SUPPLEMENTARY APPENDIX 1

Main analysis – interaction model

Likelihood ratio test *p*-value = 0.0804

Variable	Adjusted odds Ratio	95% CI for Adjusted Odds Ratio	*P* value
Secukinumab 150 mg	9.001	(0.07,1142.83)	0.3740
Secukinumab 300 mg	9.162	(0.10, 873.78)	0.3408
Patient age (years)	1.014	(0.97, 1.06)	0.4905
BMI (25kg/m^2/ 25kg/m^2≤ x≤30kg/m^2)	1.262	(0.49, 3.27)	0.6324
BMI ≥30kg/m^2	0.741	(0.25, 2.15)	0.5817
Smoking status : former	0.766	(0.23, 2.25)	0.6608
Smoking status : current	0.374	(0.12,1.12)	0.0797
Sex : male	0.366	(0.15, 0.89)	0.0261
C-reactive protein (mg/L)	1.004	(0.97, 1.04)	0.7819
Berlin MRI score for spine	0.924	(0.72, 1.19)	0.5456
Berlin MRI score for SIJ	0.964	(0.83, 1.12)	0.6362
Total back pain score (VAS)	0.999	(0.97, 1.03)	0.9690
Time since first axial signs and symptoms (years)	1.001	(0.96, 1.05)	0.9688
Number of swollen joints	0.993	(0.95. 1.04)	0.7493
Psoriatic nail dystrophy: yes	0.616	(0.26, 1.46)	0.2700
Radiographic evidence of juxta-articular bone formation: yes	1.056	(0.36, 3.13)	0.9219
Secukinumab 150 mg x Patient age (years)	0.992	(0.94, 1.05)	0.7754

Secukinumab 300 mg x Patient age (years)	0.973	(0.92, 1.03)	0.3250
Secukinumab 150 mg x BMI (25kg/m^2/ 25kg/m^2≤ x≤30kg/m^2)	0.411	(0.09, 1.85)	0.2473
Secukinumab 300 mg x BMI (25kg/m^2/ 25kg/m^2≤ x≤30kg/m^2)	0.576	(0.15, 2.16)	0.4133
Secukinumab 150 mg x BMI ≥30kg/m^2	0.301	(0.06, 1.47)	0.1384

As the likelihood ratio test *p*-value was < 0.2 (0.0804), so proceeded with interaction model.

Adjusted odds ratio: The Odds ratio effect comparing patients treated with secukinumab 150mg and patients treated with the placebo, while being identical in all the other predictors. Not interpreted in presence of interaction effects.

SUPPLEMENTARY APPENDIX 2

Statistical methods

The exploratory research question posed was to identify baseline predictors of ASAS20 response at week 12 in patients treated with secukinumab 150 mg or 300 mg. Placebo group patients were also be included in the analysis, to allow us to test whether the variables included in the model were independent predictors on treatment.

The sample size for this analysis has not been calculated to achieve a pre-defined level of power, caution must therefore be taken when interpreting the results. Primary analyses indicated a total of 254 ASAS20 responders versus 219 non-responders across all three groups at week 12. In the context of a logistic regression model, the category with the lower number of events limits the number of parameters that can be fitted to the model without yielding biased estimates of regression coefficients. The 'one-in-ten' rule of thumb allows for one parameter to be estimated for every 10 events. Since the current analysis was exploratory in nature, this rule of thumb can be relaxed, allowing approximately 219/5 = 43 parameters to be examined.

The research hypothesis was that the odds ratio associated with the effect of treatment on ASAS20 responder status at Week 12 will be different dependent on the following predictor variables: age, Body Mass Index (BMI), smoking status (tobacco and e-cigarettes), sex, C-Reactive Protein (CRP), Berlin MRI score for the spine and the sacroiliac joints (SIJ), total back pain score (BASDAI question 2), time since first axial signs and symptoms, number of swollen joints, psoriatic nail dystrophy and radiographic evidence of juxta-articular new bone formation.

A two-model approach was carried out as follows:

Main effects model 1: A logistic regression model was fitted to the data, which included a term for treatment group as well as terms for each of the predictor variables mentioned above. This was a no-interaction logit-additive model that assumes constancy of treatment odds ratios.

Interaction model 2: A second logistic regression model was fitted to the data, which included all terms from model 1 and also included interaction terms between treatment group and all other predictors.

The log-likelihood of the two models was compared using a chi-squared test to determine whether the effects of treatment depends on any of the other predictors in the model. If this test provided evidence

against the null hypothesis of no interaction at an alpha level of 20% (i.e. p-value <= 0.20) then we rejected model 1, and considered model 2 a better fit to the data and proceed with this model.

If the p-value for this comparison was >0.20, we failed to reject the null hypothesis of no interaction and considered model 1 a better fit to the data and proceed with this model. The less stringent alpha level threshold of 20% allowed easier identification of true independent predictor effects at the expense of an increase in false positive findings. Only the final model including coefficients and associated 95% confidence intervals was displayed. Additionally, a forest plot of the model coefficients was presented. In the event that interaction model was presented, an analysis of variance (ANOVA) table was also presented. Treatment contrast plots were also presented for predictor variables found to significantly interact with treatment. The present analyses do not consider the presence of three-way interactions.

In the event that a model does not converge, we ran diagnostics in order to assess potential causes. Potential solutions in the event of non-convergence may include the merging of factor levels, or the removal of variables from the model informed by their prior ranking of clinical importance.

The same two-model approach described above was followed in the subsets of patients with X-ray and HLA-B27 data, respectively, at baseline. Due to the reduced sample size in each of these subsets, the main effects model included terms for treatment and either X-ray or HLA-B27 status only, while the interaction models included these terms along with the interaction term between the two.

The effect of all variables was presented regardless of the magnitude of their individual p-values. The failure to replicate experimental results was of major concern.[1] This analysis was exploratory in nature and hypothesis generating. A prespecified statistical analysis plan documented clinical direction regarding which variables to assess, the restriction of the number of variables to examine and the analysis approach. Admittedly the analysis was under powered and the results should be treated with caution. Nail Dystrophy was a selected variable to evaluate for potential of differential treatment effects to support the nail-entheseal concept as one of the axial PsA phenotypes but was not the only variable examined

Reference:

1. Ioannidis JP. Why Most Published Research Findings Are False. *PLoS Med.* 2005;2(8):e124.