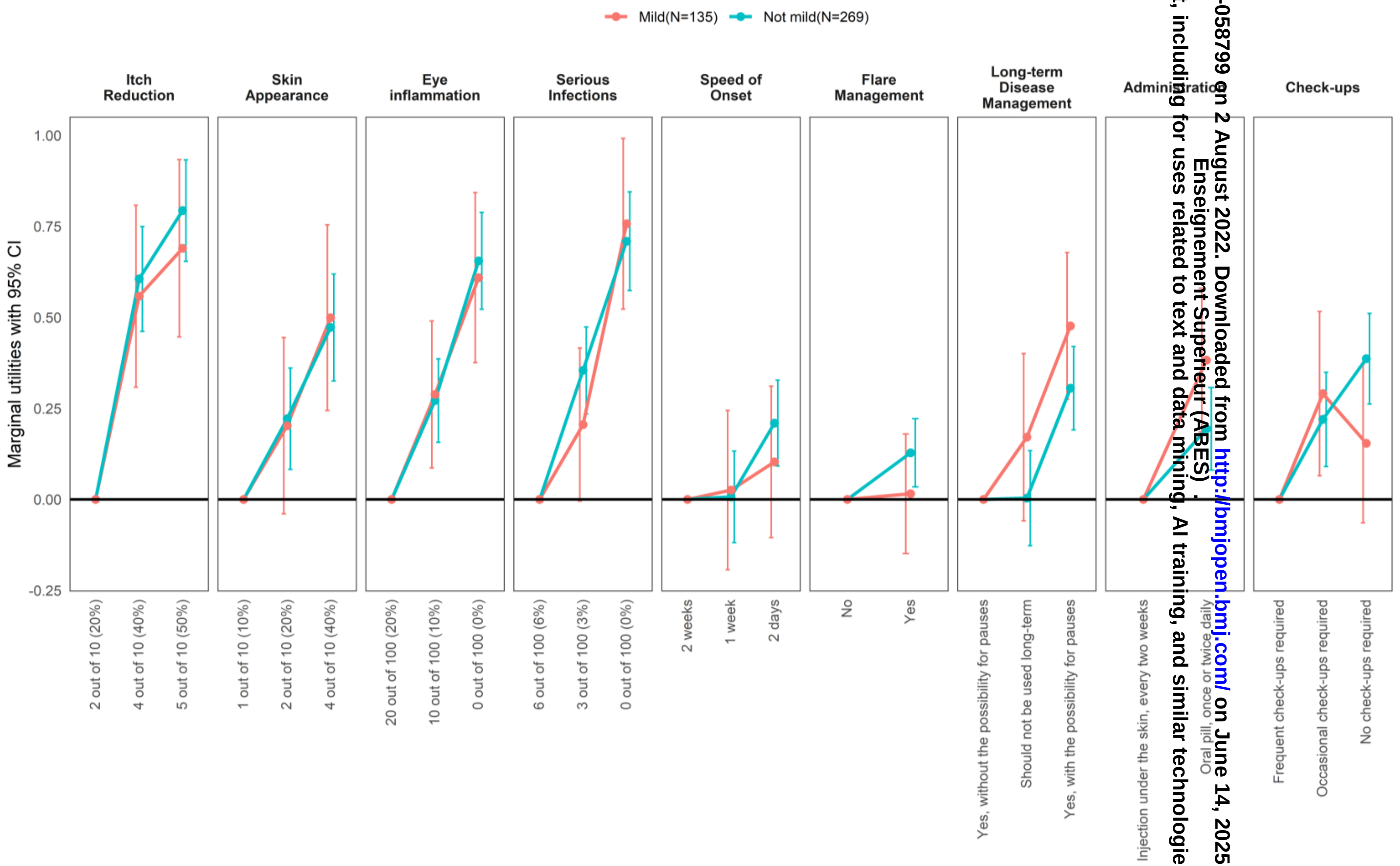


Supplemental Figure 3. MNL results by gender
Abbreviation: CI, confidence interval



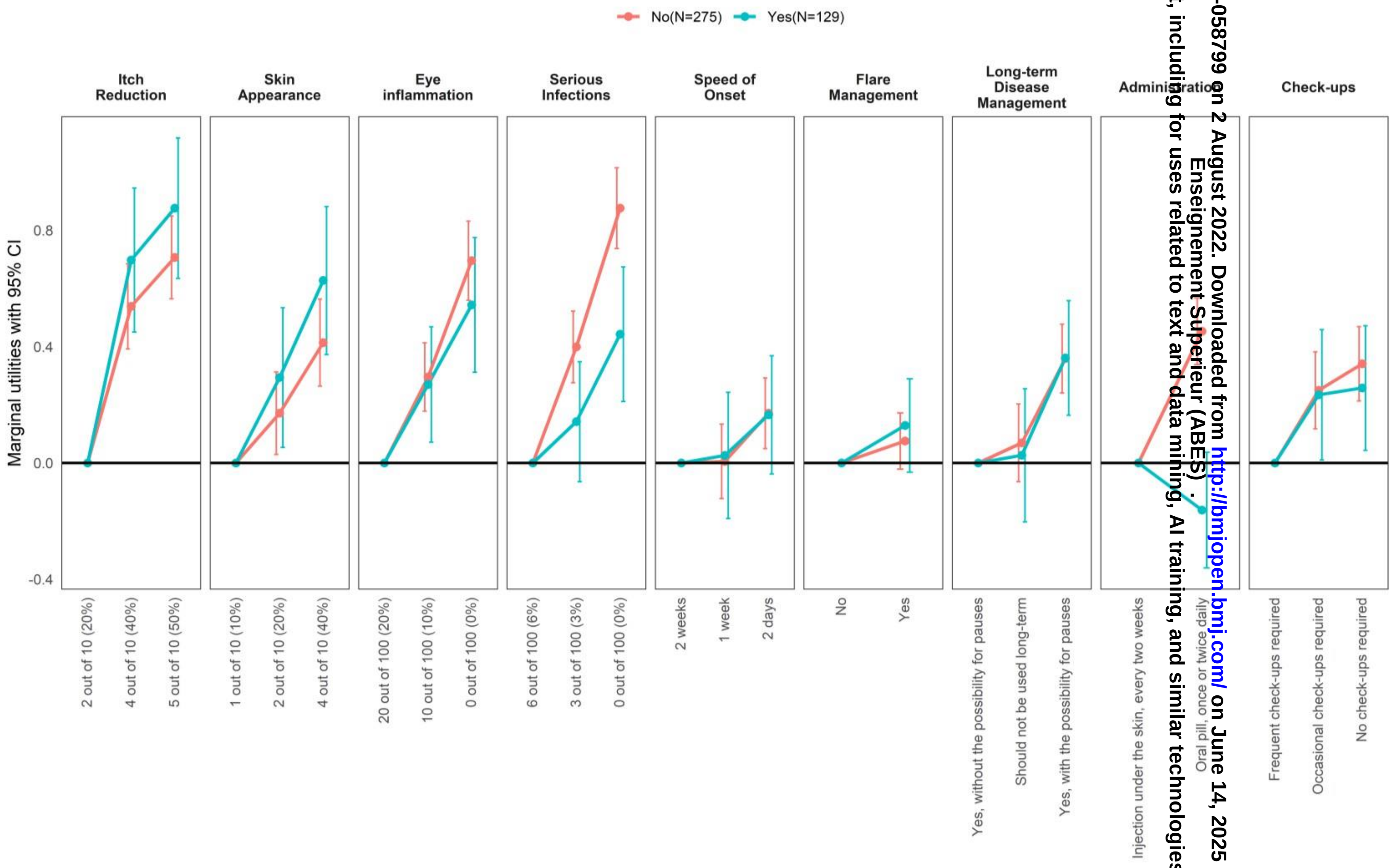
Supplemental Figure 4. MNL results by Patient Oriented Eczema Measure (POEM) overall score. Clear/Mild: 0–7; Moderate: 8–16; Severe: 17–28.

Abbreviation: CI, confidence interval



Supplemental Figure 5. MNL results by self-reported eczema severity. Mild: very mild/mild; Not mild: moderate/severe/very severe.

Abbreviation: CI, confidence interval



Supplemental Figure 6. MNL results by experience self-injecting
Abbreviation: CI, confidence interval

Checklist	Covered in manuscript	Page or section
1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?		
1.1 Were a well-defined research question and a testable hypothesis articulated?	Yes	p. 5
1.2 Was the study perspective described, and was the study placed in a particular decision-making or policy context?	Yes	p. 4-5
1.3 What is the rationale for using conjoint analysis to answer the research question?	Yes	p. 5
2. Was the choice of attributes and levels supported by evidence?		
2.1 Was attribute identification supported by evidence (literature reviews, focus groups, or other scientific methods)?	Yes (literature review)	p. 5
2.2 Was attribute selection justified and consistent with theory?	Yes	p. 5, 9-10
2.3 Was level selection for each attribute justified by the evidence and consistent with the study perspective and hypothesis?	Yes, via a literature review	p. 5
3. Was the construction of tasks appropriate?		
3.1 Was the number of attributes in each conjoint task justified (that is, full or partial profile)?	Yes, participants were surveyed for relevant attributes and no missing attributes were identified. Full choice profiles were used and patients had no issues with the number of attributes	p. 11
3.2 Was the number of profiles in each conjoint task justified?	Yes (3 profiles: A vs B vs old treatment)	p. 10-11

3.3 Was (should) an opt-out or a status-quo alternative (be) included?	Yes	p. 11
4. Was the choice of experimental design justified and evaluated?		
4.1 Was the choice of experimental design justified? Were alternative experimental designs considered?	Yes, D-efficient design assessed against good experimental design properties	p. 11
4.2 Were the properties of the experimental design evaluated?	Yes	p. 11
4.3 Was the number of conjoint tasks included in the data-collection instrument appropriate?	Yes, the number of tasks (questions) was 12 per person (36 in total)	p. 11
5. Were preferences elicited appropriately, given the research question?		
5.1 Was there sufficient motivation and explanation of conjoint tasks?	Yes	p. 11-12
5.2 Was an appropriate elicitation format (that is, rating, ranking, or choice) used? Did (should) the elicitation format allow for indifference?	Yes, the elicitation task was a choice task. The format did not allow indifference	p. 11-12
5.3 In addition to preference elicitation, did the conjoint tasks include other qualifying questions (for example, strength of preference, confidence in response, and other methods)?	Yes, validity assessments	p. 12
6. Was the data collection instrument designed appropriately?		
6.1 Was appropriate respondent information collected (such as sociodemographic, attitudinal, health history or status, and treatment experience)?	Yes	Table 2
6.2 Were the attributes and levels defined, and was any	Yes	Table 1

contextual information provided?		
6.3 Was the level of burden of the data-collection instrument appropriate? Were respondents encouraged and motivated?	Yes, this was assessed in cognitive pilot interviews and with data quality measures	p21
7. Was the data-collection plan appropriate?		
7.1 Was the sampling strategy justified (for example, sample size, stratification, and recruitment)?	Yes	p. 10
7.2 Was the mode of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)?	Yes	p. 5, 10
7.3 Were ethical considerations addressed (for example, recruitment, information and/or consent, compensation)?	Yes	p. 10
8. Were statistical analyses and model estimations appropriate?		
8.1 Were respondent characteristics examined and tested?	Yes	p. 14-16
8.2 Was the quality of the responses examined (for example, rationality, validity, reliability)?	Yes (validity and reliability)	p. 11-12, 16
8.3 Was model estimation conducted appropriately? Were issues of clustering and subgroups handled appropriately?	Yes	p. 14
9. Were the results and conclusions valid?		
9.1 Did study results reflect testable hypotheses and account for statistical uncertainty?	Yes, confidence intervals are presented	
9.2 Were study conclusions supported by the evidence and	Yes	p. 20

compared with existing findings in the literature?		
9.3 Were study limitations and generalizability adequately discussed?	Yes	p. 21
10. Was the study presentation clear, concise, and complete?		
10.1 Was study importance and research context adequately motivated?	Yes	p. 4
10.2 Were the study data-collection instrument and methods described?	Yes	p. 11
10.3 Were the study implications clearly stated and understandable to a wide audience?	Yes	p. 20-22

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058799.R1
Article Type:	Original research
Date Submitted by the Author:	28-Apr-2022
Complete List of Authors:	Thomas, Caitlin; Evidera Raibouaa, Afaf; Eli Lilly and Company Wollenberg, Andreas ; University Hospital of Ludwig-Maximilian University of Munich, Department of Dermatology and Allergy Capron, Jean-Philippe; Eli Lilly and Company Krucien, Nicolas; Evidera Karn, Hayley; Evidera Tervonen, T; Evidera
Primary Subject Heading:	Dermatology
Secondary Subject Heading:	Patient-centred medicine, Health economics, Evidence based practice
Keywords:	DERMATOLOGY, Eczema < DERMATOLOGY, THERAPEUTICS, Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Caitlin Thomas [1], Afaf Raibouaa [2], Andreas Wollenberg [3], Jean-Philippe Capron [2], Nicolas Krucien [1], Hayley Karn [1], Tommi Tervonen [1]

[1] Patient-Centered Research, Evidera, London, UK

[2] Eli Lilly & Co., Indianapolis, IN, USA

[3] Department of Dermatology and Allergy, University Hospital, Ludwig-Maximilian University of Munich, Germany

Corresponding author: Tommi Tervonen, Evidera, The Ark, 201 Talgarth Rd, London W6 8DL, UK. Phone: +44 (0)7788 237 412. Email: tommi.tervonen@evidera.com

4554 words

ABSTRACT

Objectives We aimed to quantify patient preferences for efficacy, safety, and convenience features of atopic dermatitis (AD) treatments.

Design and setting Online discrete choice experiment (DCE) survey.

Participants Adults in the UK, France, and Spain who had used AD treatments during the past 2 years.

Primary and secondary outcome measures Preferences for attributes were analysed using a multinomial logit model. Willingness to make trade-offs was expressed as the maximum acceptable decrease (MAD) in the probability of achieving clear/almost clear skin at week 16.

Results The survey was completed by 404 patients (44.1±12.0 years; 65% female; 64% moderate/severe eczema). Most patients (68%) had no prior experience of using self-injectable treatments for AD or any other illness. Participants most valued increasing the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%, followed by reducing the risks of serious infections from 6% to 0% and of eye inflammation from 20% to 0%. Participants were willing to accept a decrease in the possibility of achieving clear/almost clear skin to obtain a treatment that can be paused (MAD = 24.1%), requires occasional check-ups (MAD = 16.1%) or no check-ups (MAD = 20.9%) over frequent check-ups, is administered as a once- or twice-daily oral pill versus a subcutaneous injection every 2 weeks (MAD = 16.6%), has a 2-day over 2-week onset of action (MAD = 11.3%), and can be used for flare management (MAD = 5.8%).

Conclusions Although patients with AD most valued treatment benefits and risks, they were willing to tolerate reduced efficacy to obtain a rapid onset, oral

administration, less frequent monitoring, and a treatment that can be paused. Understanding patients' preferences for AD therapies, including new targeted therapies, can aid shared decision-making between clinicians and patients and support health technology assessments.

Keywords: Dermatology, Eczema, Health Economics, Therapeutics

Strengths and limitations of this study:

- This study used a discrete choice experiment to elicit preferences of patients in the UK, France, and Spain for attributes of atopic dermatitis treatments.
- Patients with AD most valued treatment benefits and reducing risks but were willing to accept a decrease in efficacy, as measured by the possibility of obtaining clear or almost clear skin at week 16, to obtain an oral treatment with a rapid onset of action.
- Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age, sex, or disease severity.
- This sample was an adult population from the UK, France, and Spain, and a high proportion of patients had a university education. Therefore, the study may not be generalisable to children, patients in other countries, or those with lower levels of education. In addition, patients had predominantly moderate to severe AD, and these findings may not apply to the wider AD adult population, including those with mild or very severe AD.

INTRODUCTION

Atopic dermatitis (AD) is mostly treated using emollients and moisturizers, topical corticosteroids and calcineurin inhibitors, and, for severe cases, systemic immunosuppressants.[1, 2] However, emollients and moisturisers may not be sufficiently effective, and conventional systemic immunosuppressants have many potential side effects and are not generally recommended for long-term maintenance of AD.[3, 4] New targeted therapies for treating AD are now available. Dupilumab, a subcutaneously administered human monoclonal antibody inhibiting interleukin-4 and interleukin-13 signalling, was licensed in the US and the European Union in 2017 for the treatment of AD.[5] Baricitinib and upadacitinib, oral small-molecule inhibitors of Janus kinases, were recently licensed in the European Union for the treatment of moderate-to-severe AD in patients who are candidates for systemic therapy.[6, 7]

Several additional targeted therapies are in development, including a variety of monoclonal antibodies inhibiting interleukin signalling.[1, 2, 8] These new targeted therapies have different efficacy, risks, and non-clinical attributes, especially the mode of administration. In other chronic diseases, some patients prefer oral over parenteral treatment because they perceive some barriers to parenteral administration, which may lead to reduced adherence.[9-11] Because non-health benefits cannot be captured in traditional cost-effectiveness analysis, understanding to what extent they are valued by patients can help guide health technology assessment discussions[12-16] and inform shared decision-making at the point of care.[17]

Preferences for different treatment attributes, such as their benefits, risks, mode of administration, and convenience features, can be elicited from patients using discrete choice experiments (DCEs).[18] In DCEs, participants are presented with a series of tasks where they have to select between different hypothetical treatment options, each of which is composed of one level from each attribute in such a way that they are forced to make trade-offs, such as a higher risk of an adverse event but improved efficacy. DCEs have the advantage that the results can be used to quantify to what extent participants value each of the different attributes and estimate the trade-offs they would be willing to make. We hypothesized that patients with AD would not value all attributes relevant for their treatment choices equally. In the current study, we used a DCE to elicit the preferences of patients for key efficacy, safety, and convenience attributes of targeted AD therapies and examine the trade-offs they are willing to make between them.

MATERIALS AND METHODS

An online DCE survey was conducted between October and December 2019 in adults with AD living in the UK, France, or Spain. In the DCE survey, participants completed a series of choice tasks in which they selected between hypothetical treatment options described by a set of attributes with different levels. Treatment attributes and levels included in the DCE were identified through a targeted literature review of Embase and MEDLINE for quantitative and qualitative preference studies and a review of product labels for AD treatments (search conducted 10th September 2018; see **Online Supplemental Methods** and **Online Supplemental Table 1** for details). The attribute levels included in the DCE (e.g. likelihood of achieving clear or almost clear skin at week 16) were informed by clinical data from product labels for AD treatments (where available), including both baricitinib and dupilumab, reflecting the range of potential experiences that patients may have.[19, 20] Attributes included the following: chance of achieving clear or almost clear skin at week 16, chance of achieving a meaningful reduction in itch at week 16, risk of eye inflammation, risk of serious infections, administration, flare management, long-term disease management, monitoring, and speed of onset (**Table 1**). In order to reduce the cognitive burden of the survey, we grouped attributes as benefits, risks, and other. Prior research has found that grouping benefits and risks, and randomising the order of the groups and attributes within the groups, reduces the cognitive burden on participants, thereby reducing ordering effects and increasing choice certainty and the precision of preference estimates.[21]

Table 1. Treatment attributes and levels included in the main discrete choice experiment

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Itch reduction	Eczema (Atopic Dermatitis) causes your skin to itch. Treatments for Eczema (Atopic Dermatitis) increase the probability of achieving a meaningful reduction in itch severity.	**2 out of 10 (20%):** There is a 20% chance of achieving a meaningful reduction in itch severity (reference level) **4 out of 10 (40%):** There is a 40% chance of achieving a meaningful reduction in itch severity **5 out of 10 (50%):** There is a 50% chance of achieving a meaningful reduction in itch severity

Treatment attribute	**Description of the treatment attribute presented to participants**	**Levels**
Skin appearance	Eczema (Atopic Dermatitis) affects the way your skin looks due to flaking, redness, swelling, oozing, crusting, bleeding. Treatment for Eczema (Atopic Dermatitis) may improve your skin condition, but different treatments have different impacts. In this survey, we will ask you to consider the chance of achieving clear skin after 16 weeks starting the treatment.	**1 out of 10 (10%):** After taking treatment for 16 weeks, there is a 10% chance you will have clear/almost-clear skin (reference level) **2 out of 10 (20%):** After taking treatment for 16 weeks, there is a 20% chance you will have clear/almost-clear skin **4 out of 10 (40%):** After taking treatment for 16 weeks, there is a 40% chance you will have clear/almost-clear skin
Eye inflammation	All treatments have some risk of negative side effects. Some treatments can cause minor eye infections. You may have swollen eyelids, feel sensitivity to light, feel itching or burning in your eyes, or have pink discoloration of the white in your eyes. This can be treated but may require interruption to treatment. Other treatments do not increase your risk of getting an eye inflammation.	**0 out of 100 (0%):** Your treatment does not increase the chance of an eye inflammation **10 out of 100 (10%):** There is a 10% chance of experiencing an eye inflammation **20 out of 100 (20%):** There is a 20% chance of experiencing an eye inflammation (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Serious infections	All treatments have some risk of negative side effects. Some treatments reduce your immune system's effectiveness at fighting off illness and can result in serious infections, such as pneumonia or blood poisoning, that may require treatment and hospitalisation; you may be hospitalised for around one week. There is always a very low risk of serious infection and this low risk may be increased.	**0 out of 100 (0%):** Your treatment does not increase the risk of serious infection **3 out of 100 (3%):** 3 out of 100 people will experience a serious infection **6 out of 100 (6%):** 6 out of 100 people will experience a serious infection (reference level)
Speed of onset	All medications for Eczema (Atopic Dermatitis) take some time to start working. Some medications will start to work in 2 days, but others can take 1 or 2 weeks.	**2 days:** Your medication will begin to work 2 days after starting the treatment **1 week:** Your medication will begin to work one week after starting the treatment **2 weeks:** Your medication will begin to work two weeks after starting the treatment (reference level)
Flare management	For some treatments, your doctor can increase your dose if your symptoms get worse (flare-ups). After the flare is controlled, reducing the dose again may also be an option. However, other treatments cannot be adjusted in this way and you will remain on a fixed dose, even if your symptoms change.	**Yes:** Your doctor can increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves **No:** Your doctor cannot increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Long-term disease management	Some treatments for Eczema (Atopic Dermatitis) need to be used continuously, without the option to stop and restart therapy when you want. Interruption of treatment, also known as a treatment holiday, can lead to a loss of efficacy over time. This means the therapy may not work as well when you restart treatment. These treatments must be used continuously and cannot be paused. Other treatments can be stopped and restarted (treatment holiday), with no impact on how effective the treatment is. Some treatments should not be used for the long-term, as they can have life threatening side effects, if used for a long period of time.	**Yes, with the possibility for pauses:** Treatment can be taken long term, and can be paused with no impact on how effective the treatment is **Yes, without the possibility for pauses:** Treatment can be taken long term, but must be taken continuously for there to be no impact on how effective the treatment is **Should not be used long-term:** You can pause the treatment, but using for the long-term may result in life threatening side effects (reference level)
Administration	Treatments are not all given/taken in the same way; for instance, some are pills, others are injections or topical creams. In this study we will only be considering pills and injections.	**Oral pill, once or twice daily** **Injection under the skin, every 2 weeks:** This is a subcutaneous injection, below the skin, but above muscle, usually injected into the thigh/stomach area. You can administer the injection yourself or a health care professional can administer it. If you choose to administer it yourself, you may need to be trained by a nurse on the injection technique. Treatment is once every two weeks. (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Check-ups	Some treatments require periodic blood tests taken by your doctor, because although you may not feel any symptoms, some Eczema (Atopic Dermatitis) medications can have a negative impact on your body.	**Frequent check-ups required:** Blood tests every 2 weeks during the initial 3 months of therapy and then monthly if the patient is stable (reference level). **Occasional check-ups required:** Blood tests at beginning of treatment, after 12 weeks, and then routinely, as determined by your doctor, while on treatment. **No check-ups required**

In each choice task, participants were asked to choose between different treatment options, each composed of one level from each of the attributes. Sensitivity of participants to changes in levels for each attribute were measured relative to the reference level, which is the level that patients least prefer. For example the reference level for risks is the highest level and for efficacy the reference level is the lowest level.

Cognitive pilot Interviews

To ensure the feasibility and robustness of the DCE, cognitive pilot interviews were conducted in the UK, France, and Spain (n=5 per country). The interviews involved a total of 15 patients, who were recruited using the same eligibility criteria as the main study. Patients were recruited through a number of routes, including HCP referrals, social media, and patient databases. The interviews examined whether the chosen attributes and levels were relevant, tradeable, and understandable to participants.[22] In addition, the cognitive pilot interviews assessed the complexity and clarity of the overall questionnaire. Each interview lasted approximately 60 min. Participants were provided a description of the study and completed the initial version of the study survey instrument online while sharing their screen with an interviewer and thinking aloud about the rationale behind their choices. While participants completed the DCE, interviewers probed them using a semi-structured discussion guide. At the end of the interview moderators assessed whether all attributes had been considered, and the overall relevance and plausibility of attributes and levels included in the survey; these assessments were interviewer observed and based on the patients' rationale behind decision making during the interview.

The cognitive pilot interviews were conducted in two waves, with roughly half the participants in each wave. Updates were made after wave 1 and the revised survey was subsequently tested in wave 2. The textual updates after wave 1 were largely minor wording updates to improve the understandability of the survey. However, the presentation of the task and the denominator of serious infections was updated to be consistent with the other risk attribute (eye inflammation). In wave 1, attributes were not initially grouped as benefits, risks, and other. The visualisation of the DCE was adjusted after wave 1 as some participants were misinterpreting the benefit/risk of a treatment. The updated survey grouped and labelled the attributes by category (benefits, risks, other). In wave 2, participants did not have problems understanding the benefits and risks of treatments and found it easier to consider a wider range of attributes. Patients were also asked if they thought any attributes were missing that they would want to know about when selecting a treatment. No missing attributes were identified.

The online DCE survey was initially tested in 29 to 30 participants per country. Minor updates were made to the visual presentation of the survey. Recruitment targets were to include an additional 115 participants in the UK, 115 in Spain, and 85 in France.

Ethics approval

The study was conducted according to good practice for stated preference research[16] and was approved by Ethical & Independent Review Services (Independence, MO, USA; study number 19100-01). In addition, the study was conducted in accordance with International Council on Harmonisation Guidelines for Good Clinical Practice, the ethical principles of the Declaration of Helsinki, the European Union General Data Protection Regulation, and all local laws and regulations.

Participants

Participants were recruited via recruiter databases, social media, patient associations, and online patient panels. Adults (≥18 years) living in the UK, France, or Spain with a self-reported diagnosis of AD for ≥ 12 months were eligible if they had received a topical or systemic therapy for AD in the past 2 years. Participants also had to be able to speak, read, and write the official language of the respective country. Potential participants were excluded if they had a diagnosis of psoriasis, acne, lupus erythematosus, skin cancer, or any other condition that could interfere with participation in and completion of the interview. To account for the possibility that preferences differ between participants with and without self-injectable experience, the study was initially designed to include a target of 40% of participants with prior self-injectable experience, although this was reduced to 30% during the study to allow enough participants to be recruited.

All participants provided online informed consent before participating. Participants in the cognitive pilot consented to being audio-recorded. Participants were remunerated for completing the study.

DCE survey

The DCE was generated using Ngene software v1.2.1 (ChoiceMetrics, Sydney, Australia) using a D-efficient design that was assessed against good experimental design properties. The design was optimized for the estimation of a multinomial logit (MNL) model, and, where appropriate, directional priors. The experimental design of the DCE included 36 experimental choice tasks split into three blocks, such that each participant would complete only 12 experimental choice tasks. Participants in the pilot interviews did not struggle with the number of attributes in the choice tasks. Full profiles (where no attributes were fixed to a set level to simplify the design) were therefore used. In each choice task, participants were asked to choose between two hypothetical treatment options (A and B) and an opt-out of staying with their “old treatment”, wherein each treatment option was composed of one level from each of the attributes (**Figure 1**). If a participant selected the “old treatment” option, they answered a follow-up question asking them to choose between treatment options A and B. We utilised a recommended status-quo opt-out option,[23] which remained fixed throughout the survey (while treatment A and B varied). For methodological reasons, to not overestimate patients’ willingness to accept risks, the risk of adverse events was set to 0% for both eye inflammation and serious infections. Since this would not reflect patients varied current treatments, the opt-out option was referred to as ‘old treatment’. The order of the 12 experimental choice tasks and of the attribute groups (benefits, risks, other) within the choice options was randomised across participants to minimise the influence of ordering effects.[24, 25] In addition to the 12 experimental choice tasks, participants answered two choice tasks to assess internal validity.[26] Task 13 was a repeat of the third experimental choice task seen by the participant and was intended to check the stability of their choices. Task 14 was a dominated-choice test in which one treatment option was as good as or better than the other option for all attributes and was intended to test attendance to the tasks.

In addition to the DCE, participants completed a sociodemographic/clinical questionnaire, indicated their willingness (on a 5 point scale form not willing to very willing) to have a medication that required a subcutaneous injection for each dose, and completed the Set of Brief Screening Questions to assess health literacy[27] and five of the seven items from the Numeracy Scale to assess numeracy[28] to assess their ability to understand the attributes and levels presented and their engagement in the survey.

Validity assessments

For the dominance test, which presented one treatment option with higher levels of benefits and lower levels of risks, the number of patients selecting the superior (dominating) option as their preferred treatment was recorded; selecting the superior option indicated the survey sufficiently engaged participants. The number of patients selecting the same choices in the initial and repeated tasks was also recorded; selecting the same option in both questions indicated choice stability. A respondent was classified as a serial non-participant if they chose the same treatment option for all 12 experimental choice tasks. Decision-making was considered dominated when the respondent chose their preferred treatment option based on a single attribute in all 12 experimental choice tasks. For each choice task, response times in the lower 10% of the corresponding distribution were classified as fast and those in the upper 10% as slow. Attendance to the DCE survey was classified as inadequate if ≥80% of a participant's responses for the 12 experimental choice tasks were classified as too fast or too slow.

Statistical analysis

Statistical analysis was performed using R version 3.6.1 (R Foundation, Vienna, Austria). DCE preference data were analysed using a MNL model within the random utility maximization framework[29] (see **Online Supplemental Methods** for details). This model assumed that respondents chose the alternative that resulted in the highest utility (a measure of desirability) based on the included attributes and up to a random error.[30] The main results from this model were part-worth utility estimates, which reflect participants' sensitivities to changes in the treatment attributes. A dummy coding strategy was implemented to estimate preferences for discrete changes in the treatment attributes. In addition, the MNL model included two alternative-specific constants, one that captured left-right bias (tendency to select the option presented on the left of the choice tasks) and one that captured a preference for the old treatment option.

A second MNL model with linearly coded attributes for the skin appearance attribute was also estimated to support the computation of the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. The acceptability of the underlying hypothesis of linearity in preferences for changes in the skin appearance attribute was first verified (see **Online Supplemental Methods** for details). The MAD analysis measured the percentage decrease in the chance of achieving clear or almost clear skin at week 16 a respondent was willing to accept for changes in other attributes. The 95% confidence intervals for the MAD in achieving clear or almost clear skin at week 16 were obtained using the Delta method.[31]

Subgroup analyses were performed according to country (France, Spain, UK), age (<40, 40–50, and >50 years), gender (female, male), Patient Oriented Eczema

Measure (POEM) overall score (0–7 [clear or almost clear/mild], 8–16 [moderate], severe/very severe [17–28]),[32] and self-reported eczema severity (very mild/mild, moderate/severe/very severe).

Model selection

A number of different analyses were conducted as part of model selection. Given the DCE was conducted in different countries and the initial version of the survey was developed in the English language, the first analysis was related to the possibility of combining choice data from the different countries. The translation of the survey into different languages might have induced a translation effect, which could have resulted in systematic differences in the quality of the choice data across the countries. The results of this analysis indicated that differences in observed choices across countries could not be fully explained by potential changes in the underling quality of the choice data (**Online Supplemental Methods**); as such, it was decided to pool country data and treat country of residence as a potential driver of heterogeneity in preferences alongside other personal characteristics.

The second analysis aimed to determine whether the standard MNL model would be appropriate to quantify average sample preferences. The MNL model was first compared with a mixed logit (MXL) model allowing for unobserved heterogeneity in preferences. Being the most flexible choice model, the MXL model was expected to statistically outperform the MNL model, but the objective of this analysis was to determine whether using a simpler model would lead to a biased measurement of sample preferences. The comparison of preference estimates between the two models showed a very high level of agreement (i.e., very similar preferences identified with both models) (**Online Supplemental Methods and Online Supplemental Figure 1**).

The MNL model was also compared with a nested logit (NL) model to determine whether the opt-out option “old treatment” required different treatment to the other treatment alternatives. The NL model relaxed the hypothesis of independence of irrelevant alternatives, which is a core assumption of the MNL model and implies that all three treatment options were equally substitutable. Again, the comparison of preference estimates showed a high level of agreement between the MNL and NL models (**Online Supplemental Methods and Online Supplemental Figure 2**). These results indicated that the MNL model provided an acceptable approximation of sample preferences.

Patient and public involvement

Cognitive pilot interviews were held with 15 patients to test understandability of the DCE survey. Other than participating in the DCE survey as respondents, patients were not involved in recruitment or study conduct. Investigators were blinded to the identities of the study participants, so the results of the study were not directly disseminated to them.

RESULTS

Participants

The DCE survey included 404 participants (114 in France, 145 in Spain, and 145 in the UK) who were recruited between October and December 2019. Given recruitment for the quantitative online survey used patient panels and databases, 157,553 initial invites were sent, with a 4% (n=6,287) response rate. The majority of the interested potential participants completed the screening questionnaire but were not eligible to participate, largely due to not having AD; 541 patients were eligible to participate, with 75% of those eligible completing the survey. Most participants were female (65%) with an average age of 44.1 years (**Table 2**). Most participants were employed full time (56%) and had completed university education or higher (58%). The majority of participants had moderate-to-very severe AD according to POEM scores (62%) and self-reported eczema severity (67%) but good-to-excellent self-reported overall health (69%). Topical corticosteroids (66%) were the most frequently used class of medications at the time of the survey, followed by systemic immunosuppressant therapies (27%) and biologics (18%). Topical betamethasone (29%) and hydrocortisone (24%) were the most frequent currently used individual medications. Most patients (68%) had no prior experience of using self-injectable treatments for AD or any other illness.

Table 2. Participant characteristics

Characteristic	**N=404**
Sex, n (%)	
Male	142 (35)
Female	262 (65)
Age, mean (SD)	44.1 (12.0)
Employment status	
Full time	227 (56)
Part time	75 (19)
Homemaker/housewife	21 (5)
Student	10 (2)
Unemployed	30 (7)
Retired	35 (9)
Disabled	12 (3)
Other	2 (0)
Education, n (%)	
No formal qualifications	1 (0)
Primary school or secondary education	38 (9)
College or some university	43 (11)
Completed vocational or professional certification	83 (21)
Completed university degree	148 (37)
Completed doctorate, post-doctorate, or equivalent	88 (22)
Other	3 (1)
Overall health, n (%)	
Excellent	20 (5)
Very good	96 (24)
Good	161 (40)
Fair	98 (24)
Poor	29 (7)

Characteristic	N=404
Prior experience with self-injectables (any)*	
Yes	129 (32)
No	275 (68)
Self-rated eczema severity, n (%)	
Very mild	19 (5)
Mild	116 (29)
Moderate	212 (52)
Severe	45 (11)
Very severe	12 (3)
POEM overall score, n (%)	
Clear or almost clear (0–2)	32 (8)
Mild eczema (3–7)	121 (30)
Moderate eczema (8–16)	192 (48)
Severe eczema (17–24)	47 (12)
Very severe eczema (25–28)	12 (3)
Class of AD medication currently used, n (%)†	
Topical corticosteroids	265 (66)
Topical calcineurin inhibitors	32 (8)
Phototherapy/UV treatment	20 (5)
Systemic immunosuppressant therapies	109 (27)
Biologics	72 (18)
Most frequently used current AD medications, n (%)†	
Betamethasone	119 (29)
Hydrocortisone	97 (24)
Prednisone	61 (15)
Clobetasol propionate	46 (11)

*Participants were not asked whether their prior use of self-injectables was for AD.
†Not mutually exclusive.
Abbreviations: AD, atopic dermatitis; POEM, Patient Oriented Eczema Measure; SD, standard deviation

Validity assessments

Overall, the survey sufficiently engaged participants: 89% selected the superior treatment option in the dominance test, 64% chose the same answers in the repeated choice task, and 97% spent an adequate amount of time on the choice tasks (**Online Supplemental Table 2**). Also, for 90% of participants, decisions were not dominated by a single attribute, and only 5% always chose the opt-out old treatment option. Participants were not excluded based on responses to the validity tests, following best practise recommendations,[33] as the preferences of patients may be valid and removal may induce selection bias.

Overall preferences for treatment attributes

The DCE dataset had no missing values, as patients could not proceed in the survey without answering each question or item. If participants did not complete the survey they were not remunerated or included in the dataset. Of the treatment attributes included in the DCE survey, participants most valued improving symptoms and reducing the risk of side effects (**Figure 2** and **Online Supplemental Table 3**). The most valued change was an improvement from 20% to 50% in the chance of achieving a meaningful reduction in itch at week 16, although preferences did not significantly differ between an improvement to a 40% or 50% chance of achieving a meaningful reduction in itch. The next-most valued changes, in descending order, were a decrease in the risk of serious infections from 6% to 0%, a decrease in the risk of eye inflammation from 20% to 0%, and an improvement in the chance of achieving clear or almost clear skin from 10% to 40%.

Participants also valued changes in the non-clinical attributes. The most valued change was switching from a treatment that can be used long-term but cannot be paused without affecting efficacy to one that can be used long-term with the possibility for pauses, without affecting efficacy.

An oral pill once or twice daily was preferred over a subcutaneous treatment every 2 weeks, and a 2-day onset of action was preferred over a 2-week onset of action, although participants did not have a significant preference for a 1-week over a 2-week onset of action. Participants also preferred a treatment that can manage flares by modifying the dose according to symptoms over one that cannot be used to manage flares, although this was less important than changes in other non-clinical attributes.

Subgroup analyses

Results were similar for the three included countries (UK, Spain, and France) (**Online Supplemental Figure 3**), by age (**Online Supplemental Figure 4**), by gender (**Online Supplemental Figure 5**), by POEM overall score (**Online Supplemental Figure 6**), and by self-reported eczema severity (**Online Supplemental Figure 7**). However, those aged over 50 cared more about receiving an oral pill relative to those aged 40-50 years, for whom we did not detect a significant preference for administration.

Participants who had experience of self-injecting a treatment for any illness (32%) were more willing to accept a treatment that required a subcutaneous injection and placed less importance on reducing the risk of serious infections than those who did not have experience self-injecting a treatment for any illness (**Online Supplemental Figure 8**).

Willingness to make trade-offs between treatment attributes

Participants would be willing to tolerate reduced efficacy to obtain changes in other treatment attributes. Specifically, they would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 50.1% (95% CI, 38.5%–61.8%) to increase the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%; 48.6% (95% CI, 35.2%–62.0%) to reduce the risk of serious infections from 6% to 0%; and 42.3% (95% CI, 30.0%–54.5%) to reduce the risk of eye inflammation from 20% to 0% (**Table 3**). They would also be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 24.1% (95% CI, 16.5%–31.6%) to switch from a treatment that can be used long-term but cannot be paused without losing efficacy to one that can be paused without losing efficacy;

16.6% (95% CI, 9.2%–24.0%) to switch from a subcutaneous treatment every 2 weeks to an oral pill once or twice daily; and 5.8% (95% CI, 0.5%–11.1%) to obtain a treatment whose dosage can be modified to manage flares over one that cannot. Further, participants would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 20.9% (95% CI, 12.3%–29.5%) to switch from a treatment that requires frequent check-ups to one that does not require check-ups; and 16.1% (95% CI, 8.7%–23.5%) to switch from a treatment that requires frequent check-ups to one that requires occasional check-ups.

Table 3. Maximum acceptable decrease in the probability of achieving clear or almost clear skin at week 16

Attribute/level	Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)
Itch reduction	
2 out of 10 (20%)	Reference
4 out of 10 (40%)	38.7 (28.8, 48.6)
5 out of 10 (50%)	50.1 (38.5, 61.8)
Eye inflammation	
20 out of 100 (20%)	Reference
10 out of 100 (10%)	17.9 (10.5, 25.4)
0 out of 100 (0%)	42.3 (30.0, 54.5)
Serious infections	
6 out of 100 (6%)	Reference
3 out of 100 (3%)	20.6 (12.7, 28.6)
0 out of 100 (0%)	48.6 (35.2, 62.0)
Speed of onset	
2 weeks	Reference
1 week	0.2 (−6.5, 6.9)
2 days	11.3 (4.4, 18.2)
Flare management	
No	Reference
Yes	5.8 (0.5, 11.1)
Long-term disease management	
Yes, without the possibility for pauses	Reference
Should not be used long-term	4.3 (−2.7, 11.3)
Yes, with the possibility for pauses	24.1 (16.5, 31.6)
Administration	
Injection under the skin every 2 weeks	Reference
Oral pill once or twice daily	16.6 (9.2, 24.0)
Check-ups	

Attribute/level	Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)
Frequent check-ups required	Reference
Occasional check-ups required	16.1 (8.7, 23.5)
No check-ups required	20.9 (12.3, 29.5)

Abbreviations: CI, confidence interval

DISCUSSION

The current study, which included 404 participants across the UK, France, and Spain, found that adults with AD who had recently been treated with topical and/or systemic therapy most valued increasing the benefits and reducing the risks of their treatments, although attributes specific to new targeted therapies, such as mode of administration and long-term disease management, also had a significant effect on choices. Participants were willing to tolerate a significant decrease in the possibility of achieving clear or almost clear skin to obtain a treatment that is more convenient, including an oral pill once or twice daily in place of a subcutaneous injection every 2 weeks, the ability to pause the treatment without losing efficacy, the ability to modify the dosage to manage flares, and the possibility of requiring only occasional or no check-ups instead of frequent check-ups. Further, participants with self-injectable experience for any illness were more willing to accept self-injection than participants without self-injectable experience. However, 28% of participants were ‘not willing’ or ‘somewhat not willing’ to have a medication that required an injection for each dose. Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age or sex. In addition, preferences did not significantly differ based on disease severity, as measured using the POEM score, which is in line with prior research.[34]

Two other recent DCEs have examined the treatment preferences of patients with AD. Similar to our study, a DCE in the US including 320 adults with moderate-to-severe AD[34] found that patients preferred an oral pill over subcutaneous injection and valued a rapid onset of action and increasing the chance of achieving clear or almost clear skin at week 16. A DCE including 323 patients in Japan ≥ 15 years of age with moderate to very severe AD and 121 dermatologists treating patients with AD[35] found that, as in the current study, both groups considered benefits and adverse effects the most important attributes of injectable treatments, although preferences for some treatment attributes differed between the groups. For example, patients placed more value on efficacy of improving rashes and treatment costs than dermatologists, while dermatologists valued time until response more than patients. Patients also preferred adding new treatments to current treatments as add-ons and receiving treatments at clinics, while physicians preferred reducing the number of current treatments and having patients self-administer at home. These differences in the preferences of patients and physicians emphasize the need for studies like the current one that are specifically designed to provide insight into patients' preferences.

Internal validity of the current DCE was examined using tests of choice stability and dominance, as well as by considering response times, health literacy, and numeracy. The results were in line with existing research, including for choice stability,[26] and suggested the survey sufficiently engaged participants. A potential limitation of this study is that the attributes and levels were not identified through a separate qualitative research phase but rather through a targeted review of previous quantitative and qualitative studies of patients with AD and product labels for AD treatments. We do not expect that this influenced the results because the same attributes (onset of itch relief, probability of skin clearance, frequency or ease of administration/convenience, and safety) were also identified through the qualitative phase of the US/UK study.[34]

A potential limitation of this study is the inclusion of four probabilistic attributes, which increased the complexity of the study for participants. These were included to align with clinical data. To mitigate this, we included a thorough warm-up to the DCE with practice questions relating to the probabilistic attributes. In addition, a prior AD study included four probabilistic attributes (two probabilistic benefits and two probabilistic adverse events).[34] Another limitation of this study is that we used different denominators for probabilistic benefit and risk attributes. Different denominators were utilised to ensure participants could review all attribute information simultaneously while making their choices. However, using different denominators may have increased the study complexity and introduced a potential bias. Another potential limitation of this study is reference to the opt-out as 'old', which may have been perceived negatively. We used the terminology 'old' instead of current since we were aware that we were not presenting patients with their actual current treatments, which may have caused confusion. Due to the need to limit the participants' cognitive burden, not all potentially relevant attributes could be included in the DCE survey. However, cognitive pilot interviews of 15 patients with AD indicated that the attributes and levels were relevant and that no attributes were missing. Overall, participants also found the length and complexity of the survey acceptable. A further limitation is the inclusion of patients with non-severe AD, who would possibly not receive systemic therapies.[2] However, there is value in including these patients, because patients' disease severity may vary over time and treatment recommendations may change. Also, although few differences were found in preferences by age, sex, or country, care should be taken when generalizing to underrepresented AD populations, such as patients with very severe AD, children, or patients in lower income countries. Additionally, since it is not culturally appropriate to ask about race in some European countries, data was not collected on this. We were therefore not able to determine whether this study represents the diverse ethnic groups in the study countries. Moreover, our sample included a high proportion of participants with university education and may therefore not be fully representative of the general AD population.

In conclusion, patients with AD most valued treatment benefits and reducing risks but were willing to accept a decrease in efficacy, as measured by the possibility of obtaining clear or almost clear skin at week 16, to obtain an oral treatment with a rapid onset of action. This information may help clinicians make shared decisions with patients about the most suitable treatment for AD. It can also support reimbursement applications, ensuring that health technology assessment decisions align with the preferences of individuals living with AD.

Author contributions: C.T. contributed to conception and planning of the study, acquisition and interpretation of the data, and drafting and critical revision of the manuscript. A.R. contributed to conception and design of the study, interpretation of the data, and critical revision of the manuscript. A.W. and J.-P.C. contributed to design of the study, interpretation of the data, and critical revision of the manuscript. N.K. contributed to analysis and interpretation of the data and critical revision of the manuscript. H.K. contributed to conception and design of the study and critical revision of the manuscript. T.T. contributed to conception and planning of the study, interpretation of the data, and critical revision the manuscript. All authors approved the final version of the manuscript.

Acknowledgements: Medical writing support was provided by Philip S. Leventhal, PhD, Holly Richendrfer, PhD, and Stephen Gilliver, PhD of Evidera and was paid for by Eli Lilly & Co.

Competing Interests: C.T., N.K., H.K., and T.T. are employees of Evidera, which was paid by Eli Lilly & Co. for work related to this study. A.R. and J.-P.C. are employees of Eli Lilly & Co. A.W. received personal fees from Eli Lilly & Co for work related to this study.

Funding: This work was supported by Eli Lilly & Company. Grant/award number is not applicable

Data availability statement: The datasets generated and/or analysed during the study are not publicly available, because consent was not sought from participants to allow sharing of data with third parties.

REFERENCES

1. Fishbein AB, Silverberg JI, Wilson EJ, et al. Update on Atopic Dermatitis: Diagnosis, Severity Assessment, and Treatment Selection. *J Allergy Clin Immunol Pract* 2020;8:91-101.
2. Wollenberg A, Christen-Zach S, Taieb A, et al. ETFAD/EADV Eczema task force 2020 position paper on diagnosis and treatment of atopic dermatitis in adults and children. *J Eur Acad Dermatol Venereol* 2020;34:2717-44.
3. Wollenberg A, Barbarot S, Bieber T, et al. Consensus-based European guidelines for treatment of atopic eczema (atopic dermatitis) in adults and children: part II. *J Eur Acad Dermatol Venereol* 2018;32:850-78.

4. Sidbury R, Davis DM, Cohen DE, et al. Guidelines of care for the management of atopic dermatitis: section 3. Management and treatment with phototherapy and systemic agents. *J Am Acad Dermatol* 2014;71:327-49.

5. Seegraber M, Srour J, Walter A, et al. Dupilumab for treatment of atopic dermatitis. *Expert Rev Clin Pharmacol* 2018;11:467-74.

6. OLUMIANT® Summary of Product Characteristics: Eli Lilly Nederland B.V.; 2020. Available from: https://ec.europa.eu/health/documents/community-register/2020/20201019149480/anx_149480_en.pdf. Accessed 02 December 2020.

7. Rinvoq: Summary of opinion (post authorisation): Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency; 2021. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-rinvoq-x-06-g_en.pdf. Accessed 09 November 2021.

8. Deleanu D, Nedelea I. Biological therapies for atopic dermatitis: An update (Review). *Exp Ther Med* 2018;17:1061-67.

9. Casciano R, Malangone E, Ramachandran A, et al. A quantitative assessment of patient barriers to insulin. *Int J Clin Pract* 2011;65:408-14.

10. Edel Y, Sagy I, Pokroy-Shapira E, et al. A Cross-sectional Survey on the Preference of Patients with Rheumatoid Arthritis for Route of Administration of Disease-Modifying Anti-Rheumatic Drugs: Oral Target-Specific Versus Parenteral Biologic. *Isr Med Assoc J* 2020;22:154-59.

11. Mohr DC, Boudewyn AC, Likosky W, et al. Injectable medication for the treatment of multiple sclerosis: The influence of self-efficacy expectations and

infection anxiety on adherence and ability to self-inject. *Ann Behav Med* 2001;23:125-32.

12. Bouvy JC, Cowie L, Lovett R, et al. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. *Patient* 2020;13:145-49.
13. Huls SPI, Whichello CL, van Exel J, et al. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. *Value Health* 2019;22:1318-28.
14. Janus SI, Weernink MG, van Til JA, et al. A Systematic Review To Identify the Use of Preference Elicitation Methods in Health Care Decision Making. *Value Health* 2014;17:A515-6.
15. Mühlbacher AC, Juhnke C, Beyer AR, et al. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. *Value Health* 2016;19:734-40.
16. Johnson FR, Zhou M. Patient Preferences in Regulatory Benefit-Risk Assessments: A US Perspective. *Value Health* 2016;19:741-45.
17. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. *Patient* 2017;10:263-66.
18. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics* 2018;37:201-26.
19. Radi G, Simonetti O, Rizzetto G, et al. Baricitinib: The First Jak Inhibitor Approved in Europe for the Treatment of Moderate to Severe Atopic Dermatitis in Adult Patients. *Healthcare (Basel)* 2021;9

20. Thyssen JP, Buhl T, Fernández-Peñas P, et al. Baricitinib Rapidly Improves Skin Pain Resulting in Improved Quality of Life for Patients with Atopic Dermatitis: Analyses from BREEZE-AD1, 2, and 7. *Dermatol Ther (Heidelb)* 2021;11:1599-611.

21. Heidenreich S, Phillips-Beyer A, Flamion B, et al. Benefit-Risk or Risk-Benefit Trade-Offs? Another Look at Attribute Ordering Effects in a Pilot Choice Experiment. *Patient* 2021;14:65-74.

22. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Econ* 2009;18:321-36.

23. Determann D, Gyrd-Hansen D, de Wit GA, et al. Designing Unforced Choice Experiments to Inform Health Care Decision Making: Implications of Using Opt-Out, Neither, or Status Quo Alternatives in Discrete Choice Experiments. *Med Decis Making* 2019;39:681-92.

24. Carlsson F, Mørkbak MR, Olsen SB. The first time is the hardest: A test of ordering effects in choice experiments. *J Choice Model* 2012;5:19-37.

25. Heidenreich S, Phillips-Beyer A, Flamion B, et al. Benefit–risk or risk–benefit trade-offs? Another look at attribute ordering effects in a pilot choice experiment. *Patient* 2021;14:65-74.

26. Johnson FR, Yang J-C, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. *Value Health* 2019;22:157-60.

27. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:588-94.

28. Lipkus IM, Samsa G, Rimer BK. General Performance on a Numeracy Scale among Highly Educated Samples. *Med Decis Making* 2016;21:37-44.

29. Manski CF. The structure of random utility models. *Theory Decis* 1977;8:229-54.

30. Thurstone LL. A law of comparative judgment. *Psychol Rev* 1927;34:273-86.

31. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ* 2007;16:827-40.

32. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. *Arch Dermatol* 2004;140:1513-9.

33. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ* 2006;15:797-811.

34. Boeri M, Sutphin J, Hauber B, et al. Quantifying patient preferences for systemic atopic dermatitis treatments using a discrete-choice experiment. *J Dermatolog Treat* 2020:1-10.

35. Okubo Y, Ho K-A, Fifer S, et al. Patient and physician preferences for atopic dermatitis injection treatments in Japan. *J Dermatolog Treat* 2019;31:821-30.

FIGURE LEGENDS

Figure 1. Example choice task

Figure 2. Multinomial logit results: part-worth utilities

Dr Tommi Tervonen
Evidera
London, UK
tommi.tervonen@evidera.com
+44 (0) 77 88 237 412

Adrian Aldcroft
Editor in Chief
BMJ Open

Dear Adrian Aldcroft,

We would like to submit revision of our manuscript “Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment” for further consideration of publication in *BMJ Open*.

We have carefully revised the manuscript according to all reviewer comments, which truly helped us to improve the paper. We hope you and the reviewers enjoy reading the revised version.

Thank you for considering our manuscript for publication in *BMJ Open*. We look forward to your response.

Sincerely,

Dr Tommi Tervonen
Director of Patient Preferences, Evidera, London

Table of Contents

Cover letter 2

Response to reviewers 4

Main manuscript with track changes 20

Main manuscript clean version 54

Main manuscript figures 86

Supplement with track changes 88

Supplement clean version 106

Supplemental figures 126

ISPOR checklist 134

Point-by-point response

Editorial requests:

- Please work on improving the strengths and limitations section after the abstract. What are the study's methodological strengths?

Author response: We have revised this section. The amended strengths and limitations are below:

- *"This study used a discrete choice experiment to elicit preferences of patients in the UK, France, and Spain for attributes of atopic dermatitis treatments.*
- *Patients with AD most valued treatment benefits and reducing risks but were willing to accept a decrease in efficacy, as measured by the possibility of obtaining clear or almost clear skin at week 16, to obtain an oral treatment with a rapid onset of action.*
- *Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age, sex, or disease severity.*
- *This sample was an adult population from the UK, France, and Spain, and a high proportion of patients had a university education. Therefore, the study may not be generalisable to children, patients in other countries, or those with lower levels of education. In addition, patients had predominantly moderate to severe AD, and these findings may not apply to the wider AD adult population, including those with mild or very severe AD."*

- Please clarify what reporting checklist has been completed in your submission. Is the checklist endorsed by EQUATOR?

Author response: We used the ISPOR Conjoint Checklist (ESTIMATE). This is not endorsed by EQUATOR, but there is no appropriate or more applicable checklist endorsed by EQUATOR for this type of research.

Reviewer: 1

Dr. Steven Feldman, Wake Forest University Comments to the Author:
This is a well done, interesting, timely study with clinically relevant findings.

Author response: We thank the reviewer for their appreciation of our work.

1. Words like shown, known, demonstrated, proven, etc, are signs that writing can be more concise. For example, "Studies in other chronic diseases have shown that" can be shortened to "in other chronic diseases"

Author response: Thank you for the suggestion. We have reviewed and adjusted wording throughout.

2. I would delete all claims of being "first" as such claims are of no scientific relevance and are impossible to prove would be true at the time of publication.

Author response: We have removed any mention of being the "first".

3. Phrases ending in "that" can generally be deleted from the beginning of sentences (for example, "It showed that" is useless verbiage).

Author response: Thank you for the suggestion. We have reviewed and adjusted the wording throughout.

Reviewer: 2

Dr. Stephanie Lax, University of Nottingham Comments to the Author:
I was very interested to read this paper aiming to better understand eczema treatment preferences, contributing to a culture of shared decision-making with people with eczema alongside healthcare professionals. The reporting was detailed and clear, with helpful supplementary material. Overall, I think this was a well conducted study with only minor limitations.

Author response: We thank the reviewer for their appreciation of our work.

Abstract:
1. In the abstract, I stumbled over the statement "68% naïve to self-injecting". I think this summary of demographic details in the abstract is too brief to be clear; I'd prefer the information in parentheses to be either expanded or cut.

Author response: We have revised this and have added an additional sentence on this in the abstract results and also in the participants section of the results. Please see below:

> *"Most patients (68%) had no prior experience of using self-injectable treatments for AD or any other illness"*

2. Where the MAD results are discussed, the following is omitted: MAD=20.9% if no check-ups required. Is there a reason for this?

Author response: These results are reported in the main body (willingness to make trade-offs between treatment attributes; page 18). We have now added additional text to the abstract calling out the result that patients would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 20.9% to switch from frequent to no check-ups required. The revised text in the abstract is below with the new text underlined:

> *"Participants were willing to accept a decrease in the possibility of achieving clear/almost clear skin to obtain a treatment that can be paused (MAD = 24.1%), requires occasional check-ups (MAD = 16.1%) or no check-ups (MAD = 20.9%) over frequent check-ups, is administered as a once- or twice-daily oral pill versus a subcutaneous injection every 2 weeks (MAD = 16.6%), has a 2-day over 2-week onset of action (MAD = 11.3%), and ~~the ability to use a treatment~~ can be used for flare management (MAD = 5.8%).*

Methods:
1. The authors conducted a review of the literature and identified quantitative and qualitative studies from which to draw an appropriate, succinct set of attributes. The review is described as

targeted, rather than systematic, however please can you consider adding more details in line with PRISMA, particularly regarding the search and including a search date so the reader can see how current it is?

Author response: We have revised the text in the methods in the manuscript to include the search date, underlined below:

> *"Treatment attributes and levels included in the DCE were identified through a targeted literature review of Embase and MEDLINE for quantitative and qualitative preference studies and a review of product labels for AD treatments (search conducted 10th September 2018; see Online Supplemental Methods and Online Supplemental Table 1 for details)."*

We have also added the targeted literature review search terms (Online Supplemental Table 1) and a PRISMA-based description of the literature screening (Supplemental Methods) to the supplementary materials.

2. Levels were described relative to a worst-case scenario; however, I would welcome clarity in the model specification on how the alternatives were chosen (e.g., why a maximum of 50% chance of achieving a meaningful reduction in itch severity? Or a 40% chance of clear/almost clear skin?). Does this reflect the range of potential experiences?

Author response: Yes, the attribute level ranges were informed by clinical data for atopic dermatitis treatments and reflect the range of potential experiences. We conducted a product label review of treatments looking at the efficacy and safety of treatments used to treat atopic dermatitis and the range utilised per attribute covered the maximum identified across the treatments.

For example, Dupilumab[1] reported 36% (trial 1) and 41% (trial 2) achieving a meaningful reduction in itch (NRS >= 4 improvement at week 16) and 38% (trial 1) and 36% (trial 2) achieving clear or almost clear skin (IGA 0 or 1 at week 16). Baricitinib reported a meaningful reduction in itch of 19% to 48.8% across trials with varied dosages (NRS >= 4 improvement at week 16).[2] Baricitinib reported 13.8% - 30.6% achieving clear or almost clear skin (IGA 0 or 1 at week 16) across their trials of varying dosages.[3] These numbers fall within our included attribute ranges.

We have added some additional information to the 'materials and methods' on the treatment data and how this guided the level ranges included in the study:

> *"Treatment attributes and levels included in the DCE were identified through a targeted literature review of Embase and MEDLINE for quantitative and qualitative preference studies and a review of product labels for AD treatments (search conducted 10th September 2018; see Online Supplemental Methods and Online Supplemental Table 1 for details). The attribute levels included in the DCE (e.g. likelihood of achieving clear or almost clear skin at week 16) were informed by clinical data from product labels for AD treatments (where available), including both baricitinib and dupilumab, reflecting the range of potential experiences that patients may have [19,20]"*

3. Would the authors consider making available their R scripts so that others might be better able to reproduce the work on similar data?

Author response: Unfortunately, our analyses contain partly proprietary code which cannot be shared. We also do not think the code would be very useful for other researchers as we used standard functions from existing packages for fitting the models.

Study limitations:
The authors discuss the limitations of the work well; they cover most key issues, explain mitigating steps taken, and indicate whether they were likely to have influenced the results. However, please could the following be considered:

1. Please could you comment on whether the findings of this study adequately represent the diverse ethnic groups in France, Spain, and the UK?

Author response: Thank you for this question. Since it is not culturally appropriate to ask about race in some European countries, like France, data was not collected on this. We have noted an additional limitation in the discussion that we do not know whether this sample is representative of the diverse ethnic groups of the included study countries.

> *"Additionally, since it is not culturally appropriate to ask about race in some European countries, data was not collected on this. We were therefore not able to determine whether this study represents the diverse ethnic groups in the study countries."*

2. Supplementary Table 1 indicates only about 60% choice stability; is this reasonable and could the authors comment on the implications of this on their results?

Author response: While 64% providing consistent responses in the repeated question may sound low, this is in line with existing research, and we do not think this has implications for the results. We have moved a reference in the discussion (Johnson et al, 2019) where we note the results are in line with existing research. Johnson et al, 2019 found that across 55 DCE data sets the average proportion of patients who did not provide consistent responses was 30% (± 26%) with a range of 0%-81%. In addition, a large proportion of preference studies do not include validity tests or report on them. Please see the revised text in the discussion below:

> *"Internal validity of the current DCE was examined using tests of choice stability and dominance, as well as by considering response times, health literacy, and numeracy.~~[21]~~ The results were in line with existing research, including for choice stability, [~~30~~21] and suggested ~~that~~ the survey sufficiently engaged participants ~~paid adequate attention to the survey~~."*

Other:
1. I am concerned about the wording of the validity assessments contents in terms of respondent pass or failure. In a study of patient perspectives, I think care should be taken not to undermine the patient voice where it could be some feature of the survey that results in failure. For example, I would favour "Where respondents were asked to complete repeated tasks, selection of different choices indicated a failure of stability."

Author response: Thank you for the suggestion. We agree with your comments and have revised the text in the methods accordingly to remove discussion of pass and failure.

> *"For the dominance test, which presented one treatment option with higher levels of benefits and lower levels of risks, the number of patients selecting the superior (dominating) option as their preferred treatment was recorded; selecting the superior option indicated ~~that~~ the survey sufficiently engaged participants. The number of patients selecting the same choices in the initial and repeated tasks was also recorded; selecting the same option in both questions indicated choice stability."*

We have also updated the discussion of results to say:

> *" Overall, the survey sufficiently engaged participants ~~appeared to have paid adequate attention to the DCE choice tasks~~: 89% ~~passed the~~ selected the superior treatment option in the dominance test, 64% chose the same answers in the repeated choice task, and 97% spent an adequate amount of time on the choice tasks (Online Supplemental Table 2). Also, for 90~~63~~% of participants, decisions were not dominated by a single attribute, and only 5% always chose the opt-out old therapy option"*

2. As above, in the discussion "The results were in line with existing research[31] and suggested that participants paid adequate attention to the survey" could be reframed as "the survey sufficiently engaged participants" or similar.

Author response: Thank you for the suggestion.

We have revised the discussion to say:

> *"Internal validity of the current DCE was examined using tests of choice stability and dominance, as well as by considering response times, health literacy, and numeracy.~~[21]~~ The results were in line with existing research, including for choice stability,[26] and suggested that the survey sufficiently engaged participants ~~paid adequate attention to the survey~~."*

References:
Mangham LJ, Hanson K, McPake B. How to do (or not to do) … Designing a discrete choice experiment for application in a low-income country Health Policy and Planning 2009; 24(2):151–158. https://doi.org/10.1093/heapol/czn047
The PRISMA 2020 statement: an updated guideline for reporting systematic reviews BMJ 2021; 372. https://doi.org/10.1136/bmj.n71 Sharma, A., Palaniappan, L. Improving diversity in medical research Nature Reviews Disease Primers 2021; 7:74. https://doi.org/10.1038/s41572-021-00316-8
Block J. Improving value for patients with eczema Value in Health 2018; 21:380-385. https://doi.org/10.1016/j.jval.2018.01.014.

I would like to thank Professor Kim S Thomas (Centre of Evidence Based Dermatology, University of Nottingham, UK) for critical reading of this peer review.

Reviewer: 3

Dr. Marco Boeri, RTI Health Solutions
Comments to the Author:
This paper presents the results from a discrete choice experiment (DCE) administered in the UK, France and Spain to explore preferences for atopic dermatitis medications.
Knowing, the previous work of the authors, I am surprised by the lack of discussion regarding a few major issues I have identified in both design of the study and the analysis of the data. I am convinced that the authors can explain their approach and discuss further the decisions taken during the study.

1.
Firstly, the study includes a number of attributes and levels (9 attributes seven of which have 3 levels) higher than the average study. Furthermore, 4 of these attributes are probabilistic. Guidelines for DCE design suggest to include 6-8 attributes and no more than 1 or 2 probabilistic. In addition, from the example of the DCE question included in page 30, two of the risk grid are expressed out of 10 and two are expressed out of 100. This is also not the standard approach to designing DCE questionnaires (normally guidelines suggest that it is best to use the same denominator for each probabilistic attribute included in the study).
I know the authors mention pretest interviews (5 in each country which to me seem not enough to test for the complexity of the attribute table, but I am willing to listen to the justification given by the authors), but I would like to know more about how the pretest went and in particular how interviewers established that respondents could interpret and consider 9 attributes of which 4 probabilistic with two different denominators.

Author response: Thank you for your review and important questions. We have revised a number of sections, noted below, based on your comments. Every attribute in the study was found to have a significant influence on preferences (from best to worst level), suggesting that patients considered all attributes. As such we believe these strategies were effective at reducing the cognitive burden for patients whilst ensuring key information was not missing and that the study aligned with clinical data.

We have added additional text to the methods, underlined below:

> *"Attributes included the following: chance of achieving clear or almost clear skin at week 16, chance of achieving a meaningful reduction in itch at week 16, risk of eye inflammation, risk of serious infections, administration, flare management, long-term disease management, monitoring, and speed of onset (Table 1). In order to reduce the cognitive burden of the survey, we grouped attributes as benefits, risks, and other. Prior research has found that grouping benefits and risks, and randomising the order of the groups and attributes within the groups, reduces the cognitive burden on participants, thereby reducing ordering effects and increasing choice certainty and the precision of preference estimates.[21]"*

We have separated out the cognitive pilot interviews into a new section in the methods with additional information:

> ***"Cognitive Pilot Interviews***

> *To ensure the feasibility and robustness of the DCE, cognitive pilot interviews were conducted in the UK, France, and Spain (n=5 per country). The interviews involved a total of 15 patients, who were recruited using the same eligibility criteria as the main study. Patients were recruited through a number of routes, including HCP referrals, social media, and patient databases. The interviews examined whether the chosen attributes and levels were relevant, tradeable, and understandable to participants.[22] In addition, the cognitive pilot interviews assessed the complexity and clarity of the overall questionnaire. Each interview lasted approximately 60 min. Participants were provided a description of the study and completed the initial version of the study survey instrument online while sharing their screen with an interviewer and thinking aloud about the rationale behind their choices. While participants completed the DCE, interviewers probed them using a semi-structured discussion guide. At the end of the interview moderators assessed whether all attributes had been considered, and the overall relevance and plausibility of attributes and levels included in the survey; these assessments were interviewer observed and based on the patients' rationale behind decision making during the interview.*
>
> *The cognitive pilot interviews were conducted in two waves, with roughly half the participants in each wave. Updates were made after wave 1 and the revised survey was subsequently tested in wave 2. The textual updates after wave 1 were largely minor wording updates to improve the understandability of the survey. However, the presentation of the task and the denominator of serious infections was updated to be consistent with the other risk attribute (eye inflammation). In wave 1, attributes were not initially grouped as benefits, risks, and other. The visualisation of the DCE was adjusted after wave 1 as some participants were misinterpreting the benefit/risk of a treatment. The updated survey grouped and labelled the attributes by category (benefits, risks, other). In wave 2, participants did not have problems understanding the benefits and risks of treatments and found it easier to consider a wider range of attributes. Patients were also asked if they thought any attributes were missing that they would want to know about when selecting a treatment. No missing attributes were identified.*

Additional text has also been added to the limitations regarding the number of probabilistic attributes:

> *"A potential limitation of this study is the inclusion of four probabilistic attributes, which increased the complexity of the study for participants. These were included to align with clinical data. To mitigate this, we included a thorough warm-up to the DCE with practice questions relating to the probabilistic attributes. In addition, a prior AD study included four probabilistic attributes (two probabilistic benefits and two probabilistic adverse events).[34] Another limitation of this study is that we used different denominators for probabilistic benefit and risk attributes. Different denominators were utilised to ensure participants could review all attribute information simultaneously while making their choices. However, using different denominators may have increased the study complexity and introduced a potential bias."*

2.
How was the opt-out ("your old treatment") defined? As mentioned below in my review, the study considers respondents from different countries and different severities, how did you define the old treatment? Could "current" instead of "old" be a better word (old seems to carry a negative meaning).

In particular, I am interested in how the levels were selected for each respondent. And how the authors take into consideration these differences in their model.

Author response: We have added additional text to the methods in the DCE survey section on the opt-out:

> *"We utilised a recommended status-quo opt-out option [23], which remained fixed throughout the survey (whilst treatment A and B varied). For methodological reasons, to not overestimate patients' willingness to accept risks, the risk of adverse events was set to 0% for both eye inflammation and serious infections. Since this would not reflect patients varied current treatments, the opt-out option was referred to as 'old treatment'."*

We have added additional text to the limitations:

> *"Another potential limitation of this study is reference to the opt-out as 'old', which may have been perceived negatively. We used the terminology 'old' instead of current since we were aware that we were not presenting patients with their actual current treatments, which may have caused confusion."*

3.

One thing that struck me was the sentence that 63% of responses were not dominated by a single attribute – that means 37% were. That seems high to me, so I would like to ask which attributes people dominated on and whether the authors think that is a sign that there were too many attributes.

Author response: Thank you for highlighting this. Apologies, this is an error. The data quality numbers have been updated in the manuscript. Only 10% of choices were dominated by an attribute (5% administration, 2% serious infections, 1% eye inflammation, 1% itch reduction, <1% (n=1) skin appearance, and <1% (n=1) flare management).

4.

The study recruited respondents from multiple countries (namely UK, France and Spain), but there is no mention of possible problems related to cultural and demographic differences across countries. Preferences are normally explored by country and when data are used together from multiple countries it is good practice to test for whether it is possible to pull the data (test for differences in preferences and scale). I would like to see the results from the test proposed by Swait and Louviere in 1993 to see whether you cannot reject the hypothesis of same preferences across countries (and therefore pull the data) and whether you should adjust for scale heterogeniety across countries.

The procedure is simple: estimate separate MNL for each country, estimate an MNL with all data pulled and estimate a heteroschedastic MNL model with scale fully explained by a dummy for each country, then execute the two step test explained in Swait and Louviere (1993).

Swait, J., & Louviere, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. Journal of Marketing Research, 30(3), 305–314. https://doi.org/10.2307/3172883

Author response: We have estimated an HMNL model allowing for scale differences between countries. The likelihood ratio test (LRT) indicated that this model performed significantly better than the standard

MNL model (D=11.45 (P=0.003)). However, we also estimated an extended version of the MNL model allowing for interaction effects between the country of residence and the attributes' levels and this interacted MNL (IMNL) model also appeared to significantly outperform the standard MNL model (D=66.44 (P=0.001)).

Using the scale estimates from the HMNL model, we applied a scale correction to the dataset and then re-estimated the IMNL model to determine whether the interaction effects which were found to be significant in the initial IMNL model would remain significant after having accounted for potential scale differences between countries. This was the case, indicating thus that differences in choice behaviours between countries can't be fully explained as the consequence of a change in underlying utility scale. This is the reason why we decided to pool the data together and treat "country of residence" as any other potential source of heterogeneity in preferences (alongside other personal characteristics).

We have added additional text to the manuscript in a new section:

> *"**Model selection***
>
> *A number of different analyses were conducted as part of model selection. Given the DCE was conducted in different countries and the initial version of the survey was developed in the English language, the first analysis was related to the possibility of combining choice data from the different countries. The translation of the survey into different languages might have induced a translation effect, which could have resulted in systematic differences in the quality of the choice data across the countries. The results of this analysis indicated that differences in observed choices across countries could not be fully explained by potential changes in the underling quality of the choice data (**Online Supplemental Methods**); as such, it was decided to pool country data and treat country of residence as a potential driver of heterogeneity in preferences alongside other personal characteristics.*
>
> *The second analysis aimed to determine whether the standard MNL model would be appropriate to quantify average sample preferences. The MNL model was first compared with a mixed logit (MXL) model allowing for unobserved heterogeneity in preferences. Being the most flexible choice model, the MXL model was expected to statistically outperform the MNL model, but the objective of this analysis was to determine whether using a simpler model would lead to a biased measurement of sample preferences. The comparison of preference estimates between the two models showed a very high level of agreement (i.e., very similar preferences identified with both models) (**Online Supplemental Methods and Online Supplemental Figure 1**).*
>
> *The MNL model was also compared with a nested logit (NL) model to determine whether the opt-out option "old treatment" required different treatment to the other treatment alternatives. The NL model relaxed the hypothesis of independence of irrelevant alternatives, which is a core assumption of the MNL model and implies that all three treatment options were equally substitutable. Again, the comparison of preference estimates showed a high level of agreement between the MNL and NL models (**Online Supplemental Methods and Online Supplemental Figure 2**). These results indicated that the MNL model provided an acceptable approximation of sample preferences."*

5.
The study collects data from a wide range of severity (from mild to severe) and I am concerned differences are not taken into consideration when modelling the data. The author mention this fact in the limitations, but I suggest to run the same test proposed by Swait and Louviere (1993) across different severities, at least separating severe from the rest and see if you can pull the data (conditionally to the fact that you can pull across countries).

Author response: We initially decided not to run an analysis of heteroscedasticity based on other personal characteristics than country of residence because in line with other studies[4-6] *we considered that the HMNL model should be used to explore "survey mode" effects (e.g., change in underlying experimental design; translation of survey into different languages). Of course, differences in personal characteristics of the respondents could also cause some scaling effects, however this should become apparent when running interacted analyses (all the interaction effects either reinforcing or attenuating marginal utilities would indicate likely existence of a scaling effect). Given that it is mathematically impossible to simultaneously separate scaling and trade-off effects, we initially decided to reserve the personal characteristics for the more general analysis of heterogeneity in preferences. However, we decided to follow the reviewer's suggestion and run an exploratory analysis of severity effect on choice behaviours following same approach as before (see Comment #1). Both the HMNL and IMNL models appeared to significantly outperform the standard MNL model (HMNL: D=66.44 (P=0.001); IMNL: D=68.1 (P<0.001)) such that it was impossible to conclude that differences in preferences between the subsamples would be entirely driven by a scaling effect. The Swait-Louviere procedure reached the same conclusion with combined performance of the separate models being significantly higher than performance of the pooled model (D=84.71 (P<0.001)).*

As we report in the manuscript, preferences did not differ significantly based on disease severity as measured using the POEM score. This is in line with prior research. We have added additional information on this to the discussion:

> *"Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age or sex. In addition, preferences did not significantly differ based on disease severity, as measured using the POEM score, which is in line with prior research [34]"*

6a.
Preference heterogeneity is not considered in the model (the authors run a MNL and not a RPL or LC analysis, which could cause major bias in the results as well as doubts from a reader perspective - e.g., are the more advanced models unidentified due to the construction of the sample and the study?). I would like to see evidence that advanced models considering preference heterogeneity were tested and considered not needed to analyse this data (although considering the many sources of heterogeneity identified so far, I doubt MNL can be the best model for the data).

Author response: We estimated a MXL model allowing all parameters to be independently and normally distributed (i.e., diagonal covariance matrix of random effects). Whilst the MXL model appeared to significantly outperform its MNL counterpart (LRT: D=678.39 (P<0.001)), we purposively decided to report results from the simpler MNL model because (i) results from this model are robust to technical changes in optimization procedure (unlike MXL results which are known to be sensitive to changes in *starting values, type of draws, number of draws and choice of the statistical distribution for the random effects); and (ii) a comparison of estimates between the two models showed a high level of agreement. We fitted a linear regression line through the set of (MNL; MXL) coordinates and the coefficient of determination was close to 100%. More importantly, the intercept (which can be interpreted as a measure of bias associated with use of MNL estimates instead of MXL ones) was almost null (0.012) and non-significant (P=0.462). However, the slope (1.172), which can be interpreted as a measure of scale, was significantly different from 1 (P<0.001), indicating thus that MXL model measured the same preference effects but on a higher (more precise) utility scale. Given that the research objectives of our study were to quantify trade-offs between attributes, and more specifically the maximum acceptable decrease in the probability of achieving clear/almost clear skin at week 16, a change in utility scaling is irrelevant.*

Please see the response to question 4 for additional text included on model selection.

6b.

As a discussion subpoint of the previous, since the study includes an opt-out (old treatment), I wonder why the authors do not use an error component model to account for both preference heterogeneity and the fact that the choice is possibly nested (experimentally designed alternatives from one side and opt-out – or "old treatment" on the other side).

If the authors prefer to defend a model without preference heterogeneity, I would like to suggest the use of a nested logit model.

Author response: We initially decided not to run a nested logit (NL) model because the nesting structure appeared to be very limited in this study with only two options (A; B) in the "New treatment" nest and the opt-out option ("Old treatment") on the other side. We followed the reviewer's suggestion and ran a NL model. The inclusive value (IV) parameter, which captures the degree of correlation in unobserved factors over alternatives within the "New treatment" nest, was significant (P=0.003) and implied a weak-to-moderate correlation (1-0.63=0.37). The LR test indicated that NL model significantly outperformed the MNL model (D=8.09 (P=0.004)), but the comparison of estimated effects between the two models showed an almost complete agreement (R2>99%) and the intercept of the linear regression line was null. Thus, the decision to switch to the more complex model would have no impact on the results of this study.

Please see the response to question 4 for additional text included on model selection.

7.

The study creates a measure (the MAD, or maximum acceptable decrease in benefit) to explore tradeoffs. To generate this measure the author need to assume that the respondents would accept a decrease in benefit to compensate a decrease in risk. This sounds strange. Furthermore, a measure that is commonly used for this type of exploration already exists and would be simpler to interpret (the minimum acceptable benefit – MAB – needed for an increase a risk of side effect). Why did the author decided to invent this new measure (MAD) rather than use what is already commonly employed for similar explorations in the literature?

Author response: We acknowledge that minimum acceptable benefit is a more common term used by preference researchers, but note that in clinical discussion this concept is sometimes not intuitive. In *discussion with the clinical expert of the study team, it was agreed that it was more intuitive to discuss how much efficacy a patient would be willing to give up to avoid risks of adverse events (as patients typically want to avoid risks) or to gain changes in administration that they desire. No changes have been made to the manuscript.*

Minor suggestions:
In the abstract:

- The first sentence does not fully belong to the objective (in particular the second part of the first sentence is a motivation or a conclusion statement,)

Author response: Thank you for the suggestion. We have revised the objectives and conclusions accordingly:

> *"**Objectives:** ~~Understanding patients' preferences for AD therapies, including new targeted therapies, can aid shared decision-making between clinicians and patients and support health technology assessments~~ We aimed to quantify patients' preferences for efficacy, safety, and convenience features of atopic dermatitis (AD) treatments.*
>
> ***Conclusions:** Although patients with AD most valued treatment benefits and risks, they were willing to tolerate reduced efficacy to obtain a rapid onset, oral administration, less frequent monitoring, and a treatment that can be paused. Understanding patients' preferences for AD therapies, including new targeted therapies, can aid shared decision-making between clinicians and patients and support health technology assessments."*

In the discussion:

- The study reference number 29 was conducted with 320 respondents in the US (not the US and UK)

Author response: Thank you. We have revised this text in the discussion accordingly.

> *"Similar to our study, a DCE in the US ~~and UK~~ including 320 adults with moderate-to-severe AD found that patients preferred an oral pill over subcutaneous injection and valued a rapid onset of action and increasing the chance of achieving clear or almost clear skin at week 16."*

Reviewer: 4

Dr. Lavanya Diwakar, University of Birmingham Comments to the Author:

Thank you for asking me to review this interesting paper. I feel that there should be more efforts in general to elicit patient / end user preferences for treatment and this study is a welcome addition to the literature in AD treatment.

Author response: We thank the reviewer for their appreciation of our work.

I have a few comments regarding the study:

- There should be more clarity regarding the pilot study carried out for this project: Who was interviewed? were they patients or volunteers? how were they chosen? was the questionnaire changed after this study? if so, how?

Author response: Please see the response to question 1 from reviewer 3 where we have noted the additional text added on the cognitive pilot interviews.

- There should similarly more detail given regarding the survey- how many patients attempted the survey? what was the response rate?

Author response: We have added additional data on this to the participants section of the results. Given recruitment for the quantitative online survey used patient panels and databases the response rate appears low, but this is in line with other studies using such recruitment strategies.

> *"The DCE survey included 404 participants (114 in France, 145 in Spain, and 145 in the UK) who were recruited between October and December 2019. Given that recruitment for the quantitative online survey used patient panels and databases, 157,553 initial invites were sent, with a 4% (n=6,287) response rate. The majority of the interested potential participants completed the screening questionnaire but were not eligible to participate, largely due to not having AD; 541 patients were eligible to participate, with 75% of those eligible completing the survey."*

- Were there any incomplete responses? How did you deal with missing data?

Author response: We have added additional information on this into the manuscript in the section 'Overall preferences for treatment attributes':

> *" The DCE dataset had no missing values, as patients could not proceed in the survey without answering each question or item. If participants did not complete the survey they were not remunerated or included in the dataset."*

- It appears that those who failed the quality checks were in fact included in the final analysis. This should be explained. This can compromise the results, in my opinion.

Author response: We have added additional text to the manuscript at the end of the validity assessments section:

> *"Participants were not excluded based on responses to the validity tests, following best practise recommendations[33], as the preferences of patients may be valid and removal may induce selection bias"*

- In the questionnaire, one of the options is current treatment. It is not clear which treatment the attributes for this are based on. Are the attribute values the same for this option in every question? This should be explained better please.

Author response: Please see the response to reviewer 3 question 2 where we have provided additional information on the opt-out.

- It is not clear how the attributes for the study were chosen. The reason to choose particular value levels for each attribute should also be explained.

Author response: Please see the response to reviewer 2, question 4 (question 2 in methods) where we have provided additional information on the targeted literature review and product label review and how this informed selection of attributes and levels.

- The systematic review (supplementary material) mentions 3 papers (1 quantitative and 2 qualitative) but I could not find these in the references

Author response: These references are included in the supplementary materials reference list rather than the main reference list.

- The referencing is either erroneous or incomplete. Please check.

Author response: We have reviewed and revised the referencing throughout, as needed

- Some of the results need further explanation, I think. Younger patients seem to value speed of onset differently from older patients, there is a difference in preferences between those with mild/moderate/ severe reaction; also the preference for injection treatment in those who are in fact receiving those treatments is also interesting.

Author response: We did not detect significant differences in speed of onset, by age, or any significant differences by disease severity (the confidence intervals overlap) and as such these results has not been mentioned. However, there was a difference in preferences for administration between age groups and we have now added this.

We have revised the wording in the subgroup analyses section, with underline indicating new text:

> *"**Subgroup analyses***
> *Results were similar for the three included countries (UK, Spain, and France) (Online Supplemental Figure 3), by age (Online Supplemental Figure 4), by gender (Online Supplemental Figure 5), by POEM overall score (Online Supplemental Figure 6), and by self-reported eczema severity (Online Supplemental Figure 7). However, those aged over 50 cared more about receiving an oral pill relative to those aged 40-50 years, for whom we did not detect a significant preference for administration.*
>
> *Participants who had experience self-injecting a treatment for any illness (32%) were more willing to accept a treatment that required a subcutaneous injection and placed less importance on reducing the risk of serious infections than those who did not have experience self-injecting a treatment for any illness (Online Supplemental Figure 8)."*

We've also added to the discussion on the influence of self-injectable experience:

> *Further, participants with self-injectable experience for any illness were more willing to use a subcutaneous treatment than participants without self-injectable experience. However, 28% participants were 'not willing' or 'somewhat not willing' to have a medication that required an injection for each dose.*

Since this question was not previously mentioned in the methods we have also now noted it in the methods.

> *"In addition to the DCE, participants completed a sociodemographic/clinical questionnaire, indicated their willingness (on a 5 point scale from not willing to very willing) to have a medication that required a subcutaneous injection for each dose and completed the Set of Brief Screening Questions to assess health literacy [27] and five of the seven items from the Numeracy Scale to assess numeracy [28] to assess their ability to understand the attributes and levels presented and their engagement in the survey."*

- The statistical method used (MNL Regression) is appropriate and very well explained

Author response: Thank you

New Online Supplementary Materials on Model Selection:

Combination of choice data from different countries

We estimated a heteroscedastic MNL (HMNL) model allowing for scale differences between countries. The likelihood ratio test (LRT) indicated that this model performed significantly better than the standard MNL model (D=11.45, P=0.003). We also estimated an extended version of the MNL model allowing for interaction effects between country of residence and the attributes' levels. This interacted MNL (IMNL) model also significantly outperformed the standard MNL model (D=66.44, P=0.001).

Using the scale estimates from the HMNL model, we applied a scale correction to the dataset and then re-estimated the IMNL model (RIMNL) to determine whether the interaction effects found to be significant in the initial IMNL model would remain significant after accounting for potential scale differences between countries. This was the case, indicating that differences in choice behaviours between countries could not be fully explained as the consequence of a change in underlying utility scale (**Online Supplemental Table 4**).

Accounting for unobserved heterogeneity in preferences

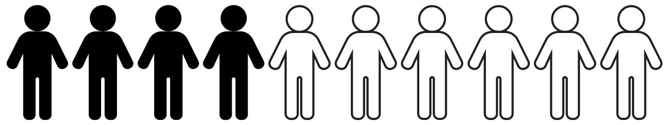
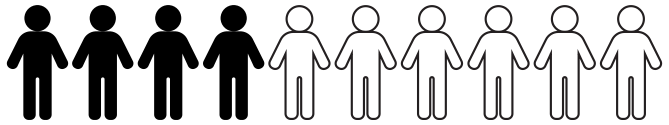
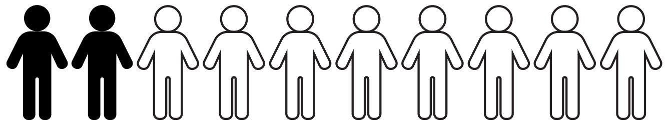
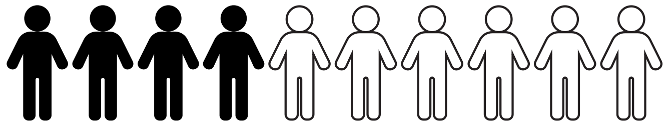
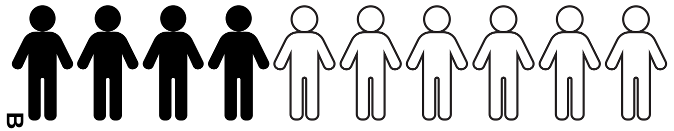
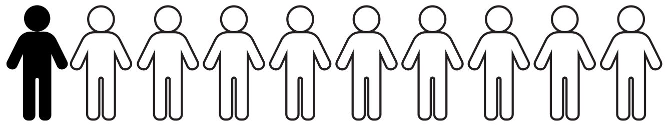
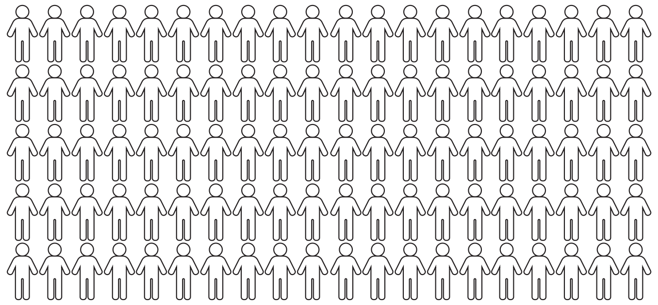
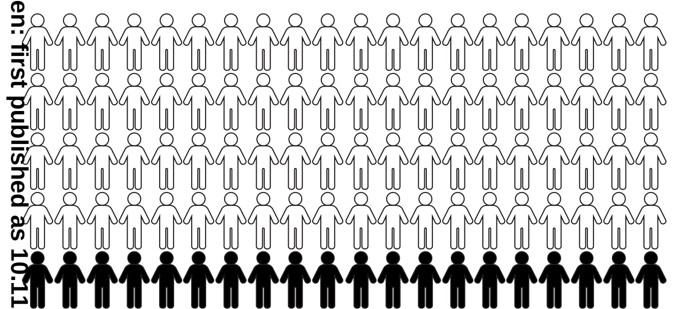
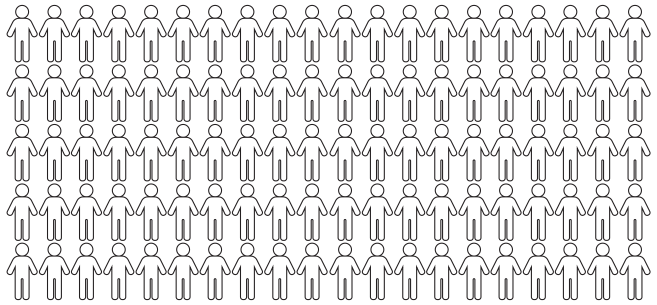
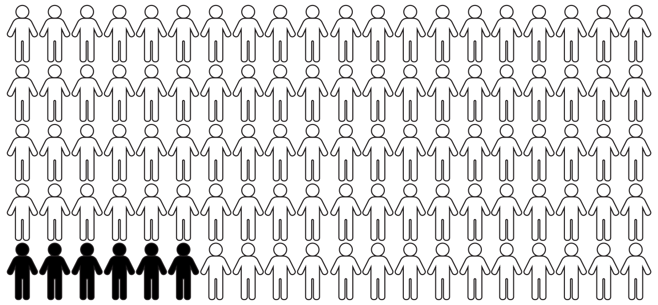
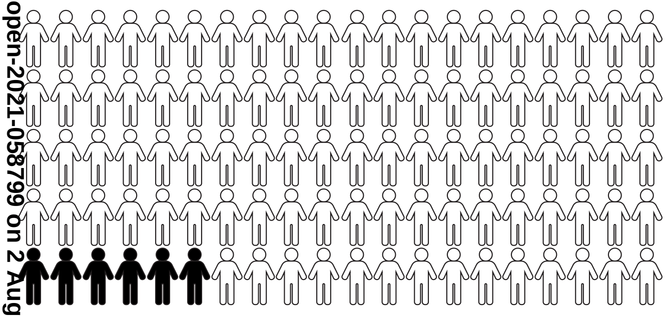
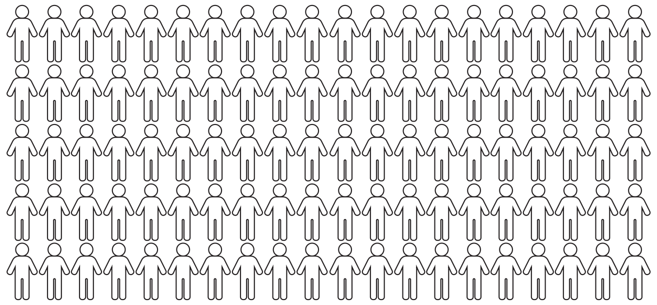
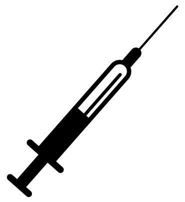
We estimated an MXL model allowing all parameters to be independently and normally distributed (i.e., diagonal covariance matrix of random effects). The MXL model significantly outperformed its MNL counterpart (LRT: D=678.39, P<0.001), but a comparison of estimates between the two models showed a high level of agreement (**Online Supplemental Figure 1**). We fitted a linear regression line through the set of coordinates (MNL; MXL) and the coefficient of determination was close to 100%. The intercept, which can be interpreted as a measure of bias associated with use of MNL estimates instead of MXL ones, was close to zero (0.012) and non-significant (P=0.462). However, the slope (1.172), which can be interpreted as a measure of scale, was significantly different from 1 (P<0.001), indicating that the MXL model measured the same preference effects but on a higher (more precise) utility scale. Given the research objectives of our study were to quantify trade-offs between attributes, and more specifically the MAD in the probability of achieving clear/almost clear skin at week 16, this change in utility scaling was deemed irrelevant.

Independence of treatment options relative to old treatment

A nested logit (NL) model was estimated to allow for a repartition of the choice options in two different nests: treatments A and B in a “New treatment” nest and the opt-out option in an “Old treatment” nest. The inclusive value (IV) parameter, which captures the degree of correlation in unobserved factors over alternatives within the "New treatment" nest, was significant (P=0.003) and implied a weak-to-moderate correlation (1-0.63=0.37). The LRT indicated that the NL model significantly outperformed the MNL model (D=8.09, P=0.004). However, a comparison of estimated effects between the two models showed a high level of agreement (r^2>99%) and the intercept of the linear regression line was null (**Online Supplemental Figure 2**).

REFERENCES

1. Food and Drug Administration. DUPIXENT® (dupilumab), https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/761055s016lbl.pdf (Revised: 2021, accessed March 29, 2022).
2. Thyssen JP, Buhl T, Fernández-Peñas P, et al. Baricitinib Rapidly Improves Skin Pain Resulting in Improved Quality of Life for Patients with Atopic Dermatitis: Analyses from BREEZE-AD1, 2, and 7. *Dermatol Ther (Heidelb)* 2021; 11: 1599-1611. 2021/07/19. DOI: 10.1007/s13555-021-00577-x.
3. Radi G, Simonetti O, Rizzetto G, et al. Baricitinib: The First Jak Inhibitor Approved in Europe for the Treatment of Moderate to Severe Atopic Dermatitis in Adult Patients. *Healthcare (Basel)* 2021; 9 2021/11/28. DOI: 10.3390/healthcare9111575.
4. Carlsson F, Mørkbak MR and Olsen SB. The first time is the hardest: A test of ordering effects in choice experiments. *J Choice Model* 2012; 5: 19-37. DOI: 10.1016/s1755-5345(13)70051-4.
5. Caussade S, Ortúzar JdD, Rizzi LI, et al. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological* 2005; 39: 621-640. DOI: https://doi.org/10.1016/j.trb.2004.07.006.
6. DeShazo JR and Fermo G. Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. *Journal of Environmental Economics and Management* 2002; 44: 123-143. DOI: https://doi.org/10.1006/jeem.2001.1199.

		[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	4 out of 10 (40%) 	4 out of 10 (40%) 	2 out of 10 (20%) 
	[illegible]	4 out of 10 (40%) 	4 out of 10 (40%) 	1 out of 10 (10%) 
[illegible]	[illegible]	0 out of 100 (0%) 	20 out of 100 (20%) 	0 out of 100 (0%) 
	[illegible]	6 out of 100 (6%) 	6 out of 100 (6%) 	0 out of 100 (0%) 
[illegible]	[illegible]	1 week	2 weeks	2 weeks
	[illegible]	✓	✗	✓
	[illegible]	Yes, with the possibility of pauses	Yes, without the possibility of pauses	Should not be used long-term
	[illegible]	Oral pill Once or twice daily	Injection under the skin  Every two weeks	Oral pill Once or twice daily
	[illegible]	Occasional check-ups required	No check-ups required	Frequent check-ups required
	[illegible]	☐	☐	☐
	[*if old treatment is chosen*] [illegible]	☐	☐	

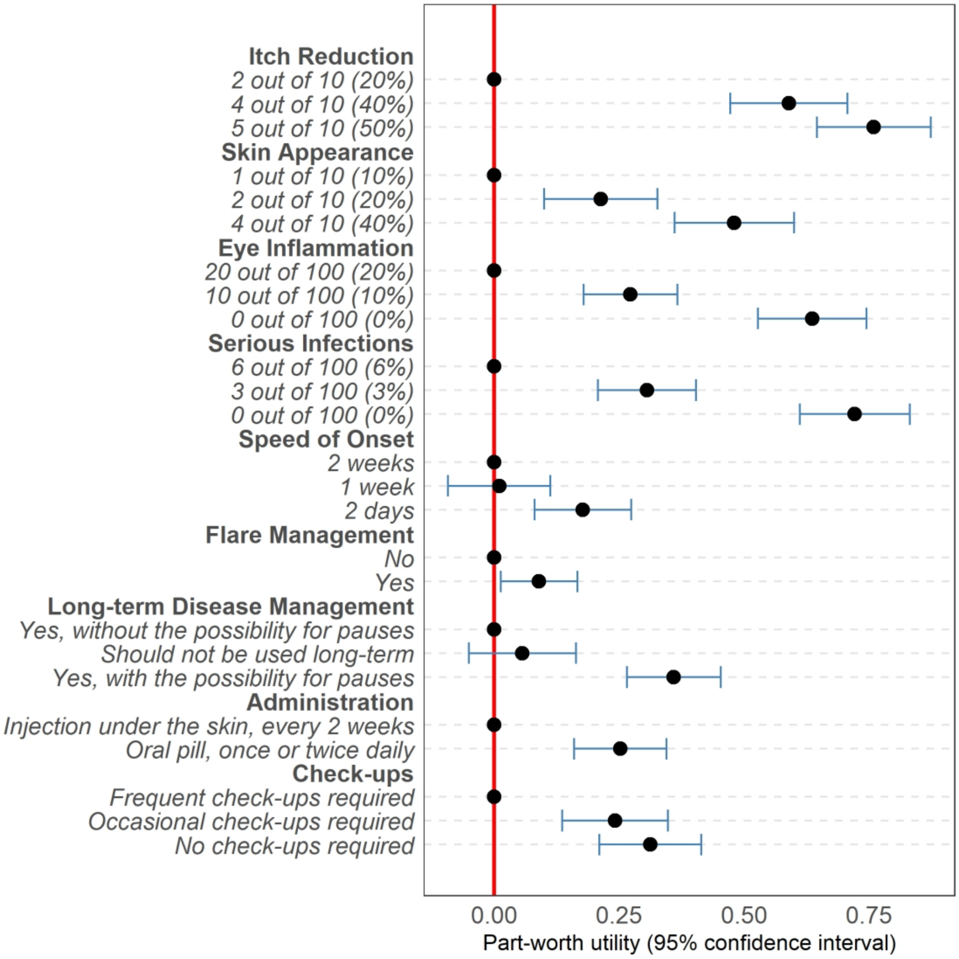


Figure 2. Multinomial logit results: part-worth utilities

177x177mm (300 x 300 DPI)

□□□□□□□□□□□□□□M□□□□□□M□□□□□D□□

□□□□□□□□□□□□□□□r□□□□□□□□□d□□□□□□

The aim of the literature review was to determine the key attributes and levels to be included in the DCE. This involved both a targeted literature review and a product label review.

The targeted literature review involved separate searches in major medical literature databases (Embase and MEDLINE) (□□□□□□□□□□□□ □□□□□□□□□□); a search for qualitative studies that considered the patient perspective on AD treatments; and a search for patient preference-specific studies, which considered AD treatments. Once key themes within the literature review were identified, the attributes were classified into corresponding categories. All abstracts were double screened, and disagreements were reviewed by a senior member of the research team.

The search strategy for the qualitative studies focused on studies that conducted interviews or focus groups, which mentioned AD or eczema, and their available treatments, as well as quality of life or patient preferences. The search excluded any non-adult studies, animal studies, clinical trials, and editorial notes. The search strategy for patient preference-specific studies sought studies that were explicitly patient preference in design, such as those utilising DCEs. Additionally, the studies had to mention AD or eczema.

The targeted literature search identified 33 potential studies. No duplicates were found, and all 33 were screened for eligibility. The abstracts were screened sequentially by two reviewers, and a third reviewer compared the rationale for inclusion and exclusion of studies to obtain the final list of full texts to screen. ~~Fifteen~~ Seven studies were excluded because they did not involve adult patients, ~~thirteen~~13

~~studies were excluded~~ because they weren’t about ~~s with~~ AD, ~~one~~ six because ~~it~~ they ~~did not mention outcomes of interest~~~~were~~did not have the study design of interest, ~~;~~ and ~~seven~~ four because ~~they were other study types not focusing on patient preferences~~no full text was available. ~~.~~

~~Of the seven remaining studies, four were excluded because a full text was not available.~~ The ~~final~~ remaining three studies included one quantitative[1] and two qualitative studies.[2, 3] In the quantitative study, the most important treatment attribute was the appearance of eczema (dryness/flakiness). In the two qualitative studies, itch reduction (symptom control), monitoring of symptoms, flexibility of treatment regimens to control flares, appearance (dryness/flakiness), and skin pain were identified themes.

Additionally, a product label search was conducted. Ten product labels for medications indicated for use in AD were reviewed in detail, including baricitinib (Olumiant®), dupilumab (Dupixent®), clobetasol propionate (Clobex®), tacrolimus (Protopic®), prednisone (Rayos®), cyclosporin (Neoral®), methotrexate, azathioprine (Imuran®), mycophenolate mofetil (CellCept®), and phototherapy. Itch reduction was most commonly reported as the percentage of patients achieving a meaningful (≥4-point reduction in the itch numerical rating scale) reduction in itch at week 16. Skin appearance was most commonly measured by the proportion of patients achieving clear or almost clear skin at week 16 (Investigator's Global Assessment scores of 0 or 1). The review of product labels also identified conjunctivitis as a differentiating and common side-effect of dupilumab that is not associated with other systemic therapies. Risk of serious infections were associated with other treatments, such as baricitinib and cyclosporine. The product label review also highlighted different modes and frequency of administration for systemic treatments, which included daily oral medication or subcutaneous administration every 2 weeks. Monitoring was also required for baricitinib and cyclosporine, but not for dupilumab.

Model specification

The analysis of all DCE responses followed random utility theory.[4-6] The model assumes that each respondent (*n*) chooses the alternative (*j*) in every DCE question (*t*) that results in the highest utility (a measure of desirability) of all available alternatives. Utility in a random utility model is defined as:

$$u(\boldsymbol{x}_{jnt}) = v(\boldsymbol{x}_{jnt}) + \varepsilon_{jnt}$$

Here the systematic utility component $v\left(\boldsymbol{x}_{jnt}\right)$ is a function of the DCE attributes and ε_{jnt} is a type 1 extreme value distributed random error. Two models are presented: a dummy-coded MNL model and an MNL model with skin appearance coded linearly, which is required to estimate the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. For the former, the utility function was defined as:

$$\begin{aligned} u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} + \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\ + \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3 \text{20\%\_skin\_appearance}_{jnt} \\ + \beta_4 \text{40\%\_skin\_appearance}_{jnt} + \beta_5 \text{10\%\_eye\_inflammation}_{jnt} \\ + \beta_6 \text{0\%\_eye\_inflammation}_{jnt} + \beta_7 \text{3\%\_serious\_infections}_{jnt} \\ + \beta_8 \text{0\%\_serious\_infections}_{jnt} + \beta_9 \text{1\_week\_onset}_{jnt} + \beta_{10} \text{2\_days\_onset}_{jnt} \\ + \beta_{11} \text{flare\_management}_{jnt} + \beta_{12} \text{long\_term\_no}_{jnt} \\ + \beta_{13} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{14} \text{oral\_admin}_{jnt} + \beta_{15} \text{no\_check\_ups}_{jnt} \\ + \beta_{16} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt} \end{aligned}$$

The constants $\alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}}$ controlled for potential bias to select the left option (Treatment A), and the Old Treatment, $\beta_1$to β_{16} were the estimated marginal utilities (i.e., estimated preference parameters), ε_{jnt} was an extreme value type I distributed error that allowed the function to be estimated in a logit model.[6] All attributes were dummy-coded. The reference level was the assumed worst-case option. Each of the estimated marginal utilities measured respondents' sensitivity to deviations from the reference level of the corresponding attribute. The sign (+ or –) of a marginal utility denotes whether patients valued this deviation positively or negatively. Only the initial choices (A vs. B vs. old ~~therapy~~treatment) were considered for the analysis of preferences. The initial and follow-up choices can be combined to allow for a more precise measurement of preferences. However, it is appropriate to combine these two types of choices only when they generate approximately the same information about participants' preferences. This condition was verified in two ways. Two MNL models were separately estimates for the initial (4,848 observations) and follow-up choices (1,126 observations), and then their preference estimates were compared. The Pearson correlation coefficient between the two sets of estimates was relatively low (0.32) as was the coefficient of determination for the linear regression (0.104), indicating poor agreement between the sets of estimates. A third MNL model was estimated on the combined initial and follow-up choices (5,974 observations), and its statistical performance was compared with the MNL model based on initial choices only. The adjusted McFadden pseudo-R^2 was lower for the model based on combined choices (7.3%) than for the initial model (8.3%), indicating that combining the initial and follow-up choices had a detrimental effect on the explanatory power of the model.

The linear coding of skin appearance was required to derive meaningful MAD measures. This measure was obtained by estimating the baseline utility function with skin appearance being coded as linear (i.e., one marginal utility is estimated instead of β_3 and β_4 for skin appearance). The utility function was defined as:

$$\begin{aligned}u_{jnt} = {} & \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} + \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\ & + \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3\text{\_skin\_appearance}_{jnt} \\ & + \beta_4 \text{10\%\_eye\_inflammation}_{jnt} + \beta_5 \text{0\%\_eye\_inflammation}_{jnt} \\ & + \beta_6 \text{3\%\_serious\_infections}_{jnt} + \beta_7 \text{0\%\_serious\_infections}_{jnt} \\ & + \beta_8 \text{1\_week\_onset}_{jnt} + \beta_9 \text{2\_days\_onset}_{jnt} + \beta_{10} \text{flare\_management}_{jnt} \\ & + \beta_{11} \text{long\_term\_no}_{jnt} + \beta_{12} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{13} \text{oral\_admin}_{jnt} \\ & + \beta_{14} \text{no}_{\text{check}_{\text{ups}}}{}_{jnt} + \beta_{15} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt}\end{aligned}$$

Each marginal utility was then divided by the marginal utility for skin appearance:

$$\text{MAD}_k = \frac{\widehat{\beta}_k}{\widehat{\beta}_3}$$

Such linear encoding is based on the underlying assumption of linearity in preferences, wherein a one-unit change in the attribute has a constant effect on respondents' choices and does not depend on the absolute value of the attribute level (e.g., a one-unit change from 15% to 16% increase in the chance of achieving clear or almost clear skin at week 16 has the same effect as the change from 20% to 21%). The validity of the assumption of linearity in preferences was tested by analysing the trend in risk estimates from the dummy-coded MNL model. Estimates were obtained for every attribute level in the dummy-coded MNL model (i.e., 3 levels for skin appearance). The linearity of skin appearance was tested by fitting a linear regression and evaluating its coefficient of determination. The assumption of linearity in skin appearance was accepted with a coefficient of 0.81, which exceeds the threshold of 0.7 to verify linearity.

[illegible]

We ~~have~~ estimated a~~n~~ heteroscedastic ~~H~~MNL (HMNL) model allowing for scale differences between countries. The likelihood ratio test (LRT) indicated that this model performed significantly better than the standard MNL model (D=11.45, ~~(~~P=0.003~~)~~). We also estimated an extended version of the MNL model allowing for interaction effects between ~~the~~ country of residence and the attributes' levels. ~~and~~

~~t~~This interacted MNL (IMNL) model also ~~appeared to~~ significantly outperformed the standard MNL model (D=66.44, ~~(~~P=0.001~~)~~).

Using the scale estimates from the HMNL model, we applied a scale correction to the dataset and then re-estimated the IMNL model (RIMNL) to determine whether the interaction effects ~~which were~~ found to be significant in the initial IMNL model would remain significant after ~~having~~ accountinged for potential scale differences between countries. This was the case, indicating ~~thus~~ that differences in choice behaviours between countries ~~cannot~~could not be fully explained as the consequence of a change in underlying utility scale (□ □□□□□□□□□□ □□□□□□□□□□□).

□ □□□□□□□□□□r □□□□□□□r □□d □□□□r □□□□□□□□□□□□r □□□r □□□□□□

We estimated an MXL model allowing all parameters to be independently and normally distributed (i.e., diagonal covariance matrix of random effects). The MXL model ~~appeared to~~ significantly outperformed its MNL counterpart (LRT: D=678.39, ~~(~~P<0.001~~)~~), but a comparison of estimates between the two models showed a high level of agreement (□ □□□□□□□□□□□ □□□□□□□□□r □□□). We fitted a linear regression line through the set of ~~(MNL; MXL)~~ coordinates (MNL; MXL) and the coefficient of determination was close to 100%. The intercept, ~~(~~which can be interpreted as a measure of bias associated with use of MNL estimates instead of MXL ones,~~)~~ was close to zero (0.012) and non-significant (P=0.462). However, the slope (1.172), which can be interpreted as a measure of scale, was significantly different from 1 (P<0.001), indicating ~~thus~~ that the MXL model measured the same preference effects but on a higher (more precise) utility scale. Given ~~that~~ the research objectives of our study were to quantify trade-offs between attributes, and more specifically the ~~maximum acceptable decrease~~MAD in the probability of achieving clear/almost clear skin at week 16, this change in utility scaling was deemed irrelevant.

□□d□□□□d□□□□□□□r□□□□ □□□□□□□□□r□□□□□□□□□□d□r□□□□ □□□□

A nested logit (NL) model was estimated to allow for a repartition of the choice options in two different nests: ~~Two options (~~treatments A and~~;~~ B~~)~~ in ~~the~~a “New treatment” nest and the opt-out option in an ~~(~~“Old treatment”~~) on the other side~~ nest. The inclusive value (IV) parameter, which captures the degree of correlation in unobserved factors over alternatives within the "New treatment" nest, was significant (P=0.003) and implied a weak-to-moderate correlation (1-0.63=0.37). The LRT~~ test~~ indicated that the NL model significantly outperformed the MNL model (D=8.09, ~~(~~P=0.004~~)~~)~~, but the~~. However, a comparison of estimated effects between the two models showed a high level of agreement (~~R~~r^2>99%) and the intercept of the linear regression line was null (□ □□□□□□□□□□ □□□□□□□□r□□□).

R□□□r□□□□□□

1. Hauber AB, Mohamed AF, Gonzalez JM, et al. Benefit-risk tradeoff preferences for chronic hand eczema treatments. *Journal of Dermatological Treatment* 2016;28:40-46.
2. Howells LM, Chalmers JR, Cowdell F, et al. ‘When it goes back to my normal I suppose’: a qualitative study using online focus groups to explore perceptions of ‘control’ among people with eczema and parents of children with eczema in the UK. *BMJ Open* 2017;7:e017731.
3. DiBenedetti D, Baranowski E. Assessing patient and physician experiences with severe chronic hand eczema. *J Am Acad Dermatol* 2015;72:AB77.
4. Marschak J. Binary choice constraints on random utility indicators Mathematical methods in the social sciences. Mathematical methods in the social sciences. Palo Alto, CA, USA: Stanford University Press 1959:312-29.
5. Manski CF. The structure of random utility models. *Theory and Decision* 1977;8:229-54.
6. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. Frontiers in econometrics. New York: Academic Press 1973:105-42.

[illegible]

[illegible]

No.	Query	Results	Date
#6	#1 AND (#2 AND #3 OR (#4 AND #5))	33	10-Sep-18
#5	((('qualitative research'/exp OR 'nursing methodology research'/exp OR ethnograph*:ti,ab OR lived) AND experience*:ti,ab OR narrative) AND analysis:ti,ab OR grounded) AND interview*:ti,ab OR themes:ab,ti	80104	10-Sep-18
#4	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1743076	10-Sep-18
#3	'quantitative study'/exp OR 'discrete choice' OR 'dce':ab,ti OR 'discrete choice experiment*':ab,ti OR 'choice experiment*':ab,ti OR 'conjoint':ab,ti OR 'conjoint analysis':ab,ti OR 'bws':ab,ti OR 'benefit risk':ab,ti OR 'thresholding':ab,ti OR 'multiple criteria decision analysis':ab,ti OR 'benefit-risk':ab,ti OR 'tradeoff':ab,ti OR 'best-worst scaling':ab,ti OR 'ahp':ab,ti OR 'analytic hierarchy':ab,ti OR 'swing weighting':ab,ti OR 'threshold technique':ab,ti OR 'risk benefit analysis':ab,ti	68917	10-Sep-18
#2	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1370306	10-Sep-18
#1	'eczema'/exp OR 'atopic dermatitis'/exp	61560	10-Sep-18

d

		r		
Choice stability, n (%)				
Passed the test	260 (64)	71 (62)	94 (65)	95 (66)
Failed the test	144 (36)	43 (38)	51 (35)	50 (34)
Choice dominance [a], n (%)				
Passed the test	359 (89)	109 (96)	130 (90)	120 (83)
Failed the test	45 (11)	5 (4)	15 (10)	25 (17)
Serial non-participation [b], n (%)				
Never select the same option	384 (95)	108 (95)	136 (94)	140 (97)
Always select treatment A	0 (0)	0 (0)	0 (0)	0 (0)
Always select treatment B	1 (0)	1 (1)	0 (0)	0 (0)
Always select old ~~therapy~~treatment	19 (5)	5 (4)	9 (6)	5 (3)
Dominated decision making [c], n (%)				
Itch reduction	6 (1)	1 (1)	2 (1)	3 (2)
Skin appearance	1 (<1)	0 (0)	1 (1)	0 (0)
Eye inflammation	3 (1)	1 (1)	1 (1)	1 (1)
Serious infections	8 (2)	3 (3)	3 (2)	2 (1)
Speed of onset	0 (0)	0 (0)	0 (0)	0 (0)
Flare management	1 (<1)	0 (0)	0 (0)	1 (1)
Long-term disease management	0 (0)	0 (0)	0 (0)	0 (0)
Administration	21 (5)	8 (7)	8 (6)	5 (3)
Check-ups	2 (<1)	1 (1)	0 (0)	1 (1)
None	362 (90)	100 (88)	130 (90)	132 (91)
Response time for DCE choice task section only [d], n (%)				
Adequate	391 (97)	111 (97)	143 (99)	137 (95)
Inadequate	13 (3)	3 (3)	2 (1)	8 (5)
Time to complete DCE choice task section only, n (%)				
<5 min	236 (58)	64 (56)	93 (64)	79 (54)
5-10 min	123 (30)	38 (33)	38 (26)	47 (32)
10-15 min	28 (7)	7 (6)	9 (6)	12 (8)
15-20 min	4 (1)	1 (1)	2 (1)	1 (1)
>20 min	13 (3)	4 (4)	3 (2)	6 (4)

Abbreviations: DCE, discrete choice experiment

[a] A respondent was considered to have failed the test if they chose the inferior (dominated) option as their preferred treatment.

[b] A respondent was classified as a serial non-participant if they choose the same option for all 12 experimental choice tasks.

[c] Decision making was considered dominated when the respondent choses the best option on one attribute in all 12 experimental tasks.

[d] Response times in the lower 10% of the distribution were classed as too fast, and those in the upper 10% of the distribution as too slow. A participant was considered to have had an adequate response time if <80% of choice tasks were answered too fast or too slow.

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
Alternative specific constant	Old ~~therapy~~treatment	1.46 (0.12)***	[1.23; 1.69]
	Option A	-0.04 (0.04)	[-0.11; 0.03]
Itch reduction	2 out of 10 (20%)	Reference	-
	4 out of 10 (40%)	0.59 (0.06)***	[0.47; 0.71]
	5 out of 10 (50%)	0.76 (0.06)***	[0.65; 0.87]
Skin appearance	1 out of 10 (10%)	Reference	-
	2 out of 10 (20%)	0.21 (0.06)***	[0.10; 0.33]
	4 out of 10 (40%)	0.48 (0.06)***	[0.36; 0.60]
Eye inflammation	20 out of 100 (20%)	Reference	-
	10 out of 100 (10%)	0.27 (0.05)***	[0.18; 0.37]
	0 out of 100 (0%)	0.64 (0.06)***	[0.53; 0.75]
Serious infections	6 out of 100 (6%)	Reference	-
	3 out of 100 (3%)	0.31 (0.05)***	[0.21; 0.40]
	0 out of 100 (0%)	0.72 (0.06)***	[0.61; 0.83]
Speed of onset	2 weeks	Reference	-
	1 week	0.01 (0.05)	[-0.09; 0.11]
	2 days	0.18 (0.05)***	[0.08; 0.27]
Flare management	No	Reference	-
	Yes	0.09 (0.04)*	[0.01; 0.17]
Long-term disease management	Yes, without the possibility for pauses	Reference	-
	Should not be used long-term	0.06 (0.05)	[-0.05; 0.16]
	Yes, with the possibility for pauses	0.36 (0.05)***	[0.27; 0.45]
Administration	Injection under the skin, every 2 weeks	Reference	-
	Oral pill, once or twice daily	0.25 (0.05)***	[0.16; 0.35]
Check-ups	Frequent check-ups required	Reference	-
	Occasional check-ups required	0.24 (0.05)***	[0.14; 0.35]
	No check-ups required	0.31 (0.05)***	[0.21; 0.41]
Number of observations		4848	

Model log-likelihood at convergence	-4867
Adjusted pseudo R^2	0.08
Bayesian information criterion	9887

Abbreviations: CI, confidence interval; MLE, maximum likelihood estimate; SE, standard error

[illegible]

		MLE (SE)			
Attributes and levels	Sample	MNL	HMNL	IMNL	RIMNL
1. Preferences					
Alternative Specific Constant					
Old ~~therapy~~treatment	Overall	1.458 (0.115)***	1.643 (0.139)***	1.392 (0.200)***	1.392 (0.182)***
Option A	Overall	-0.038 (0.037)	-0.042 (0.042)	-0.007 (0.061)	-0.007 (0.062)
Itch Reduction					
2 out of 10 (20%)	Overall	Reference	-	-	-
4 out of 10 (40%)	Overall	0.590 (0.060)***	0.671 (0.073)***	0.651 (0.101)***	0.651 (0.098)***
5 out of 10 (50%)	Overall	0.760 (0.058)***	0.858 (0.072)***	0.733 (0.100)***	0.733 (0.095)***
Skin Appearance					
2 out of 10 (20%)	Overall	0.214 (0.058)***	0.246 (0.066)***	0.243 (0.098)*	0.243 (0.096)*
4 out of 10 (40%)	Overall	0.481 (0.061)***	0.554 (0.072)***	0.606 (0.105)***	0.607 (0.100)***
1 out of 10 (10%)	Overall	Reference	-	-	-
Eye inflammation					
20 out of 100 (20%)	Overall	Reference	-	-	-
10 out of 100 (10%)	Overall	0.273 (0.048)***	0.317 (0.056)***	0.398 (0.080)***	0.398 (0.079)***
0 out of 100 (0%)	Overall	0.637 (0.056)***	0.723 (0.068)***	0.676 (0.092)***	0.677 (0.092)***

Serious Infections					
0 out of 100 (0%)	Overall	0.722 (0.056)***	0.800 (0.067)***	0.522 (0.093)***	0.523 (0.093)***
6 out of 100 (6%)	Overall	Reference	-	-	-
3 out of 100 (3%)	Overall	0.306 (0.050)***	0.339 (0.057)***	0.197 (0.083)*	0.197 (0.082)*
Speed of Onset					
2 weeks	Overall	Reference	-	-	-
1 week	Overall	0.010 (0.052)	0.011 (0.059)	0.019 (0.088)	0.019 (0.086)
2 days	Overall	0.178 (0.049)***	0.205 (0.057)***	0.217 (0.083)**	0.217 (0.082)**
Flare Management					
No	Overall	Reference	-	-	-
Yes	Overall	0.090 (0.039)*	0.109 (0.045)*	0.161 (0.065)*	0.161 (0.064)*
Long-term Disease Management					
Yes, without the possibility for pauses	Overall	Reference	-	-	-
Should not be used long-term	Overall	0.057 (0.054)	0.056 (0.062)	-0.012 (0.093)	-0.012 (0.091)
Yes, with the possibility for pauses	Overall	0.360 (0.048)***	0.399 (0.056)***	0.297 (0.080)***	0.297 (0.079)***
Administration					
Injection under the skin, every two weeks	Overall	Reference	-	-	-
Oral pill, once or twice daily	Overall	0.253 (0.047)***	0.294 (0.055)***	0.322 (0.078)***	0.322 (0.079)***
Check-ups					
Frequent check-ups required	Overall	Reference	-	-	-

Occasional check-ups required	Overall	0.242 (0.054)***	0.286 (0.063)***	0.328 (0.090)***	0.328 (0.091)***
No check-ups required	Overall	0.312 (0.052)***	0.366 (0.061)***	0.417 (0.086)***	0.417 (0.086)***
2. Interaction effects					
Alternative Specific Constant					
Old ~~therapy~~treatment	France	-	-	0.118 (0.311)	0.358 (0.257)
Old ~~therapy~~treatment	Spain	-	-	0.104 (0.336)	0.586 (0.298)*
Option A	France	-	-	-0.066 (0.094)	-0.077 (0.103)
Option A	Spain	-	-	-0.035 (0.089)	-0.048 (0.105)
Itch Reduction					
4 out of 10 (40%)	France	-	-	-0.150 (0.156)	-0.069 (0.154)
4 out of 10 (40%)	Spain	-	-	-0.057 (0.153)	0.134 (0.163)
5 out of 10 (50%)	France	-	-	0.066 (0.155)	0.194 (0.151)
5 out of 10 (50%)	Spain	-	-	0.024 (0.151)	0.268 (0.159)
Skin Appearance					
2 out of 10 (20%)	France	-	-	0.029 (0.149)	0.072 (0.155)
2 out of 10 (20%)	Spain	-	-	-0.099 (0.143)	-0.053 (0.156)
4 out of 10 (40%)	France	-	-	-0.200 (0.162)	-0.135 (0.157)
4 out of 10 (40%)	Spain	-	-	-0.194 (0.162)	-0.062 (0.165)
Eye inflammation					
10 out of 100 (10%)	France	-	-	-0.272 (0.121)*	-0.252 (0.132)

10 out of 100 (10%)	Spain	-	-	-0.127 (0.114)	-0.040 (0.133)
0 out of 100 (0%)	France	-	-	-0.086 (0.140)	0.007 (0.153)
0 out of 100 (0%)	Spain	-	-	-0.029 (0.132)	0.179 (0.154)
Serious Infections					
0 out of 100 (0%)	France	-	-	0.343 (0.142)*	0.480 (0.152)**
0 out of 100 (0%)	Spain	-	-	0.300 (0.136)*	0.564 (0.154)***
3 out of 100 (3%)	France	-	-	0.227 (0.127)	0.294 (0.134)*
3 out of 100 (3%)	Spain	-	-	0.131 (0.121)	0.238 (0.137)
Speed of Onset					
1 week	France	-	-	-0.064 (0.135)	-0.072 (0.143)
1 week	Spain	-	-	0.022 (0.129)	0.036 (0.142)
2 days	France	-	-	-0.043 (0.127)	-0.016 (0.136)
2 days	Spain	-	-	-0.080 (0.121)	-0.035 (0.137)
Flare Management					
Yes	France	-	-	-0.085 (0.098)	-0.073 (0.106)
Yes	Spain	-	-	-0.130 (0.093)	-0.120 (0.108)
Long-term Disease Management					
Should not be used long-term	France	-	-	0.033 (0.144)	0.036 (0.149)
Should not be used long-term	Spain	-	-	0.172 (0.136)	0.224 (0.153)

Yes, with the possibility for pauses	France	-	-	0.034 (0.123)	0.087 (0.129)
Yes, with the possibility for pauses	Spain	-	-	0.153 (0.121)	0.299 (0.135)*
Administration					
Oral pill, once or twice daily	France	-	-	-0.042 (0.119)	0.002 (0.130)
Oral pill, once or twice daily	Spain	-	-	-0.152 (0.111)	-0.098 (0.132)
Check-ups					
Occasional check-ups required	France	-	-	-0.010 (0.138)	0.042 (0.148)
Occasional check-ups required	Spain	-	-	-0.223 (0.132)	-0.189 (0.153)
No check-ups required	France	-	-	-0.043 (0.130)	0.017 (0.140)
No check-ups required	Spain	-	-	-0.249 (0.124)*	-0.195 (0.144)
Country of residence					
France	Overall	-	-0.148 (0.084)	-	-
Spain	Overall	-	-0.280 (0.084)***	-	-
UK	Overall	-	Reference	-	-
4. Model information					
Parameters	-	18	20	54	54
LL	-	-4866.9	-4861.2	-4833.7	-4833.7
AIC	-	9769.8	9762.4	9775.4	9775.4
BIC	-	9886.6	9892.2	10125.7	10125.7
APR	-	8.30%	8.40%	8.20%	8.20%

Abbreviations: AIC, Akaike information criterion; APR, Adjusted McFadden Pseudo R^2; BIC, Bayesian information criterion; HMNL, heteroskedastic multinomial logit; IMNL, interacted multinomial logit; LL, log-likelihood; MLE, maximum likelihood estimate; MNL, multinomial logit; RIMNL, re-estimated interacted multinomial logit; SE, standard error

Significance: *** P-value < 0.001, ** P-value < .01, * P-value < .05

□ □ □□□ □□□ □□□□□**M**□□ □□ □□**M**□□□ □**D**□□

□ □□□□□□□□□□□□□□□□□**r** □□□□□□□□**d** □□□□□□

The aim of the literature review was to determine the key attributes and levels to be included in the DCE. This involved both a targeted literature review and a product label review.

The targeted literature review involved separate searches in major medical literature databases (Embase and MEDLINE) (□ □□□□□□□□□□□□□□ □□□□□□□□□□□□); a search for qualitative studies that considered the patient perspective on AD treatments; and a search for patient preference-specific studies, which considered AD treatments. Once key themes within the literature review were identified, the attributes were classified into corresponding categories. All abstracts were double screened, and disagreements were reviewed by a senior member of the research team.

The search strategy for the qualitative studies focused on studies that conducted interviews or focus groups, which mentioned AD or eczema, and their available treatments, as well as quality of life or patient preferences. The search excluded any non-adult studies, animal studies, clinical trials, and editorial notes. The search strategy for patient preference-specific studies sought studies that were explicitly patient preference in design, such as those utilising DCEs. Additionally, the studies had to mention AD or eczema.

The targeted literature search identified 33 potential studies. No duplicates were found, and all 33 were screened for eligibility. The abstracts were screened sequentially by two reviewers, and a third reviewer compared the rationale for inclusion and exclusion of studies to obtain the final list of full texts to screen. Seven studies were excluded because they did not involve adult patients, 13 because they weren’t about AD, six because they did not have the study design of interest, and four because no full text was available. The remaining three studies included one quantitative[1] and two qualitative studies.[2, 3] In the quantitative study, the most important treatment attribute was the appearance of eczema (dryness/flakiness). In the two qualitative studies, itch reduction (symptom control), monitoring of symptoms, flexibility of treatment regimens to control flares, appearance (dryness/flakiness), and skin pain were identified themes.

Additionally, a product label search was conducted. Ten product labels for medications indicated for use in AD were reviewed in detail, including baricitinib (Olumiant®), dupilumab (Dupixent®), clobetasol propionate (Clobex®), tacrolimus (Protopic®), prednisone (Rayos®), cyclosporin (Neoral®), methotrexate, azathioprine (Imuran®), mycophenolate mofetil (CellCept®), and phototherapy. Itch reduction was most commonly reported as the percentage of patients achieving a meaningful (≥4-point reduction in the itch numerical rating scale) reduction in itch at week 16. Skin appearance was most commonly measured by the proportion of patients achieving clear or almost clear skin at week 16 (Investigator's Global Assessment scores of 0 or 1). The review of product labels also identified conjunctivitis as a differentiating and common side-effect of dupilumab that is not associated with other systemic therapies. Risk of serious infections were associated with other treatments, such as baricitinib and cyclosporine. The product label review also highlighted different modes and frequency of administration for systemic treatments, which included daily oral medication or subcutaneous administration every 2 weeks. Monitoring was also required for baricitinib and cyclosporine, but not for dupilumab.

Model specification

The analysis of all DCE responses followed random utility theory.[4-6] The model assumes that each respondent (*n*) chooses the alternative (*j*) in every DCE question (*t*) that results in the highest utility (a measure of desirability) of all available alternatives. Utility in a random utility model is defined as:

$$u(\boldsymbol{x}_{jnt}) = v(\boldsymbol{x}_{jnt}) + \varepsilon_{jnt}$$

Here the systematic utility component $v(\boldsymbol{x}_{jnt})$ is a function of the DCE attributes and ε_{jnt} is a type 1 extreme value distributed random error. Two models are presented: a dummy-coded MNL model and an MNL model with skin appearance coded linearly, which is required to estimate the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. For the former, the utility function was defined as:

$$\begin{aligned} u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} &+ \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\ &+ \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3 \text{20\%\_skin\_appearance}_{jnt} \\ &+ \beta_4 \text{40\%\_skin\_appearance}_{jnt} + \beta_5 \text{10\%\_eye\_inflammation}_{jnt} \\ &+ \beta_6 \text{0\%\_eye\_inflammation}_{jnt} + \beta_7 \text{3\%\_serious\_infections}_{jnt} \\ &+ \beta_8 \text{0\%\_serious\_infections}_{jnt} + \beta_9 \text{1\_week\_onset}_{jnt} + \beta_{10} \text{2\_days\_onset}_{jnt} \\ &+ \beta_{11} \text{flare\_management}_{jnt} + \beta_{12} \text{long\_term\_no}_{jnt} \\ &+ \beta_{13} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{14} \text{oral\_admin}_{jnt} + \beta_{15} \text{no\_check\_ups}_{jnt} \\ &+ \beta_{16} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt} \end{aligned}$$

The constants $\alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}}$ controlled for potential bias to select the left option (Treatment A), and the Old Treatment, β_1 to β_{16} were the estimated marginal utilities (i.e., estimated preference parameters), ε_{jnt} was an extreme value type I distributed error that allowed the function to be estimated in a logit model.[6] All attributes were dummy-coded. The reference level was the assumed worst-case option. Each of the estimated marginal utilities measured respondents' sensitivity to deviations from the reference level of the corresponding attribute. The sign (+ or –) of a marginal utility denotes whether patients valued this deviation positively or negatively. Only the initial choices (A vs. B vs. old treatment) were considered for the analysis of preferences. The initial and follow-up choices can be combined to allow for a more precise measurement of preferences. However, it is appropriate to combine these two types of choices only when they generate approximately the same information about participants' preferences. This condition was verified in two ways. Two MNL models were separately estimates for the initial (4,848 observations) and follow-up choices (1,126 observations), and then their preference estimates were compared. The Pearson correlation coefficient between the two sets of estimates was relatively low (0.32) as was the coefficient of determination for the linear regression (0.104), indicating poor agreement between the sets of estimates. A third MNL model was estimated on the combined initial and follow-up choices (5,974 observations), and its statistical performance was compared with the MNL model based on initial choices only. The adjusted McFadden pseudo-R^2 was lower for the model based on combined choices (7.3%) than for the initial model (8.3%), indicating that combining the initial and follow-up choices had a detrimental effect on the explanatory power of the model.

The linear coding of skin appearance was required to derive meaningful MAD measures. This measure was obtained by estimating the baseline utility function with skin appearance being coded as linear (i.e., one marginal utility is estimated instead of β_3 and β_4 for skin appearance). The utility function was defined as:

$$
\begin{aligned}
u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} + \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\
+ \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3\text{\_skin\_appearance}_{jnt} \\
+ \beta_4 \text{10\%\_eye\_inflammation}_{jnt} + \beta_5 \text{0\%\_eye\_inflammation}_{jnt} \\
+ \beta_6 \text{3\%\_serious\_infections}_{jnt} + \beta_7 \text{0\%\_serious\_infections}_{jnt} \\
+ \beta_8 \text{1\_week\_onset}_{jnt} + \beta_9 \text{2\_days\_onset}_{jnt} + \beta_{10} \text{flare\_management}_{jnt} \\
+ \beta_{11} \text{long\_term\_no}_{jnt} + \beta_{12} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{13} \text{oral\_admin}_{jnt} \\
+ \beta_{14} \text{no}_{\text{check}_{\text{ups}}}{}_{jnt} + \beta_{15} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt}
\end{aligned}
$$

Each marginal utility was then divided by the marginal utility for skin appearance:

$$\text{MAD}_k = \frac{\widehat{\beta}_k}{\widehat{\beta}_3}□$$

Such linear encoding is based on the underlying assumption of linearity in preferences, wherein a one-unit change in the attribute has a constant effect on respondents' choices and does not depend on the absolute value of the attribute level (e.g., a one-unit change from 15% to 16% increase in the chance of achieving clear or almost clear skin at week 16 has the same effect as the change from 20% to 21%). The validity of the assumption of linearity in preferences was tested by analysing the trend in risk estimates from the dummy-coded MNL model. Estimates were obtained for every attribute level in the dummy-coded MNL model (i.e., 3 levels for skin appearance). The linearity of skin appearance was tested by fitting a linear regression and evaluating its coefficient of determination. The assumption of linearity in skin appearance was accepted with a coefficient of 0.81, which exceeds the threshold of 0.7 to verify linearity.

□□□ □□□□□□□□□□□□□□□□d□□□□r□□ d□□□r□□□□□□□□r□□□□

We estimated a heteroscedastic MNL (HMNL) model allowing for scale differences between countries. The likelihood ratio test (LRT) indicated that this model performed significantly better than the standard MNL model (D=11.45, P=0.003). We also estimated an extended version of the MNL model allowing for interaction effects between country of residence and the attributes' levels. This interacted MNL (IMNL) model also significantly outperformed the standard MNL model (D=66.44, P=0.001).

Using the scale estimates from the HMNL model, we applied a scale correction to the dataset and then re-estimated the IMNL model (RIMNL) to determine whether the interaction effects found to be significant in the initial IMNL model would remain significant after accounting for potential scale differences between countries. This was the case, indicating that differences in choice behaviours between countries could not be fully explained as the consequence of a change in underlying utility scale ([illegible]).

[illegible]

We estimated an MXL model allowing all parameters to be independently and normally distributed (i.e., diagonal covariance matrix of random effects). The MXL model significantly outperformed its MNL counterpart (LRT: D=678.39, P<0.001), but a comparison of estimates between the two models showed a high level of agreement ([illegible]). We fitted a linear regression line through the set of coordinates (MNL; MXL) and the coefficient of determination was close to 100%. The intercept, which can be interpreted as a measure of bias associated with use of MNL estimates instead of MXL ones, was close to zero (0.012) and non-significant (P=0.462). However, the slope (1.172), which can be interpreted as a measure of scale, was significantly different from 1 (P<0.001), indicating that the MXL model measured the same preference effects but on a higher (more precise) utility scale. Given the research objectives of our study were to quantify trade-offs between attributes, and more specifically the MAD in the probability of achieving clear/almost clear skin at week 16, this change in utility scaling was deemed irrelevant.

[illegible]

A nested logit (NL) model was estimated to allow for a repartition of the choice options in two different nests: treatments A and B in a “New treatment” nest and the opt-out option in an “Old treatment” nest. The inclusive value (IV) parameter, which captures the degree of correlation in unobserved factors over alternatives within the "New treatment" nest, was significant (P=0.003) and implied a weak-to-moderate correlation (1-0.63=0.37). The LRT indicated that the NL model significantly outperformed the MNL model (D=8.09, P=0.004). However, a comparison of estimated effects between the two models showed a high level of agreement (r^2>99%) and the intercept of the linear regression line was null ([illegible]).

References

1. Hauber AB, Mohamed AF, Gonzalez JM, et al. Benefit-risk tradeoff preferences for chronic hand eczema treatments. *Journal of Dermatological Treatment* 2016;28:40-46.
2. Howells LM, Chalmers JR, Cowdell F, et al. ‘When it goes back to my normal I suppose’: a qualitative study using online focus groups to explore perceptions of ‘control’ among people with eczema and parents of children with eczema in the UK. *BMJ Open* 2017;7:e017731.
3. DiBenedetti D, Baranowski E. Assessing patient and physician experiences with severe chronic hand eczema. *J Am Acad Dermatol* 2015;72:AB77.
4. Marschak J. Binary choice constraints on random utility indicators Mathematical methods in the social sciences. Mathematical methods in the social sciences. Palo Alto, CA, USA: Stanford University Press 1959:312-29.
5. Manski CF. The structure of random utility models. *Theory and Decision* 1977;8:229-54.
6. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. Frontiers in econometrics. New York: Academic Press 1973:105-42.

□ □ □□□ □□□ □ □ □ □ □M□ □ □ □ □□□□ □ □ □ □ □

□ □□□□□□ □□□□□ □□□□□□□□□□□□□r□□□d□□□r□□r□r□□□□ □□□r□□□r□ □□

No.	Query	Results	Date
#6	#1 AND (#2 AND #3 OR (#4 AND #5))	33	10-Sep-18
#5	((('qualitative research'/exp OR 'nursing methodology research'/exp OR ethnograph*:ti,ab OR lived) AND experience*:ti,ab OR narrative) AND analysis:ti,ab OR grounded) AND interview*:ti,ab OR themes:ab,ti	80104	10-Sep-18
#4	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1743076	10-Sep-18
#3	'quantitative study'/exp OR 'discrete choice' OR 'dce':ab,ti OR 'discrete choice experiment*':ab,ti OR 'choice experiment*':ab,ti OR 'conjoint':ab,ti OR 'conjoint analysis':ab,ti OR 'bws':ab,ti OR 'benefit risk':ab,ti OR 'thresholding':ab,ti OR 'multiple criteria decision analysis':ab,ti OR 'benefit-risk':ab,ti OR 'tradeoff':ab,ti OR 'best-worst scaling':ab,ti OR 'ahp':ab,ti OR 'analytic hierarchy':ab,ti OR 'swing weighting':ab,ti OR 'threshold technique':ab,ti OR 'risk benefit analysis':ab,ti	68917	10-Sep-18
#2	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1370306	10-Sep-18
#1	'eczema'/exp OR 'atopic dermatitis'/exp	61560	10-Sep-18

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Choice stability, n (%)				
Passed the test	260 (64)	71 (62)	94 (65)	95 (66)
Failed the test	144 (36)	43 (38)	51 (35)	50 (34)
Choice dominance [a], n (%)				
Passed the test	359 (89)	109 (96)	130 (90)	120 (83)
Failed the test	45 (11)	5 (4)	15 (10)	25 (17)
Serial non-participation [b], n (%)				
Never select the same option	384 (95)	108 (95)	136 (94)	140 (97)
Always select treatment A	0 (0)	0 (0)	0 (0)	0 (0)
Always select treatment B	1 (0)	1 (1)	0 (0)	0 (0)
Always select old treatment	19 (5)	5 (4)	9 (6)	5 (3)
Dominated decision making [c], n (%)				
Itch reduction	6 (1)	1 (1)	2 (1)	3 (2)
Skin appearance	1 (<1)	0 (0)	1 (1)	0 (0)
Eye inflammation	3 (1)	1 (1)	1 (1)	1 (1)
Serious infections	8 (2)	3 (3)	3 (2)	2 (1)
Speed of onset	0 (0)	0 (0)	0 (0)	0 (0)
Flare management	1 (<1)	0 (0)	0 (0)	1 (1)
Long-term disease management	0 (0)	0 (0)	0 (0)	0 (0)
Administration	21 (5)	8 (7)	8 (6)	5 (3)
Check-ups	2 (<1)	1 (1)	0 (0)	1 (1)
None	362 (90)	100 (88)	130 (90)	132 (91)
Response time for DCE choice task section only [d], n (%)				
Adequate	391 (97)	111 (97)	143 (99)	137 (95)
Inadequate	13 (3)	3 (3)	2 (1)	8 (5)
Time to complete DCE choice task section only, n (%)				
<5 min	236 (58)	64 (56)	93 (64)	79 (54)
5-10 min	123 (30)	38 (33)	38 (26)	47 (32)
10-15 min	28 (7)	7 (6)	9 (6)	12 (8)
15-20 min	4 (1)	1 (1)	2 (1)	1 (1)
>20 min	13 (3)	4 (4)	3 (2)	6 (4)

Abbreviations: DCE, discrete choice experiment

[a] A respondent was considered to have failed the test if they chose the inferior (dominated) option as their preferred treatment.

[b] A respondent was classified as a serial non-participant if they choose the same option for all 12 experimental choice tasks.

[c] Decision making was considered dominated when the respondent choses the best option on one attribute in all 12 experimental tasks.

[d] Response times in the lower 10% of the distribution were classed as too fast, and those in the upper 10% of the distribution as too slow. A participant was considered to have had an adequate response time if <80% of choice tasks were answered too fast or too slow.

M r d

r		M	
Alternative specific constant	Old treatment	1.46 (0.12)***	[1.23; 1.69]
	Option A	-0.04 (0.04)	[-0.11; 0.03]
Itch reduction	2 out of 10 (20%)	Reference	-
	4 out of 10 (40%)	0.59 (0.06)***	[0.47; 0.71]
	5 out of 10 (50%)	0.76 (0.06)***	[0.65; 0.87]
Skin appearance	1 out of 10 (10%)	Reference	-
	2 out of 10 (20%)	0.21 (0.06)***	[0.10; 0.33]
	4 out of 10 (40%)	0.48 (0.06)***	[0.36; 0.60]
Eye inflammation	20 out of 100 (20%)	Reference	-
	10 out of 100 (10%)	0.27 (0.05)***	[0.18; 0.37]
	0 out of 100 (0%)	0.64 (0.06)***	[0.53; 0.75]
Serious infections	6 out of 100 (6%)	Reference	-
	3 out of 100 (3%)	0.31 (0.05)***	[0.21; 0.40]
	0 out of 100 (0%)	0.72 (0.06)***	[0.61; 0.83]
Speed of onset	2 weeks	Reference	-
	1 week	0.01 (0.05)	[-0.09; 0.11]
	2 days	0.18 (0.05)***	[0.08; 0.27]
Flare management	No	Reference	-
	Yes	0.09 (0.04)*	[0.01; 0.17]
Long-term disease management	Yes, without the possibility for pauses	Reference	-
	Should not be used long-term	0.06 (0.05)	[-0.05; 0.16]
	Yes, with the possibility for pauses	0.36 (0.05)***	[0.27; 0.45]
Administration	Injection under the skin, every 2 weeks	Reference	-
	Oral pill, once or twice daily	0.25 (0.05)***	[0.16; 0.35]
Check-ups	Frequent check-ups required	Reference	-
	Occasional check-ups required	0.24 (0.05)***	[0.14; 0.35]
	No check-ups required	0.31 (0.05)***	[0.21; 0.41]
Number of observations		4848	

Model log-likelihood at convergence	-4867
Adjusted pseudo R^2	0.08
Bayesian information criterion	9887

Abbreviations: CI, confidence interval; MLE, maximum likelihood estimate; SE, standard error

□ □□□□□□□□□□ □□□□□□□□□□□□ □□ □□r□□□□□r□□□□□□□□r□□□□ □d□□□

		MLE (SE)			
Attributes and levels	Sample	MNL	HMNL	IMNL	RIMNL
1. Preferences					
Alternative Specific Constant					
Old treatment	Overall	1.458 (0.115)***	1.643 (0.139)***	1.392 (0.200)***	1.392 (0.182)***
Option A	Overall	-0.038 (0.037)	-0.042 (0.042)	-0.007 (0.061)	-0.007 (0.062)
Itch Reduction					
2 out of 10 (20%)	Overall	Reference	-	-	-
4 out of 10 (40%)	Overall	0.590 (0.060)***	0.671 (0.073)***	0.651 (0.101)***	0.651 (0.098)***
5 out of 10 (50%)	Overall	0.760 (0.058)***	0.858 (0.072)***	0.733 (0.100)***	0.733 (0.095)***
Skin Appearance					
2 out of 10 (20%)	Overall	0.214 (0.058)***	0.246 (0.066)***	0.243 (0.098)*	0.243 (0.096)*
4 out of 10 (40%)	Overall	0.481 (0.061)***	0.554 (0.072)***	0.606 (0.105)***	0.607 (0.100)***
1 out of 10 (10%)	Overall	Reference	-	-	-
Eye inflammation					
20 out of 100 (20%)	Overall	Reference	-	-	-
10 out of 100 (10%)	Overall	0.273 (0.048)***	0.317 (0.056)***	0.398 (0.080)***	0.398 (0.079)***
0 out of 100 (0%)	Overall	0.637 (0.056)***	0.723 (0.068)***	0.676 (0.092)***	0.677 (0.092)***

Serious Infections					
0 out of 100 (0%)	Overall	0.722 (0.056)***	0.800 (0.067)***	0.522 (0.093)***	0.523 (0.093)***
6 out of 100 (6%)	Overall	Reference	-	-	-
3 out of 100 (3%)	Overall	0.306 (0.050)***	0.339 (0.057)***	0.197 (0.083)*	0.197 (0.082)*
Speed of Onset					
2 weeks	Overall	Reference	-	-	-
1 week	Overall	0.010 (0.052)	0.011 (0.059)	0.019 (0.088)	0.019 (0.086)
2 days	Overall	0.178 (0.049)***	0.205 (0.057)***	0.217 (0.083)**	0.217 (0.082)**
Flare Management					
No	Overall	Reference	-	-	-
Yes	Overall	0.090 (0.039)*	0.109 (0.045)*	0.161 (0.065)*	0.161 (0.064)*
Long-term Disease Management					
Yes, without the possibility for pauses	Overall	Reference	-	-	-
Should not be used long-term	Overall	0.057 (0.054)	0.056 (0.062)	-0.012 (0.093)	-0.012 (0.091)
Yes, with the possibility for pauses	Overall	0.360 (0.048)***	0.399 (0.056)***	0.297 (0.080)***	0.297 (0.079)***
Administration					
Injection under the skin, every two weeks	Overall	Reference	-	-	-
Oral pill, once or twice daily	Overall	0.253 (0.047)***	0.294 (0.055)***	0.322 (0.078)***	0.322 (0.079)***
Check-ups					
Frequent check-ups required	Overall	Reference	-	-	-

Occasional check-ups required	Overall	0.242 (0.054)***	0.286 (0.063)***	0.328 (0.090)***	0.328 (0.091)***
No check-ups required	Overall	0.312 (0.052)***	0.366 (0.061)***	0.417 (0.086)***	0.417 (0.086)***
2. Interaction effects					
Alternative Specific Constant					
Old treatment	France	-	-	0.118 (0.311)	0.358 (0.257)
Old treatment	Spain	-	-	0.104 (0.336)	0.586 (0.298)*
Option A	France	-	-	-0.066 (0.094)	-0.077 (0.103)
Option A	Spain	-	-	-0.035 (0.089)	-0.048 (0.105)
Itch Reduction					
4 out of 10 (40%)	France	-	-	-0.150 (0.156)	-0.069 (0.154)
4 out of 10 (40%)	Spain	-	-	-0.057 (0.153)	0.134 (0.163)
5 out of 10 (50%)	France	-	-	0.066 (0.155)	0.194 (0.151)
5 out of 10 (50%)	Spain	-	-	0.024 (0.151)	0.268 (0.159)
Skin Appearance					
2 out of 10 (20%)	France	-	-	0.029 (0.149)	0.072 (0.155)
2 out of 10 (20%)	Spain	-	-	-0.099 (0.143)	-0.053 (0.156)
4 out of 10 (40%)	France	-	-	-0.200 (0.162)	-0.135 (0.157)
4 out of 10 (40%)	Spain	-	-	-0.194 (0.162)	-0.062 (0.165)
Eye inflammation					
10 out of 100 (10%)	France	-	-	-0.272 (0.121)*	-0.252 (0.132)

10 out of 100 (10%)	Spain	-	-	-0.127 (0.114)	-0.040 (0.133)
0 out of 100 (0%)	France	-	-	-0.086 (0.140)	0.007 (0.153)
0 out of 100 (0%)	Spain	-	-	-0.029 (0.132)	0.179 (0.154)
Serious Infections					
0 out of 100 (0%)	France	-	-	0.343 (0.142)*	0.480 (0.152)**
0 out of 100 (0%)	Spain	-	-	0.300 (0.136)*	0.564 (0.154)***
3 out of 100 (3%)	France	-	-	0.227 (0.127)	0.294 (0.134)*
3 out of 100 (3%)	Spain	-	-	0.131 (0.121)	0.238 (0.137)
Speed of Onset					
1 week	France	-	-	-0.064 (0.135)	-0.072 (0.143)
1 week	Spain	-	-	0.022 (0.129)	0.036 (0.142)
2 days	France	-	-	-0.043 (0.127)	-0.016 (0.136)
2 days	Spain	-	-	-0.080 (0.121)	-0.035 (0.137)
Flare Management					
Yes	France	-	-	-0.085 (0.098)	-0.073 (0.106)
Yes	Spain	-	-	-0.130 (0.093)	-0.120 (0.108)
Long-term Disease Management					
Should not be used long-term	France	-	-	0.033 (0.144)	0.036 (0.149)
Should not be used long-term	Spain	-	-	0.172 (0.136)	0.224 (0.153)

Yes, with the possibility for pauses	France	-	-	0.034 (0.123)	0.087 (0.129)
Yes, with the possibility for pauses	Spain	-	-	0.153 (0.121)	0.299 (0.135)*
Administration					
Oral pill, once or twice daily	France	-	-	-0.042 (0.119)	0.002 (0.130)
Oral pill, once or twice daily	Spain	-	-	-0.152 (0.111)	-0.098 (0.132)
Check-ups					
Occasional check-ups required	France	-	-	-0.010 (0.138)	0.042 (0.148)
Occasional check-ups required	Spain	-	-	-0.223 (0.132)	-0.189 (0.153)
No check-ups required	France	-	-	-0.043 (0.130)	0.017 (0.140)
No check-ups required	Spain	-	-	-0.249 (0.124)*	-0.195 (0.144)
Country of residence					
France	Overall	-	-0.148 (0.084)	-	-
Spain	Overall	-	-0.280 (0.084)***	-	-
UK	Overall	-	Reference	-	-
4. Model information					
Parameters	-	18	20	54	54
LL	-	-4866.9	-4861.2	-4833.7	4833.7
AIC	-	9769.8	9762.4	9775.4	9775.4
BIC	-	9886.6	9892.2	10125.7	10125.7
APR	-	8.30%	8.40%	8.20%	8.20%

Abbreviations: AIC, Akaike information criterion; APR, Adjusted McFadden Pseudo R^2; BIC, Bayesian information criterion; HMNL, heteroskedastic multinomial logit; IMNL, interacted multinomial logit; LL, log-likelihood; MLE, maximum likelihood estimate; MNL, multinomial logit; RIMNL, re-estimated interacted multinomial logit; SE, standard error

Significance: *** P-value < 0.001, ** P-value < .01, * P-value < .05

□□□□□□□□□□□□□M□□□□□□□□□□R□□□□□□D□□

□□□□□□□□□□□□□□□□□□□□□r□□□□□□□□r□□□□□□□□□□□□□□□□□□□□□□□□M□□□□□d M□□□□d□□□

Abbreviation: MNL, multinomial logit; MXL, mixed logit

□□□□□□□□□□□□□□□□□□□□□r□□□□□□□□r□□□□□□□□□□□□□□□□□□□□□□□□□□d M□□□□d□□□

Abbreviations: MNL, multinomial logit; NL, nested logit□

□□□□□□□□□□□□□□□□□□□□□r□□□M□□r□□□□□□□□□□□r□

Abbreviation: CI, confidence interval

□□□□□□□□□□□□□□□□□□□□□r□□□M□□r□□□□□□□□□□□

Abbreviation: CI, confidence interval

□□□□□□□□□□□□□□□□□□□□□r□□□M□□r□□□□□□□□□□□□d□r

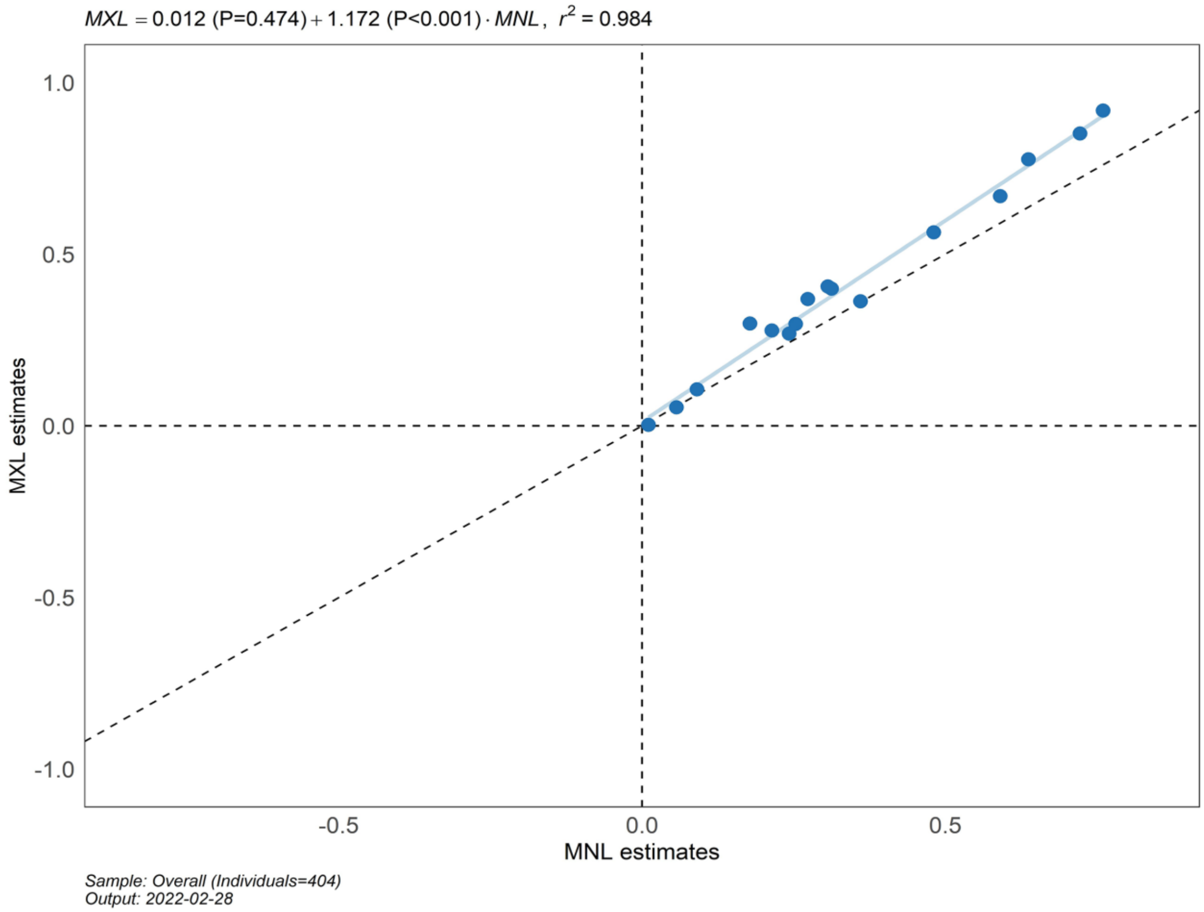
Abbreviation: CI, confidence interval

□□□□□□□□□□□□□□□□□□□□□r□□□M□□r□□□□□□□□□□□□□□□□r□□□□d□□□□□□M□□□□r□□□□□M□□□□r□□□□□□□r□□□□□r M□d□□–
□□M□d□r□□□□8–□□□□□□□r□□□□–□8□□
Abbreviation: CI, confidence interval

□□□□□□□□□□□□□□□□□□□□□r□□□M□□r□□□□□□□□□□□□r□□□r□d□□□□□□□□□□□r□□□M□d□□□r□□□□d□□d□□□□□□d□□
□□d□r□□□□□□□r□□□r□□□□□□r□□
Abbreviation: CI, confidence interval

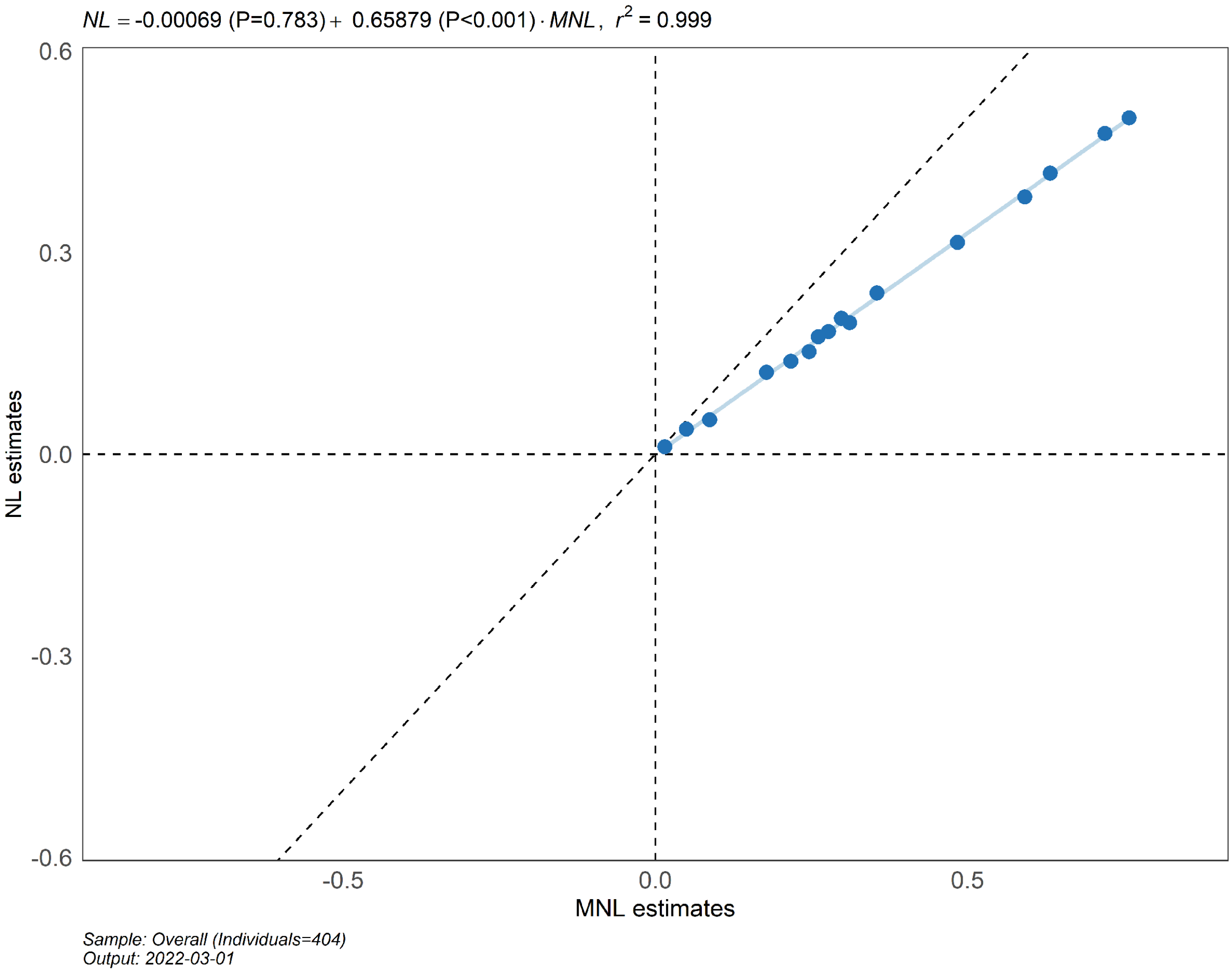
[illegible]

Abbreviation: CI, confidence interval



Supplemental Figure 1. Comparison of estimates between MXL and MNL models

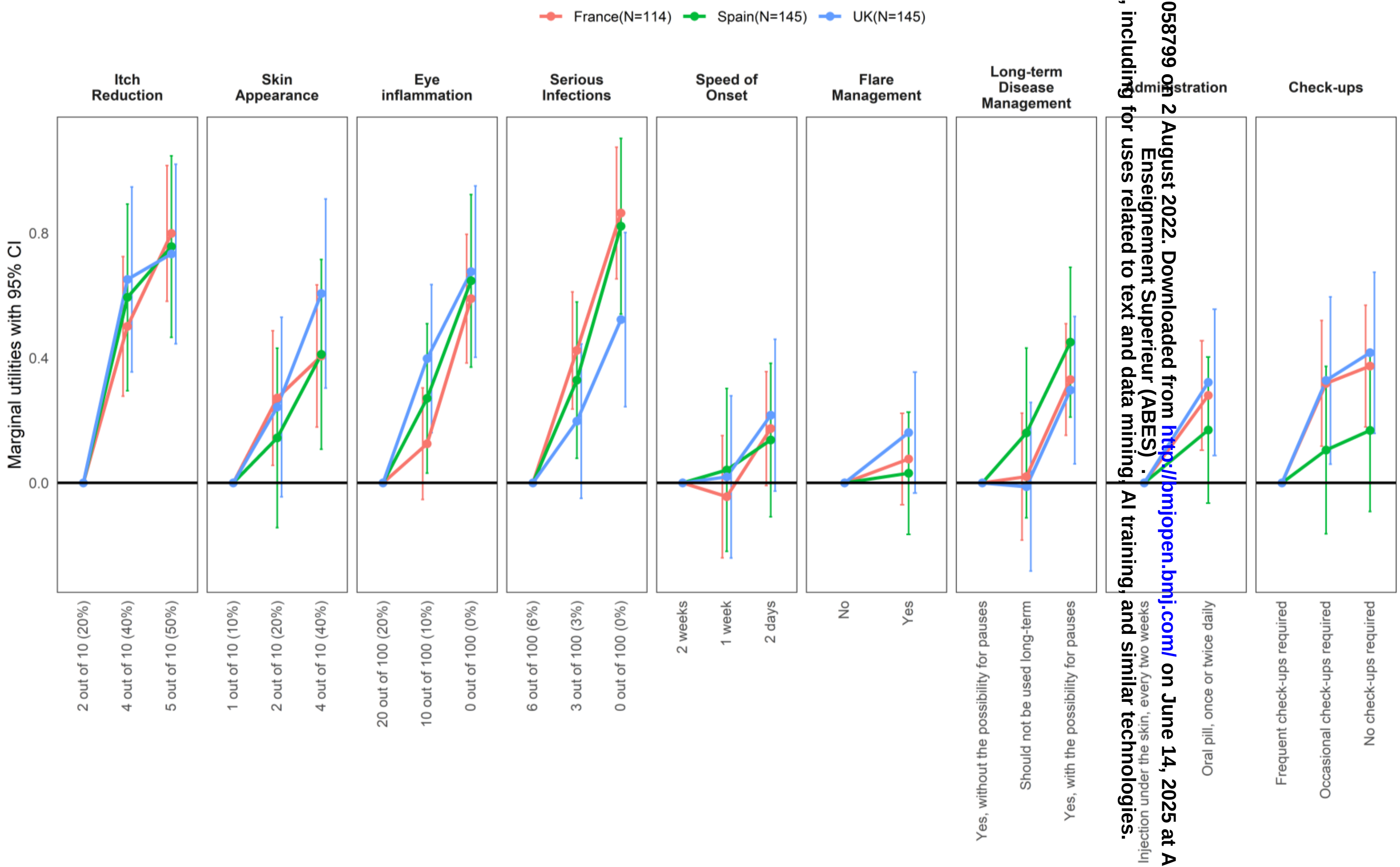
Abbreviation: MNL, multinomial logit; MXL, mixed logit



Supplemental Figure 2. Comparison of estimates between NL and MNL models

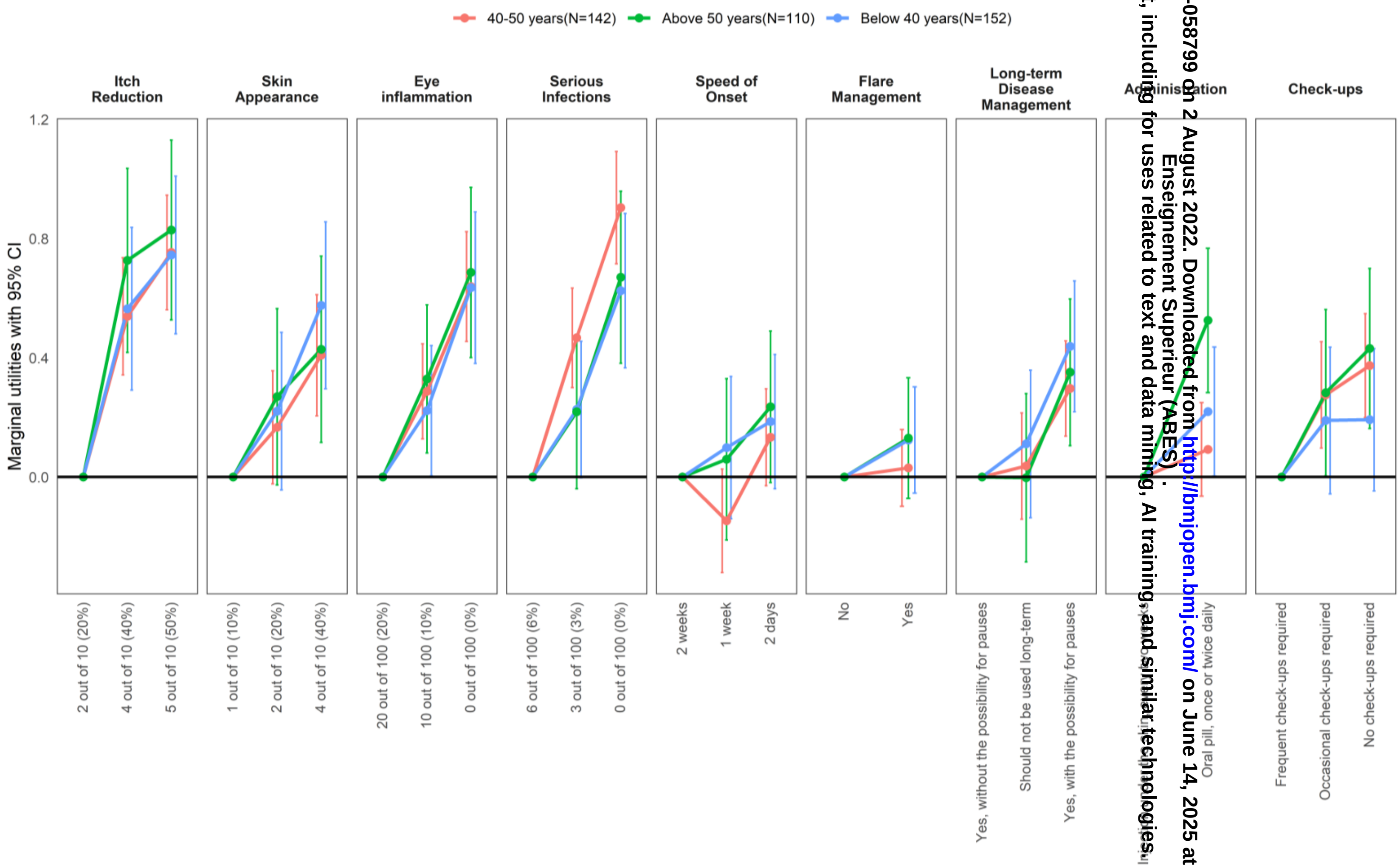
Abbreviations: MNL, multinomial logit; NL, nested logit

ew only



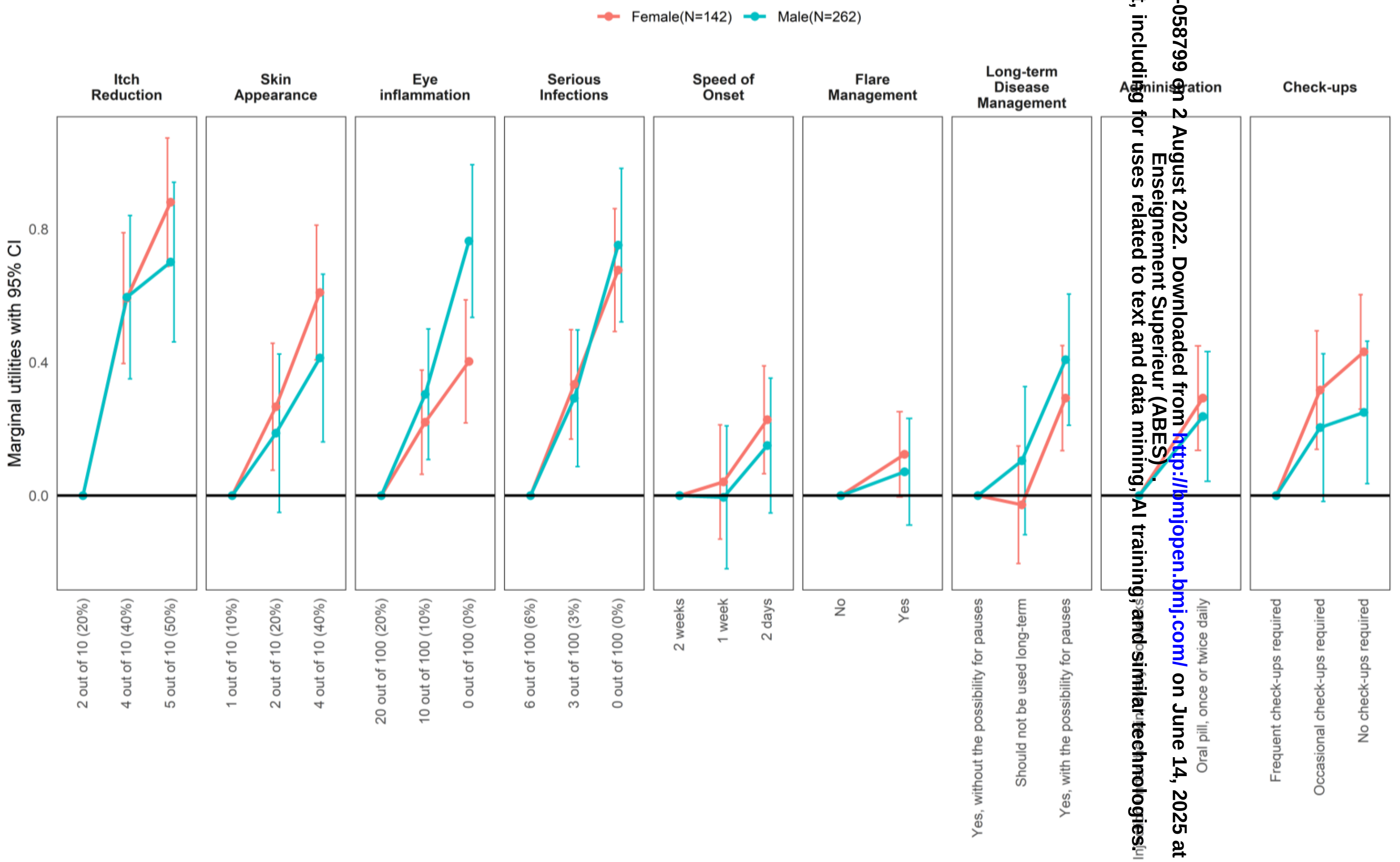
[illegible]

Abbreviation: CI, confidence interval



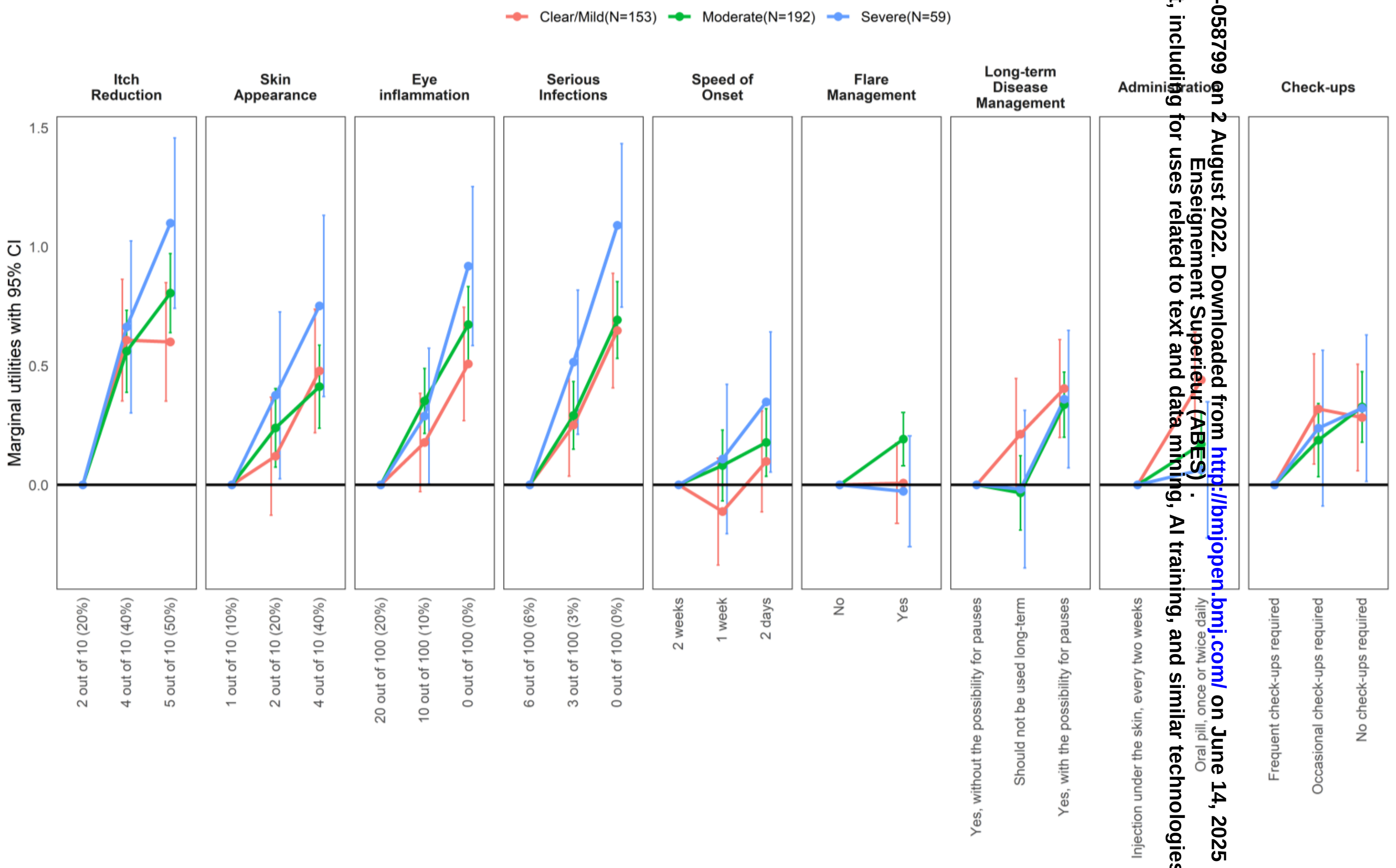
[illegible]

Abbreviation: CI, confidence interval



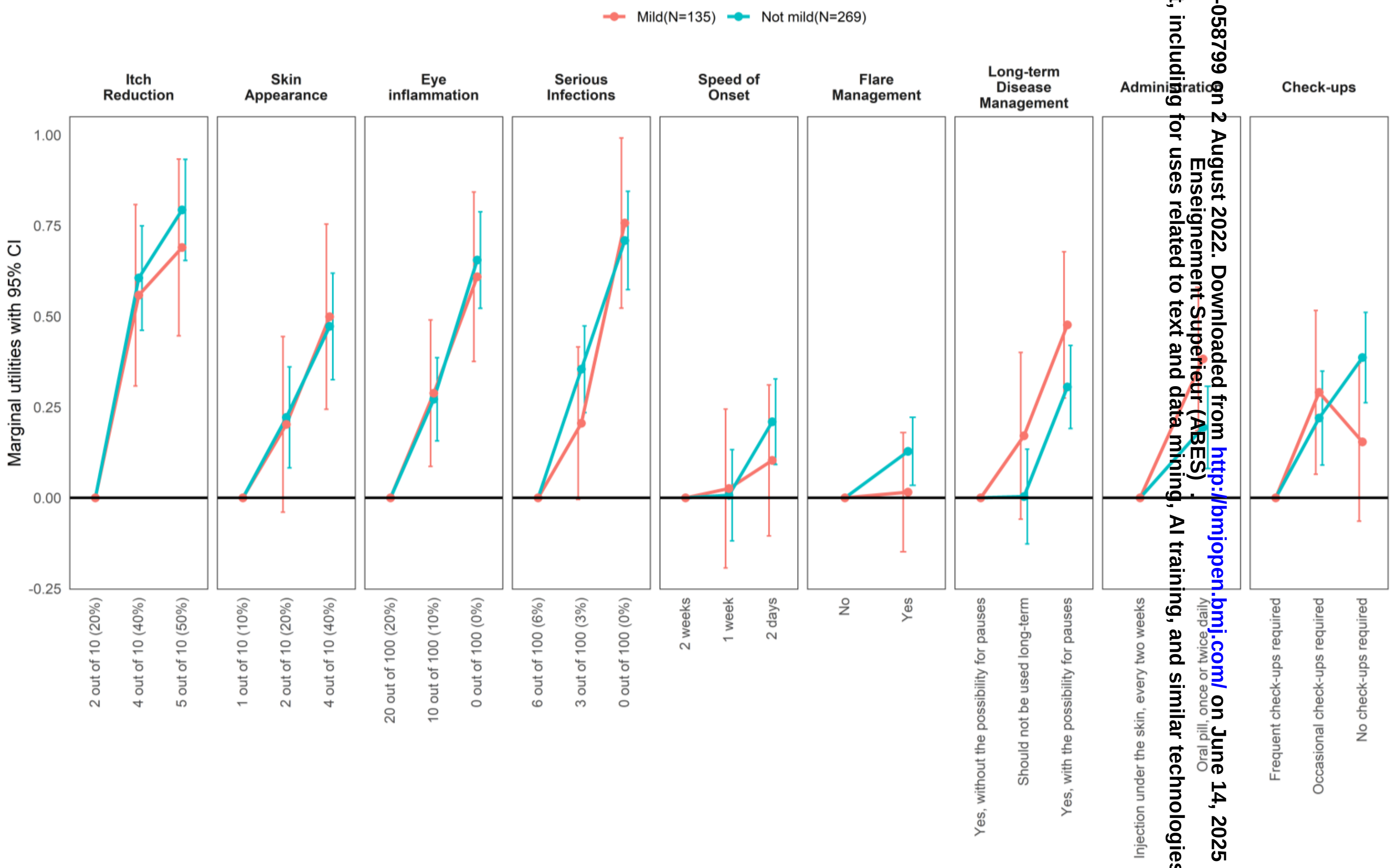
[illegible]

Abbreviation: CI, confidence interval



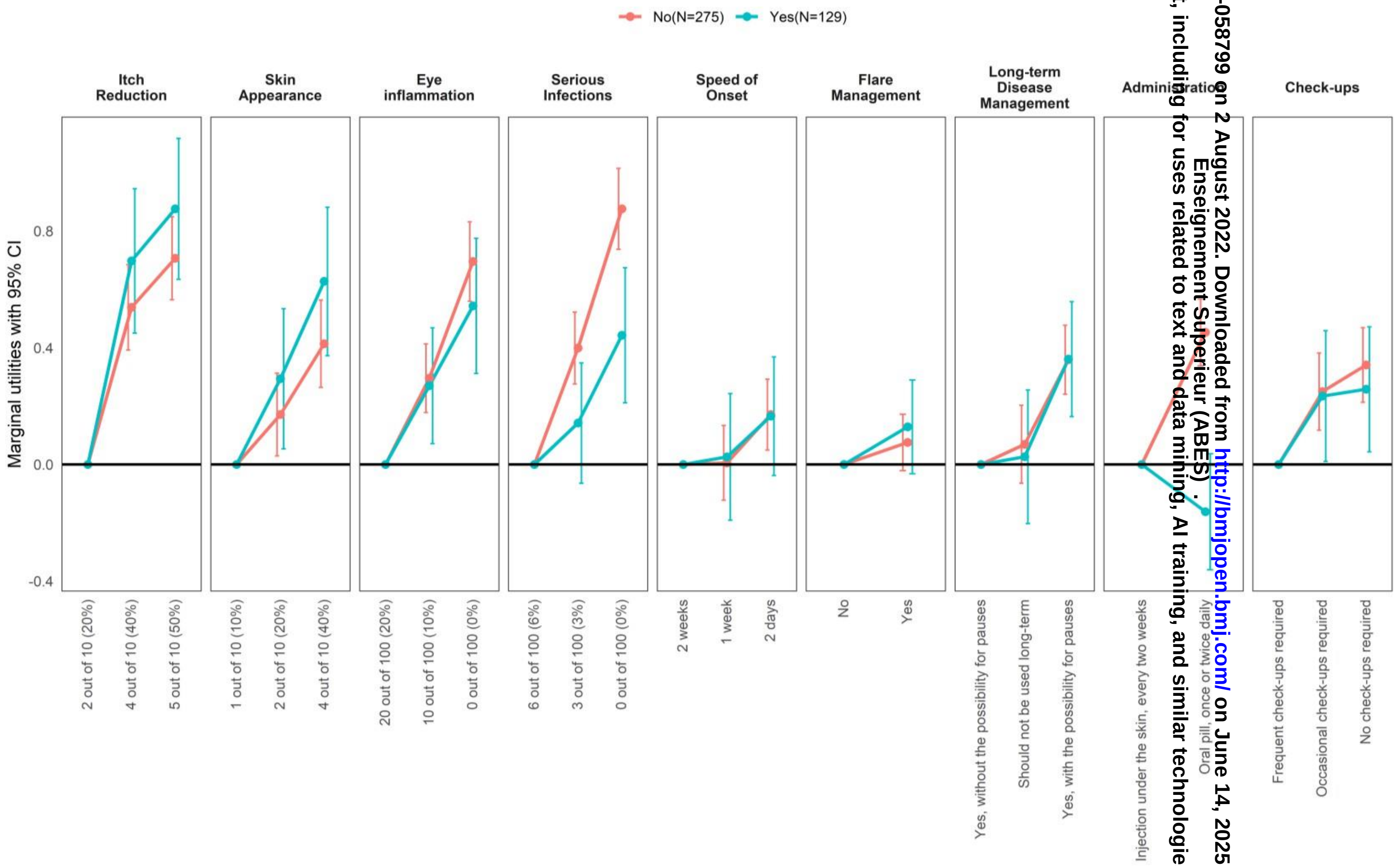
[illegible]

Abbreviation: CI, confidence interval



[illegible]

Abbreviation: CI, confidence interval



□□□□□□ □□□□□□□r□8□M□□r□□□□□□□□□r□□□□□□□□□□□□□

Abbreviation: CI, confidence interval

Checklist	Covered in manuscript	Page or section
1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?		
1.1 Were a well-defined research question and a testable hypothesis articulated?	Yes	p. 5
1.2 Was the study perspective described, and was the study placed in a particular decision-making or policy context?	Yes	p. 4-5
1.3 What is the rationale for using conjoint analysis to answer the research question?	Yes	p. 5
2. Was the choice of attributes and levels supported by evidence?		
2.1 Was attribute identification supported by evidence (literature reviews, focus groups, or other scientific methods)?	Yes (literature review)	p. 5
2.2 Was attribute selection justified and consistent with theory?	Yes	p. 5, 9-10
2.3 Was level selection for each attribute justified by the evidence and consistent with the study perspective and hypothesis?	Yes, via a literature review	p. 5
3. Was the construction of tasks appropriate?		
3.1 Was the number of attributes in each conjoint task justified (that is, full or partial profile)?	Yes, participants were surveyed for relevant attributes and no missing attributes were identified. Full choice profiles were used and patients had no issues with the number of attributes	p. 11
3.2 Was the number of profiles in each conjoint task justified?	Yes (3 profiles: A vs B vs old treatment)	p. 13

3.3 Was (should) an opt-out or a status-quo alternative (be) included?	Yes	p. 13
4. Was the choice of experimental design justified and evaluated?		
4.1 Was the choice of experimental design justified? Were alternative experimental designs considered?	Yes, D-efficient design assessed against good experimental design properties	p. 13
4.2 Were the properties of the experimental design evaluated?	Yes	p. 13
4.3 Was the number of conjoint tasks included in the data-collection instrument appropriate?	Yes, the number of tasks (questions) was 12 per person (36 in total)	p. 13
5. Were preferences elicited appropriately, given the research question?		
5.1 Was there sufficient motivation and explanation of conjoint tasks?	Yes	p. 13-14
5.2 Was an appropriate elicitation format (that is, rating, ranking, or choice) used? Did (should) the elicitation format allow for indifference?	Yes, the elicitation task was a choice task. The format did not allow indifference	p. 13-14
5.3 In addition to preference elicitation, did the conjoint tasks include other qualifying questions (for example, strength of preference, confidence in response, and other methods)?	Yes, validity assessments	p. 14
6. Was the data collection instrument designed appropriately?		
6.1 Was appropriate respondent information collected (such as sociodemographic, attitudinal, health history or status, and treatment experience)?	Yes	Table 2
6.2 Were the attributes and levels defined, and was any	Yes	Table 1

contextual information provided?		
6.3 Was the level of burden of the data-collection instrument appropriate? Were respondents encouraged and motivated?	Yes, this was assessed in cognitive pilot interviews and with data quality measures	p. 25
7. Was the data-collection plan appropriate?		
7.1 Was the sampling strategy justified (for example, sample size, stratification, and recruitment)?	Yes	p. 10
7.2 Was the mode of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)?	Yes	p. 5, 10
7.3 Were ethical considerations addressed (for example, recruitment, information and/or consent, compensation)?	Yes	p. 12
8. Were statistical analyses and model estimations appropriate?		
8.1 Were respondent characteristics examined and tested?	Yes	p. 17-19
8.2 Was the quality of the responses examined (for example, rationality, validity, reliability)?	Yes (validity and reliability)	p. 13-14, 19
8.3 Was model estimation conducted appropriately? Were issues of clustering and subgroups handled appropriately?	Yes	p. 16
9. Were the results and conclusions valid?		
9.1 Did study results reflect testable hypotheses and account for statistical uncertainty?	Yes, confidence intervals are presented	results
9.2 Were study conclusions supported by the evidence and	Yes	p. 23-26

compared with existing findings in the literature?		
9.3 Were study limitations and generalizability adequately discussed?	Yes	p. 25
10. Was the study presentation clear, concise, and complete?		
10.1 Was study importance and research context adequately motivated?	Yes	p. 4
10.2 Were the study data-collection instrument and methods described?	Yes	p. 13-16
10.3 Were the study implications clearly stated and understandable to a wide audience?	Yes	p. 23-26

BMJ Open

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-058799.R2
Article Type:	Original research
Date Submitted by the Author:	04-Jul-2022
Complete List of Authors:	Thomas, Caitlin; Evidera Raibouaa, Afaf; Eli Lilly and Company Wollenberg, Andreas ; University Hospital of Ludwig-Maximilian University of Munich, Department of Dermatology and Allergy Capron, Jean-Philippe; Eli Lilly and Company Krucien, Nicolas; Evidera Karn, Hayley; Evidera Tervonen, T; Evidera
Primary Subject Heading:	Dermatology
Secondary Subject Heading:	Patient-centred medicine, Health economics, Evidence based practice
Keywords:	DERMATOLOGY, Eczema < DERMATOLOGY, THERAPEUTICS, Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Patient preferences for atopic dermatitis medications in the United Kingdom, France, and Spain: a discrete choice experiment

Caitlin Thomas [1], Afaf Raibouaa [2], Andreas Wollenberg [3], Jean-Philippe Capron [2], Nicolas Krucien [1], Hayley Karn [1], Tommi Tervonen [1]

[1] Patient-Centered Research, Evidera, London, UK

[2] Eli Lilly & Co., Indianapolis, IN, USA

[3] Department of Dermatology and Allergy, University Hospital, Ludwig-Maximilian University of Munich, Germany

Corresponding author: Tommi Tervonen, Evidera, The Ark, 201 Talgarth Rd, London W6 8DL, UK. Phone: +44 (0)7788 237 412. Email: tommi.tervonen@evidera.com

4558 words

ABSTRACT

Objectives We aimed to quantify patient preferences for efficacy, safety, and convenience features of atopic dermatitis (AD) treatments.

Design and setting Online discrete choice experiment (DCE) survey.

Participants Adults in the UK, France, and Spain who had used AD treatments during the past 2 years.

Primary and secondary outcome measures Preferences for attributes were analysed using a multinomial logit model. Willingness to make trade-offs was expressed as the maximum acceptable decrease (MAD) in the probability of achieving clear/almost clear skin at week 16.

Results The survey was completed by 404 patients (44.1±12.0 years; 65% female; 64% moderate/severe eczema). Most patients (68%) had no prior experience of using self-injectable treatments for AD or any other illness. Participants most valued increasing the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%, followed by reducing the risks of serious infections from 6% to 0% and of eye inflammation from 20% to 0%. Participants were willing to accept a decrease in the possibility of achieving clear/almost clear skin to obtain a treatment that can be paused (MAD = 24.1%), requires occasional check-ups (MAD = 16.1%) or no check-ups (MAD = 20.9%) over frequent check-ups, is administered as a once- or twice-daily oral pill versus a subcutaneous injection every 2 weeks (MAD = 16.6%), has a 2-day over 2-week onset of action (MAD = 11.3%), and can be used for flare management (MAD = 5.8%).

Conclusions Although patients with AD most valued treatment benefits and risks, they were willing to tolerate reduced efficacy to obtain a rapid onset, oral

administration, less frequent monitoring, and a treatment that can be paused. Understanding patients’ preferences for AD therapies, including new targeted therapies, can aid shared decision-making between clinicians and patients and support health technology assessments.

Keywords: Dermatology, Eczema, Health Economics, Therapeutics

Strengths and limitations of this study:

- This study utilised a discrete choice experiment, which allowed us to quantitatively assess the trade-offs that patients with AD are willing to make between clinical and non-clinical treatment characteristics.
- Pilot testing and validity measures were performed to ensure that the target population could understand the survey and traded-off appropriately between the treatment attributes
- Study participants had predominantly self-reported moderate to severe AD (assessed with the Patient Oriented Eczema Measure), and these findings may not apply to the wider AD adult population, including those with mild or very severe AD.

INTRODUCTION

Atopic dermatitis (AD) is mostly treated using emollients and moisturizers, topical corticosteroids and calcineurin inhibitors, and, for severe cases, systemic immunosuppressants.[1, 2] However, emollients and moisturisers may not be sufficiently effective, and conventional systemic immunosuppressants have many potential side effects and are not generally recommended for long-term maintenance of AD.[3, 4] New targeted therapies for treating AD are now available. Dupilumab, a subcutaneously administered human monoclonal antibody inhibiting interleukin-4 and interleukin-13 signalling, was licensed in the US and the European Union in 2017 for the treatment of AD.[5] Baricitinib and upadacitinib, oral small-molecule inhibitors of Janus kinases, were recently licensed in the European Union for the treatment of moderate-to-severe AD in patients who are candidates for systemic therapy.[6, 7]

Several additional targeted therapies are in development, including a variety of monoclonal antibodies inhibiting interleukin signalling.[1, 2, 8] These new targeted therapies have different efficacy, risks, and non-clinical attributes, especially the mode of administration. In other chronic diseases, some patients prefer oral over parenteral treatment because they perceive some barriers to parenteral administration, which may lead to reduced adherence.[9-11] Because non-health benefits cannot be captured in traditional cost-effectiveness analysis, understanding to what extent they are valued by patients can help guide health technology assessment discussions[12-16] and inform shared decision-making at the point of care.[17]

Preferences for different treatment attributes, such as their benefits, risks, mode of administration, and convenience features, can be elicited from patients using discrete choice experiments (DCEs).[18] In DCEs, participants are presented with a series of tasks where they have to select between different hypothetical treatment options, each of which is composed of one level from each attribute in such a way that they are forced to make trade-offs, such as a higher risk of an adverse event but improved efficacy. DCEs have the advantage that the results can be used to quantify to what extent participants value each of the different attributes and estimate the trade-offs they would be willing to make. We hypothesized that patients with AD would not value all attributes relevant for their treatment choices equally. In the current study, we used a DCE to elicit the preferences of patients for key efficacy, safety, and convenience attributes of targeted AD therapies and examine the trade-offs they are willing to make between them.

MATERIALS AND METHODS

An online DCE survey was conducted between October and December 2019 in adults with AD living in the UK, France, or Spain. In the DCE survey, participants completed a series of choice tasks in which they selected between hypothetical treatment options described by a set of attributes with different levels. Treatment attributes and levels included in the DCE were identified through a targeted literature review of Embase and MEDLINE for quantitative and qualitative preference studies and a review of product labels for AD treatments (search conducted 10th September 2018; see **Online Supplemental Methods** and **Online Supplemental Table 1** for details). The attribute levels included in the DCE (e.g. likelihood of achieving clear or almost clear skin at week 16) were informed by clinical data from product labels for AD treatments (where available), including both baricitinib and dupilumab, reflecting the range of potential experiences that patients may have.[19, 20] Attributes included the following: chance of achieving clear or almost clear skin at week 16, chance of achieving a meaningful reduction in itch at week 16, risk of eye inflammation, risk of serious infections, administration, flare management, long-term disease management, monitoring, and speed of onset (**Table 1**). In order to reduce the cognitive burden of the survey, we grouped attributes as benefits, risks, and other. Prior research has found that grouping benefits and risks, and randomising the order of the groups and attributes within the groups, reduces the cognitive burden on participants, thereby reducing ordering effects and increasing choice certainty and the precision of preference estimates.[21]

Table 1. Treatment attributes and levels included in the main discrete choice experiment

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Itch reduction	Eczema (Atopic Dermatitis) causes your skin to itch. Treatments for Eczema (Atopic Dermatitis) increase the probability of achieving a meaningful reduction in itch severity.	**2 out of 10 (20%):** There is a 20% chance of achieving a meaningful reduction in itch severity (reference level) **4 out of 10 (40%):** There is a 40% chance of achieving a meaningful reduction in itch severity **5 out of 10 (50%):** There is a 50% chance of achieving a meaningful reduction in itch severity
Skin appearance	Eczema (Atopic Dermatitis) affects the way your skin looks due to flaking, redness, swelling, oozing, crusting, bleeding. Treatment for Eczema (Atopic Dermatitis) may improve your skin condition, but different treatments have different impacts. In this survey, we will ask you to consider the chance of achieving clear skin after 16 weeks starting the treatment.	**1 out of 10 (10%):** After taking treatment for 16 weeks, there is a 10% chance you will have clear/almost-clear skin (reference level) **2 out of 10 (20%):** After taking treatment for 16 weeks, there is a 20% chance you will have clear/almost-clear skin **4 out of 10 (40%):** After taking treatment for 16 weeks, there is a 40% chance you will have clear/almost-clear skin

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Eye inflammation	All treatments have some risk of negative side effects. Some treatments can cause minor eye infections. You may have swollen eyelids, feel sensitivity to light, feel itching or burning in your eyes, or have pink discoloration of the white in your eyes. This can be treated but may require interruption to treatment. Other treatments do not increase your risk of getting an eye inflammation.	**0 out of 100 (0%):** Your treatment does not increase the chance of an eye inflammation **10 out of 100 (10%):** There is a 10% chance of experiencing an eye inflammation **20 out of 100 (20%):** There is a 20% chance of experiencing an eye inflammation (reference level)
Serious infections	All treatments have some risk of negative side effects. Some treatments reduce your immune system's effectiveness at fighting off illness and can result in serious infections, such as pneumonia or blood poisoning, that may require treatment and hospitalisation; you may be hospitalised for around one week. There is always a very low risk of serious infection and this low risk may be increased.	**0 out of 100 (0%):** Your treatment does not increase the risk of serious infection **3 out of 100 (3%):** 3 out of 100 people will experience a serious infection **6 out of 100 (6%):** 6 out of 100 people will experience a serious infection (reference level)
Speed of onset	All medications for Eczema (Atopic Dermatitis) take some time to start working. Some medications will start to work in 2 days, but others can take 1 or 2 weeks.	**2 days:** Your medication will begin to work 2 days after starting the treatment **1 week:** Your medication will begin to work one week after starting the treatment **2 weeks:** Your medication will begin to work two weeks after starting the treatment (reference level)

Treatment attribute	**Description of the treatment attribute presented to participants**	**Levels**
Flare management	For some treatments, your doctor can increase your dose if your symptoms get worse (flare-ups). After the flare is controlled, reducing the dose again may also be an option. However, other treatments cannot be adjusted in this way and you will remain on a fixed dose, even if your symptoms change.	**Yes:** Your doctor can increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves **No:** Your doctor cannot increase or decrease your dose when your Eczema (Atopic Dermatitis) gets worse or improves (reference level)
Long-term disease management	Some treatments for Eczema (Atopic Dermatitis) need to be used continuously, without the option to stop and restart therapy when you want. Interruption of treatment, also known as a treatment holiday, can lead to a loss of efficacy over time. This means the therapy may not work as well when you restart treatment. These treatments must be used continuously and cannot be paused. Other treatments can be stopped and restarted (treatment holiday), with no impact on how effective the treatment is. Some treatments should not be used for the long-term, as they can have life threatening side effects, if used for a long period of time.	**Yes, with the possibility for pauses:** Treatment can be taken long term, and can be paused with no impact on how effective the treatment is **Yes, without the possibility for pauses:** Treatment can be taken long term, but must be taken continuously for there to be no impact on how effective the treatment is **Should not be used long-term:** You can pause the treatment, but using for the long-term may result in life threatening side effects (reference level)

Treatment attribute	Description of the treatment attribute presented to participants	Levels
Administration	Treatments are not all given/taken in the same way; for instance, some are pills, others are injections or topical creams. In this study we will only be considering pills and injections.	**Oral pill, once or twice daily** **Injection under the skin, every 2 weeks:** This is a subcutaneous injection, below the skin, but above muscle, usually injected into the thigh/stomach area. You can administer the injection yourself or a health care professional can administer it. If you choose to administer it yourself, you may need to be trained by a nurse on the injection technique. Treatment is once every two weeks. (reference level)
Check-ups	Some treatments require periodic blood tests taken by your doctor, because although you may not feel any symptoms, some Eczema (Atopic Dermatitis) medications can have a negative impact on your body.	**Frequent check-ups required:** Blood tests every 2 weeks during the initial 3 months of therapy and then monthly if the patient is stable (reference level). **Occasional check-ups required:** Blood tests at beginning of treatment, after 12 weeks, and then routinely, as determined by your doctor, while on treatment. **No check-ups required**

In each choice task, participants were asked to choose between different treatment options, each composed of one level from each of the attributes. Sensitivity of participants to changes in levels for each attribute were measured relative to the reference level, which is the level that patients least prefer. For example the reference level for risks is the highest level and for efficacy the reference level is the lowest level.

Cognitive pilot Interviews

To ensure the feasibility and robustness of the DCE, cognitive pilot interviews were conducted in the UK, France, and Spain (n=5 per country). The interviews involved a total of 15 patients, who were recruited using the same eligibility criteria as the main study. Patients were recruited through a number of routes, including HCP referrals, social media, and patient databases. The interviews examined whether the chosen attributes and levels were relevant, tradeable, and understandable to participants.[22] In addition, the cognitive pilot interviews assessed the complexity and clarity of the overall questionnaire. Each interview lasted approximately 60 min. Participants were provided a description of the study and completed the initial version of the study survey instrument online while sharing their screen with an interviewer and thinking aloud about the rationale behind their choices. While participants completed the DCE, interviewers probed them using a semi-structured discussion guide. At the end of the interview moderators assessed whether all attributes had been considered, and the overall relevance and plausibility of attributes and levels included in the survey; these assessments were interviewer observed and based on the patients' rationale behind decision making during the interview.

The cognitive pilot interviews were conducted in two waves, with roughly half the participants in each wave. Updates were made after wave 1 and the revised survey was subsequently tested in wave 2. The textual updates after wave 1 were largely minor wording updates to improve the understandability of the survey. However, the presentation of the task and the denominator of serious infections was updated to be consistent with the other risk attribute (eye inflammation). In wave 1, attributes were not initially grouped as benefits, risks, and other. The visualisation of the DCE was adjusted after wave 1 as some participants were misinterpreting the benefit/risk of a treatment. The updated survey grouped and labelled the attributes by category (benefits, risks, other). In wave 2, participants did not have problems understanding the benefits and risks of treatments and found it easier to consider a wider range of attributes. Patients were also asked if they thought any attributes were missing that they would want to know about when selecting a treatment. No missing attributes were identified.

The online DCE survey was initially tested in 29 to 30 participants per country. Minor updates were made to the visual presentation of the survey. Recruitment targets were to include an additional 115 participants in the UK, 115 in Spain, and 85 in France.

Ethics approval

The study was conducted according to good practice for stated preference research[16] and was approved by Ethical & Independent Review Services (Independence, MO, USA; study number 19100-01). In addition, the study was conducted in accordance with International Council on Harmonisation Guidelines for Good Clinical Practice, the ethical principles of the Declaration of Helsinki, the European Union General Data Protection Regulation, and all local laws and regulations.

Participants

Participants were recruited via recruiter databases, social media, patient associations, and online patient panels. Adults (≥18 years) living in the UK, France, or Spain with a self-reported diagnosis of AD for ≥ 12 months were eligible if they had received a topical or systemic therapy for AD in the past 2 years. Participants also had to be able to speak, read, and write the official language of the respective country. Potential participants were excluded if they had a diagnosis of psoriasis, acne, lupus erythematosus, skin cancer, or any other condition that could interfere with participation in and completion of the interview. To account for the possibility that preferences differ between participants with and without self-injectable experience, the study was initially designed to include a target of 40% of participants with prior self-injectable experience, although this was reduced to 30% during the study to allow enough participants to be recruited.

All participants provided online informed consent before participating. Participants in the cognitive pilot consented to being audio-recorded. Participants were remunerated for completing the study.

DCE survey

The DCE was generated using Ngene software v1.2.1 (ChoiceMetrics, Sydney, Australia) using a D-efficient design that was assessed against good experimental design properties. The design was optimized for the estimation of a multinomial logit (MNL) model, and, where appropriate, directional priors. The experimental design of the DCE included 36 experimental choice tasks split into three blocks, such that each participant would complete only 12 experimental choice tasks. Participants in the pilot interviews did not struggle with the number of attributes in the choice tasks. Full profiles (where no attributes were fixed to a set level to simplify the design) were therefore used. In each choice task, participants were asked to choose between two hypothetical treatment options (A and B) and an opt-out of staying with their “old treatment”, wherein each treatment option was composed of one level from each of the attributes (**Figure 1**). If a participant selected the “old treatment” option, they answered a follow-up question asking them to choose between treatment options A and B. We utilised a recommended status-quo opt-out option,[23] which remained fixed throughout the survey (while treatment A and B varied). For methodological reasons, to not overestimate patients’ willingness to accept risks, the risk of adverse events was set to 0% for both eye inflammation and serious infections. Since this would not reflect patients varied current treatments, the opt-out option was referred to as ‘old treatment’. The order of the 12 experimental choice tasks and of the attribute groups (benefits, risks, other) within the choice options was randomised across participants to minimise the influence of ordering effects.[24, 25] In addition to the 12 experimental choice tasks, participants answered two choice tasks to assess internal validity.[26] Task 13 was a repeat of the third experimental choice task seen by the participant and was intended to check the stability of their choices. Task 14 was a dominated-choice test in which one treatment option was as good as or better than the other option for all attributes and was intended to test attendance to the tasks.

In addition to the DCE, participants completed a sociodemographic/clinical questionnaire, indicated their willingness (on a 5 point scale form not willing to very willing) to have a medication that required a subcutaneous injection for each dose, and completed the Set of Brief Screening Questions to assess health literacy[27] and five of the seven items from the Numeracy Scale to assess numeracy[28] to assess their ability to understand the attributes and levels presented and their engagement in the survey.

Validity assessments

For the dominance test, which presented one treatment option with higher levels of benefits and lower levels of risks, the number of patients selecting the superior (dominating) option as their preferred treatment was recorded; selecting the superior option indicated the survey sufficiently engaged participants. The number of patients selecting the same choices in the initial and repeated tasks was also recorded; selecting the same option in both questions indicated choice stability. A respondent was classified as a serial non-participant if they chose the same treatment option for all 12 experimental choice tasks. Decision-making was considered dominated when the respondent chose their preferred treatment option based on a single attribute in all 12 experimental choice tasks. For each choice task, response times in the lower 10% of the corresponding distribution were classified as fast and those in the upper 10% as slow. Attendance to the DCE survey was classified as inadequate if ≥80% of a participant's responses for the 12 experimental choice tasks were classified as too fast or too slow.

Statistical analysis

Statistical analysis was performed using R version 3.6.1 (R Foundation, Vienna, Austria). DCE preference data were analysed using a MNL model within the random utility maximization framework[29] (see **Online Supplemental Methods** for details). This model assumed that respondents chose the alternative that resulted in the highest utility (a measure of desirability) based on the included attributes and up to a random error.[30] The main results from this model were part-worth utility estimates, which reflect participants' sensitivities to changes in the treatment attributes. A dummy coding strategy was implemented to estimate preferences for discrete changes in the treatment attributes. In addition, the MNL model included two alternative-specific constants, one that captured left-right bias (tendency to select the option presented on the left of the choice tasks) and one that captured a preference for the old treatment option.

A second MNL model with linearly coded attributes for the skin appearance attribute was also estimated to support the computation of the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. The acceptability of the underlying hypothesis of linearity in preferences for changes in the skin appearance attribute was first verified (see **Online Supplemental Methods** for details). The MAD analysis measured the percentage decrease in the chance of achieving clear or almost clear skin at week 16 a respondent was willing to accept for changes in other attributes. The 95% confidence intervals for the MAD in achieving clear or almost clear skin at week 16 were obtained using the Delta method.[31]

Subgroup analyses were performed according to country (France, Spain, UK), age (<40, 40–50, and >50 years), gender (female, male), Patient Oriented Eczema Measure (POEM) overall score (0–7 [clear or almost clear/mild], 8–16 [moderate], severe/very severe [17–28]),[32] and self-reported eczema severity (very mild/mild, moderate/severe/very severe).

Model selection

A number of different analyses were conducted as part of model selection. Given the DCE was conducted in different countries and the initial version of the survey was developed in the English language, the first analysis was related to the possibility of combining choice data from the different countries. The translation of the survey into different languages might have induced a translation effect, which could have resulted in systematic differences in the quality of the choice data across the countries. The results of this analysis indicated that differences in observed choices across countries could not be fully explained by potential changes in the underling quality of the choice data (**Online Supplemental Methods and Online Supplemental Table 2**); as such, it was decided to pool country data and treat country of residence as a potential driver of heterogeneity in preferences alongside other personal characteristics.

The second analysis aimed to determine whether the standard MNL model would be appropriate to quantify average sample preferences. The MNL model was first compared with a mixed logit (MXL) model allowing for unobserved heterogeneity in preferences. Being the most flexible choice model, the MXL model was expected to statistically outperform the MNL model, but the objective of this analysis was to determine whether using a simpler model would lead to a biased measurement of sample preferences. The comparison of preference estimates between the two models showed a very high level of agreement (i.e., very similar preferences identified with both models) (**Online Supplemental Methods and Online Supplemental Figure 1**).

The MNL model was also compared with a nested logit (NL) model to determine whether the opt-out option “old treatment” required different treatment to the other treatment alternatives. The NL model relaxed the hypothesis of independence of irrelevant alternatives, which is a core assumption of the MNL model and implies that all three treatment options were equally substitutable. Again, the comparison of preference estimates showed a high level of agreement between the MNL and NL models (**Online Supplemental Methods and Online Supplemental Figure 2**). These results indicated that the MNL model provided an acceptable approximation of sample preferences.

Patient and public involvement

Cognitive pilot interviews were held with 15 patients to test understandability of the DCE survey. Other than participating in the DCE survey as respondents, patients were not involved in recruitment or study conduct. Investigators were blinded to the identities of the study participants, so the results of the study were not directly disseminated to them.

RESULTS

Participants

The DCE survey included 404 participants (114 in France, 145 in Spain, and 145 in the UK) who were recruited between October and December 2019. Given recruitment for the quantitative online survey used patient panels and databases, 157,553 initial invites were sent, with a 4% (n=6,287) response rate. The majority of the interested potential participants completed the screening questionnaire but were not eligible to participate, largely due to not having AD; 541 patients were eligible to participate, with 75% of those eligible completing the survey. Most participants were female (65%) with an average age of 44.1 years (**Table 2**). Most participants were employed full time (56%) and had completed university education or higher (58%). The majority of participants had moderate-to-very severe AD according to POEM scores (62%) and self-reported eczema severity (67%) but good-to-excellent self-reported overall health (69%). Topical corticosteroids (66%) were the most frequently used class of medications at the time of the survey, followed by systemic immunosuppressant therapies (27%) and biologics (18%). Topical betamethasone (29%) and hydrocortisone (24%) were the most frequent currently used individual medications. Most patients (68%) had no prior experience of using self-injectable treatments for AD or any other illness.

Table 2. Participant characteristics

Characteristic	N=404
Sex, n (%)	
Male	142 (35)
Female	262 (65)
Age, mean (SD)	44.1 (12.0)
Employment status	
Full time	227 (56)
Part time	75 (19)

Characteristic	N=404
Homemaker/housewife	21 (5)
Student	10 (2)
Unemployed	30 (7)
Retired	35 (9)
Disabled	12 (3)
Other	2 (0)
Education, n (%)	
No formal qualifications	1 (0)
Primary school or secondary education	38 (9)
College or some university	43 (11)
Completed vocational or professional certification	83 (21)
Completed university degree	148 (37)
Completed doctorate, post-doctorate, or equivalent	88 (22)
Other	3 (1)
Overall health, n (%)	
Excellent	20 (5)
Very good	96 (24)
Good	161 (40)
Fair	98 (24)
Poor	29 (7)
Prior experience with self-injectables (any)*	
Yes	129 (32)
No	275 (68)
Self-rated eczema severity, n (%)	
Very mild	19 (5)
Mild	116 (29)
Moderate	212 (52)
Severe	45 (11)
Very severe	12 (3)
POEM overall score, n (%)	
Clear or almost clear (0–2)	32 (8)
Mild eczema (3–7)	121 (30)
Moderate eczema (8–16)	192 (48)
Severe eczema (17–24)	47 (12)
Very severe eczema (25–28)	12 (3)
Class of AD medication currently used, n (%)†	
Topical corticosteroids	265 (66)
Topical calcineurin inhibitors	32 (8)
Phototherapy/UV treatment	20 (5)
Systemic immunosuppressant therapies	109 (27)
Biologics	72 (18)
Most frequently used current AD medications, n (%)†	
Betamethasone	119 (29)
Hydrocortisone	97 (24)
Prednisone	61 (15)
Clobetasol propionate	46 (11)

*Participants were not asked whether their prior use of self-injectables was for AD.

†Not mutually exclusive.

Abbreviations: AD, atopic dermatitis; POEM, Patient Oriented Eczema Measure; SD, standard deviation

Validity assessments

Overall, the survey sufficiently engaged participants: 89% selected the superior treatment option in the dominance test, 64% chose the same answers in the repeated choice task, and 97% spent an adequate amount of time on the choice tasks (**Online Supplemental Table 3**). Also, for 90% of participants, decisions were not dominated by a single attribute, and only 5% always chose the opt-out old treatment option. Participants were not excluded based on responses to the validity tests, following best practise recommendations,[33] as the preferences of patients may be valid and removal may induce selection bias.

Overall preferences for treatment attributes

The DCE dataset had no missing values, as patients could not proceed in the survey without answering each question or item. If participants did not complete the survey they were not remunerated or included in the dataset. Of the treatment attributes included in the DCE survey, participants most valued improving symptoms and reducing the risk of side effects (**Figure 2** and **Online Supplemental Table 4**). The most valued change was an improvement from 20% to 50% in the chance of achieving a meaningful reduction in itch at week 16, although preferences did not significantly differ between an improvement to a 40% or 50% chance of achieving a meaningful reduction in itch. The next-most valued changes, in descending order, were a decrease in the risk of serious infections from 6% to 0%, a decrease in the risk of eye inflammation from 20% to 0%, and an improvement in the chance of achieving clear or almost clear skin from 10% to 40%.

Participants also valued changes in the non-clinical attributes. The most valued change was switching from a treatment that can be used long-term but cannot be paused without affecting efficacy to one that can be used long-term with the possibility for pauses, without affecting efficacy.

An oral pill once or twice daily was preferred over a subcutaneous treatment every 2 weeks, and a 2-day onset of action was preferred over a 2-week onset of action, although participants did not have a significant preference for a 1-week over a 2-week onset of action. Participants also preferred a treatment that can manage flares by modifying the dose according to symptoms over one that cannot be used to manage flares, although this was less important than changes in other non-clinical attributes.

Subgroup analyses

Results were similar for the three included countries (UK, Spain, and France) (**Online Supplemental Figure 3**), by age (**Online Supplemental Figure 4**), by gender (**Online Supplemental Figure 5**), by POEM overall score (**Online Supplemental Figure 6**), and by self-reported eczema severity (**Online Supplemental Figure 7**). However, those aged over 50 cared more about receiving an oral pill relative to those aged 40-50 years, for whom we did not detect a significant preference for administration.

Participants who had experience of self-injecting a treatment for any illness (32%) were more willing to accept a treatment that required a subcutaneous injection and placed less importance on reducing the risk of serious infections than those who did not have experience self-injecting a treatment for any illness (**Online Supplemental Figure 8**).

Willingness to make trade-offs between treatment attributes

Participants would be willing to tolerate reduced efficacy to obtain changes in other treatment attributes. Specifically, they would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 50.1% (95% CI, 38.5%–61.8%) to increase the chance of achieving a meaningful reduction in itch at week 16 from 20% to 50%; 48.6% (95% CI, 35.2%–62.0%) to reduce the risk of serious infections from 6% to 0%; and 42.3% (95% CI, 30.0%–54.5%) to reduce the risk of eye inflammation from 20% to 0% (**Table 3**). They would also be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 24.1% (95% CI, 16.5%–31.6%) to switch from a treatment that can be used long-term but cannot be paused without losing efficacy to one that can be paused without losing efficacy; 16.6% (95% CI, 9.2%–24.0%) to switch from a subcutaneous treatment every 2 weeks to an oral pill once or twice daily; and 5.8% (95% CI, 0.5%–11.1%) to obtain a treatment whose dosage can be modified to manage flares over one that cannot. Further, participants would be willing to tolerate a decrease in the probability of achieving clear or almost clear skin of 20.9% (95% CI, 12.3%–29.5%) to switch from a treatment that requires frequent check-ups to one that does not require check-ups; and 16.1% (95% CI, 8.7%–23.5%) to switch from a treatment that requires frequent check-ups to one that requires occasional check-ups.

Table 3. Maximum acceptable decrease in the probability of achieving clear or almost clear skin at week 16

Attribute/level	Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)
Itch reduction	
2 out of 10 (20%)	Reference
4 out of 10 (40%)	38.7 (28.8, 48.6)
5 out of 10 (50%)	50.1 (38.5, 61.8)

Attribute/level	Maximum acceptable decrease in the probability of achieving clear or almost clear skin (95% CI)
Eye inflammation	
20 out of 100 (20%)	Reference
10 out of 100 (10%)	17.9 (10.5, 25.4)
0 out of 100 (0%)	42.3 (30.0, 54.5)
Serious infections	
6 out of 100 (6%)	Reference
3 out of 100 (3%)	20.6 (12.7, 28.6)
0 out of 100 (0%)	48.6 (35.2, 62.0)
Speed of onset	
2 weeks	Reference
1 week	0.2 (−6.5, 6.9)
2 days	11.3 (4.4, 18.2)
Flare management	
No	Reference
Yes	5.8 (0.5, 11.1)
Long-term disease management	
Yes, without the possibility for pauses	Reference
Should not be used long-term	4.3 (−2.7, 11.3)
Yes, with the possibility for pauses	24.1 (16.5, 31.6)
Administration	
Injection under the skin every 2 weeks	Reference
Oral pill once or twice daily	16.6 (9.2, 24.0)
Check-ups	
Frequent check-ups required	Reference
Occasional check-ups required	16.1 (8.7, 23.5)
No check-ups required	20.9 (12.3, 29.5)

Abbreviations: CI, confidence interval

DISCUSSION

The current study, which included 404 participants across the UK, France, and Spain, found that adults with AD who had recently been treated with topical and/or systemic therapy most valued increasing the benefits and reducing the risks of their treatments, although attributes specific to new targeted therapies, such as mode of administration and long-term disease management, also had a significant effect on choices. Participants were willing to tolerate a significant decrease in the possibility of achieving clear or almost clear skin to obtain a treatment that is more convenient, including an oral pill once or twice daily in place of a subcutaneous injection every 2 weeks, the ability to pause the treatment without losing efficacy, the ability to modify the dosage to manage flares, and the possibility of requiring only occasional or no check-ups instead of frequent check-ups. Further, participants with self-injectable experience for any illness were more willing to accept self-injection than participants without self-injectable experience. However, 28% of participants were ‘not willing’ or ‘somewhat not willing’ to have a medication that required an injection for each dose. Preferences were similar between the three countries included (UK, France, and Spain) and were largely unaffected by age or sex. In addition, preferences did not significantly differ based on disease severity, as measured using the POEM score, which is in line with prior research.[34]

Two other recent DCEs have examined the treatment preferences of patients with AD. Similar to our study, a DCE in the US including 320 adults with moderate-to-severe AD[34] found that patients preferred an oral pill over subcutaneous injection and valued a rapid onset of action and increasing the chance of achieving clear or almost clear skin at week 16. A DCE including 323 patients in Japan ≥ 15 years of age with moderate to very severe AD and 121 dermatologists treating patients with AD[35] found that, as in the current study, both groups considered benefits and adverse effects the most important attributes of injectable treatments, although preferences for some treatment attributes differed between the groups. For example, patients placed more value on efficacy of improving rashes and treatment costs than dermatologists, while dermatologists valued time until response more than patients. Patients also preferred adding new treatments to current treatments as add-ons and receiving treatments at clinics, while physicians preferred reducing the number of current treatments and having patients self-administer at home. These differences in the preferences of patients and physicians emphasize the need for studies like the current one that are specifically designed to provide insight into patients' preferences.

Internal validity of the current DCE was examined using tests of choice stability and dominance, as well as by considering response times, health literacy, and numeracy. The results were in line with existing research, including for choice stability,[26] and suggested the survey sufficiently engaged participants. A potential limitation of this study is that the attributes and levels were not identified through a separate qualitative research phase but rather through a targeted review of previous quantitative and qualitative studies of patients with AD and product labels for AD treatments. We do not expect that this influenced the results because the same attributes (onset of itch relief, probability of skin clearance, frequency or ease of administration/convenience, and safety) were also identified through the qualitative phase of the US study.[34]

A potential limitation of this study is the inclusion of four probabilistic attributes, which increased the complexity of the study for participants. These were included to align with clinical data. To mitigate this, we included a thorough warm-up to the DCE with practice questions relating to the probabilistic attributes. In addition, a prior AD study included four probabilistic attributes (two probabilistic benefits and two probabilistic adverse events).[34] Another limitation of this study is that we used different denominators for probabilistic benefit and risk attributes. Different denominators were utilised to ensure participants could review all attribute information simultaneously while making their choices. However, using different denominators may have increased the study complexity and introduced a potential bias. Another potential limitation of this study is reference to the opt-out as 'old', which may have been perceived negatively. We used the terminology 'old' instead of current since we were aware that we were not presenting patients with their actual current treatments, which may have caused confusion. Due to the need to limit the participants' cognitive burden, not all potentially relevant attributes could be included in the DCE survey. However, cognitive pilot interviews of 15 patients with AD indicated that the attributes and levels were relevant and that no attributes were missing. Overall, participants also found the length and complexity of the survey acceptable. A further limitation is the inclusion of patients with non-severe AD, who would possibly not receive systemic therapies.[2] However, there is value in including these patients, because patients' disease severity may vary over time and treatment recommendations may change. Also, although few differences were found in preferences by age, sex, or country, care should be taken when generalizing to underrepresented AD populations, such as patients with very severe AD, children, or patients in lower income countries. Additionally, since it is not culturally appropriate to ask about race in some European countries, data was not collected on this. We were therefore not able to determine whether this study represents the diverse ethnic groups in the study countries. Moreover, our sample included a high proportion of participants with university education and may therefore not be fully representative of the general AD population.

In conclusion, patients with AD most valued treatment benefits and reducing risks but were willing to accept a decrease in efficacy, as measured by the possibility of obtaining clear or almost clear skin at week 16, to obtain an oral treatment with a rapid onset of action. This information may help clinicians make shared decisions with patients about the most suitable treatment for AD. It can also support reimbursement applications, ensuring that health technology assessment decisions align with the preferences of individuals living with AD.

Author contributions: C.T. contributed to conception and planning of the study, acquisition and interpretation of the data, and drafting and critical revision of the manuscript. A.R. contributed to conception and design of the study, interpretation of the data, and critical revision of the manuscript. A.W. and J.-P.C. contributed to design of the study, interpretation of the data, and critical revision of the manuscript. N.K. contributed to analysis and interpretation of the data and critical revision of the manuscript. H.K. contributed to conception and design of the study and critical revision of the manuscript. T.T. contributed to conception and planning of the study, interpretation of the data, and critical revision the manuscript. All authors approved the final version of the manuscript.

Acknowledgements: Medical writing support was provided by Philip S. Leventhal, PhD, Holly Richendrfer, PhD, and Stephen Gilliver, PhD of Evidera and was paid for by Eli Lilly & Co.

Competing Interests: C.T., N.K., H.K., and T.T. are employees of Evidera, which was paid by Eli Lilly & Co. for work related to this study. A.R. and J.-P.C. are employees of Eli Lilly & Co. A.W. received personal fees from Eli Lilly & Co for work related to this study.

Funding: This work was supported by Eli Lilly & Company. Grant/award number is not applicable

Data availability statement: The datasets generated and/or analysed during the study are not publicly available, because consent was not sought from participants to allow sharing of data with third parties.

REFERENCES

1. Fishbein AB, Silverberg JI, Wilson EJ, et al. Update on Atopic Dermatitis: Diagnosis, Severity Assessment, and Treatment Selection. *J Allergy Clin Immunol Pract* 2020;8:91-101.
2. Wollenberg A, Christen-Zach S, Taieb A, et al. ETFAD/EADV Eczema task force 2020 position paper on diagnosis and treatment of atopic dermatitis in adults and children. *J Eur Acad Dermatol Venereol* 2020;34:2717-44.
3. Wollenberg A, Barbarot S, Bieber T, et al. Consensus-based European guidelines for treatment of atopic eczema (atopic dermatitis) in adults and children: part II. *J Eur Acad Dermatol Venereol* 2018;32:850-78.
4. Sidbury R, Davis DM, Cohen DE, et al. Guidelines of care for the management of atopic dermatitis: section 3. Management and treatment with phototherapy and systemic agents. *J Am Acad Dermatol* 2014;71:327-49.
5. Seegraber M, Srour J, Walter A, et al. Dupilumab for treatment of atopic dermatitis. *Expert Rev Clin Pharmacol* 2018;11:467-74.
6. OLUMIANT® Summary of Product Characteristics: Eli Lilly Nederland B.V.; 2020. Available from: https://ec.europa.eu/health/documents/community-register/2020/20201019149480/anx_149480_en.pdf. Accessed 02 December 2020.
7. Rinvoq: Summary of opinion (post authorisation): Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency; 2021. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-rinvoq-x-06-g_en.pdf. Accessed 09 November 2021.

8. Deleanu D, Nedelea I. Biological therapies for atopic dermatitis: An update (Review). *Exp Ther Med* 2018;17:1061-67.

9. Casciano R, Malangone E, Ramachandran A, et al. A quantitative assessment of patient barriers to insulin. *Int J Clin Pract* 2011;65:408-14.

10. Edel Y, Sagy I, Pokroy-Shapira E, et al. A Cross-sectional Survey on the Preference of Patients with Rheumatoid Arthritis for Route of Administration of Disease-Modifying Anti-Rheumatic Drugs: Oral Target-Specific Versus Parenteral Biologic. *Isr Med Assoc J* 2020;22:154-59.

11. Mohr DC, Boudewyn AC, Likosky W, et al. Injectable medication for the treatment of multiple sclerosis: The influence of self-efficacy expectations and infection anxiety on adherence and ability to self-inject. *Ann Behav Med* 2001;23:125-32.

12. Bouvy JC, Cowie L, Lovett R, et al. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. *Patient* 2020;13:145-49.

13. Huls SPI, Whichello CL, van Exel J, et al. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. *Value Health* 2019;22:1318-28.

14. Janus SI, Weernink MG, van Til JA, et al. A Systematic Review To Identify the Use of Preference Elicitation Methods in Health Care Decision Making. *Value Health* 2014;17:A515-6.

15. Mühlbacher AC, Juhnke C, Beyer AR, et al. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. *Value Health* 2016;19:734-40.

16. Johnson FR, Zhou M. Patient Preferences in Regulatory Benefit-Risk Assessments: A US Perspective. *Value Health* 2016;19:741-45.

17. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. *Patient* 2017;10:263-66.

18. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics* 2018;37:201-26.

19. Radi G, Simonetti O, Rizzetto G, et al. Baricitinib: The First Jak Inhibitor Approved in Europe for the Treatment of Moderate to Severe Atopic Dermatitis in Adult Patients. *Healthcare (Basel)* 2021;9

20. Thyssen JP, Buhl T, Fernández-Peñas P, et al. Baricitinib Rapidly Improves Skin Pain Resulting in Improved Quality of Life for Patients with Atopic Dermatitis: Analyses from BREEZE-AD1, 2, and 7. *Dermatol Ther (Heidelb)* 2021;11:1599-611.

21. Heidenreich S, Phillips-Beyer A, Flamion B, et al. Benefit-Risk or Risk-Benefit Trade-Offs? Another Look at Attribute Ordering Effects in a Pilot Choice Experiment. *Patient* 2021;14:65-74.

22. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Econ* 2009;18:321-36.

23. Determann D, Gyrd-Hansen D, de Wit GA, et al. Designing Unforced Choice Experiments to Inform Health Care Decision Making: Implications of Using Opt-Out, Neither, or Status Quo Alternatives in Discrete Choice Experiments. *Med Decis Making* 2019;39:681-92.

24. Carlsson F, Mørkbak MR, Olsen SB. The first time is the hardest: A test of ordering effects in choice experiments. *J Choice Model* 2012;5:19-37.

25. Heidenreich S, Phillips-Beyer A, Flamion B, et al. Benefit–risk or risk–benefit trade-offs? Another look at attribute ordering effects in a pilot choice experiment. *Patient* 2021;14:65-74.

26. Johnson FR, Yang J-C, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. *Value Health* 2019;22:157-60.

27. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:588-94.

28. Lipkus IM, Samsa G, Rimer BK. General Performance on a Numeracy Scale among Highly Educated Samples. *Med Decis Making* 2016;21:37-44.

29. Manski CF. The structure of random utility models. *Theory Decis* 1977;8:229-54.

30. Thurstone LL. A law of comparative judgment. *Psychol Rev* 1927;34:273-86.

31. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ* 2007;16:827-40.

32. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. *Arch Dermatol* 2004;140:1513-9.

33. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ* 2006;15:797-811.

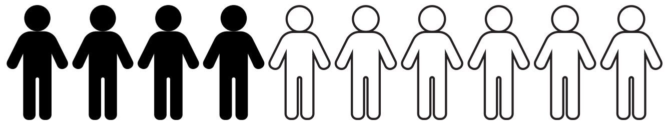
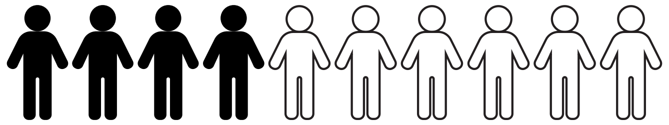
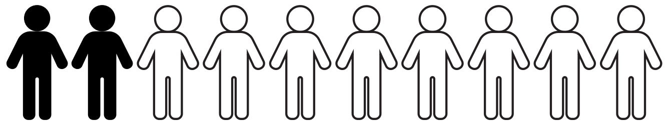
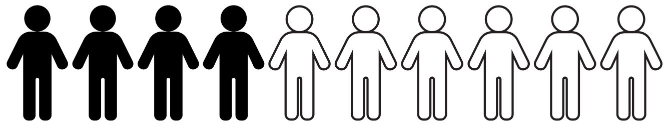
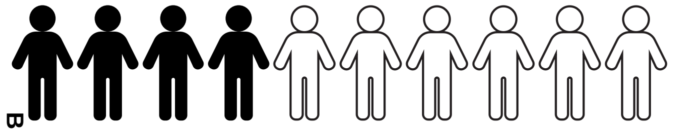
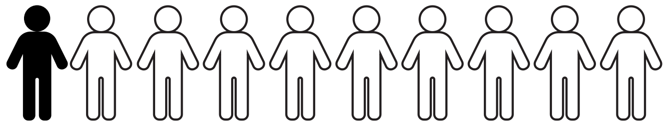
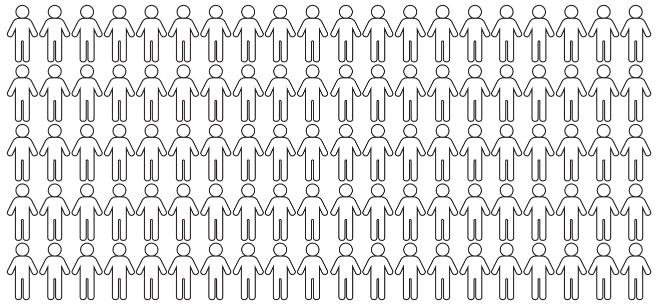
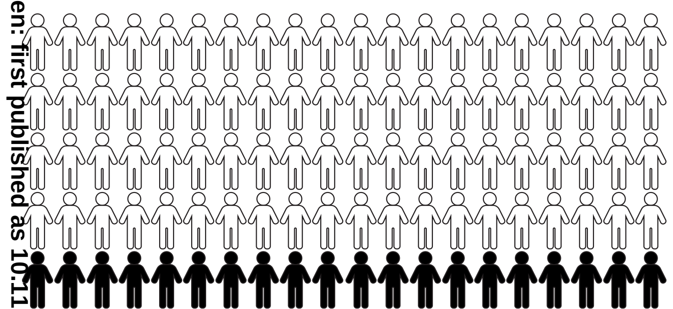
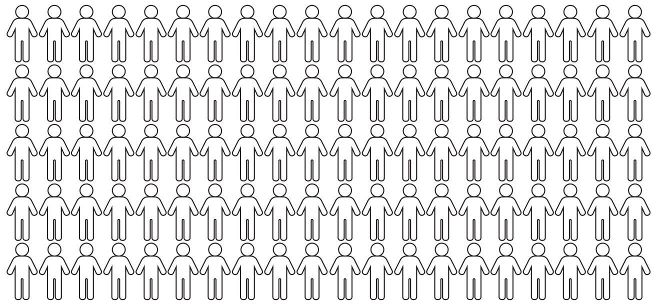
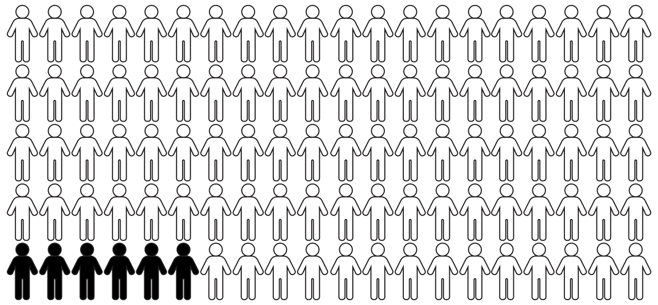
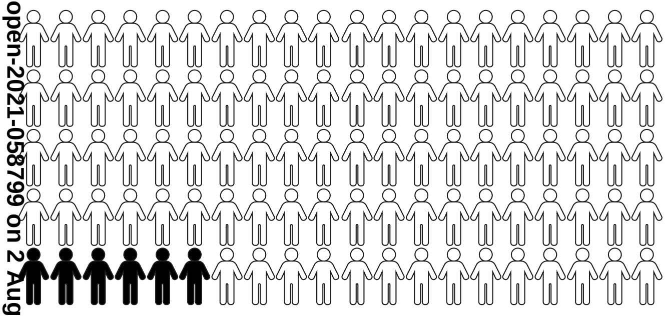
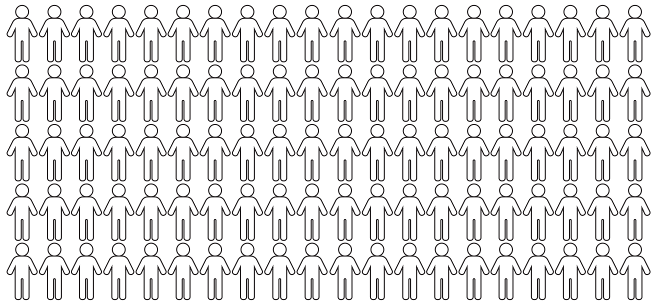
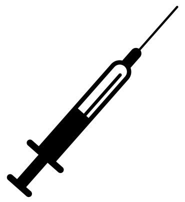
34. Boeri M, Sutphin J, Hauber B, et al. Quantifying patient preferences for systemic atopic dermatitis treatments using a discrete-choice experiment. *J Dermatolog Treat* 2020:1-10.

35. Okubo Y, Ho K-A, Fifer S, et al. Patient and physician preferences for atopic dermatitis injection treatments in Japan. *J Dermatolog Treat* 2019;31:821-30.

FIGURE LEGENDS

Figure 1. Example choice task

Figure 2. Multinomial logit results: part-worth utilities

		[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	4 out of 10 (40%) 	4 out of 10 (40%) 	2 out of 10 (20%) 
	[illegible]	4 out of 10 (40%) 	4 out of 10 (40%) 	1 out of 10 (10%) 
[illegible]	[illegible]	0 out of 100 (0%) 	20 out of 100 (20%) 	0 out of 100 (0%) 
	[illegible]	6 out of 100 (6%) 	6 out of 100 (6%) 	0 out of 100 (0%) 
[illegible]	[illegible]	1 week	2 weeks	2 weeks
	[illegible]	✓	✗	✓
	[illegible]	Yes, with the possibility of pauses	Yes, without the possibility of pauses	Should not be used long-term
	[illegible]	Oral pill Once or twice daily	Injection under the skin  Every two weeks	Oral pill Once or twice daily
	[illegible]	Occasional check-ups required	No check-ups required	Frequent check-ups required
	[illegible]	☐	☐	☐
	[*if old treatment is chosen*] [illegible]	☐	☐	

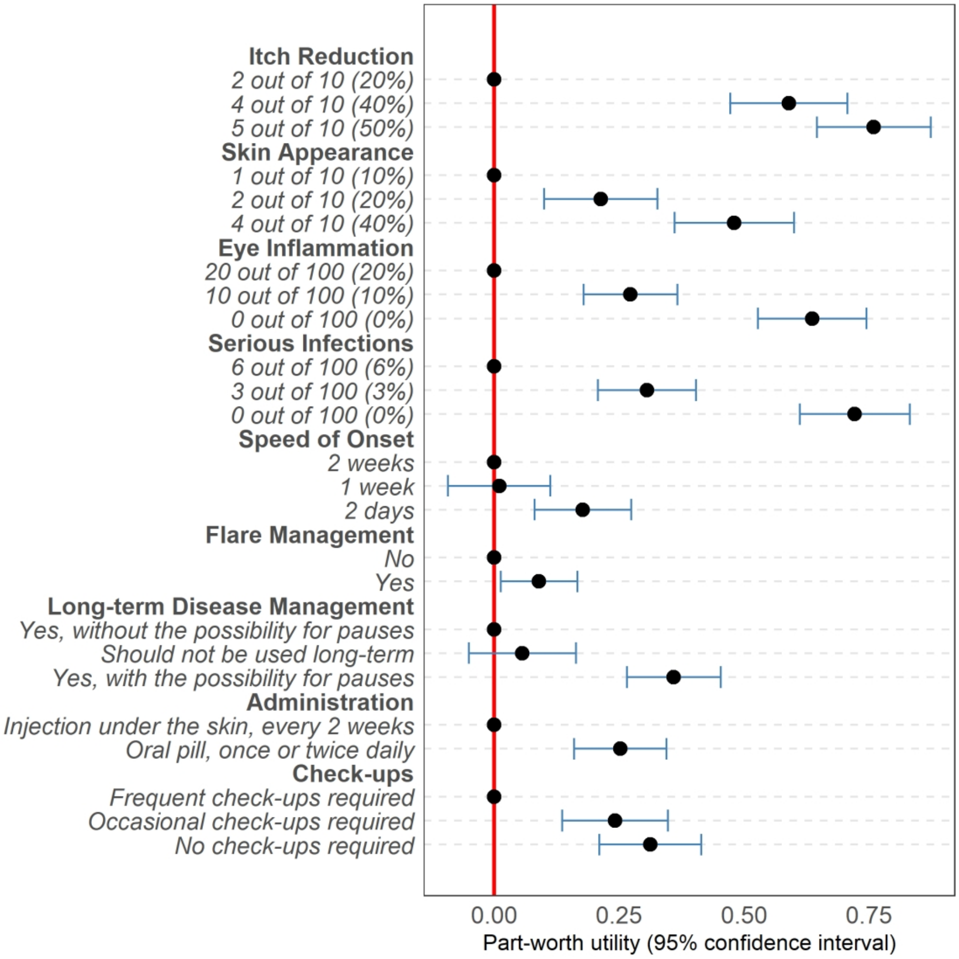


Figure 2. Multinomial logit results: part-worth utilities

177x177mm (300 x 300 DPI)

□□□□□□□□□□□□M□□□□□M□□□□D□□

□□□□□□□□□□□□□□r□□□□□□□d□□□□□□□

The aim of the literature review was to determine the key attributes and levels to be included in the DCE. This involved both a targeted literature review and a product label review.

The targeted literature review involved separate searches in major medical literature databases (Embase and MEDLINE) (□□□□□□□□□□ □□□□□□□□); a search for qualitative studies that considered the patient perspective on AD treatments; and a search for patient preference-specific studies, which considered AD treatments. Once key themes within the literature review were identified, the attributes were classified into corresponding categories. All abstracts were double screened, and disagreements were reviewed by a senior member of the research team.

The search strategy for the qualitative studies focused on studies that conducted interviews or focus groups, which mentioned AD or eczema, and their available treatments, as well as quality of life or patient preferences. The search excluded any non-adult studies, animal studies, clinical trials, and editorial notes. The search strategy for patient preference-specific studies sought studies that were explicitly patient preference in design, such as those utilising DCEs. Additionally, the studies had to mention AD or eczema.

The targeted literature search identified 33 potential studies. No duplicates were found, and all 33 were screened for eligibility. The abstracts were screened sequentially by two reviewers, and a third reviewer compared the rationale for inclusion and exclusion of studies to obtain the final list of full texts to screen. Seven studies were excluded because they did not involve adult patients, 13 because they weren't about AD, six because they did not have the study design of interest, and four because no full text was available. The remaining three studies included one quantitative[1] and two qualitative studies.[2, 3] In the quantitative study, the most important treatment attribute was the appearance of eczema (dryness/flakiness). In the two qualitative studies, itch reduction (symptom control), monitoring of symptoms, flexibility of treatment regimens to control flares, appearance (dryness/flakiness), and skin pain were identified themes.

Additionally, a product label search was conducted. Ten product labels for medications indicated for use in AD were reviewed in detail, including baricitinib (Olumiant®), dupilumab (Dupixent®), clobetasol propionate (Clobex®), tacrolimus (Protopic®), prednisone (Rayos®), cyclosporin (Neoral®), methotrexate, azathioprine (Imuran®), mycophenolate mofetil (CellCept®), and phototherapy. Itch reduction was most commonly reported as the percentage of patients achieving a meaningful (≥4-point reduction in the itch numerical rating scale) reduction in itch at week 16. Skin appearance was most commonly measured by the proportion of patients achieving clear or almost clear skin at week 16 (Investigator's Global Assessment scores of 0 or 1). The review of product labels also identified conjunctivitis as a differentiating and common side-effect of dupilumab that is not associated with other systemic therapies. Risk of serious infections were associated with other treatments, such as baricitinib and cyclosporine. The product label review also highlighted different modes and frequency of administration for systemic treatments, which included daily oral medication or subcutaneous administration every 2 weeks. Monitoring was also required for baricitinib and cyclosporine, but not for dupilumab.

Model specification ☐

The analysis of all DCE responses followed random utility theory.[4-6] The model assumes that each respondent (*n*) chooses the alternative (*j*) in every DCE question (*t*) that results in the highest utility (a measure of desirability) of all available alternatives. Utility in a random utility model is defined as:

$$u(\boldsymbol{x}_{jnt}) = v(\boldsymbol{x}_{jnt}) + \varepsilon_{jnt}$$

Here the systematic utility component $v(\boldsymbol{x}_{jnt})$ is a function of the DCE attributes and ε_{jnt} is a type 1 extreme value distributed random error. Two models are presented: a dummy-coded MNL model and an MNL model with skin appearance coded linearly, which is required to estimate the maximum acceptable decrease (MAD) in the probability of achieving clear or almost clear skin at week 16. For the former, the utility function was defined as:

$$\begin{aligned} u_{jnt} = \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} &+ \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\ &+ \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3 \text{20\%\_skin\_appearance}_{jnt} \\ &+ \beta_4 \text{40\%\_skin\_appearance}_{jnt} + \beta_5 \text{10\%\_eye\_inflammation}_{jnt} \\ &+ \beta_6 \text{0\%\_eye\_inflammation}_{jnt} + \beta_7 \text{3\%\_serious\_infections}_{jnt} \\ &+ \beta_8 \text{0\%\_serious\_infections}_{jnt} + \beta_9 \text{1\_week\_onset}_{jnt} + \beta_{10} \text{2\_days\_onset}_{jnt} \\ &+ \beta_{11} \text{flare\_management}_{jnt} + \beta_{12} \text{long\_term\_no}_{jnt} \\ &+ \beta_{13} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{14} \text{oral\_admin}_{jnt} + \beta_{15} \text{no\_check\_ups}_{jnt} \\ &+ \beta_{16} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt} \end{aligned}$$

The constants $\alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}}$ controlled for potential bias to select the left option (Treatment A), and the Old Treatment, β_1 to β_{16} were the estimated marginal utilities (i.e., estimated preference parameters), ε_{jnt} was an extreme value type I distributed error that allowed the function to be estimated in a logit model.[6] All attributes were dummy-coded. The reference level was the assumed worst-case option. Each of the estimated marginal utilities measured respondents' sensitivity to deviations from the reference level of the corresponding attribute. The sign (+ or –) of a marginal utility denotes whether patients valued this deviation positively or negatively. Only the initial choices (A vs. B vs. old treatment) were considered for the analysis of preferences. The initial and follow-up choices can be combined to allow for a more precise measurement of preferences. However, it is appropriate to combine these two types of choices only when they generate approximately the same information about participants' preferences. This condition was verified in two ways. Two MNL models were separately estimates for the initial (4,848 observations) and follow-up choices (1,126 observations), and then their preference estimates were compared. The Pearson correlation coefficient between the two sets of estimates was relatively low (0.32) as was the coefficient of determination for the linear regression (0.104), indicating poor agreement between the sets of estimates. A third MNL model was estimated on the combined initial and follow-up choices (5,974 observations), and its statistical performance was compared with the MNL model based on initial choices only. The adjusted McFadden pseudo-R^2 was lower for the model based on combined choices (7.3%) than for the initial model (8.3%), indicating that combining the initial and follow-up choices had a detrimental effect on the explanatory power of the model.

The linear coding of skin appearance was required to derive meaningful MAD measures. This measure was obtained by estimating the baseline utility function with skin appearance being coded as linear (i.e., one marginal utility is estimated instead of β_3 and β_4 for skin appearance). The utility function was defined as:

$$
\begin{aligned}
u_{jnt} = {} & \alpha_{\text{Treatment A}} + \alpha_{\text{Old Treatment}} + \beta_1 \text{40\%\_itch\_reduction}_{jnt} \\
& + \beta_2 \text{50\%\_itch\_reduction}_{jnt} + \beta_3\text{\_skin\_appearance}_{jnt} \\
& + \beta_4 \text{10\%\_eye\_inflammation}_{jnt} + \beta_5 \text{0\%\_eye\_inflammation}_{jnt} \\
& + \beta_6 \text{3\%\_serious\_infections}_{jnt} + \beta_7 \text{0\%\_serious\_infections}_{jnt} \\
& + \beta_8 \text{1\_week\_onset}_{jnt} + \beta_9 \text{2\_days\_onset}_{jnt} + \beta_{10} \text{flare\_management}_{jnt} \\
& + \beta_{11} \text{long\_term\_no}_{jnt} + \beta_{12} \text{long\_term\_yes\_pauses}_{jnt} + \beta_{13} \text{oral\_admin}_{jnt} \\
& + \beta_{14} \text{no}_{\text{check}_{\text{ups}}}{}_{jnt} + \beta_{15} \text{occassional\_check\_ups}_{jnt} + \varepsilon_{jnt}
\end{aligned}
$$

Each marginal utility was then divided by the marginal utility for skin appearance:

$$\text{MAD}_k = \frac{\widehat{\beta}_k}{\widehat{\beta}_3}$$

Such linear encoding is based on the underlying assumption of linearity in preferences, wherein a one-unit change in the attribute has a constant effect on respondents' choices and does not depend on the absolute value of the attribute level (e.g., a one-unit change from 15% to 16% increase in the chance of achieving clear or almost clear skin at week 16 has the same effect as the change from 20% to 21%). The validity of the assumption of linearity in preferences was tested by analysing the trend in risk estimates from the dummy-coded MNL model. Estimates were obtained for every attribute level in the dummy-coded MNL model (i.e., 3 levels for skin appearance). The linearity of skin appearance was tested by fitting a linear regression and evaluating its coefficient of determination. The assumption of linearity in skin appearance was accepted with a coefficient of 0.81, which exceeds the threshold of 0.7 to verify linearity.

□□□ □□□□□□□□□□□□□□□□□□□□□d□□□□r□□ d□□□r□□□□□□□□r□□□□

We estimated a heteroscedastic MNL (HMNL) model allowing for scale differences between countries. The likelihood ratio test (LRT) indicated that this model performed significantly better than the standard MNL model (D=11.45, P=0.003). We also estimated an extended version of the MNL model allowing for interaction effects between country of residence and the attributes' levels. This interacted MNL (IMNL) model also significantly outperformed the standard MNL model (D=66.44, P=0.001).

Using the scale estimates from the HMNL model, we applied a scale correction to the dataset and then re-estimated the IMNL model (RIMNL) to determine whether the interaction effects found to be significant in the initial IMNL model would remain significant after accounting for potential scale differences between countries. This was the case, indicating that differences in choice behaviours between countries could not be fully explained as the consequence of a change in underlying utility scale (□ □□□□□□□□□□□□□ □□□□□□□□□□□□).

□ □□□□□□□□□□□□□r□□□□□□□□r□□d□□□□r□□□□□□□□□□□□□r□□□r□□□□□□

We estimated an MXL model allowing all parameters to be independently and normally distributed (i.e., diagonal covariance matrix of random effects). The MXL model significantly outperformed its MNL counterpart (LRT: D=678.39, P<0.001), but a comparison of estimates between the two models showed a high level of agreement (□ □□□□□□□□□□□□□ □□□□□□□□□□r□□□). We fitted a linear regression line through the set of coordinates (MNL; MXL) and the coefficient of determination was close to 100%. The intercept, which can be interpreted as a measure of bias associated with use of MNL estimates instead of MXL ones, was close to zero (0.012) and non-significant (P=0.462). However, the slope (1.172), which can be interpreted as a measure of scale, was significantly different from 1 (P<0.001), indicating that the MXL model measured the same preference effects but on a higher (more precise) utility scale. Given the research objectives of our study were to quantify trade-offs between attributes, and more specifically the MAD in the probability of achieving clear/almost clear skin at week 16, this change in utility scaling was deemed irrelevant.

[illegible]

A nested logit (NL) model was estimated to allow for a repartition of the choice options in two different nests: treatments A and B in a “New treatment” nest and the opt-out option in an “Old treatment” nest. The inclusive value (IV) parameter, which captures the degree of correlation in unobserved factors over alternatives within the "New treatment" nest, was significant (P=0.003) and implied a weak-to-moderate correlation (1-0.63=0.37). The LRT indicated that the NL model significantly outperformed the MNL model (D=8.09, P=0.004). However, a comparison of estimated effects between the two models showed a high level of agreement (r^2>99%) and the intercept of the linear regression line was null ([illegible]).

References

1. Hauber AB, Mohamed AF, Gonzalez JM, et al. Benefit-risk tradeoff preferences for chronic hand eczema treatments. *Journal of Dermatological Treatment* 2016;28:40-46.
2. Howells LM, Chalmers JR, Cowdell F, et al. ‘When it goes back to my normal I suppose’: a qualitative study using online focus groups to explore perceptions of ‘control’ among people with eczema and parents of children with eczema in the UK. *BMJ Open* 2017;7:e017731.
3. DiBenedetti D, Baranowski E. Assessing patient and physician experiences with severe chronic hand eczema. *J Am Acad Dermatol* 2015;72:AB77.
4. Marschak J. Binary choice constraints on random utility indicators Mathematical methods in the social sciences. Mathematical methods in the social sciences. Palo Alto, CA, USA: Stanford University Press 1959:312-29.
5. Manski CF. The structure of random utility models. *Theory and Decision* 1977;8:229-54.
6. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. Frontiers in econometrics. New York: Academic Press 1973:105-42.

□ □ □□ □□ □□□□M□□ □□ □□□ □ □□□□

□ □□□□□□□□□ □□□□□□□□□□□□r□□□d□□□r□□r□r□□□□ □□□r□□□r□ □□

No.	Query	Results	Date
#6	#1 AND (#2 AND #3 OR (#4 AND #5))	33	10-Sep-18
#5	((('qualitative research'/exp OR 'nursing methodology research'/exp OR ethnograph*:ti,ab OR lived) AND experience*:ti,ab OR narrative) AND analysis:ti,ab OR grounded) AND interview*:ti,ab OR themes:ab,ti	80104	10-Sep-18
#4	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1743076	10-Sep-18
#3	'quantitative study'/exp OR 'discrete choice' OR 'dce':ab,ti OR 'discrete choice experiment*':ab,ti OR 'choice experiment*':ab,ti OR 'conjoint':ab,ti OR 'conjoint analysis':ab,ti OR 'bws':ab,ti OR 'benefit risk':ab,ti OR 'thresholding':ab,ti OR 'multiple criteria decision analysis':ab,ti OR 'benefit-risk':ab,ti OR 'tradeoff':ab,ti OR 'best-worst scaling':ab,ti OR 'ahp':ab,ti OR 'analytic hierarchy':ab,ti OR 'swing weighting':ab,ti OR 'threshold technique':ab,ti OR 'risk benefit analysis':ab,ti	68917	10-Sep-18
#2	'treatment attribute*':ab,ti OR 'attributes':ab,ti OR 'preference*' OR 'trade off':ab,ti OR value:ab,ti OR 'patient decision making':ab,ti OR 'treatment satisfaction':ab,ti OR 'patient experience':ab,ti OR perception*:ab,ti OR attitude*:ab,ti OR 'patient preference':ab,ti	1370306	10-Sep-18
#1	'eczema'/exp OR 'atopic dermatitis'/exp	61560	10-Sep-18

[illegible]

Attributes and levels	Sample	MLE (SE)			
		MNL	HMNL	IMNL	RIMNL
1. Preferences					
Alternative Specific Constant					
Old treatment	Overall	1.458 (0.115)***	1.643 (0.139)***	1.392 (0.200)***	1.392 (0.182)***
Option A	Overall	-0.038 (0.037)	-0.042 (0.042)	-0.007 (0.061)	-0.007 (0.062)
Itch Reduction					
2 out of 10 (20%)	Overall	Reference	-	-	-
4 out of 10 (40%)	Overall	0.590 (0.060)***	0.671 (0.073)***	0.651 (0.101)***	0.651 (0.098)***
5 out of 10 (50%)	Overall	0.760 (0.058)***	0.858 (0.072)***	0.733 (0.100)***	0.733 (0.095)***
Skin Appearance					
2 out of 10 (20%)	Overall	0.214 (0.058)***	0.246 (0.066)***	0.243 (0.098)*	0.243 (0.096)*
4 out of 10 (40%)	Overall	0.481 (0.061)***	0.554 (0.072)***	0.606 (0.105)***	0.607 (0.100)***
1 out of 10 (10%)	Overall	Reference	-	-	-
Eye inflammation					
20 out of 100 (20%)	Overall	Reference	-	-	-
10 out of 100 (10%)	Overall	0.273 (0.048)***	0.317 (0.056)***	0.398 (0.080)***	0.398 (0.079)***
0 out of 100 (0%)	Overall	0.637 (0.056)***	0.723 (0.068)***	0.676 (0.092)***	0.677 (0.092)***
Serious Infections					

0 out of 100 (0%)	Overall	0.722 (0.056)***	0.800 (0.067)***	0.522 (0.093)***	0.523 (0.093)***
6 out of 100 (6%)	Overall	Reference	-	-	-
3 out of 100 (3%)	Overall	0.306 (0.050)***	0.339 (0.057)***	0.197 (0.083)*	0.197 (0.082)*
Speed of Onset					
2 weeks	Overall	Reference	-	-	-
1 week	Overall	0.010 (0.052)	0.011 (0.059)	0.019 (0.088)	0.019 (0.086)
2 days	Overall	0.178 (0.049)***	0.205 (0.057)***	0.217 (0.083)**	0.217 (0.082)**
Flare Management					
No	Overall	Reference	-	-	-
Yes	Overall	0.090 (0.039)*	0.109 (0.045)*	0.161 (0.065)*	0.161 (0.064)*
Long-term Disease Management					
Yes, without the possibility for pauses	Overall	Reference	-	-	-
Should not be used long-term	Overall	0.057 (0.054)	0.056 (0.062)	-0.012 (0.093)	-0.012 (0.091)
Yes, with the possibility for pauses	Overall	0.360 (0.048)***	0.399 (0.056)***	0.297 (0.080)***	0.297 (0.079)***
Administration					
Injection under the skin, every two weeks	Overall	Reference	-	-	-
Oral pill, once or twice daily	Overall	0.253 (0.047)***	0.294 (0.055)***	0.322 (0.078)***	0.322 (0.079)***
Check-ups					
Frequent check-ups required	Overall	Reference	-	-	-

Occasional check-ups required	Overall	0.242 (0.054)***	0.286 (0.063)***	0.328 (0.090)***	0.328 (0.091)***
No check-ups required	Overall	0.312 (0.052)***	0.366 (0.061)***	0.417 (0.086)***	0.417 (0.086)***
2. Interaction effects					
Alternative Specific Constant					
Old treatment	France	-	-	0.118 (0.311)	0.358 (0.257)
Old treatment	Spain	-	-	0.104 (0.336)	0.586 (0.298)*
Option A	France	-	-	-0.066 (0.094)	-0.077 (0.103)
Option A	Spain	-	-	-0.035 (0.089)	-0.048 (0.105)
Itch Reduction					
4 out of 10 (40%)	France	-	-	-0.150 (0.156)	-0.069 (0.154)
4 out of 10 (40%)	Spain	-	-	-0.057 (0.153)	0.134 (0.163)
5 out of 10 (50%)	France	-	-	0.066 (0.155)	0.194 (0.151)
5 out of 10 (50%)	Spain	-	-	0.024 (0.151)	0.268 (0.159)
Skin Appearance					
2 out of 10 (20%)	France	-	-	0.029 (0.149)	0.072 (0.155)
2 out of 10 (20%)	Spain	-	-	-0.099 (0.143)	-0.053 (0.156)
4 out of 10 (40%)	France	-	-	-0.200 (0.162)	-0.135 (0.157)
4 out of 10 (40%)	Spain	-	-	-0.194 (0.162)	-0.062 (0.165)
Eye inflammation					
10 out of 100 (10%)	France	-	-	-0.272 (0.121)*	-0.252 (0.132)

10 out of 100 (10%)	Spain	-	-	-0.127 (0.114)	-0.040 (0.133)
0 out of 100 (0%)	France	-	-	-0.086 (0.140)	0.007 (0.153)
0 out of 100 (0%)	Spain	-	-	-0.029 (0.132)	0.179 (0.154)
Serious Infections					
0 out of 100 (0%)	France	-	-	0.343 (0.142)*	0.480 (0.152)**
0 out of 100 (0%)	Spain	-	-	0.300 (0.136)*	0.564 (0.154)***
3 out of 100 (3%)	France	-	-	0.227 (0.127)	0.294 (0.134)*
3 out of 100 (3%)	Spain	-	-	0.131 (0.121)	0.238 (0.137)
Speed of Onset					
1 week	France	-	-	-0.064 (0.135)	-0.072 (0.143)
1 week	Spain	-	-	0.022 (0.129)	0.036 (0.142)
2 days	France	-	-	-0.043 (0.127)	-0.016 (0.136)
2 days	Spain	-	-	-0.080 (0.121)	-0.035 (0.137)
Flare Management					
Yes	France	-	-	-0.085 (0.098)	-0.073 (0.106)
Yes	Spain	-	-	-0.130 (0.093)	-0.120 (0.108)
Long-term Disease Management					
Should not be used long-term	France	-	-	0.033 (0.144)	0.036 (0.149)
Should not be used long-term	Spain	-	-	0.172 (0.136)	0.224 (0.153)

Yes, with the possibility for pauses	France	-	-	0.034 (0.123)	0.087 (0.129)
Yes, with the possibility for pauses	Spain	-	-	0.153 (0.121)	-0.299 (0.135)*
Administration					
Oral pill, once or twice daily	France	-	-	-0.042 (0.119)	0.002 (0.130)
Oral pill, once or twice daily	Spain	-	-	-0.152 (0.111)	-0.098 (0.132)
Check-ups					
Occasional check-ups required	France	-	-	-0.010 (0.138)	0.042 (0.148)
Occasional check-ups required	Spain	-	-	-0.223 (0.132)	-0.189 (0.153)
No check-ups required	France	-	-	-0.043 (0.130)	0.017 (0.140)
No check-ups required	Spain	-	-	-0.249 (0.124)*	-0.195 (0.144)
Country of residence					
France	Overall	-	-0.148 (0.084)	-	-
Spain	Overall	-	-0.280 (0.084)***	-	-
UK	Overall	-	Reference	-	-
4. Model information					
Parameters	-	18	20	54	54
LL	-	-4866.9	-4861.2	-4833.7	-4833.7
AIC	-	9769.8	9762.4	9775.4	9775.4
BIC	-	9886.6	9892.2	10125.7	10125.7
APR	-	8.30%	8.40%	8.20%	8.20%

Abbreviations: AIC, Akaike information criterion; APR, Adjusted McFadden Pseudo R^2; BIC, Bayesian information criterion; HMNL, heteroskedastic multinomial logit; IMNL, interacted multinomial logit; LL, log-likelihood; MLE, maximum likelihood estimate; MNL, multinomial logit; RIMNL, re-estimated interacted multinomial logit; SE, standard error

Significance: *** P-value < 0.001, ** P-value < .01, * P-value < .05

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Choice stability, n (%)				
Passed the test	260 (64)	71 (62)	94 (65)	95 (66)
Failed the test	144 (36)	43 (38)	51 (35)	50 (34)
Choice dominance [a], n (%)				
Passed the test	359 (89)	109 (96)	130 (90)	120 (83)
Failed the test	45 (11)	5 (4)	15 (10)	25 (17)
Serial non-participation [b], n (%)				
Never select the same option	384 (95)	108 (95)	136 (94)	140 (97)
Always select treatment A	0 (0)	0 (0)	0 (0)	0 (0)
Always select treatment B	1 (0)	1 (1)	0 (0)	0 (0)
Always select old treatment	19 (5)	5 (4)	9 (6)	5 (3)
Dominated decision making [c], n (%)				
Itch reduction	6 (1)	1 (1)	2 (1)	3 (2)
Skin appearance	1 (<1)	0 (0)	1 (1)	0 (0)
Eye inflammation	3 (1)	1 (1)	1 (1)	1 (1)
Serious infections	8 (2)	3 (3)	3 (2)	2 (1)
Speed of onset	0 (0)	0 (0)	0 (0)	0 (0)
Flare management	1 (<1)	0 (0)	0 (0)	1 (1)
Long-term disease management	0 (0)	0 (0)	0 (0)	0 (0)
Administration	21 (5)	8 (7)	8 (6)	5 (3)
Check-ups	2 (<1)	1 (1)	0 (0)	1 (1)
None	362 (90)	100 (88)	130 (90)	132 (91)
Response time for DCE choice task section only [d], n (%)				
Adequate	391 (97)	111 (97)	143 (99)	137 (95)
Inadequate	13 (3)	3 (3)	2 (1)	8 (5)
Time to complete DCE choice task section only, n (%)				
<5 min	236 (58)	64 (56)	93 (64)	79 (54)
5-10 min	123 (30)	38 (33)	38 (26)	47 (32)
10-15 min	28 (7)	7 (6)	9 (6)	12 (8)
15-20 min	4 (1)	1 (1)	2 (1)	1 (1)
>20 min	13 (3)	4 (4)	3 (2)	6 (4)

Abbreviations: DCE, discrete choice experiment

[a] A respondent was considered to have failed the test if they chose the inferior (dominated) option as their preferred treatment.

[b] A respondent was classified as a serial non-participant if they choose the same option for all 12 experimental choice tasks.

[c] Decision making was considered dominated when the respondent choses the best option on one attribute in all 12 experimental tasks.

[d] Response times in the lower 10% of the distribution were classed as too fast, and those in the upper 10% of the distribution as too slow. A participant was considered to have had an adequate response time if <80% of choice tasks were answered too fast or too slow.

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
Alternative specific constant	Old treatment	1.46 (0.12)***	[1.23; 1.69]
	Option A	-0.04 (0.04)	[-0.11; 0.03]
Itch reduction	2 out of 10 (20%)	Reference	-
	4 out of 10 (40%)	0.59 (0.06)***	[0.47; 0.71]
	5 out of 10 (50%)	0.76 (0.06)***	[0.65; 0.87]
Skin appearance	1 out of 10 (10%)	Reference	-
	2 out of 10 (20%)	0.21 (0.06)***	[0.10; 0.33]
	4 out of 10 (40%)	0.48 (0.06)***	[0.36; 0.60]
Eye inflammation	20 out of 100 (20%)	Reference	-
	10 out of 100 (10%)	0.27 (0.05)***	[0.18; 0.37]
	0 out of 100 (0%)	0.64 (0.06)***	[0.53; 0.75]
Serious infections	6 out of 100 (6%)	Reference	-
	3 out of 100 (3%)	0.31 (0.05)***	[0.21; 0.40]
	0 out of 100 (0%)	0.72 (0.06)***	[0.61; 0.83]
Speed of onset	2 weeks	Reference	-
	1 week	0.01 (0.05)	[-0.09; 0.11]
	2 days	0.18 (0.05)***	[0.08; 0.27]
Flare management	No	Reference	-
	Yes	0.09 (0.04)*	[0.01; 0.17]
Long-term disease management	Yes, without the possibility for pauses	Reference	-
	Should not be used long-term	0.06 (0.05)	[-0.05; 0.16]
	Yes, with the possibility for pauses	0.36 (0.05)***	[0.27; 0.45]
Administration	Injection under the skin, every 2 weeks	Reference	-
	Oral pill, once or twice daily	0.25 (0.05)***	[0.16; 0.35]
Check-ups	Frequent check-ups required	Reference	-
	Occasional check-ups required	0.24 (0.05)***	[0.14; 0.35]
	No check-ups required	0.31 (0.05)***	[0.21; 0.41]
Number of observations		4848	

Model log-likelihood at convergence	-4867
Adjusted pseudo R^2	0.08
Bayesian information criterion	9887

Abbreviations: CI, confidence interval; MLE, maximum likelihood estimate; SE, standard error

[illegible]

[illegible]

Abbreviations: MNL, multinomial logit; MXL, mixed logit

[illegible]

Abbreviations: MNL, multinomial logit; NL, nested logit

[illegible]

Abbreviation: CI, confidence interval

[illegible]

Abbreviation: CI, confidence interval

[illegible]

Abbreviation: CI, confidence interval

[illegible]

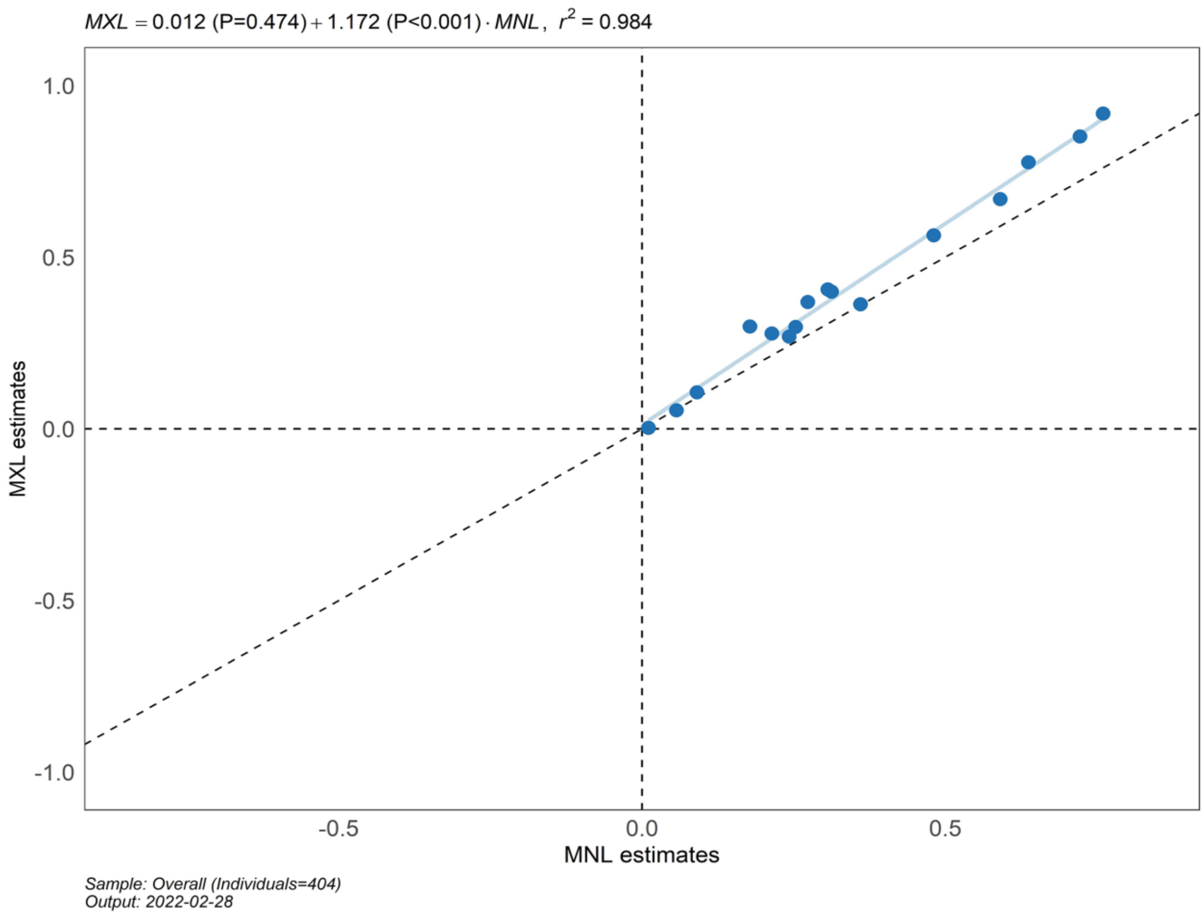
Abbreviation: CI, confidence interval

[illegible]

Abbreviation: CI, confidence interval

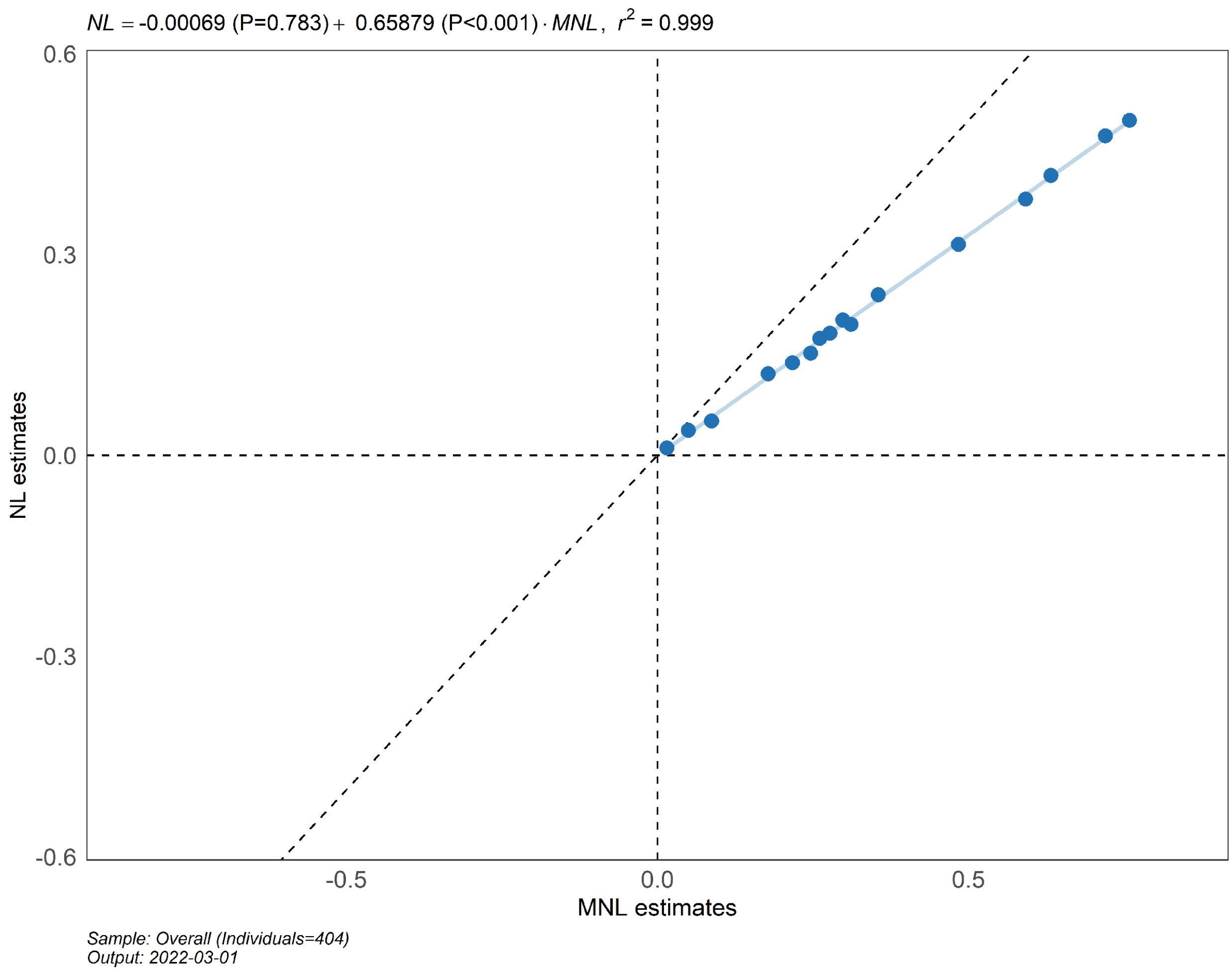
□ □□□□□□□□□□ □□□□□□□□r□8□M□□r□□□□□□□□□□□□r□□□□□□□□□□□□□□□

Abbreviation: CI, confidence interval



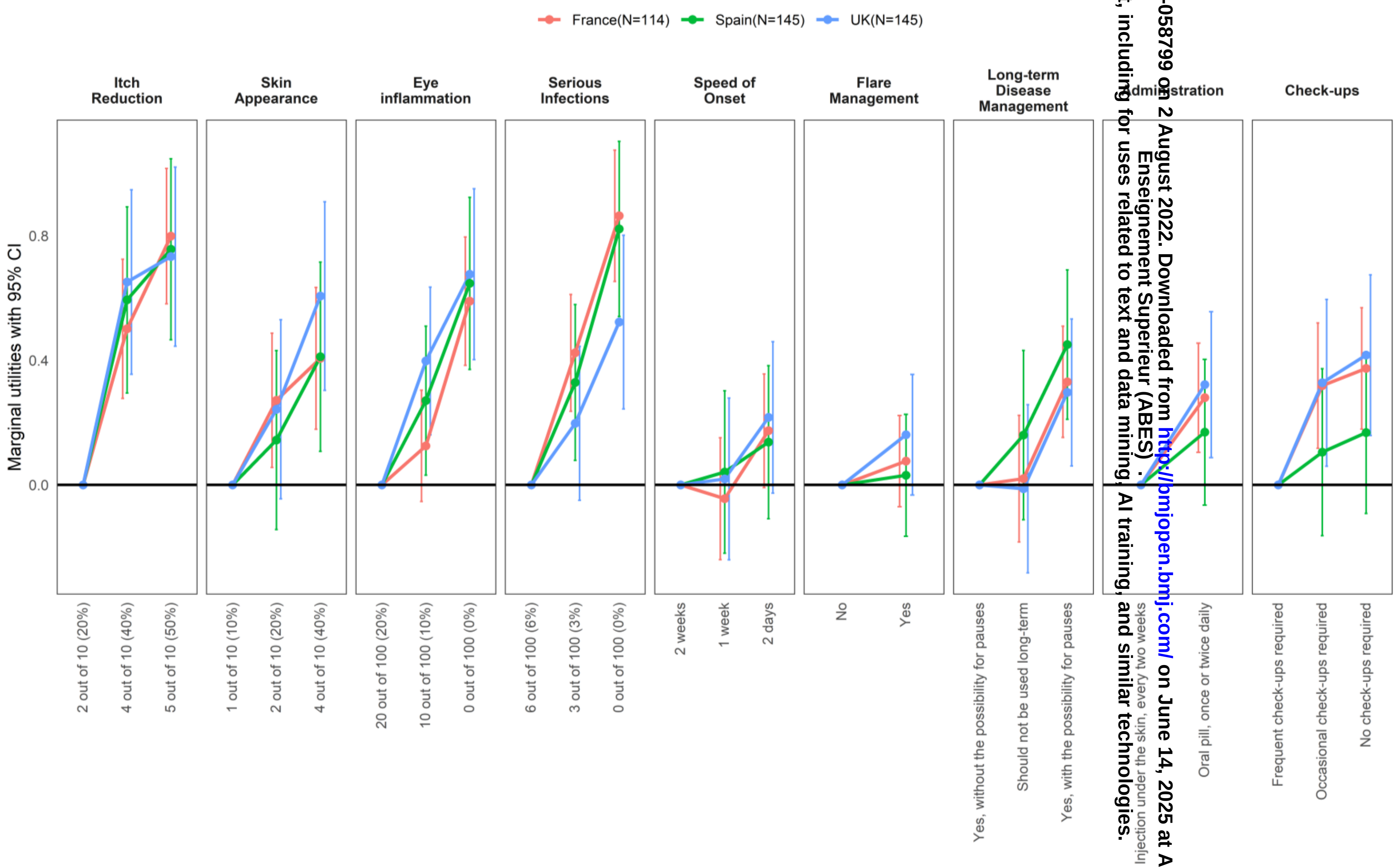
Supplemental Figure 1. Comparison of estimates between MXL and MNL models
Abbreviation: MNL, multinomial logit; MXL, mixed logit

iew only



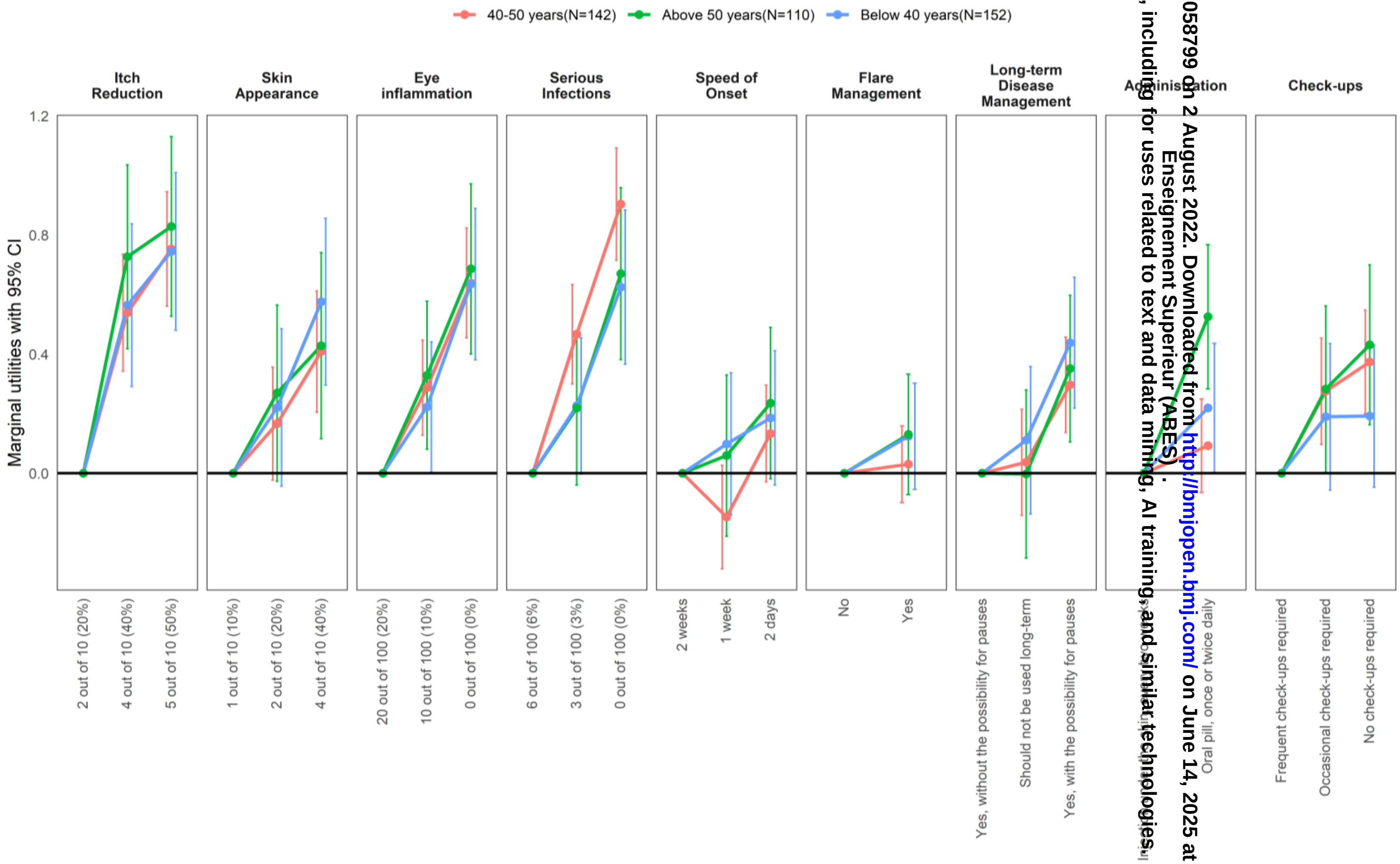
Supplemental Figure 2. Comparison of estimates between NL and MNL models

Abbreviations: MNL, multinomial logit; NL, nested logit



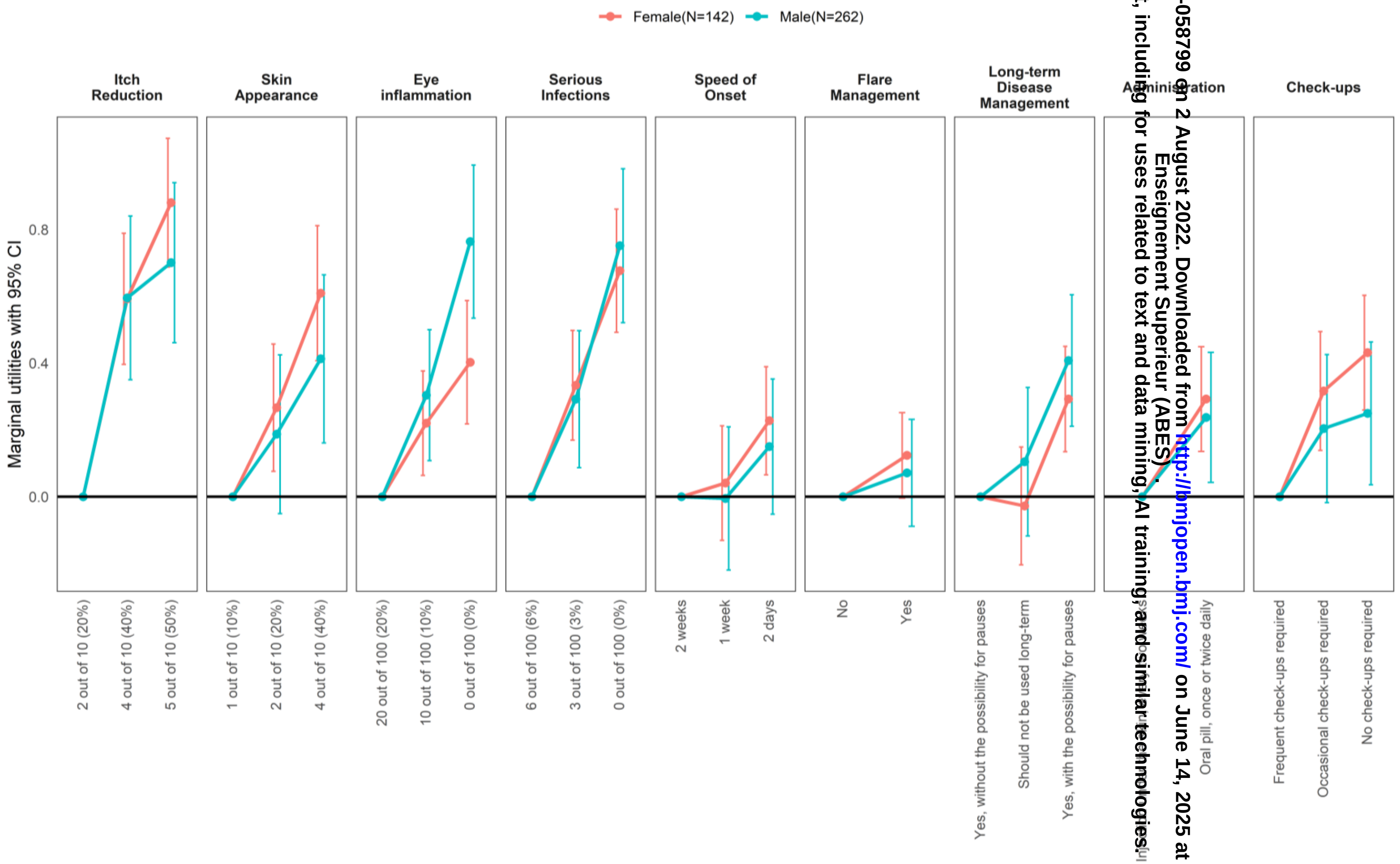
□□□□□□ □□□□□□□r□□□M□□r□□□□□□□□□□r□□

Abbreviation: CI, confidence interval



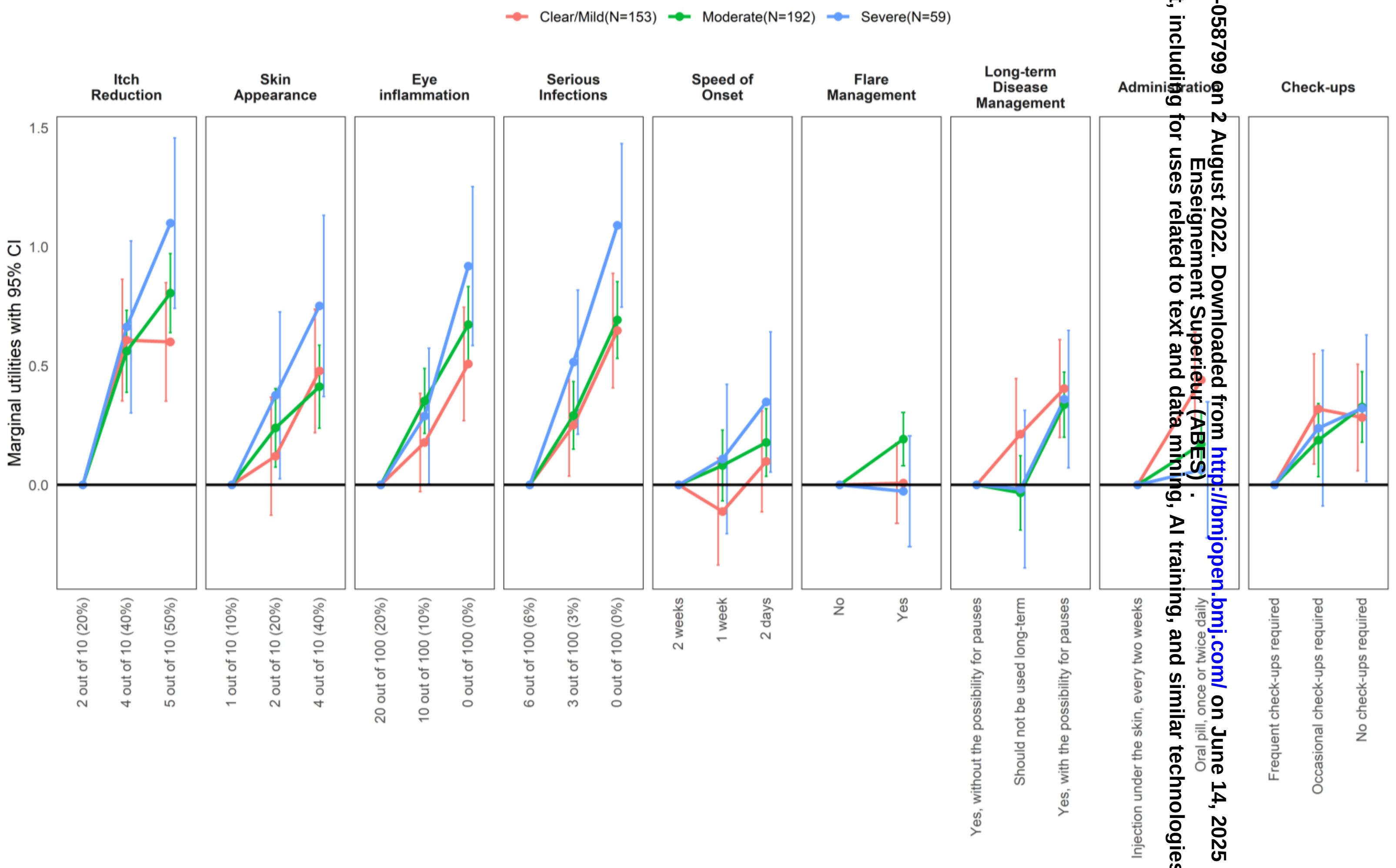
□□□□□□□ □□□□□□□□□□r□□□□M□□r□□□□□□□□□□□□□

Abbreviation: CI, confidence interval

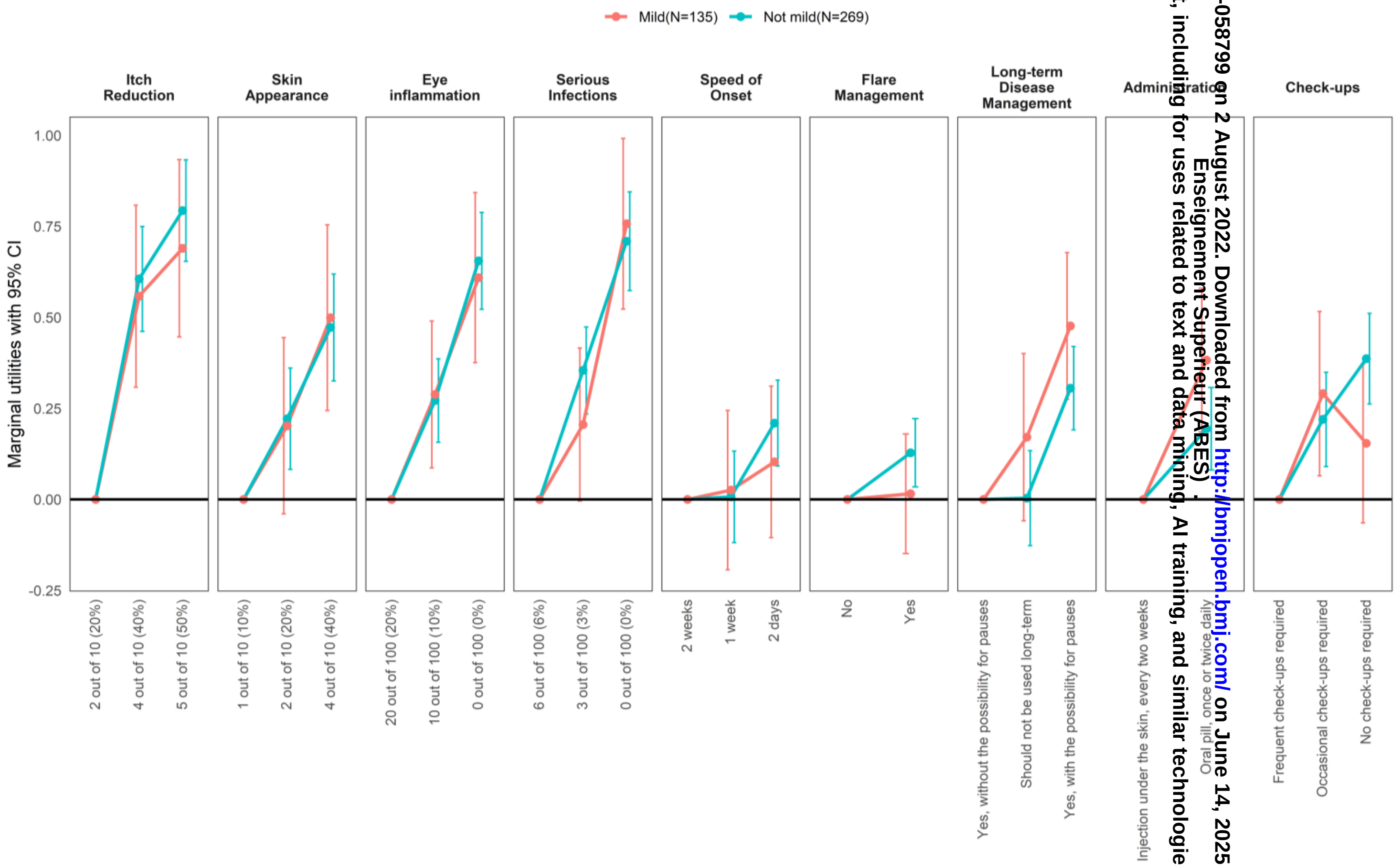


[illegible]

Abbreviation: CI, confidence interval

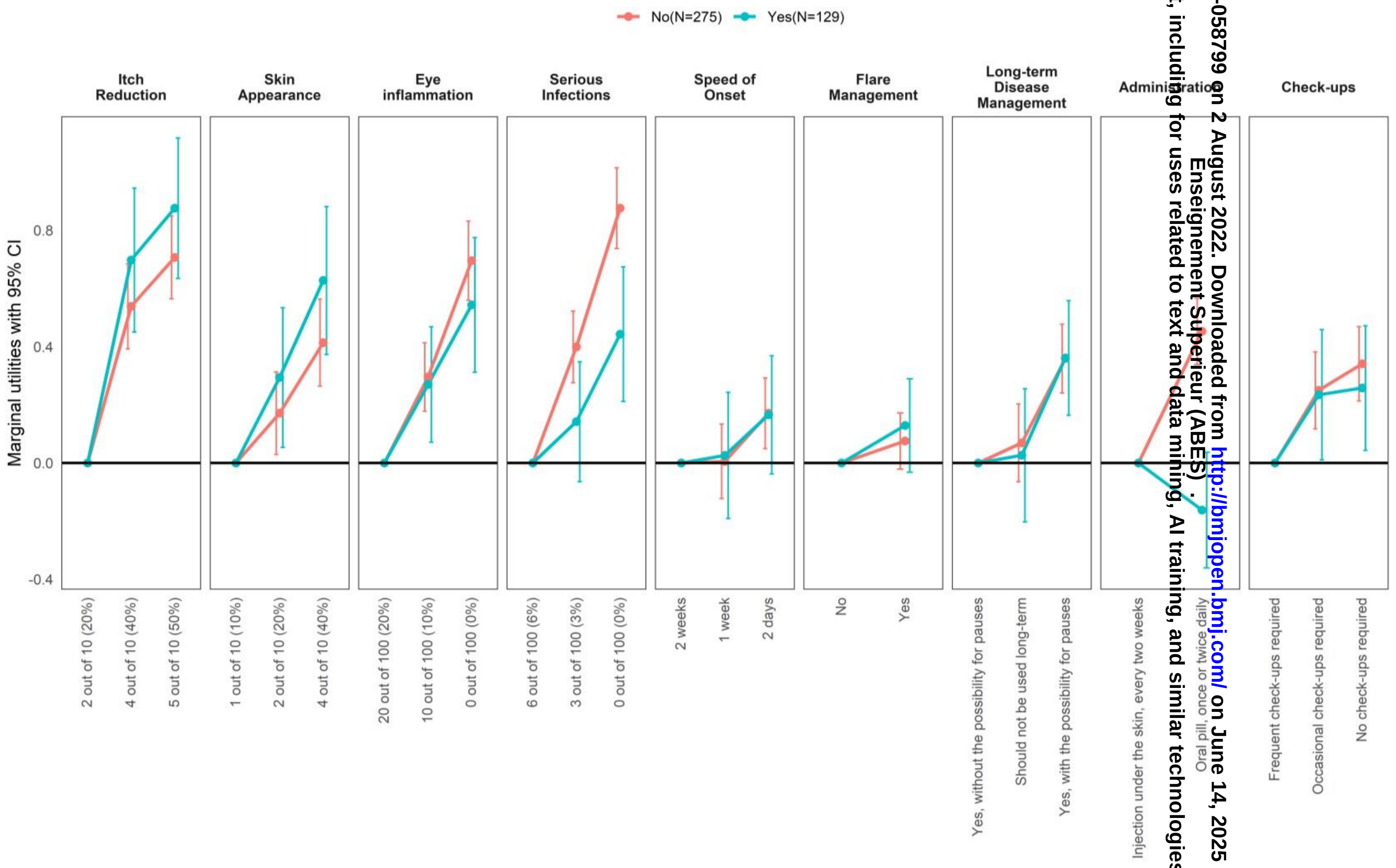


[illegible]
[illegible]
[illegible]
Abbreviation: CI, confidence interval



[illegible]

Abbreviation: CI, confidence interval



□□□□□□ □□□□□□□r□8□M□□r□□□□□□□□□□□□r□□□□□□□□□□□□□□

Abbreviation: CI, confidence interval

Checklist	Covered in manuscript	Page or section
1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?		
1.1 Were a well-defined research question and a testable hypothesis articulated?	Yes	p. 5
1.2 Was the study perspective described, and was the study placed in a particular decision-making or policy context?	Yes	p. 4-5
1.3 What is the rationale for using conjoint analysis to answer the research question?	Yes	p. 5
2. Was the choice of attributes and levels supported by evidence?		
2.1 Was attribute identification supported by evidence (literature reviews, focus groups, or other scientific methods)?	Yes (literature review)	p. 5
2.2 Was attribute selection justified and consistent with theory?	Yes	p. 5, 9-10
2.3 Was level selection for each attribute justified by the evidence and consistent with the study perspective and hypothesis?	Yes, via a literature review	p. 5
3. Was the construction of tasks appropriate?		
3.1 Was the number of attributes in each conjoint task justified (that is, full or partial profile)?	Yes, participants were surveyed for relevant attributes and no missing attributes were identified. Full choice profiles were used and patients had no issues with the number of attributes	p. 11
3.2 Was the number of profiles in each conjoint task justified?	Yes (3 profiles: A vs B vs old treatment)	p. 13

3.3 Was (should) an opt-out or a status-quo alternative (be) included?	Yes	p. 13
4. Was the choice of experimental design justified and evaluated?		
4.1 Was the choice of experimental design justified? Were alternative experimental designs considered?	Yes, D-efficient design assessed against good experimental design properties	p. 13
4.2 Were the properties of the experimental design evaluated?	Yes	p. 13
4.3 Was the number of conjoint tasks included in the data-collection instrument appropriate?	Yes, the number of tasks (questions) was 12 per person (36 in total)	p. 13
5. Were preferences elicited appropriately, given the research question?		
5.1 Was there sufficient motivation and explanation of conjoint tasks?	Yes	p. 13-14
5.2 Was an appropriate elicitation format (that is, rating, ranking, or choice) used? Did (should) the elicitation format allow for indifference?	Yes, the elicitation task was a choice task. The format did not allow indifference	p. 13-14
5.3 In addition to preference elicitation, did the conjoint tasks include other qualifying questions (for example, strength of preference, confidence in response, and other methods)?	Yes, validity assessments	p. 14
6. Was the data collection instrument designed appropriately?		
6.1 Was appropriate respondent information collected (such as sociodemographic, attitudinal, health history or status, and treatment experience)?	Yes	Table 2
6.2 Were the attributes and levels defined, and was any	Yes	Table 1

contextual information provided?		
6.3 Was the level of burden of the data-collection instrument appropriate? Were respondents encouraged and motivated?	Yes, this was assessed in cognitive pilot interviews and with data quality measures	p. 25
7. Was the data-collection plan appropriate?		
7.1 Was the sampling strategy justified (for example, sample size, stratification, and recruitment)?	Yes	p. 10
7.2 Was the mode of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)?	Yes	p. 5, 10
7.3 Were ethical considerations addressed (for example, recruitment, information and/or consent, compensation)?	Yes	p. 12
8. Were statistical analyses and model estimations appropriate?		
8.1 Were respondent characteristics examined and tested?	Yes	p. 17-19
8.2 Was the quality of the responses examined (for example, rationality, validity, reliability)?	Yes (validity and reliability)	p. 13-14, 19
8.3 Was model estimation conducted appropriately? Were issues of clustering and subgroups handled appropriately?	Yes	p. 16
9. Were the results and conclusions valid?		
9.1 Did study results reflect testable hypotheses and account for statistical uncertainty?	Yes, confidence intervals are presented	results
9.2 Were study conclusions supported by the evidence and	Yes	p. 23-26

compared with existing findings in the literature?		
9.3 Were study limitations and generalizability adequately discussed?	Yes	p. 25
10. Was the study presentation clear, concise, and complete?		
10.1 Was study importance and research context adequately motivated?	Yes	p. 4
10.2 Were the study data-collection instrument and methods described?	Yes	p. 13-16
10.3 Were the study implications clearly stated and understandable to a wide audience?	Yes	p. 23-26