

Improvement in latent variable indirect response modeling of multiple categorical clinical endpoints: application to modeling of guselkumab treatment effects in psoriatic patients

Chuanpu Hu¹ · Bruce Randazzo² · Amarnath Sharma¹ · Honghui Zhou¹

Received: 22 March 2017 / Accepted: 9 June 2017 / Published online: 20 June 2017
© Springer Science+Business Media, LLC 2017

Abstract Exposure–response modeling plays an important role in optimizing dose and dosing regimens during clinical drug development. The modeling of multiple endpoints is made possible in part by recent progress in latent variable indirect response (IDR) modeling for ordered categorical endpoints. This manuscript aims to investigate the level of improvement achievable by jointly modeling two such endpoints in the latent variable IDR modeling framework through the sharing of model parameters. This is illustrated with an application to the exposure–response of guselkumab, a human IgG1 monoclonal antibody in clinical development that blocks IL-23. A Phase 2b study was conducted in 238 patients with psoriasis for which disease severity was assessed using Psoriasis Area and Severity Index (PASI) and Physician’s Global Assessment (PGA) scores. A latent variable Type I IDR model was developed to evaluate the therapeutic effect of guselkumab dosing on 75, 90 and 100% improvement of PASI scores from baseline and PGA scores, with placebo effect empirically modeled. The results showed that the joint model is able to describe the observed data better with fewer parameters

compared with the common approach of separately modeling the endpoints.

Keywords Population pharmacokinetic/pharmacodynamic modeling · NONMEM · Clinical drug development

Introduction

Exposure–response (E–R) modeling of clinical endpoints is important in drug development for facilitating informative dosing selection. A widely used class of E–R models includes the indirect response (IDR) models [1]. These models are most often used to describe pharmacodynamics endpoints with the mechanism of delay. However, many clinical trial endpoints are based on disease scores that are not physiological variables. For example, two types of commonly used efficacy measurements in psoriasis are the Psoriasis Area and Severity Index (PASI) score, ranged 0–72 with 0.1 increments, and Physician’s Global Assessment (PGA) scores, a 6-point scale measuring disease severity (0 = cleared, 1 = minimal, 2 = mild, 3 = moderate, 4 = marked, and 5 = severe) [2]. Clinical trial endpoints typically include proportions of patients achieving various criteria including the following: PASI 75, PASI 90, and PASI 100, representing 75, 90, and 100% improvement in PASI score from baseline, respectively; PGA score = 0, or PGA score ≤ 1. Applications of IDR models to categorical clinical endpoints have emerged in the last decade via the latent variable approach [3].

E–R modeling using PASI scores as continuous variables have been conducted before [4–7], however evidence of their ability to accurately predict the PASI criteria appears lacking. PGA scores are most effectively analyzed as an ordered clinical endpoint [8]. The PASI criteria (Pc),

Electronic supplementary material The online version of this article (doi:10.1007/s10928-017-9531-3) contains supplementary material, which is available to authorized users.

✉ Chuanpu Hu
CHu25@its.jnj.com

¹ Global Clinical Pharmacology, Janssen Research & Development, LLC, 1400 McKean Road, PO Box 776, Spring House, PA 19477, USA

² Clinical Immunology, Janssen Research & Development, LLC, 1400 McKean Road, PO Box 776, Spring House, PA 19477, USA

namely PASI 75, PASI 90, and PASI 100, can be combined into one ordered categorical endpoint P_c having four possible outcomes: $P_c = 0$, if achieving PASI 100; $P_c = 1$, if achieving PASI 90 but not PASI 100; $P_c = 2$, if achieving PASI 75 but not PASI 90; and $P_c = 3$, if not achieving PASI 75. Conceivably, because P_c and PGA both measure the same disease activity, their E–R characteristics should be similar, thus jointly modeling them should allow better integration of information. When modeling a continuous and a categorical endpoint measuring the same disease activity, Hu et al. [9] recently showed that joint modeling of endpoints could be more parsimonious and yet better describe the individual endpoints, compared with separately modeling the endpoints. This report addresses whether this type of improvement is possible when modeling two categorical endpoints. The answer is by no means a priori clear, due to the difference in nature between continuous and categorical endpoints. Additionally, with two categorical endpoints, the optimal choice of a common latent variable may not be obvious.

Psoriasis is a chronic immune-mediated skin disorder [10–12]. Interleukin (IL)-12 and -23 have been implicated in the pathogenesis of psoriasis [13–15], and agents that block IL-12 and IL-23 have demonstrated efficacy in the treatment of moderate-to-severe plaque psoriasis [14]. Guselkumab is a monoclonal antibody that specifically blocks IL-23. Using data from a Phase 2b dose-ranging clinical trial of guselkumab in psoriasis [2], this report investigates the potential benefit of jointly modeling two categorical clinical endpoints and the source of potential improvement, and discusses the impact of latent variable choice.

Methods

Study design

A Phase 2, randomized, double-blind, parallel, dose-ranging study was conducted in patients with moderate-to-severe plaque psoriasis. Approximately 240 patients were randomly assigned to treatment with subcutaneous injection of guselkumab 5, 50, or 200 mg at Weeks 0 and 4 followed by every-12-week (q12w) dosing, or 15 or 100 mg, with q8w dosing, or placebo. The placebo group crossed-over to the 100 mg q8w dosing at Week 16. The last dose was given at Week 40. Data from the Week 40 database lock was used for analysis. The detailed study design has been previously published [2].

Guselkumab serum concentration, antibodies, PASI, and PGA measurements

Serum samples of guselkumab, along with PASI and PGA scores, were collected q4w during Weeks 0–40. At visits when study patients received the study agent, blood samples were collected prior to study agent administration. A validated electrochemiluminescence immunoassay with a lower limit of quantification (LLOQ) of 0.01 $\mu\text{g/mL}$ at a minimum required 1:10 dilution was used to measure serum guselkumab concentrations. A small number (5.3% in total) of post-dose pharmacokinetic (PK) measurements were below LLOQ and excluded from analysis. Serum samples for the evaluation of antibodies to guselkumab were collected at Weeks 0, 16, and 40. Antibodies were detected using a validated sensitive and drug-tolerant electrochemiluminescence immunoassay method using the MSD platform. The observed sensitivity of this anti-drug antibody (ADA) assay was 3.1 ng/mL for antibodies to guselkumab in human serum that did not contain guselkumab; as validated, 15 ng/mL of antibodies to guselkumab could be detected in the presence of up to 3125 ng/mL of guselkumab in human serum samples. Study patients were classified as having a positive antibody status if antibodies to guselkumab were detected in the sample at any visit after guselkumab treatment. The final dataset contained 238 patients with 2014 PK measurements, 2220 post-baseline PASI scores, and 2456 PGA scores.

Population pharmacokinetics model

The population PK analysis was conducted for guselkumab to generate individual parameters that adequately describe patient PK profiles to facilitate E–R modeling. Based on earlier experience [5], a confirmatory population PK analysis [16, 17] was implemented using a one-compartment model with first-order absorption and first-order elimination [apparent clearance (CL/F), apparent volume of distribution (V/F), and absorption rate constant (k_a)]. Between-patient random effects on CL/F , V/F and k_a were included using lognormal distributions. Correlation between BSVs on CL/F and V/F was also included, along with baseline body weight effects on them using a power model standardized to the median baseline body weight of 90 kg. An additive-plus-proportional error model was used, with the standard deviation (SD) of the additive component fixed at approximately 0.0029 based on the LLOQ value of 0.01 and assuming a uniform distribution of $U(0, \text{LLOQ})$. In our experience, this usually will result in similar data likelihood, nearly identical PK model parameter estimate, and slightly shortened run time, and could occasionally

stabilize parameter estimation. Individual Bayesian PK parameter estimates for patients were obtained for the E–R model development.

Latent variable indirect response model framework

The latent variable approach presumes an underlying latent variable such that the endpoint occurs when the latent variable crosses certain thresholds. More precisely, let n be the number of categories of an ordered categorical variable Y , $L(t)$ be the latent variable, and α_k be the thresholds where $k = 1, 2, \dots, n-1$, such that:

$$ACR \leq k \Leftrightarrow L(t) < \alpha_k$$

model $L(t)$ as:

$$L(t) = M(t) + \sigma \varepsilon \tag{1}$$

where $M(t)$ is the model predictor, ε is distributed with mean 0 and variance 1, and σ is the error standard deviation. Assuming that ε follows the standard normal distribution, then:

$$\begin{aligned} \text{prob}[Y \leq k] &= \text{prob}[L(t) < \alpha_k] = \text{prob}[\varepsilon < (\alpha_k - M(t))/\sigma] \\ &= \Phi[(\alpha_k - M(t))/\sigma] \end{aligned}$$

In this setting of latent variables, σ is not identifiable and may be assumed to be equal to 1; this gives:

$$\Phi^{-1}[\text{prob}(Y \leq k)] = \alpha_k - M(t) \tag{2}$$

which corresponds to probit regression. Assuming ε follows a logistic distribution leads to logit regression. In the context of E–R modeling, this derivation was first given in Hutmacher et al. [18]. To stabilize parameter estimation, α_k are typically better re-parameterized; e.g., for $n = 3$, as (α_1, d_0, d_2) with $d_0, d_2 > 0$ such that $\alpha_0 = \alpha_1 - d_0$ and $\alpha_2 = \alpha_1 + d_2$.

The latent variable representation Eq. 2 allows mechanism-based models to be used for $M(t)$. Between-subject variability is typically modeled at the intercept level with an additive normal distribution $\eta \sim N(0, \omega^2)$. Modeling the total treatment effect as the sum of placebo effect $f_p(-t)$ and drug effect $f_d(t)$, this leads to the mixed-effect probit regression, as follows:

$$\Phi^{-1}[\text{prob}(Y \leq k)] = \alpha_k + f_p(t) + f_d(t) + \eta \tag{3}$$

The placebo effect may typically be modeled empirically, e.g., with an exponential function:

$$f_p(t) = -F_p \exp(-r_p t) \tag{4}$$

where F_p is the maximum placebo effect and r_p is the rate of onset. The drug effect was modeled using a latent variable $R(t)$, governed by:

$$\frac{dR(t)}{dt} = k_{in} \left(1 - \frac{C_p}{IC_{50} + C_p} \right) - k_{out} R(t) \tag{5}$$

where C_p is drug concentration, and k_{in} , IC_{50} , and k_{out} are parameters in a Type I IDR model. It was further assumed that at baseline $R(0) = 1$, yielding $k_{in} = k_{out}$. The reduction of $R(t)$ was assumed to drive the drug effect through:

$$f_d(t) = DE[1 - R(t)] \tag{6}$$

where DE is a parameter to be estimated that determines the magnitude of drug effect.

Theoretically, the representation of drug effect in Eqs. 3–6 has been shown to be equivalent to a change-from-baseline latent-variable IDR model [19], under which k_{out} may be interpreted as the rate of drug effect onset and offset, and DE may be interpreted as the baseline of the latent variable prior to normalization [3]. Theoretical characteristics of general change-from-baseline IDR models, which have one less parameter than their corresponding IDR models, have been derived [19, 20]. Change-from-baseline latent IDR models are needed in the modeling of categorical endpoints because the latent variable is determined only up to a constant and therefore needs to be normalized [3, 18]. For more details on the theoretical characteristics of latent variable IDR models, see [3].

No covariate effects were explored for the E–R model due to the small sample size.

E–R modeling of PASI criteria and PGA scores

Equations 3–6 were first fitted to Pc and PGA data separately, and then simultaneously with BSV correlation and shared parameters explored. Practically, maximum sharing could occur if the underlying latent variables for the two endpoints differ by only a scale factor, in which case only one parameter Sc could be used to jointly model Pc and PGA, as follows:

$$\Phi^{-1}[\text{prob}(Pc \leq k)] = \alpha_{k,Pc} + f_p(t) + f_d(t) + \eta \tag{7}$$

$$\Phi^{-1}[\text{prob}(PGA \leq k)] = \alpha_{k,PGA} + Sc[f_p(t) + f_d(t) + \eta] \tag{8}$$

with $f_p(t)$ and $f_d(t)$ given by Eqs. 4–6. In theory, sharing on the intercept parameters could also occur, which would imply similarities between the Pc and PGA category separations. However this type of similarity may be unlikely in practice.

Fitting Eqs. 3–8 simultaneously to the Pc and PGA data reduces the total number of fixed and random effect parameters by 4 and 1, respectively, compared with separately modeling the endpoints.

Model estimation and evaluation

The sequential PK/PD modeling approach was used by first fixing the individual empirical Bayesian PK parameter estimates [18]. NONMEM version 7.3 was used for all modeling [21]. The LAPLACE estimation option was used for population PK parameter estimation, and the Importance estimation option was used for E–R modeling. While most models were pre-specified in this analysis, a decrease in the NONMEM minimum objective function value (OFV) of 10.83, corresponding to a nominal p value of 0.001, was considered the threshold criterion of whether including an additional model parameter improves the model fit in certain assessment of E–R modeling. Visual predictive check (VPC) was used for model evaluation by simulating 500 replicates [22].

Results

Demographics and baseline characteristics

Baseline body weight, the only influential PK covariate, ranged between 45 and 175 kg, with a mean (SD) of 91 (22) kg. Fourteen patients had positive ADA status, nearly evenly spread across the six treatment groups. More detailed demographics and baseline covariates were reported previously [2].

Guselkumab population pharmacokinetic modeling

Parameter estimates of the confirmatory analysis are given in Table 1. Standard goodness-of-fit diagnostics (shown in Figure S1 in Supplementary Material) indicated no anomalies. PK parameter estimates and their standard errors appeared within expectations and were comparable with those of Phase 1 [3, 23]. Estimating the standard

deviation of the additive component of the residual error (instead of fixing at approximately 0.0029) resulted in an estimate of 0.0016 with an associated RSE of near 100%, nearly identical NONMEM OFV, and nearly identical estimates for the remaining parameters. This supported fixing this parameter based on the theoretical consideration. Figure 1 shows the VPC results. The observed 95% percentiles for the 5 mg group were abnormally high, which might be partly due to the fact that two patients incorrectly received higher doses. While VPC accounted for the effect of the incorrect doses, the added variability might have not been fully accounted for. The exact reason is unclear and could be due to sample variation. Overall, the model reasonably described the observed data. It is noted that the number of patients with positive ADA status was below 20, the pre-specified threshold of allowing reliable estimation in the confirmatory analysis [16, 17], and therefore excluded from the covariate inclusion. With over eight PK samples per patient, any potential ADA effect was considered unlikely to substantially affect the quality of the individual empirical Bayesian parameter estimates, whose generation for subsequent E–R modeling was the main objective of this PK analysis.

PASI response model

Equations 3–6 were fitted to the PASI response data. Estimation of the model was stable, and model parameter estimates are given in Table 2. Estimation precision was reasonable. VPC results are shown in Fig. 2, and the model reasonably described the observed data.

PGA model

The number of PGA scores of 4 and 5 were relatively small, totaling 209, and therefore, they were combined with

Table 1 Population pharmacokinetic model parameter estimates

Parameter	Unit	Description	Estimate (relative standard error, %)
CL/F	L	Apparent clearance	0.567 (3.7)
V/F	L/day	Apparent volume of distribution	14.3 (3.4)
K_a	1/day	Absorption rate constant	4.93 (4.9)
BWT_{CL}		Baseline body weight effect on CL/F	1.16 (11.1)
BWT_V		Baseline body weight effect on V/F	1.04 (13.1)
ω_{CL}^2		Variance of between-subject variability for CL/F	0.307 (23.3)
ω_V^2		Variance of between-subject variability for V/F	0.229 (26.0)
$\omega_{CL,V}$		Correlation between ω_{CL}^2 and ω_V^2	0.252 (27.6)
σ^2		Variance of proportional residual error	0.0757 (13.9)

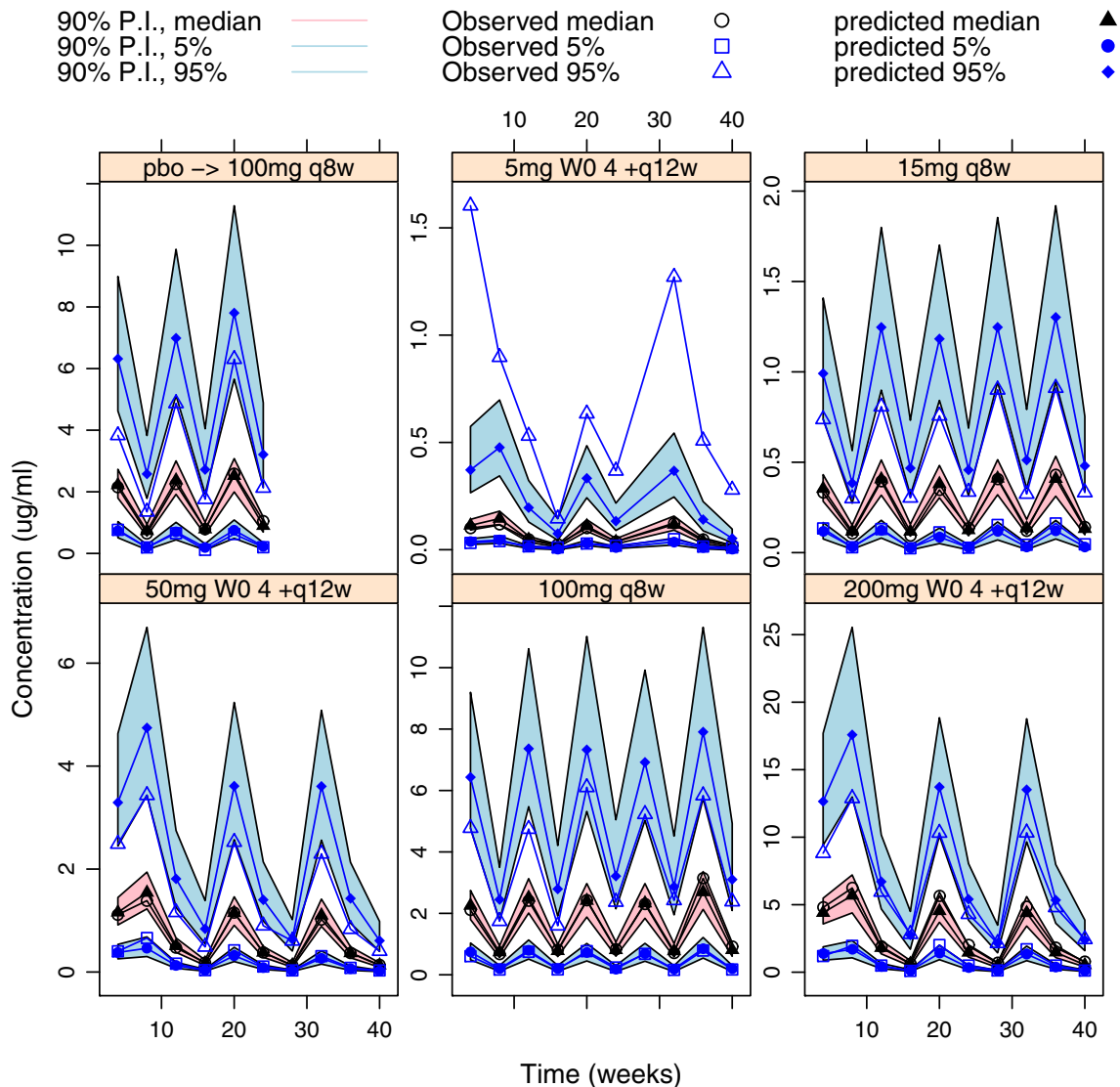


Fig. 1 Visual predictive check results of the guselkumab population pharmacokinetics model by treatment group

PGA scores of 3. Consequently, $n = 4$ was used in Eqs. 3–6 to model the probability of achieving PGA scores of $k = 0, 1, 2,$ or 3 . The model parameter estimates are given in Table 2. Estimation precision was reasonable. VPC results are shown in Fig. 3, and the model reasonably described observed data.

PASI—PGA correlation model

Because Pc and PGA are both measures of disease activity, patients responding to one type of measure may be expected to also respond to the other type; therefore their corresponding BSV terms (η_{Pc} and η_{PGA} in Table 2) may be expected to be correlated. To investigate this, a BSV

correlation model was employed by fitting the Pc and PGA data simultaneously, applying Eqs. 3–6 to Pc and PGA with distinct fixed and random effect parameters, and incorporating an additional correlation effect between the η_{Pc} and η_{PGA} using a bivariate normal distribution. This resulted in a NONMEM OFV decrease of over 300, indicating a significant improvement of the fit. The model parameter estimates are given in Table 2. The parameters for individual Pc and PGA components were similar to those obtained with the separate models. The correlation between η_{Pc} and η_{PGA} was estimated as 0.94. While positive correlation was expected, the high correlation magnitude suggested that one common BSV term could be sufficient.

Table 2 PASI criteria and PGA exposure–response model parameter estimates

Parameter	Unit	Model component	Description	Separate model (% RSE)	Correlation model (% RSE)	Joint model (% RSE)
$\alpha_{1,Pc}$		PASI 75/90/ 100	Intercept	−6.73 (6.1)	−6.63 (7.49)	−6.41 (4.37)
$d_{0,Pc}$			Intercept	1.48 (4.16)	1.48 (4.13)	1.44 (4.15)
$d_{2,Pc}$			Intercept	1.49 (4.37)	1.47 (4.55)	1.44 (4.36)
$F_{p,Pc}$			Maximum placebo effect	2.28 (23.8)	2.39 (19.1)	1.94 (15.7)
$r_{p,Pc}$	1/day		Rate of placebo effect onset	0.0193 (25.5)	0.0175 (22.6)	0.0165 (15.5)
$IC_{50,Pc}$	μg/ mL		Potency	0.07 (31.3)	0.0656 (27)	0.0663 (21.3)
DE_{Pc}			Drug effect	6.34 (6.61)	5.94 (6.68)	6.24 (4.93)
$k_{out,Pc}$	1/day		Rate of drug effect onset	0.0193 (8.73)	0.0198 (8.29)	0.0212 (5.69)
$\text{Var}(\eta_{Pc})$			Variance of between-subject variability	3.18 (12)	3.16 (11.8)	2.97 (11.7)
$\alpha_{2,PGA}$		PGA	Intercept	−3.29 (4.36)	−3.29 (5.09)	−3.09 (4.06)
$d_{1,PGA}$			Intercept	1.49 (3.42)	1.51 (3.38)	1.48 (3.35)
$d_{2,PGA}$			Intercept	1.4 (4.24)	1.4 (4.39)	1.36 (4.1)
$d_{3,PGA}$			Intercept	1.82 (4.75)	1.78 (5)	1.7 (4.53)
$F_{p,PGA}$			Maximum placebo effect	1.17 (17.8)	1.08 (18.3)	
$r_{p,PGA}$	1/day		Rate of placebo effect onset	0.0179 (21.2)	0.0193 (23.8)	
$IC_{50,PGA}$	μg/ mL		Potency	0.0996 (22)	0.0945 (20.1)	
DE_{PGA}			Drug effect	4.07 (5.13)	4.15 (4.72)	
$k_{out,PGA}$	1/day		Rate of drug effect onset	0.0226 (7.33)	0.022 (7.27)	
$\text{Var}(\eta_{PGA})$			Variance of between-subject variability	1.11 (11.4)	1.13 (11.3)	
$\text{cor}(\eta_{Pc}, \eta_{PGA})$		Correlation model	Between-subject variability correlation		1.78 (10.7)	
Sc		Joint model	Scale parameter of latent variable			0.601 (4.1)

Parameters are indexed for Pc and PGA respectively

Pc psoriasis area severity index criteria, PGA physician's global assessment, RSE relative standard error

Joint PASI—PGA model

Modeling Pc and PGA data together also provided a framework to assess the similarity of fixed effects between endpoints along with that of random effects. It could be hypothesized that, based on binding, a single latent variable could govern both endpoints through IDR models. Indeed, comparing the separate Pc and PGA model parameter estimates in Table 2 shows that the rate parameters r_p and k_{out} were relatively similar between the endpoints, along with the potency parameter IC_{50} . Furthermore, the maximum effect parameters F_p and DE appeared to differ in a relatively narrow range of 1.5–2, and the standard deviations of BSV differed similarly as well. This suggested that the underlying latent variables for the two endpoints could indeed differ by only a scale factor, and motivated the use of only one parameter Sc to jointly model Pc and PGA as in Eqs. 7–8.

The final joint model of fitting Eqs. 4–8 simultaneously to the Pc and PGA data resulted in a NONMEM OFV increase of 64 compared with the correlation model. The joint model used five fewer fixed-effect parameters and two fewer random-effect parameters, indicating a significant improvement of the fit. Comparing with the separate model, the joint model used five fewer fixed-effect parameters and one fewer random-effect parameter, and yet with a decrease in OFV of over 200. This showed a similar nature of improvement in fit compared with the joint analysis of continuous and ordered categorical data [9]. The joint model parameter estimates are given in Table 2. For Pc, the joint model parameter estimates were generally similar to those obtained with the separate model. Estimation precision improved, more notably for those parameters with RSE > 20% in the separate model. This may be expected, as the precision for those parameters less precisely estimated may benefit more by including

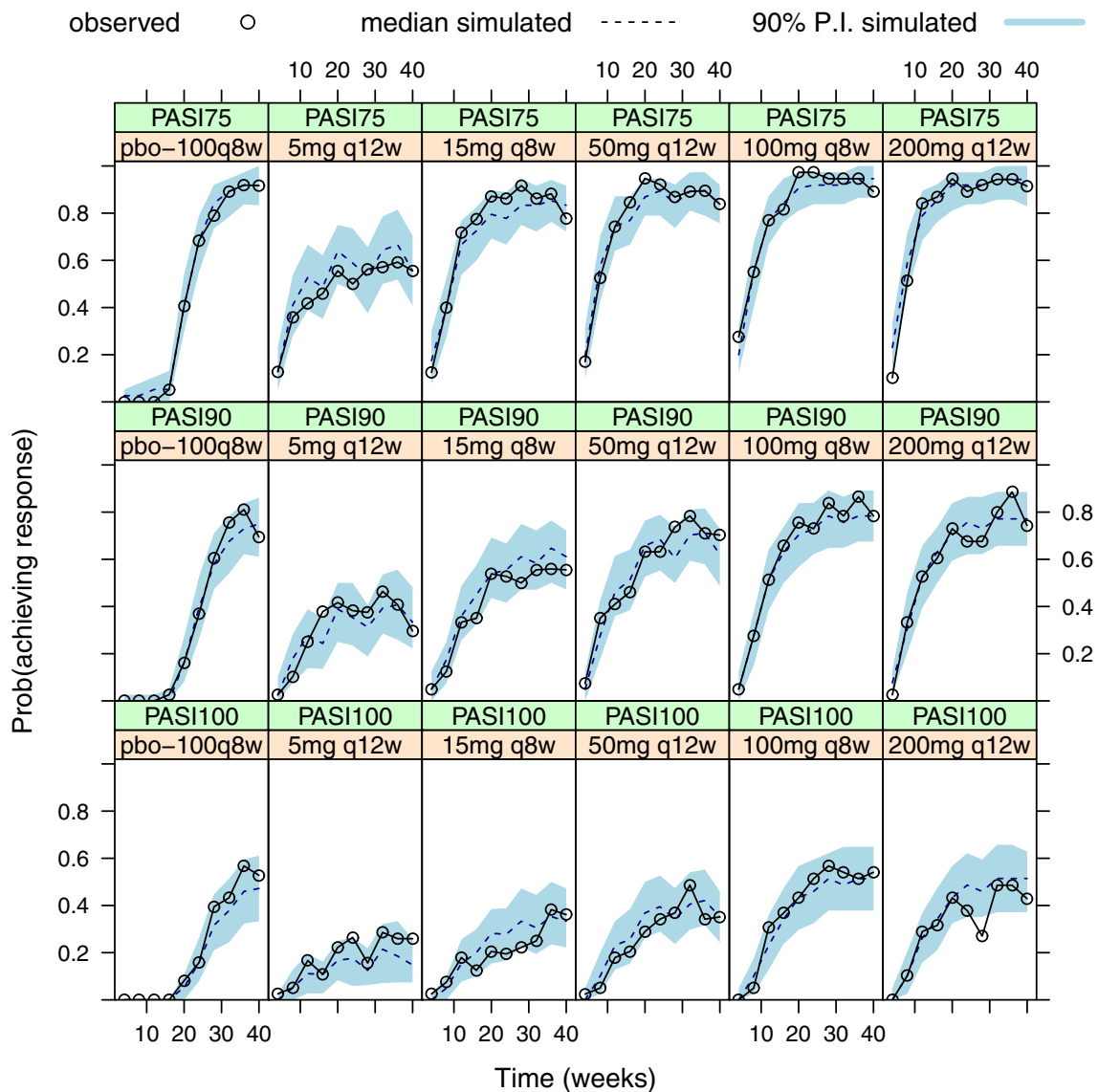


Fig. 2 Median model predictions at planned observation times and 90% prediction intervals (PI), in overlay with observed Psoriasis Area and Severity Index (PASI) response frequencies, for the separate

model. PASI75, PASI90, PASI100: proportion of subjects achieving 75, 90, or 100% PASI reduction from baseline

additional information (from PGA data). VPC results of the joint model are given in Figs. 4, 5, which showed a similarly reasonable description of the data as Figs. 2, 3. This is consistent with the reasonable RSE magnitudes and the similarity between the parameter estimates of the joint and the separate models.

Discussion

A parsimonious population E–R model for two ordered categorical endpoints, i.e., PASI criteria (PASI 75, PASI 90, and PASI 100) and PGA scores, was developed based on the Phase 2 dose-ranging study of guselkumab. This was

motivated by the previous development of joint modeling of a continuous and an ordered categorical endpoint [9]. Similarly to the previous scenario [9], the joint model achieved significant improvement in model fit in terms of NONMEM OFV, while using fewer parameters than the separate model. The correlation model investigation demonstrated that the improved fit of the joint model is due to the BSV correlation. However, unlike in the previous scenario [9] where additional BSV terms for the categorical endpoint could be estimated by leveraging information from the continuous endpoint, the joint model for two categorical endpoints had essentially the same number of parameters. Hence, its description of individual endpoints apparently did not improve in terms of the VPCs. In

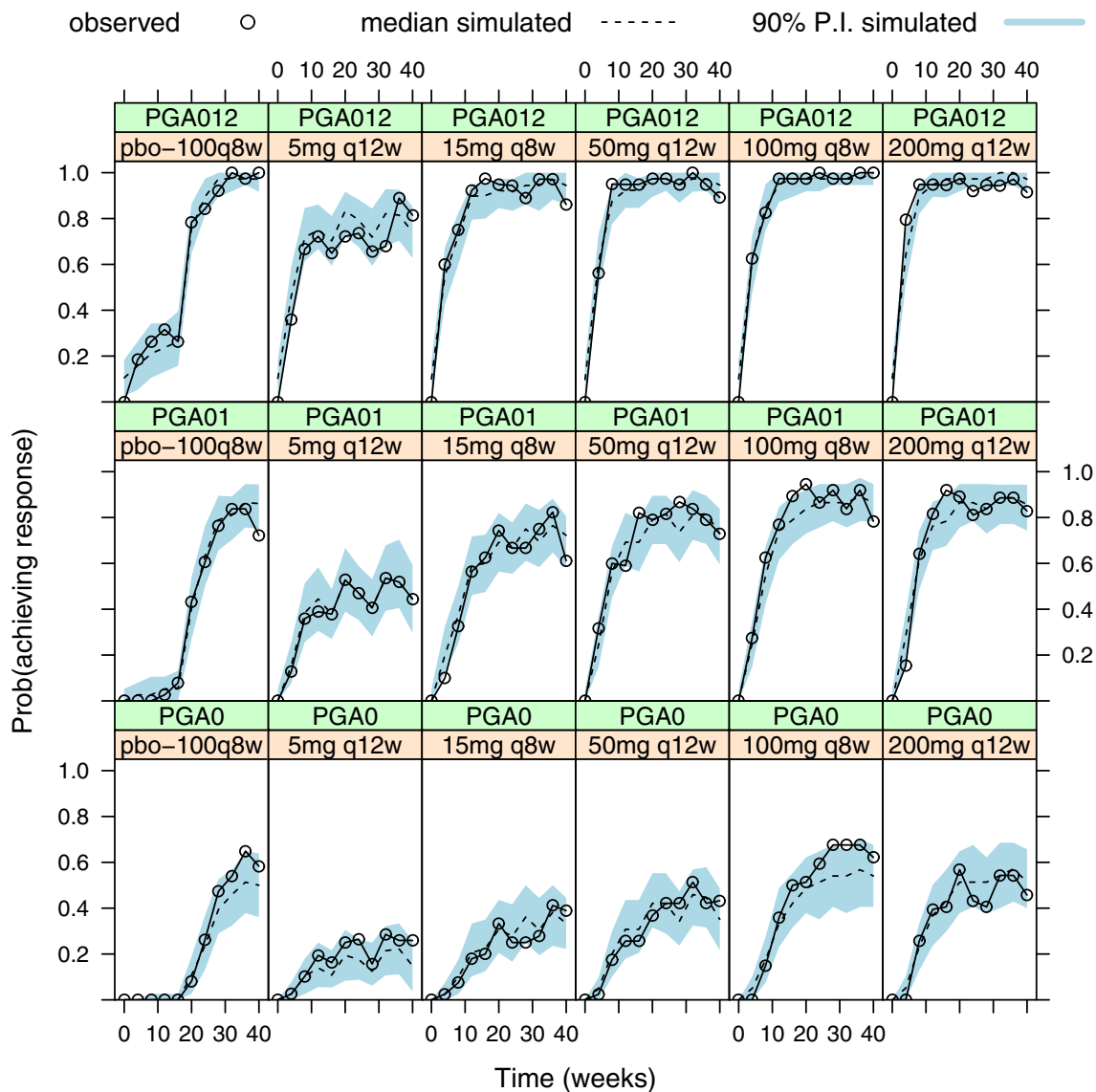


Fig. 3 Median model predictions at planned observation times and 90% prediction intervals (PI), in overlay with observed Physician's Global Assessment (PGA) response frequencies, for the separate model. PGA0, PGA = 0; PGA01, PGA = 0 or 1; PGA012, PGA = 0, 1, or 2

principle, separate modeling of the endpoints should result in their unbiased estimation, provided it is reasonably supported by the data. Joint modeling cannot improve on estimation accuracy; actually, accuracy could deteriorate if the joint mechanism is inappropriately assumed. Where the joint modeling stands to gain is in the precision of estimation, as evidenced by the improvement of RSEs in Table 1, achieved through the use of all relevant information. Improvement in precision is important for predictions and decision making. In terms of drug development, improving precision allows the trial objectives to be achieved with fewer subjects and reduced costs.

As endpoints, Pc is a measure of change from baseline by nature while PGA is an absolute measure. This raises the conceptual question of why their analyses could be

pooled. From a theoretical perspective, this may be due to the fact that the categories analyzed represent relatively large improvements in efficacy, thus making the conceptual difference smaller. Indeed in psoriasis drug development, Pc and PGA are used somewhat interchangeably as efficacy measures. Furthermore, when endpoints are measured at the same time, the residual level correlation may also be modeled [24]; this is not expected to affect the separate endpoint predictions but can be important in the predictions of achieving joint criteria using both endpoints [25].

Since the categorical endpoint Pc is derived from PASI scores, it is natural to ask whether PASI score could be used as a latent variable to model Pc. Indeed, previous PASI-related longitudinal E-R modeling [4–7], including that used to guide the design of the Phase 2 dose-ranging

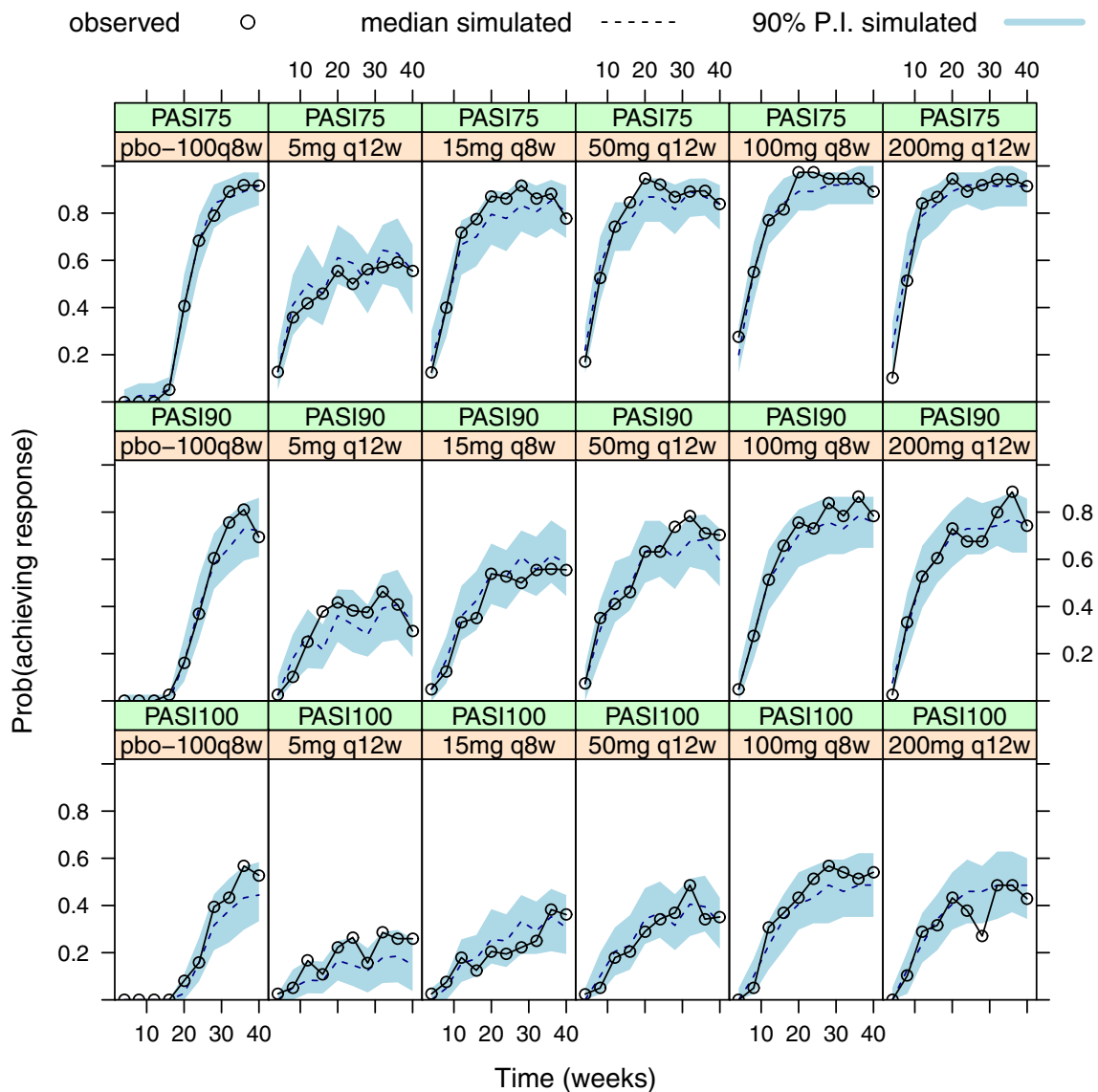


Fig. 4 Median model predictions at planned observation times and 90% prediction intervals (PI), in overlay with observed Psoriasis Area and Severity Index (PASI) response frequencies, for the joint model.

PASI75, PASI90, PASI100: proportion of subjects achieving 75, 90, or 100% PASI reduction from baseline

study of guselkumab, first embellished an E–R model of PASI score as a continuous variable, and then used the model to predict the various categories of Pc. To investigate this issue, it is helpful to first examine the predictive ability of the earlier model [5], which is of interest by itself. For this purpose, an external VPC of the earlier model was conducted with the Phase 2 PASI data by treatment, shown in Fig. S2 in Supplementary Material. The model predicted the median PASI scores very well, which in part explained why the Phase 2 study reached its exact objectives by achieving the targeted treatment separations. On the other hand, the continuous PASI score model could not sufficiently predict Pc. As Fig. S2 shows, the model under-predicted the 5% percentiles, which could

not be improved even after updating the model with Phase 2 data. To our knowledge, no continuous PASI score models have yet been published that accurately describe Pc. This could be explained by two reasons. Firstly, most previous model development used the additive-plus-proportional residual error model [4–7], which is ill-behaved in this case, as the likelihood conceptually could become arbitrarily large due to the fact that PASI = 0 may be observed. Secondly, PASI scores are Bounded Outcome Scores (BOS), which report a discrete set of restrictive values on a finite range [26]. BOS data often demonstrate non-standard, e.g., J- or U- shaped, distributions, and data transformations aiming to achieve normality can be much more difficult than many other types of skewed

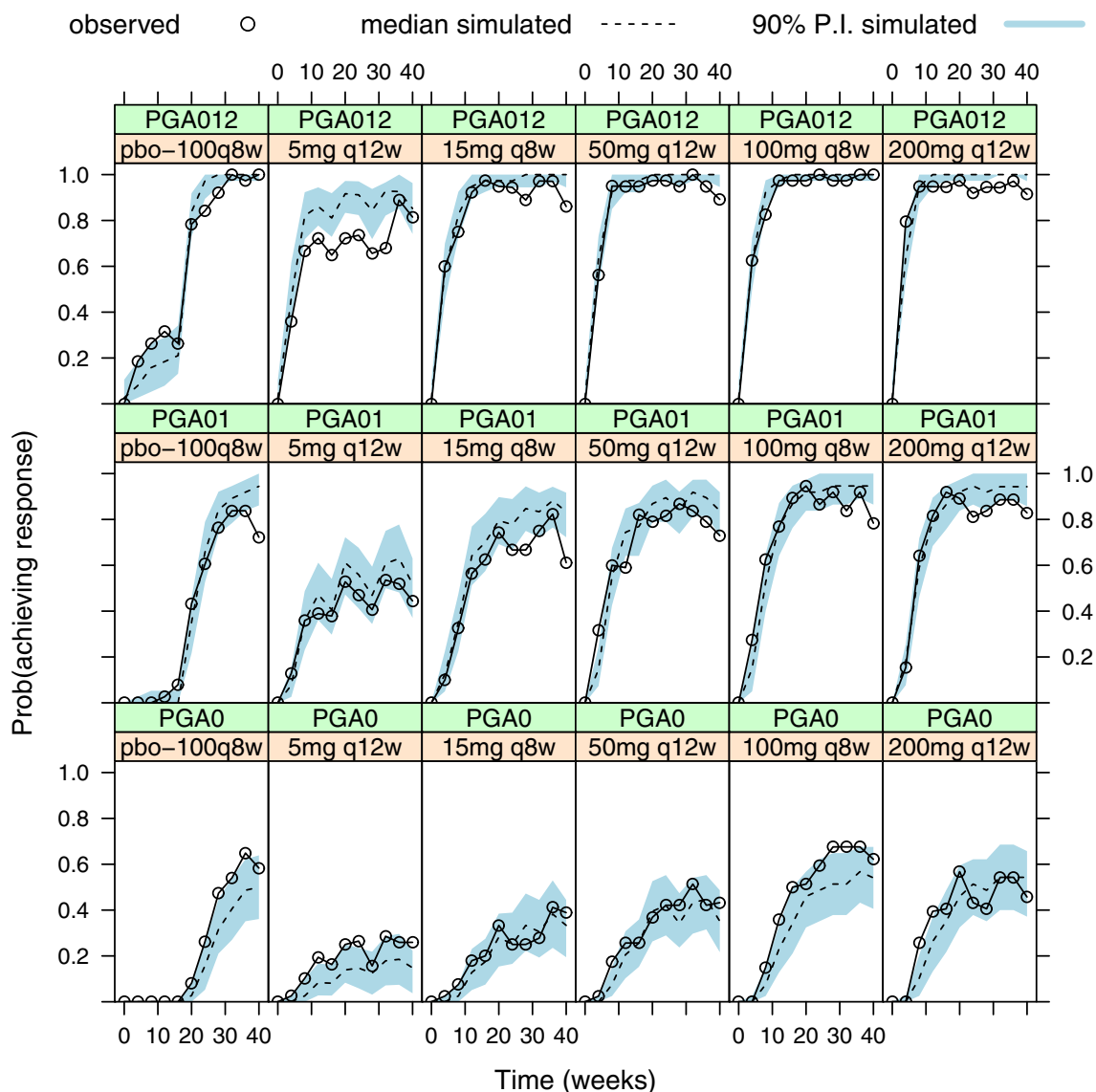


Fig. 5 Median model predictions at planned observation times and 90% prediction intervals (PI), in overlay with observed Physician's Global Assessment (PGA) response frequencies, for the joint model. PGA0, PGA = 0; PGA01, PGA = 0 or 1; PGA012, PGA = 0,1, or 2

distributions. For example, beta-regression has often been used for BOS modeling using a transformation of original scores to the open interval (0, 1), by shrinking values at the boundary by an apparently innocuous small factor of 0.01 [27, 28]. Theoretical examination of the effect of this small factor shows that reducing its value sufficiently could make the influence of boundary values arbitrarily large, thus rendering the interpretation of the results dubious. Other approaches explicitly dealing with the nature of BOS have been developed, including Estimating Transformations [29], Coarsened Grid with or without transformations [26, 30], and a nonparametric approach [31]. Each of these explicit BOS approaches has conceptual appeal; indeed, extensive efforts applying these approaches lead to various degrees of improvement in describing the PASI score

distributions. Unfortunately, none resulted in any improvement in describing P_c . This appears to be due to the fact that, in order to accurately describe P_c , the continuous PASI score model must accurately predict not only the mean data, but also the entire distribution. This is extremely difficult for a clinical endpoint with highly skewed distributions. The high efficacy of guselkumab resulted in just such a data distribution with most scores on the low end, near or achieving 0. For this reason, modeling P_c directly as an ordered categorical variable is likely an effective approach by modeling a presumed unobservable latent variable, instead of using the “true” latent variable even if it is observed (and thus no longer “latent”).

While the Phase 1 model had limited ability to predict P_c , it led to a Phase 2 study that, in light of its outcome,

was considered well-designed. This shows that the model does not have to be perfect to make correct decisions. Indeed, if the model were perfect, there would be no need for additional clinical trials. The drug development paradigm does demand increased precision as the stages progress, as later stage decisions incur increased costs. That is, levels of precision that are acceptable at early stages may not suffice for later stages, and improvements should be consistently sought as information accumulates.

A general issue with ordered categorical data is how many levels should be modeled. Increasing the number of categories modeled increases the use of available information in the data, provided that estimation of the intercept parameters could be supported. On the other hand, all levels are often not of similar interest. For example, achieving $\text{PGA} = 0$ and $\text{PGA} \leq 1$ are considered more relevant efficacy markers. During the development of the separate PGA model, we initially attempted to further combine PGA level 2 with those ≥ 3 . However this resulted in model non-convergence. Detailed investigation showed that having the PGA level 3 modeled is important for the estimation of placebo effect, which also affects the entire model estimation [3]. In Fig. 3, the placebo effect for PGA is visually discernable only in the $\text{PGA} \leq 2$ group. This supported the relevance of retaining this level in the model even though it is not of direct clinical interest. While in theory placebo effect estimation could be supported by the relative differences among active treatments groups and thus not always require observing a trend in the placebo group, confounding may occur and precise estimation may become difficult in such situations, especially when the effect is small. It is of interest to note that estimation of the separate PASI Pc model could be supported without the need of including additional levels, e.g., PASI 50. Unnecessarily including such levels would not only deviate from the main interest of modeling, but also cause lack of fit due to the skewness of the PASI score distribution discussed above.

While the importance of model precision may seem obvious, tendencies may exist that sacrifice utilization of relevant information for the sake of convenience. For example, there is a view that landmark E–R analysis is often sufficient even when longitudinal data are available [32]. This optimistic view may be due to the fact that decisions often focus on predictions based on only point estimates, and the uncertainties are ignored. This may in part have contributed to the fact that few publications report observations that confirm earlier phase model predictions, especially for clinical efficacy modeling based on proof-of-concept trials where the uncertainty may be most pronounced [5]. Recent IDR model developments have made longitudinal modeling more adaptive to clinical efficacy endpoints of various types, and relevant

information should be used to increase the E–R analysis precision whenever possible.

Acknowledgements The authors thank Kristin Ruley Sharples, PhD, for her editorial assistance, and two anonymous reviewers for their constructive suggestions.

Funding This research was funded by Janssen Research and Development, LLC.

References

- Sharma A, Jusko WJ (1996) Characterization of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm* 24(6):611–635
- Gordon KB, Duffin KC, Bissonnette R, Prinz JC, Wasfi Y, Li S, Shen YK, Szapary P, Randazzo B, Reich K (2015) A phase 2 trial of guselkumab versus adalimumab for plaque psoriasis. *N Engl J Med* 373(2):136–144. doi:10.1056/NEJMoa1501646
- Hu C (2014) Exposure-response modeling of clinical end points using latent variable indirect response models. *CPT* 3:e117. doi:10.1038/psp.2014.15
- Zhou H, Hu C, Zhu Y, Lu M, Liao S, Yeilding N, Davis HM (2010) Population-based exposure-efficacy modeling of ustekinumab in patients with moderate to severe plaque psoriasis. *J Clin Pharmacol* 50(3):257–267. doi:10.1177/0091270009343695
- Hu C, Wasfi Y, Zhuang Y, Zhou H (2014) Information contributed by meta-analysis in exposure-response modeling: application to phase 2 dose selection of guselkumab in patients with moderate-to-severe psoriasis. *J Pharmacokinet Pharmacodyn* 41(3):239–250. doi:10.1007/s10928-014-9360-6
- Salinger DH, Endres CJ, Martin DA, Gibbs MA (2014) A semi-mechanistic model to characterize the pharmacokinetics and pharmacodynamics of brodalumab in healthy volunteers and subjects with psoriasis in a first-in-human single ascending dose study. *Clin Pharmacol Drug Dev* 3(4):276–283. doi:10.1002/cpdd.103
- Tham LS, Tang CC, Choi SL, Satterwhite JH, Cameron GS, Banerjee S (2014) Population exposure-response model to support dosing evaluation of ixekizumab in patients with chronic plaque psoriasis. *J Clin Pharmacol* 54(10):1117–1124. doi:10.1002/jcph.312
- Hu C, Szapary PO, Yeilding N, Zhou H (2011) Informative dropout modeling of longitudinal ordered categorical data and model validation: application to exposure-response modeling of physician's global assessment score for ustekinumab in patients with psoriasis. *J Pharmacokinet Pharmacodyn* 38(2):237–260
- Hu C, Zhou H (2016) Improvement in latent variable indirect response joint modeling of a continuous and a categorical clinical endpoint in rheumatoid arthritis. *J Pharmacokinet Pharmacodyn* 43(1):45–54
- Koo J (1996) Population-based epidemiologic study of psoriasis with emphasis on quality of life assessment. *Dermatol Clin* 14(3):485–496
- Krueger GG, Duvic M (1994) Epidemiology of psoriasis: clinical issues. *J Invest Dermatol* 102(6):14S–18S
- Schon MP, Boehncke WH (2005) Psoriasis. *N Engl J Med* 352(18):1899–1912
- Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, McClanahan T, Kastelein RA, Cua DJ (2005) IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J Exp Med* 201(2):233–240

14. Rozenblit M, Lebwohl M (2009) New biologics for psoriasis and psoriatic arthritis. *Dermatol Ther* 22(1):56–60
15. Yawalkar N, Karlen S, Hunger R, Brand CU, Braathen LR (1998) Expression of interleukin-12 is increased in psoriatic skin. *J Invest Dermatol* 111(6):1053–1057. doi:[10.1046/j.1523-1747.1998.00446.x](https://doi.org/10.1046/j.1523-1747.1998.00446.x)
16. Hu C, Zhang J, Zhou H (2011) Confirmatory analysis for phase III population pharmacokinetics. *Pharm Stat* 10(7):812–822
17. Hu C, Zhou H (2008) An improved approach for confirmatory phase III population pharmacokinetic analysis. *J Clin Pharmacol* 48(7):812–822. doi:[10.1177/0091270008318670](https://doi.org/10.1177/0091270008318670)
18. Hutmacher MM, Krishnaswami S, Kowalski KG (2008) Exposure-response modeling using latent variables for the efficacy of a JAK3 inhibitor administered to rheumatoid arthritis patients. *J Pharmacokinet Pharmacodyn* 35:139–157
19. Hu C, Xu Z, Mendelsohn A, Zhou H (2013) Latent variable indirect response modeling of categorical endpoints representing change from baseline. *J Pharmacokinet Pharmacodyn* 40(1):81–91
20. Woo S, Pawaskar D, Jusko WJ (2009) Methods of utilizing baseline values for indirect response models. *J Pharmacokinet Pharmacodyn* 36:381–405
21. Beal SL, Sheiner LB, Boeckmann A, Bauer RJ (2009) NONMEM User's Guides (1989–2009). Icon Development Solutions, Ellicott City, MD
22. Karlsson MO, Holford NHG (2008) A tutorial on visual predictive checks. www.page-meeting.org/?abstract=1434
23. Zhuang Y, Calderon C, Marciniak SJ Jr, Bouman-Thio E, Szapary P, Yang TY, Schantz A, Davis HM, Zhou H, Xu Z (2016) First-in-human study to assess guselkumab (anti-IL-23 mAb) pharmacokinetics/safety in healthy subjects and patients with moderate-to-severe psoriasis. *Eur J Clin Pharmacol* 72(11):1303–1310. doi:[10.1007/s00228-016-2110-5](https://doi.org/10.1007/s00228-016-2110-5)
24. Laffont CM, Vandemeulebroecke M, Concordet D (2014) Multivariate analysis of longitudinal ordinal data with mixed effects models, with application to clinical outcomes in osteoarthritis. *J Am Stat Assoc* 109(507):955–966
25. Hu C, Szapary PO, Mendelsohn AM, Zhou H (2014) Latent variable indirect response joint modeling of a continuous and a categorical clinical endpoint. *J Pharmacokinet Pharmacodyn* 41(4):335–349. doi:[10.1007/s10928-014-9366-0](https://doi.org/10.1007/s10928-014-9366-0)
26. Lesaffre E, Rizopoulos D, Tsonaka R (2007) The logistic transform for bounded outcome scores. *Biostatistics* 8(1):72–85
27. Smithson MV, Verkuilen J (2006) A better lemon squeezer? Maximumlikelihood regression with beta-distributed dependent variables. *Psychol Methods* 11(1):54–71
28. Xu XS, Samtani M, Yuan M, Nandy P (2014) Modeling of bounded outcome scores with data on the boundaries: application to disability assessment for dementia scores in Alzheimer's disease. *AAPS J* 16(6):1271–1281. doi:[10.1208/s12248-014-9655-y](https://doi.org/10.1208/s12248-014-9655-y)
29. Hutmacher MM, French JL, Krishnaswami S, Menon S (2011) Estimating transformations for repeated measures modeling of continuous bounded outcome data. *Stat Med* 30(9):935–949. doi:[10.1002/sim.4155](https://doi.org/10.1002/sim.4155)
30. Hu C, Yeilding N, Davis HM, Zhou H (2011) Bounded outcome score modeling: application to treating psoriasis with ustekinumab. *J Pharmacokinet Pharmacodyn* 38(4):497–517
31. Parsons NR (2013) Proportional-odds models for repeated composite and long ordinal outcome scales. *Stat Med* 32(18):3181–3191. doi:[10.1002/sim.5756](https://doi.org/10.1002/sim.5756)
32. Overgaard RV, Ingwersen SH, Tornøe CW (2015) Establishing good practices for exposure-response analysis of clinical endpoints in drug development. *CPT* 4(10):565–575. doi:[10.1002/psp4.12015](https://doi.org/10.1002/psp4.12015)