ORIGINAL RESEARCH ARTICLE

Moving Beyond Boundaries: Utilization of Longitudinal Exposure–Response Model for Bounded Outcome Score to Inform Decision Making in the Accelerated Drug Development Paradigm

Yeamin Huh[1] · Jessica Wojciechowski[1] · Vivek S. Purohit[1]

Abstract

Background and Objectives As drug development scientists strive to accelerate availability of therapies for patients, model-informed drug development (MIDD) plays an important role in contextualizing existing information and facilitating decision making. This paper describes an example of MIDD, where modeling and simulation informed decision making in the circumstance of a combined phase 2b and single pivotal study for ritlecitinib (JAK3/TEC family kinases inhibitor).

Methods Longitudinal exposure–response (ER) modeling was conducted to describe ritlecitinib efficacy in alopecia areata patients. The Severity of Alopecia Tool (SALT) score (a continuous bounded outcome [CBO] score [0–100]) was used as the efficacy response. The average concentration during the time interval between two adjacent SALT scores was used as the exposure metric driving efficacy.

Results The developed model well described the longitudinal SALT profile of ritlecitinib as well as the frequency of boundary data. The CBO model indicated tested doses in the phase 2b/3 clinical trial are in the ascending region of ER and contextualized a loading dose effect that impacted onset of efficacy without long-term benefit. It also identified disease severity as the only covariate impacting efficacy. The model-based simulation further informed impact of treatment interruption on the loss of efficacy in the absence of a dedicated treatment withdrawal study. Results indicated temporary treatment interruption $\leq$ 6 weeks is not expected to result in significant loss of efficacy.

Conclusion The CBO modeling approach and simulation supported the single pivotal trial strategy and guided dose selection in the accelerated drug development program of ritlecitinib, which can be applied to many indications where efficacy is measured on a bounded scale.

1 Introduction

Model-informed drug development (MIDD) is a powerful approach that uses quantitative modeling and simulation tools to support drug development and decision making. Since the MIDD concept was introduced in the 2004 Critical Path Initiative by the Food and Drug Administration (FDA) [1], application of MIDD has been expanded to include various stages of drug development, and the role of MIDD approaches has been increasingly recognized to help reduce the time and cost of drug development [2]. Especially in drug development programs with compressed timelines, the role of MIDD becomes critical for contextualizing existing and accruing new information, filling in knowledge gaps, and facilitating decision making.

Ritlecitinib is an orally bioavailable, small molecule drug that irreversibly inhibits Janus kinase 3 (JAK3) and tyrosine kinase expressed in hepatocellular carcinoma (TEC) family kinases [3]. Treatment with ritlecitinib is predicted to inhibit the inflammatory pathways mediated by interleukin (IL)-7, IL-15, and IL-21, which have been implicated in the pathogenic pathways of alopecia areata (AA) [4]. The proof of concept for ritlecitinib in the treatment of AA has been demonstrated in a phase 2a study (NCT02974868) [5]. Without a separate dose-ranging phase 2b study, the development program moved to a combined phase 2b/3 study that served both as a dose-ranging and single pivotal study. Therefore, understanding of dose/exposure–response (ER) relationships, dose and regimen optimization, and demonstration of long-term treatment benefit were all evaluated within a single trial.

✉ Yeamin Huh
yeamin.huh@pfizer.com

1 Worldwide Development, Research and Medical, Pfizer Inc, 280 Shennecossett Rd, Groton, CT 06340, USA

△ Adis

Key Points

To inform decision making in the ritlecitinib (an inhibitor of JAK3/TEC family kinases) program, where dose-ranging and confirmatory evidence of efficacy were generated in a single clinical trial, longitudinal exposure–response (ER) modeling was conducted to evaluate ritlecitinib efficacy in patients with alopecia areata, based on Severity of Alopecia Tool score data using a continuous bounded outcome (CBO) modeling approach.

The CBO modeling and simulation addressed contextualization of the ritlecitinib ER relationship and loading dose effect to guide dose selection, understanding of variability in efficacy due to covariates, and the impact of temporary treatment interruption on ritlecitinib efficacy.

This CBO modeling approach can be applied to many indications where efficacy is measured on a bounded scale to enable decision making in the accelerated drug development paradigm.

Here, an example of MIDD is provided, where modeling and simulation helped facilitate decision making in an accelerated drug development program. A longitudinal ER analysis was conducted to evaluate the efficacy of ritlecitinib in patients with AA, using the absolute Severity of Alopecia Tool (SALT) score data from all phase 2 and 3 studies (B7931005, B7981015, and B7981032). The SALT score is measured on a bounded scale (0–100) and therefore presents a technical challenge in model development due to the highly skewed distribution of the data, including a significant proportion of boundary data points [6]. This paper discusses an application of pharmacometric methodology to describe the continuous bounded outcome (CBO) data of SALT scores in the ritlecitinib program and utilization of CBO modeling and simulation for drug development decision making.

2 Methods

2.1 Study Design

This analysis was conducted on pooled data from three clinical studies (B7931005, B7981015, and B7981032). B7931005 was a phase 2a study to investigate ritlecitinib and brepocitinib in adults with AA with 50% or greater scalp hair loss (NCT02974868). The study consisted of three periods: a 24-week double-blind treatment period, an up to 48-week single-blind extension (SBE) period, and a 24-week

extension period. Only the initial 24-week period and SBE period of ritlecitinib and placebo data were included in the analysis. In the first period, ritlecitinib 200 mg once daily (QD) for 4 weeks followed by ritlecitinib 50 mg QD for 20 weeks and matching placebo were administered. The SBE period started after a 4-week drug holiday. In the SBE period, responders with a ≥ 30% change from baseline (CFB) received placebo until they met the re-treatment criterion (% CFB at week 24 − % CFB post week 24 > 30%) and then were re-treated with ritlecitinib 200 mg QD for 4 weeks, followed by 50 mg QD. Non-responders received another course of ritlecitinib 200 mg QD for 4 weeks, followed by 50 mg QD for 20 weeks.

B7981015 was a phase 2b/3 study to investigate the efficacy and safety of ritlecitinib in adults and adolescents with AA (≥ 12 years) with 50% or greater scalp hair loss (NCT03732807). The treatment period comprised a 24-week placebo-controlled period followed by a 24-week extension phase. Eligible participants were randomized in a 2:2:2:2:1:1:1 manner to the following treatments: (1) ritlecitinib 200 mg QD for 4 weeks, followed by ritlecitinib 50 mg QD for 44 weeks; (2) ritlecitinib 200 mg QD for 4 weeks, followed by ritlecitinib 30 mg QD for 44 weeks; (3) ritlecitinib 50 mg QD for 48 weeks; (4) ritlecitinib 30 mg QD for 48 weeks; (5) ritlecitinib 10 mg QD for 48 weeks; (6) placebo for 24 weeks → ritlecitinib 200 mg QD for 4 weeks, followed by ritlecitinib 50 mg QD for 20 weeks; and (7) placebo for 24 weeks → ritlecitinib 50 mg QD for 24 weeks.

B7981032 was a phase 3 long-term study to evaluate the safety and efficacy of ritlecitinib in adults and adolescents with AA (≥ 12 years) with 25% or greater scalp hair loss (NCT04006457). All eligible participants enrolled in study B7981032 after participation in either B7931005 or B7981015 received ritlecitinib 50 mg QD. All de novo participants received ritlecitinib 200 mg QD for 4 weeks, followed by ritlecitinib 50 mg QD. Since the B7981032 study was not completed at the time of this analysis, the only available datacut at the time of modeling analysis was included in the analysis dataset.

All the studies were conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice. The final protocols were approved by the institutional review boards.

2.2 Study Assessment

Sparse pharmacokinetics (PK) samples were collected in all studies. The exposure metric was derived from empirical Bayes estimates (EBEs) of the final population PK model [7]. Average drug concentration (C_{avg}) during the time interval between previous SALT record and the

current SALT record was calculated based on the patient's dosing diary, to be used as the exposure metric for ER analysis.

SALT is a quantitative assessment of AA severity that captures percentage hair loss and was consistently collected as an efficacy endpoint in all studies [8]. The SALT score assessment schedule for each study is available in Supplementary Table S1 (see the electronic supplementary material). The score ranges from 0 to 100, where 100 represents complete hair loss and 0 represents no hair loss.

2.3 ER Analysis

The pharmacometric methodology of CBO analysis proposed by Hutmacher et al. was adapted and modified [9]. The non-boundary data were first scaled between 0 and 1, and a transformation family of Aranda-Ordaz functions was applied to help normalize the skewed distribution of CBO data as follows:

$$
\begin{aligned}
z &= \frac{(100 - y) - 0}{100 - 0}, z \in (0, 1) \\
x &= h(y, \alpha) = \log\left[\frac{(1 - z)^{-\alpha} - 1}{\alpha}\right], \alpha \neq 0,
\end{aligned}
\tag{1}
$$

where y is the observation in the SALT scale and x is the observation in the transformed scale. The transformation factor α was estimated during model fitting.

A general nonlinear mixed-effects model was then constructed based on the transformed response:

$$
x|\eta = h(y, \alpha)|\eta = \mu(\eta) + \sigma \cdot \varepsilon,
\tag{2}
$$

where $x|\eta$ represents x conditioning on a vector of subject-specific random effects η, $\mu(\eta)$ is a conditional mean, σ is the residual error magnitude, and ε is the residual random error, which is assumed to be normally distributed with mean 0 and variance 1. The conditional mean $\mu(\eta)$ was modeled as:

$$
\mu(\eta) = f_b(\eta) - f_{\text{placebo}}(t) - f_{\text{drug}}(t),
\tag{3}
$$

where f_b is the baseline as a function of fixed effects (BASE) and random effects (η) [BASE + η], $f_{\text{placebo}}(t)$ is the placebo effect function, and $f_{\text{drug}}(t)$ is the drug effect function.

The placebo effect model was developed first, and then the drug effect model was added to the selected placebo effect model. For both placebo and drug effect functions, an indirect response model was considered using a latent variable approach to handle the delayed onset and offset of the response [10]:

$$
\begin{aligned}
f_{\text{placebo}}(t) &= \text{PBO}(t) - \frac{k_{\text{in1}}}{k_{\text{out1}}}, f_{\text{drug}}(t) = E(t) - \frac{k_{\text{in2}}}{k_{\text{out2}}} \\
\frac{d\text{PBO}(t)}{dt} &= k_{\text{in1}} \cdot \left[1 + I_{\text{PBO}} \cdot P_{\max}\right] - k_{\text{out1}} \cdot \text{PBO}(t) \\
\frac{dE(t)}{dt} &= k_{\text{in2}} \cdot \left[1 + \frac{E_{\max} \cdot C_{\text{avg}}(t)}{\text{EC}_{50} + C_{\text{avg}}(t)}\right] - k_{\text{out2}} \cdot E(t) \\
\text{PBO}(t = 0) &= \frac{k_{\text{in1}}}{k_{\text{out1}}}, E(t = 0) = \frac{k_{\text{in2}}}{k_{\text{out2}}},
\end{aligned}
\tag{4}
$$

where $\text{PBO}(t)$ and $E(t)$ are latent variables, I_{PBO} is an indicator variable that equals 1 if treatment was given and equals 0 otherwise, $P_{\max}$ is the maximum placebo effect, $E_{\max}$ is the maximum effect, k_{in} and k_{out} are rate constants determining a delay between placebo or drug treatment and response, and EC_{50} represents the C_{avg} yielding half of $E_{\max}$. k_{in} and k_{out} were parameterized as $\frac{\ln 2}{t_{1/2}}$, such that the rate constant of k_{in} and k_{out} can be viewed in the unit of time ($t_{1/2}$ was estimated instead of k_{in} or k_{out}).

Data on the boundaries, 0 or 100, were treated as censored data when constructing the likelihood. The interpretation of censoring is similar to that in PK and pharmacodynamic (PD) assays in that if a more sensitive way of measurement were available, values of 0 or 100 would not have been observed. The likelihoods of 0 observations being less than the minimum observable non-zero value and 100 observations being greater than the maximum observable non-hundred value were estimated, such that the likelihood for all the data is maximized in the model development.

Inter-individual variability (IIV) was incorporated in BASE, $E_{\max}$, and $k_{\text{in2}}/k_{\text{out2}}$ using a multiplicative exponential error model ($P_i = P_{\text{pop}} \bullet \exp(\eta_i)$ for ith individual) and $P_{\max}$ with an additive model ($P_i = P_{\text{pop}} + \eta_i$) to allow both disease worsening and improving.

The covariates tested were effects of sex, weight, age, race, region, disease severity (alopecia totalis [AT]/alopecia universalis [AU] status), and prior treatment on BASE; sex, weight, age, race, region, disease severity, prior treatment, AA duration since first diagnosis (DURF), duration of current AA episode (DURC), and baseline SALT score on $E_{\max}$; and age, weight, region, disease severity, prior treatment, DURF, and DURC on $k_{\text{in2}}/k_{\text{out2}}$. Stepwise covariate analysis was performed using both forward addition ($p < 0.05$) and backward elimination ($p < 0.001$).

Model adequacy was evaluated through changes in objective function value (OFV), visual inspection of diagnostic plots, precision of the parameter estimates, and decreases in IIV and residual variability. The final model

was further evaluated for its predictive performance by visual predictive check (VPC).

ER analyses were performed using NONMEM version 7.5.0. Exploratory analyses, diagnostic plots, post-processing of NONMEM output, and simulations were performed with R version 4.0.3. Perl-speaks-NONMEM (PsN) version 5.2.6 was used for performing sampling importance resampling (SIR). The NONMEM analyses were conducted using the Laplace estimation method with interaction and ADVAN13 with TOL = 6. The stochastic approximation expectation maximization (SAEM) method with importance sampling (IMP) was used for the estimation algorithm.

2.4 Clinical Trial Simulation to Understand Full Dose-Response Based on Limited Dose Range Data

To better understand translation of the established ER relationship in the transformed scale into the original SALT scale, clinical trial simulation was performed for various QD doses. One thousand datasets of longitudinal SALT scores for 130 participants, an average sample of the single arms of the B7981015 study ($n = 129$–132), were simulated for placebo, 30-, 50-, 100-, 200-, 400-, and 600-mg ritlecitinib QD dosage regimens. The demography of the 130 participants was assumed to be identical to that of the 50-mg arm of study B7981015. For each trial simulation, PK profiles were simulated first using EBEs for the 50-mg arm from the PK model, and SALT profiles were then simulated using a parameter set randomly drawn from a multivariate normal distribution using the population estimates and corresponding variance–covariance matrix of the estimates from the final CBO model. The SALT scores were simulated based on the transformed scale first and then back-transformed into the original SALT scale. Both PK concentrations and SALT scores including residual variability were simulated every week up to week 48, and placebo-adjusted responder rates for SALT ≤ 20 at week 24 and week 48 were collected for each simulated trial. The median and 95% prediction interval (PI) of the placebo-adjusted responder rates from 1000 simulated trials were summarized for each dosage regimen.

2.5 Clinical Trial Simulation to Understand Loading Dose Effect

The purpose of this simulation was to delineate the loading dose effect on clinical onset of response as well as overall outcome. The simulation scheme was identical to that of the previous simulation except the explored dosing regimens. One thousand datasets of longitudinal SALT scores for 130 participants were simulated for placebo, 30 mg QD, 50 mg

QD, and 200 mg QD for 4 weeks, followed by 30 mg QD and 200 mg QD for 4 weeks, followed by 50-mg QD dosage regimens. The SALT score was simulated for every week up to week 48 to correctly capture the onset of response. In study B7981015, the clinical onset was defined as the time when the responder rate for SALT ≤ 20 separated from placebo based on 95% confidence interval (CIs). Therefore, simulation results were summarized based on the placebo-adjusted SALT ≤ 20 responder rate, to evaluate when the lower bound of the 95% CI for placebo-adjusted responder rate became > 0. Since clinical onset based on this definition would be influenced by sample size, clinical trial simulation was conducted using the same sample size for the single arm of B7981015 ($n = 130$).

2.6 Clinical Trial Simulation to Evaluate Treatment Interruption Effect

Simulation was conducted to assess the impact of treatment interruption on loss of efficacy, SALT ≤ 20, after patients achieve a stable response. The study participants for the simulation were assumed to be identical to participants in study B7981015 ($n = 715$), and longitudinal SALT scores for each individual were simulated based on EBEs of final model parameters. In this simulation, all the participants were treated with 50 mg QD until week 96 to ensure the SALT response had reached a plateau, and SALT scores were simulated for every week up to week 144 (48 weeks after treatment withdrawal) to capture any changes of SALT score after treatment interruption. For responders, defined as participants achieving SALT ≤ 20 at week 96, time to lose SALT ≤ 20 response was collected and the proportions of responders losing SALT ≤ 20 response at various treatment interruptions with durations of 4, 6, 8, 10, 12, 14, 16, 24, 36, and 48 weeks were summarized.

3 Results

3.1 Observed Data Summary

The final analysis dataset included 11,857 observations from 1268 patients. The baseline characteristics are summarized in Table 1. Continuous covariates are further summarized by disease severity. The distribution of continuous covariates was relatively similar between patients with non-AT/AU (the "non-AT/AU group") and patients with AT/AU (the "AT/AU group") except the baseline SALT score and disease duration for the current episode. The AT/AU group had baseline SALT scores of 100 by definition of disease classification, whereas the non-AT/AU group had baseline scores < 100. The AT/AU group also had a relatively longer duration of

disease for the current episode compared with the non-AT/AU group.

3.2 Final ER Model

The final placebo model included three transit compartments in addition to the initial indirect response model in Eq. 4 to describe the delayed response. The entire model development history is available in Supplementary Table S2 (see the electronic supplementary material). Separate k_{in1} and k_{out1} estimation was not supported; hence, k_{in1} was assumed to be the same as k_{out1}. Since baseline SALT scores were different between non-AT/AU and AT/AU, disease severity (non-AT/AU vs AT/AU) was incorporated as a structural covariate on BASE.

The final drug effect model was an E_{max}/EC_{50} model with two transit compartments to describe the delayed response. Separate k_{in2} and k_{out2} estimation was not supported; hence, k_{in2} was assumed to be the same as k_{out2}. The model run separately estimating k_{in2} and k_{out2} was completed without producing a covariance step, indicating that the model may be overparameterized. Study B7981032 effect on BASE was incorporated to address differences in inclusion criteria between studies (baseline SALT $\geq$ 50 for B7931005 and B7981015 vs baseline SALT $\geq$ 25 for B7981032). AT/AU effects on P_{max} and k_{out2} were additionally included as structural covariates to better describe AT/AU response.

The selected base model with both placebo and drug effect includes IIV on BASE, P_{max}, E_{max}, or k_{out2} parameters. After the base model selection, covariate effects on the base ER model parameters were graphically examined with inter-individual random effect on BASE, E_{max}, or k_{out2} vs covariate plots. There were no strong correlations observed on any of the plots (Figs. S1–S3, see the electronic supplementary material). In line with this, the forward addition step of covariate analysis did not identify any important covariates. Therefore, the final base model was the selected final model. The schematic diagram of the final model is presented in Fig. 1.

Parameter estimates and SIR results of the final model are presented in Table 2. Asymptotic 95% CI and SIR 95% CI were very similar and demonstrated all the parameters were estimated with acceptable precision. Baseline values were different based on the disease severity and study, with a population mean of 1.92 (87.4% in original scale) for the non-AT/AU group in B7931005 and B7981015, 0.68 (67.3% in original scale) for the non-AT/AU group in B7981032, and 11.6 (100%) for the AT/AU group. The P_{max} parameter was not precisely estimated for the non-AT/AU group and therefore was fixed to zero, whereas a small positive placebo effect was estimated for the AT/AU group. The P_{max} estimate of 2.75 in the transformed scale

Table 1 Summary of baseline characteristics in the analysis data

Covariate	Summary
Protocol, n (%)	
B7931005	95 (7.5)
B7981015	715 (56.4)
B7981032	458 (36.1)
Sex, n (%)	
Female	805 (63.5)
Male	463 (36.5)
Race, n (%)	
African American	46 (3.6)
Asian	286 (22.6)
Other	47 (3.7)
White	889 (70.1)
Disease severity, n (%)	
Non-AT/AU	736 (58.0)
AT/AU	532 (42.0)
Prior pharmacological treatment for AA, n (%)	
No	22 (1.7)
Yes	499 (39.4)
Unknown	747 (58.9)
Age group, n (%)	
Adolescent (12 to < 18 years)	170 (13.4)
Adult ($\geq$ 18 years)	1098 (86.6)
Age, years	
Mean (SD)	
Total	33.8 (14.2)
Non-AT/AU	33.3 (14.3)
AT/AU	34.3 (14.0)
Median (Min, Max)	
Total	32 (12, 72)
AA	31 (12, 72)
AT/AU	33 (12, 72)
Weight, kg	
Mean (SD)	
Total	70.7 (17.6)
Non-AT/AU	70.2 (18.2)
AT/AU	71.4 (16.8)
Median (Min, Max)	
Total	68.4 (29.6, 200.0)
Non-AT/AU	67.8 (29.6, 200.0)
AT/AU	69.3 (35.0, 131.0)
Baseline SALT, %	
Mean (SD)	
Total	84.9 (21.4)
Non-AT/AU	74.0 (22.5)
AT/AU	100 (0)
Median (Min, Max)	
Total	98.6 (25.2, 100)

Table 1 (continued)

Covariate	Summary
Non-AT/AU	77.9 (25.2, 99.8)
AT/AU	100 (100, 100)
AA duration since first diagnosis, years	
Mean (SD)	
Total	10.0 (10.5)
Non-AT/AU	10.0 (10.3)
AT/AU	10.0 (10.8)
Median (Min, Max)	
Total	6.67 (0.04, 60.1)
Non-AT/AU	6.84 (0.04, 58.2)
AT/AU	6.43 (0.20, 60.1)
Duration of current AA episode, years	
Mean (SD)	
Total	3.22 (2.82)
Non-AT/AU	2.79 (2.74)
AT/AU	3.80 (2.82)
Median (Min, Max)	
Total	2.25 (0.02, 29.5)
Non-AT/AU	1.72 (0.02, 29.5)
AT/AU	3.08 (0.04, 10.7)

AA alopecia areata, *AT* alopecia totalis, *AU* alopecia universalis, *Max* maximum, *Min* minimum, *n* number of patients, *SALT* Severity of Alopecia Tool, *SD* standard deviation

did not induce any changes in SALT score for the AT/AU group, due to their large baseline value.

The E_{max} parameter estimate was 15.8 in the transformed scale for both non-AT/AU and AT/AU groups, which is translated into a complete recovery in SALT score (SALT score of 0 for non-AT/AU group and 0.33 for AT/AU group). The k_{out2} estimate in half-life was smaller for the AT/AU group (3.11 weeks for AT/AU vs 7.80 weeks for non-AT/AU, which translates to mean transit time ($\frac{\#\,of\,transit+1}{k_{out2}}$) of 13.5 weeks vs 33.8 weeks, respectively), but the quicker onset in the AT/AU group was offset by their higher baseline, which requires a large change in SALT score to reach meaningful clinical response. The EC_{50} estimate was a C_{avg} of 53.6 ng/mL, similar to the C_{avg} of 50 mg QD (52 ng/mL).

The model diagnostics are presented in Fig. 2. The transformation factor (α) was estimated to be 1.19 (Table 2), which adjusted the skewed distribution of the data to be normally distributed (Fig. 2A). The percentages of the boundary score were 5.39% for SALT = 0 and 24.9% for SALT = 100 in observed data, and 5.74% (95% CI 4.84–6.67) for SALT = 0 and 24.6% (95% CI 23.0–26.2) for SALT = 100 in the simulated data, indicating that the censored model was able to predict the percentage of boundary data adequately (Fig. 2B). Lastly, the model was developed for raw SALT score, but the VPC plot (Fig. 2C) showed generally good agreement between observed data and simulated data based on responder rate for SALT ≤ 20, which was the primary endpoint of the pivotal study.

3.3 Clinical Trial Simulation to Understand Full Dose-Response Based on Limited Dose Range Data

The simulation results for the placebo-adjusted SALT ≤ 20 responder rate at various QD doses are presented in Fig. 3. Simulation was conducted for a dosage range of 30–600 mg QD to illustrate where the tested dosages in B7981015 are located on the ER curve. Based on simulation results from the ER relationship established in the CBO model, higher efficacy is expected at doses greater than 50 mg, with dosages of 400 mg QD approaching the maximum efficacy.
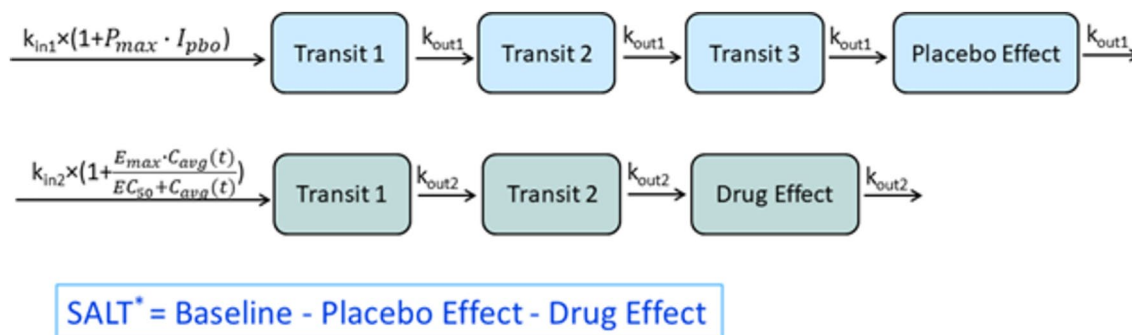

Fig. 1 Schematic diagram of longitudinal exposure–response model for SALT score. k_{in} and k_{out} are rate constants determining a delay between placebo or drug treatment and response, P_{max} is the maximum placebo effect, I_{PBO} is an indicator variable that = 1 if treatment was given and = 0 otherwise, E_{max} is the maximum drug effect, EC_{50} represents the PK exposure ($C_{avg}(t)$) yielding half of E_{max}, and SALT* represents SALT score in transformed scale. C_{avg} average drug concentration, *PK* pharmacokinetics, *SALT* Severity of Alopecia Tool

Table 2 Parameter estimates and SIR results of the final model

Parameter	Estimate[a]	RSE (%)	95% CI	SIR median	SIR 95% CI
BASE for non-AT/AU	1.92	4.04	(1.77, 2.07)	1.92	(1.77, 2.07)
BASE for AT/AU	11.6	2.83	(10.9, 12.2)	11.6	(10.9, 12.1)
B7981032 effect on BASE for non-AT/AU	− 0.645	5.76	(− 0.718, − 0.572)	− 0.644	(− 0.717, − 0.574)
P_{max} for non-AT/AU	0 fix	-	-	-	-
P_{max} for AT/AU	2.75	8.52	(2.29, 3.21)	2.77	(2.28, 3.25)
k_{out1} half-life (wk)[b]	1.93 fix	-	-	-	-
E_{max}	15.8	6.23	(13.9, 17.8)	15.8	(13.8, 17.6)
k_{out2} half-life (wk)	7.80	7.61	(6.64, 8.96)	7.71	(6.51, 9.01)
AT/AU effect on k_{out2} half-life (wk)	− 0.601	5.10	(− 0.661, − 0.541)	− 0.598	(− 0.660, − 0.542)
EC_{50} (ng/mL)	53.6	9.83	(43.3, 64.0)	53.8	(43.6, 65.4)
Residual error	1.18	1.51	(1.14, 1.21)	1.18	(1.14, 1.21)
Transformation factor	1.19	2.73	(1.12, 1.25)	1.19	(1.13, 1.25)
ω^2_{BASE}	0.225	6.54	(0.196, 0.254)	0.224	(0.196, 0.254)
ω^2_{Pmax}	7.62	8.27	(6.39, 8.86)	7.61	(6.42, 8.94)
Covariance—E_{max} and P_{max}	− 0.442	19.5	(− 0.611, − 0.273)	− 0.455	(− 0.618, − 0.299)
ω^2_{Emax}	0.754	12.2	(0.573, 0.934)	0.769	(0.596, 0.955)
Covariance—E_{max} and k_{out2}	0.549	12.2	(0.418, 0.681)	0.551	(0.426, 0.685)
ω^2_{Kout2}	0.973	9.14	(0.799, 1.15)	0.963	(0.791, 1.14)

AT alopecia totalis, *AU* alopecia universalis, *BASE* baseline as a function of fixed effects, C_{avg} average drug concentration, *CI* confidence interval, EC_{50} represents the PK exposure ($C_{avg}(t)$) yielding half of E_{max}, E_{max} maximum drug effect, k_{out} represents determinants of loss, *PK* pharmacokinetics, P_{max} maximum placebo effect, *RSE* relative standard error, *SIR* sampling importance resampling, ω^2 represents the variance estimate, *wk* week

[a]Parameter estimates are on a transformed scale

[b]k_{out1} half-life was fixed from the final placebo model. Condition number was 72. Shrinkages for intersubject variability were 16% for BASE, 33% for P_{max}, 37% for E_{max}, and 37% for k_{out2}. Epsilon shrinkage was 25%

3.4 Clinical Trial Simulation to Understand Loading Dose Effect

The simulation results are visually presented in Fig. 4 for loading (200/30 mg or 200/50 mg QD) and non-loading (30 mg or 50 mg QD) dosage regimens. For the comparison, the PK profiles of those regimens are also presented in the top panel of Fig. 4. As expected from the small half-life of ritlecitinib (~2 h), PK exposures quickly reached the new steady state after the loading dose was switched to the maintenance dose at week 4. In contrast, the loading dose effect stayed longer in clinical efficacy, thereby resulting in greater efficacy for the loading dose regimen than the non-loading dose regimen at week 24, when the primary endpoint was measured in study B7981015. The simulation results indicated that the loading dose achieved the clinical onset of SALT ≤ 20 seven weeks faster for the 30-mg QD group (6 vs 13 weeks) and 3 weeks faster for the 50-mg QD group (6 vs 9 weeks). However, the initial quicker response was not sustained and did not translate into higher response rates long term, such that the 95% CIs of responder rates were largely overlapped between loading and non-loading dose regimens at week 48.

3.5 Clinical Trial Simulation to Evaluate Treatment Interruption Effect

The analysis data included limited information from the SBE period of study B7931005 that could demonstrate SALT score change after treatment withdrawal and re-treatment. Randomly selected individual model fits for the SBE period data of study B7931005 are presented in Fig. 5A, demonstrating that the final model well described all the initial response to treatment, treatment withdrawal, and re-treatment data.

Based on this finding, SALT score profiles after treatment withdrawal were predicted using EBEs of the final model for all B7981015 study patients, as seen in the example profiles of Fig. 5B. The proportion of responders losing SALT ≤ 20 response after treatment withdrawal was further summarized according to the duration of treatment interruption (Table 3). Results indicated that the risk of losing SALT ≤ 20 response is very low (i.e., < 10%) with treatment interruption of ≤ 6 weeks, and 70% of patients are predicted to lose SALT ≤ 20 response with 48 weeks of treatment interruption.

A Data Distribution in Original Scale · Data Distribution in Transformed Scale
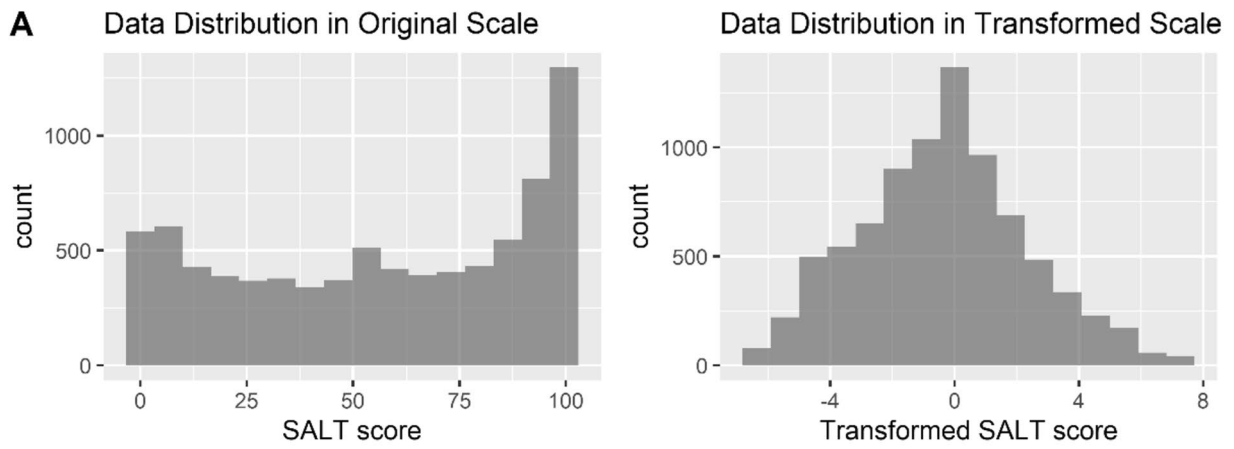
B

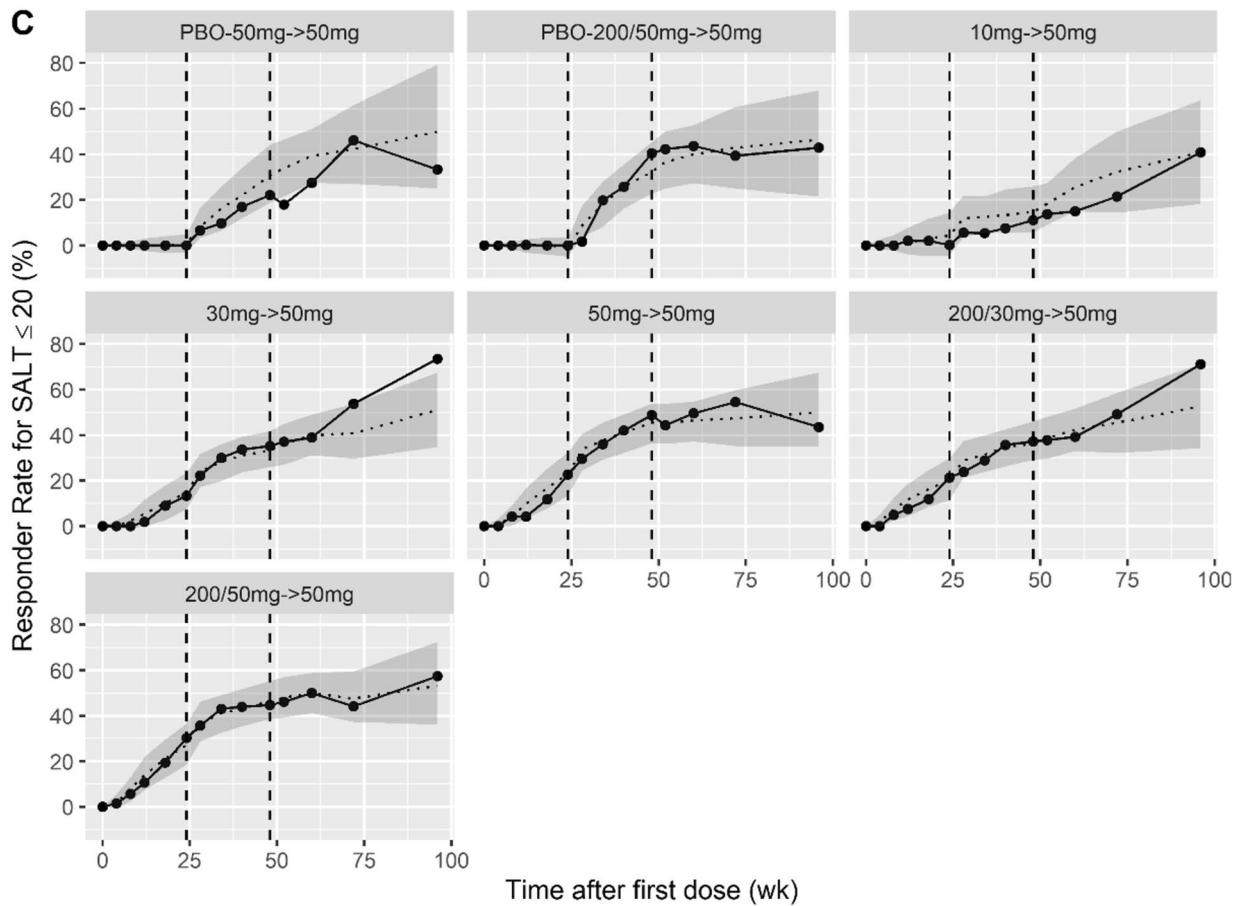| Observed proportion of 0 (%)	Observed proportion of 100 (%)	Simulated proportion of 0 with 95% CI (%)	Simulated proportion of 100 with 95% CI (%)
5.39	24.9	5.74 (4.87,6.69)	24.6 (23.0,26.2)

C

◀**Fig. 2** Diagnostic plots. **A** Distribution of observed data excluding boundary values in original vs transformed scale. **B** Percentage of bounded outcomes for observed vs simulated data. **C** Visual predictive check of the final model in terms of responder rate for SALT ≤ 20. The *black line* represents the observed placebo-adjusted responder rate up to week 24 and observed responder rate after week 24, and the *gray shaded region* represents the 95% prediction interval from the final model for the corresponding endpoint. The *dotted lines* represent median prediction from the final model for the corresponding endpoint. The *dashed lines* represent when the non-placebo-controlled period (week 24) and B7981032 study (week 48) started. *CI* confidence interval, *PBO* placebo, *SALT* Severity of Alopecia Tool, *wk* week

4 Discussion

The analysis in this paper characterized the longitudinal ER relationship of ritlecitinib efficacy in patients with AA using CBO modeling analysis. The description of CBO modeling strategy and how the established model informed decision making in the accelerated drug development program are discussed below.

4.1 Application of Pharmacometrics Methodology to Describe CBO Data

Bounded scales are commonly used to assess disease status and in turn used as clinical efficacy endpoints in clinical trials. Examples of such scales are the Psoriasis Area and Severity Index (0–72) [11], Disability Assessment for Dementia (0–100) [12], and Functional Assessment Questionnaire (0–30) [13]. Several pharmacometric modeling methodologies have been proposed to handle CBO data [9, 14–16]. In the current analysis, a censoring approach with transformation of non-boundary data proposed by Hutmacher et al. [9] was utilized due to its flexibility in handling boundary values, which consist of almost 30% of the analysis data. The developed model adequately adjusted the skewed distribution of the data and well described the longitudinal SALT score profile of ritlecitinib as well as the frequency of the boundary data.

The developed CBO model also utilized a latent variable indirect response modeling approach. The indirect response model is a widely used semi-mechanistic framework to link PD responses to PK exposure when there is a delay between PK exposure and PD responses. However, if the PD endpoint is a categorical or bounded outcome variable, the PD responses cannot be used as a raw scale for a dependent variable of PKPD modeling. To be applicable for such cases, the latent variable approach was developed, which links PD responses with drug exposure through an unobservable latent variable in the indirect response model framework [17]. In this modeling, the drug effect is assumed to be driven by a fluctuation of unobservable latent variable.
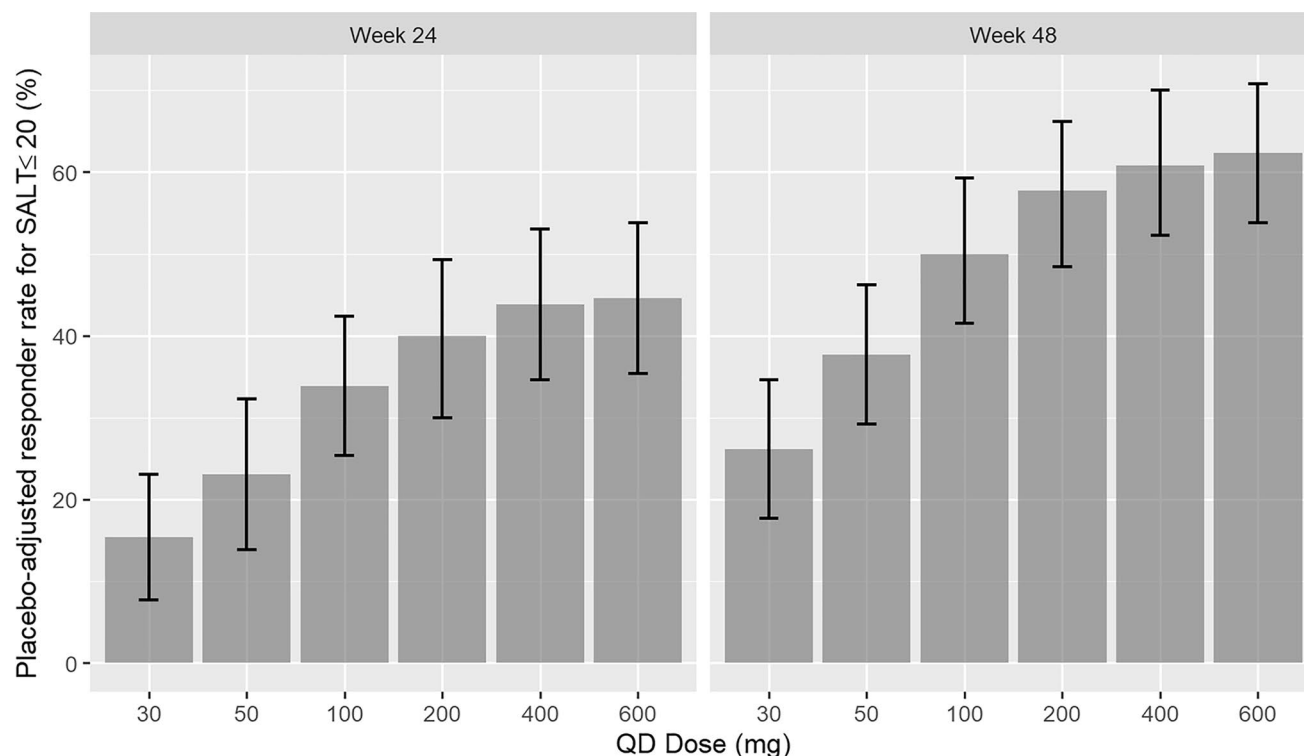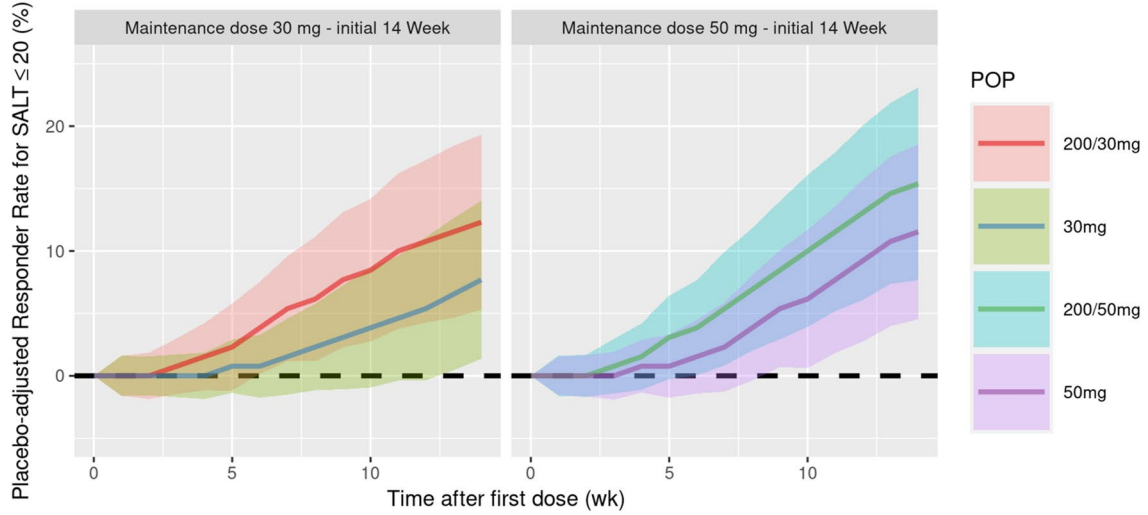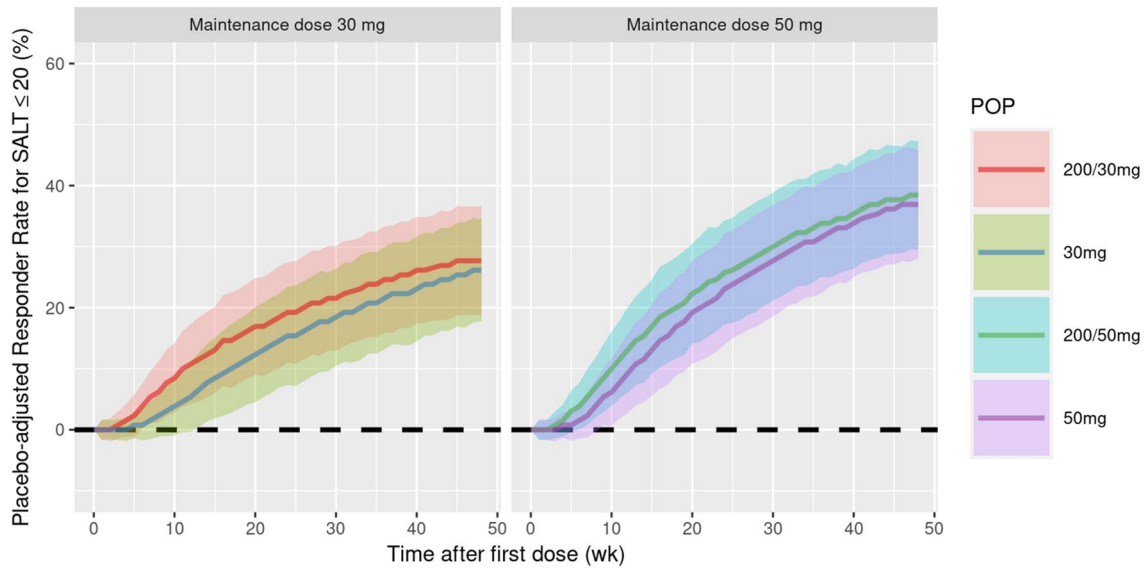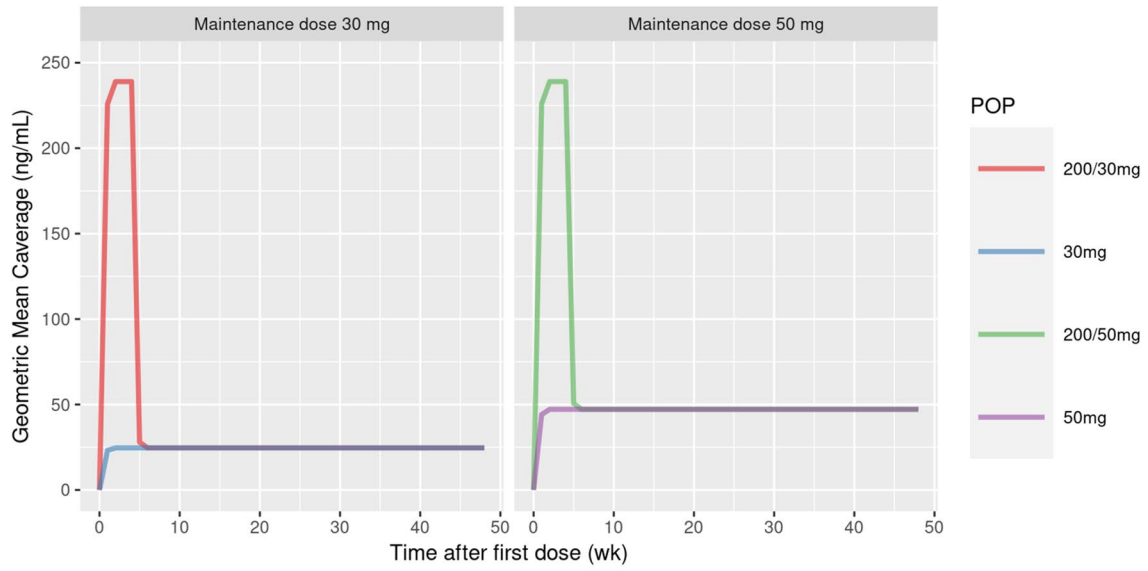

Fig. 3 Placebo-adjusted responder rate for SALT ≤ 20 at week 24 and week 48 for various QD doses. *Error bars* represent 95% prediction intervals. *QD* once a day, *SALT* Severity of Alopecia Tool

◄**Fig. 4** Evaluation of the loading dose effect on the clinical onset of efficacy when PD half-life is long. *Top figures* represent the PK profile of loading vs non-loading dosing regimens as geometric mean $C_{average}$, and *middle figures* represent the efficacy profile of loading vs non-loading regimens as median placebo-adjusted responder rate with 95% confidence intervals. *Bottom figures* represent the efficacy profile of loading vs non-loading regimens for the initial 24 weeks. $C_{average}$ average drug concentration, *PD* pharmacodynamic, *PK* pharmacokinetics, *POP* population, *wk* week

Ritlecitinib irreversibly inhibits JAK3/TEC family kinases [3], and the improvement in SALT score is the gradual outcome of their modulation on the downstream signaling pathways. Therefore, the latent variable can be considered as a summation of all of ritlecitinib's downstream effects, which eventually affect hair follicles to exhibit scalp hair growth.

The C_{avg} during the time interval between previous SALT score and the current SALT score was used as the PK exposure metric driving efficacy in the current analysis. The individual post-hoc concentration estimates could not be used due to the impractical long run time (> 150 h), and the summary measure of PK exposure was considered instead to complete the analysis in time. The C_{avg} has been widely used as a predictive PK exposure metric of PD response, unless the PD effect is acute (e.g., maximum concentration [C_{max}] is the key driver for drug-induced QT prolongation) or sustained inhibition of the target determines the PD effect (minimum concentration is the key driver of efficacy for antibiotic drugs) [18]. Given that ritlecitinib irreversibly inhibits JAK3/TEC family kinases [3] and the improvement in SALT score is the gradual outcome of their modulation on the downstream signaling pathways, it is reasonable to assume that the efficacy is not sensitive to the short-term fluctuation in exposure such as C_{max} and is rather influenced by average exposure over a period of time (i.e., C_{avg}) [19]. The selected PK exposure metric, C_{avg} during the time interval of the adjacent two SALT scores, has an advantage over steady-state C_{avg} calculated solely based on dose and clearance, which is often considered as the first choice of exposure metric in ER analyses. This metric can account for PK exposure fluctuation due to, e.g., treatment interruptions, and therefore can be considered as an acceptable surrogate for concentration when technical challenge is expected with using concentration for ER analyses.

### 4.2 Utilization of CBO Analysis for Drug Development Decision Making

Understanding of the ER relationship based on CBO analysis justified the single pivotal trial strategy. The CBO analysis results showed the EC_{50} estimate was 53.6 ng/mL, similar to the C_{avg} of 50 mg QD, suggesting that the tested dose range is in the ascending part of the entire dose-response relationship. By establishing EC_{50} estimates based on a totality of information from all the patient studies, the analysis demonstrated ritlecitinib exposure-dependent efficacy increase, which provides a causal confirmation of treatment effect of ritlecitinib in patients with AA. The causal evidence of effectiveness supports utilization of a single pivotal trial, in addition to empiric confirmation of treatment effect at a stringent α level of 0.00125 in study B7981015 [20, 21].

The CBO analysis guided dose selection by contextualizing study findings. The loading dose of 200 mg was evaluated in study B7981015 based on the hypothesis that maximal inhibition of the immunomodulatory pathways at initiation of treatment can accelerate clinical response, which can be maintained by subsequent lower maintenance doses. The loading dose is usually considered when the PK half-life of a drug is long and there is a need to accelerate the time to reach the target drug concentration. In inflammatory diseases, a loading dose was considered to address high inflammatory load in the active initial phase even though translation of high exposure into better clinical outcome has not been clearly established [22]. The CBO model-based evaluation based on study data indicates that the loading dose effect was on accelerating the onset of efficacy, with no long-term benefit, due to a combination of long PD half-life and ascending region of dose-response for the tested dose range. Both loading and non-loading dose regimens eventually reached their maintenance dose–level efficacy, which is in line with pharmacological first principles of a concentration-driven pharmacological response. Given that AA is not a life-threatening disease and need for both loading and maintenance doses creates complexity in prescribing, dispensing, and administration, contextualization of a loading dose using a CBO model aided selection of a non-loading dose regimen, 50 mg QD, as the proposed registrational dose.

The disease severity was the only important covariate identified in the current analysis, with no other covariates that could explain the variability in the efficacy. The DURC has been hypothesized to influence the treatment response, where patients whose current AA episode has lasted > 10 years are less likely to respond to treatment [23]. However, no obvious trend was observed in the DURC and the efficacy parameters. The study inclusion criteria for B7981015 and B7981032 restricted the duration for the current episode of hair loss to ≤ 10 years, which may have contributed to the above observation. Overall, ritlecitinib efficacy was lower in patients with AT/AU than in patients without AT/AU, consistent with what was reported in multiple other studies [24, 25]. The model explained this efficacy difference with large baseline difference between the non-AT/AU and AT/AU groups (1.92 vs 11.6) in the transformed scale, which cannot be captured in the SALT scale (87.4% vs 100%). The drug

A B7931005 Individual Model Fit
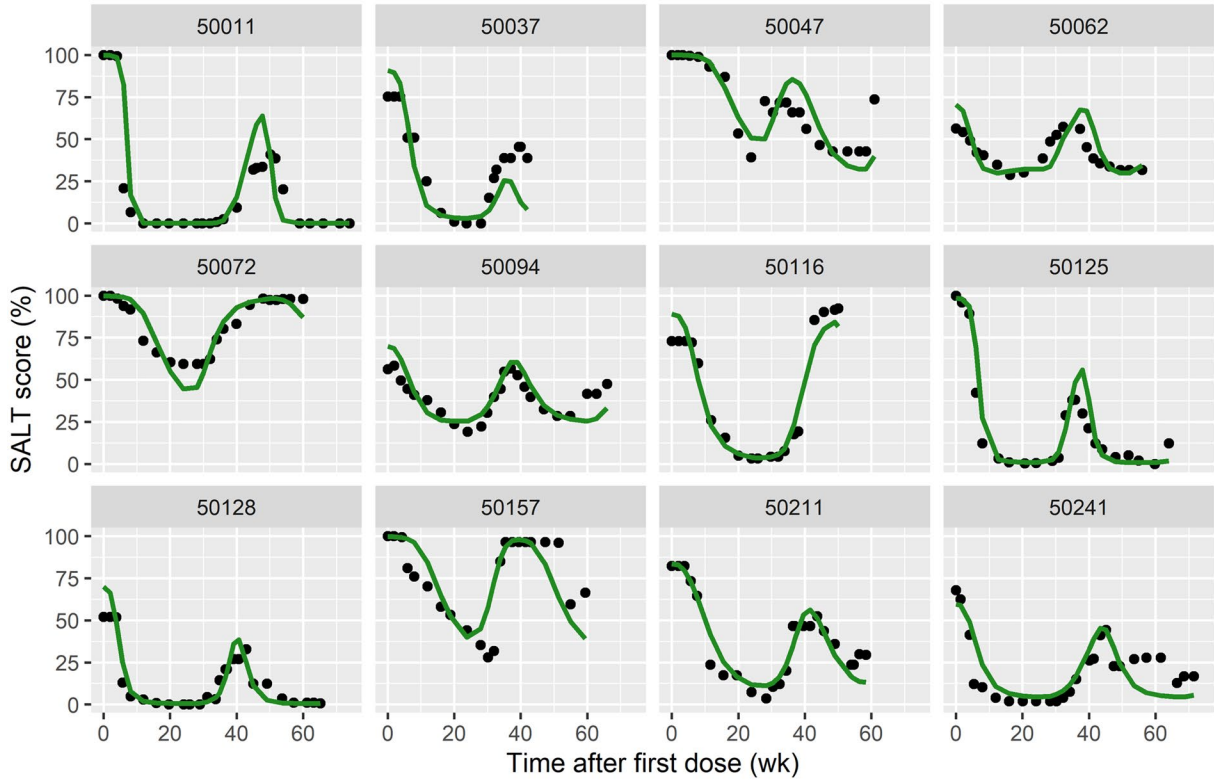
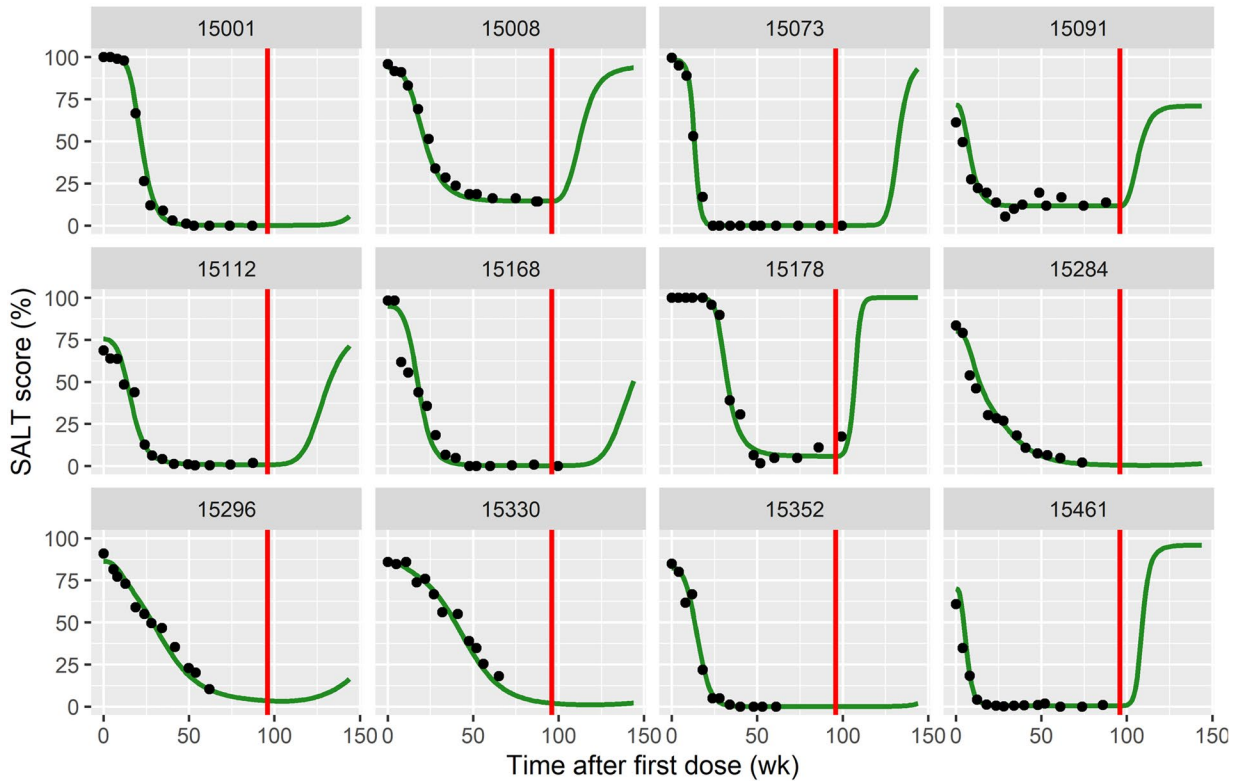**B** 50mg QD Dose Interruption Simulation Examples

◄Fig. 5 Randomly selected individual model fit to support treatment interruption simulation. **A** Individual model fit for study B7931005 where both drug onset and offset information is available. **B** Example of predicted individual SALT score profiles after treatment interruption. The *black circles* are observations, and the *green lines* are model predictions. The *red lines* represent the week 96 time point when treatment interruption starts. *QD* once a day, *SALT* Severity of Alopecia Tool, *wk* week

effect is additive to the baseline based on the transformed scale in the model, and therefore, the same PK exposure is expected to result in different efficacy between the non-AT/AU and AT/AU groups according to their baseline values. The maximum possible score on the SALT scale is limited to 100. However, there may be a pathophysiological mechanism in patients with AT/AU that is not adequately reflected in the SALT scale, which could drive the lower efficacy in the AT/AU group.

The CBO model-based simulation further informed the impact of treatment interruption on the loss of efficacy in the absence of a dedicated treatment withdrawal study. The results indicated that temporary treatment interruption for ≤ 6 weeks is not expected to result in a loss of SALT≤ 20 response in the majority of patients. However, treatment interruptions of a longer period may lead to significant loss of regrown scalp hair, as represented by almost 70% of SALT ≤ 20 responders losing SALT ≤ 20 response by 48 weeks after treatment withdrawal. Given challenges in a formal randomized treatment withdrawal study due to patients' psychological burden with new hair loss, the model-based simulation was considered as an alternative approach to provide a quantitative assessment of the impact of treatment interruption on treatment outcome.

Table 3 Predicted proportion of responders losing SALT ≤ 20 response for various dose interruption durations

Treatment interruption duration (≤ weeks)	Proportion of responders losing SALT ≤ 20 response (%)
4	3.6
6	8.5
8	15
10	20
12	25
14	30
16	33
24	44
36	59
48	70

SALT Severity of Alopecia Tool

5 Conclusion

The longitudinal ER relationship of ritlecitinib was successfully characterized using a CBO modeling approach. The developed CBO model was utilized to contextualize phase 3 study findings, support the single pivotal trial strategy, and guide dose selection in the ritlecitinib program. The concept of CBO modeling based on raw score and subsequent simulations can be applied to any indication where efficacy is measured on a bounded scale to facilitate decision making in the accelerated drug development paradigm.

Clinical Trial Registration Numbers NCT02974868, NCT03732807, and NCT04006457.

Declarations

Conflict of Interest/Competing Interests The authors are employees and shareholders of Pfizer Inc.

Ethics Approval/Consent to Participate Protocols from all studies were reviewed and approved by the institutional review boards or ethics committees of the participating institutions. The studies were conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects (Council for International Organizations of Medical Sciences 2002), the International Council for Harmonization Guideline for Good Clinical Practice, and the Declaration of Helsinki. All participants provided informed consent.

Consent for Publication/Availability of Data/Material Code Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author and Pfizer upon request, and subject to review. Subject to certain criteria, conditions, and exceptions, Pfizer may also provide access to the related individual de-identified participant data. See https://www.pfizer.com/science/clinical-trials/trial-data-and-results for more information.

References

1. Kim TH, Shin S, Shin BS. Model-based drug development: application of modeling and simulation in drug development. J Pharm Investig. 2018;48(4):431–41.

2. Madabushi R, Seo P, Zhao L, Tegenge M, Zhu H. Review: role of model-informed drug development approaches in the lifecycle of drug development and regulatory decision-making. Pharm Res. 2022;39(8):1669–80.

3. Xu H, Jesson MI, Seneviratne UI, Lin TH, Sharif MN, Xue L, et al. PF-06651600, a dual JAK3/TEC family kinase inhibitor. ACS Chem Biol. 2019;14(6):1235–42.

4. Suárez-Fariñas M, Ungar B, Noda S, Shroff A, Mansouri Y, Fuentes-Duculan J, et al. Alopecia areata profiling shows TH1, TH2, and IL-23 cytokine activation without parallel TH17/TH22 skewing. J Allergy Clin Immunol. 2015;136(5):1277–87.

5. King B, Guttman-Yassky E, Peeva E, Banerjee A, Sinclair R, Pavel AB, et al. A phase 2a randomized, placebo-controlled study to evaluate the efficacy and safety of the oral Janus kinase inhibitors ritlecitinib and brepocitinib in alopecia areata: 24-week results. J Am Acad Dermatol. 2021;85(2):379–87.

6. Hu C. On the comparison of methods in analyzing bounded outcome score data. AAPS J. 2019;21(6):102.

7. Wojciechowski J, Purohit VS, Huh Y, Banfield C, Nicholas T. Evolution of ritlecitinib population pharmacokinetic models during clinical drug development. Clin Pharmacokinet. 2023;62(12):1765–79.

8. Olsen EA. Investigative guidelines for alopecia areata. Dermatol Ther. 2011;24(3):311–9.

9. Hutmacher MM, French JL, Krishnaswami S, Menon S. Estimating transformations for repeated measures modeling of continuous bounded outcome data. Stat Med. 2011;30(9):935–49.

10. Hu C, Xu Z, Mendelsohn AM, Zhou H. Latent variable indirect response modeling of categorical endpoints representing change from baseline. J Pharmacokinet Pharmacodyn. 2013;40(1):81–91.

11. Mattei PL, Corey KC, Kimball AB. Psoriasis Area Severity Index (PASI) and the Dermatology Life Quality Index (DLQI): the correlation between disease severity and psychological burden in patients treated with biological therapies. J Eur Acad Dermatol Venereol. 2014;28(3):333–7.

12. Sánchez-Pérez A, López-Roig S, Pampliega Pérez A, Peral Gómez P, Pastor M, Hurtado-Pomares M. Translation and adaptation of the Disability Assessment for Dementia scale in the Spanish population. Med Clin (Barc). 2017;149(6):248–52.

13. Ito K, Hutmacher MM, Corrigan BW. Modeling of Functional Assessment Questionnaire (FAQ) as continuous bounded data from the ADNI database. J Pharmacokinet Pharmacodyn. 2012;39(6):601–18.

14. Xu XS, Samtani M, Yuan M, Nandy P. Modeling of bounded outcome scores with data on the boundaries: application to disability assessment for dementia scores in Alzheimer's disease. AAPS J. 2014;16(6):1271–81.

15. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods. 2006;11(1):54–71.

16. Wellhagen GJ, Kjellsson MC, Karlsson MO. A bounded integer model for rating and composite scale data. AAPS J. 2019;21(4):74.

17. Hutmacher MM, Krishnaswami S, Kowalski KG. Exposure-response modeling using latent variables for the efficacy of a JAK3 inhibitor administered to rheumatoid arthritis patients. J Pharmacokinet Pharmacodyn. 2008;35(2):139–57.

18. Maurer TS, Smith D, Beaumont K, Di L. Dose predictions for drug design. J Med Chem. 2020;63(12):6423–35.

19. Lamba M, Hutmacher MM, Furst DE, Dikranian A, Dowty ME, Conrado D, et al. Model-informed development and registration of a once-daily regimen of extended-release tofacitinib. Clin Pharmacol Ther. 2017;101(6):745–53.

20. Peck CC, Rubin DB, Sheiner LB. Hypothesis: a single clinical trial plus causal evidence of effectiveness is sufficient for drug approval. Clin Pharmacol Ther. 2003;73(6):481–90.

21. Fisher LD. One large, well-designed, multicenter study as an alternative to the usual FDA paradigm. Drug Inform J. 1999;33(1):265–71.

22. Geurts-Voerman GE, Verhoef LM, van den Bemt BJF, den Broeder AA. The pharmacological and clinical aspects behind dose loading of biological disease modifying anti-rheumatic drugs (bDMARDs) in auto-immune rheumatic diseases (AIRDs): rationale and systematic narrative review of clinical evidence. BMC Rheumatol. 2020;4:37.

23. Liu LY, Craiglow BG, Dai F, King BA. Tofacitinib for the treatment of severe alopecia areata and variants: a study of 90 patients. J Am Acad Dermatol. 2017;76(1):22–8.

24. Tosti A, Bellavista S, Iorizzo M. Alopecia areata: a long term follow-up study of 191 patients. J Am Acad Dermatol. 2006;55(3):438–41.

25. Burroway B, Griggs J, Tosti A. Alopecia totalis and universalis long-term outcomes: a review. J Eur Acad Dermatol Venereol. 2020;34(4):709–15.