CrossMark

ORIGINAL PAPER

Improvement in latent variable indirect response joint modeling of a continuous and a categorical clinical endpoint in rheumatoid arthritis

Chuanpu Hu[1] · Honghui Zhou[1]

Abstract Improving the quality of exposure–response modeling is important in clinical drug development. The general joint modeling of multiple endpoints is made possible in part by recent progress on the latent variable indirect response (IDR) modeling for ordered categorical endpoints. This manuscript aims to investigate, when modeling a continuous and a categorical clinical endpoint, the level of improvement achievable by joint modeling in the latent variable IDR modeling framework through the sharing of model parameters for the individual endpoints, guided by the appropriate representation of drug and placebo mechanism. This was illustrated with data from two phase III clinical trials of intravenously administered mAb X for the treatment of rheumatoid arthritis, with the 28-joint disease activity score (DAS28) and 20, 50, and 70 % improvement in the American College of Rheumatology (ACR20, ACR50, and ACR70) disease severity criteria were used as efficacy endpoints. The joint modeling framework led to a parsimonious final model with reasonable performance, evaluated by visual predictive check. The results showed that, compared with the more common approach of separately modeling the endpoints, it is possible for the joint model to be more parsimonious and yet better describe the individual endpoints. In particular, the joint model may better describe one endpoint through subject-specific random effects that would not have been estimable from data of this endpoint alone.

Introduction

Exposure–response (E–R) modeling has become more impactful in drug development by allowing optimal dosing selection and potentially providing better understanding of drug's mechanism of action. Clinical trial endpoints are often disease scores that are either continuous or ordered categorical. For example, in rheumatoid arthritis (RA), two types of commonly used efficacy endpoints are the 28-joint disease activity score using CRP (DAS28) and 20, 50, and 70 % improvement in the American College of Rheumatology disease severity criteria (ACR20, ACR50, and ACR70) [1]. Parsimonious and informative E–R modeling of such endpoints is important in clinical development, especially at late stage in which appropriate decisions rely on accurate predictions that often require the use of models more complex than landmark analyses using direct correlation between exposure metrics and response [2].

A widely used class of E–R models includes the Types I–IV indirect response (IDR) models [3]. For continuous clinical endpoints, these models could be directly applied even though their origin of use was to describe physiological variables, and their presumed consistency with the mechanism of drug action lends confidence to the model predictions. For discrete clinical endpoints, IDR models may be applied by means of the latent variable approach, most notably applied for ACR20, ACR50, and ACR70. Hutmacher et al. [4] initiated the latent variable IDR model for ACR20. Hu et al. [5] extended the latent variable approach to simultaneously model ACR20, ACR50, and

✉ Chuanpu Hu
chu25@its.jnj.com

[1] Quantitative Sciences, Janssen Research & Development, LLC, PO Box 776, 1400 McKean Road, Spring House, PA 19477, USA

ACR70, while making the model more parsimonious [6]. The latter approach allows the placebo effect terms to be estimated separately using only placebo data, which facilitates the ease of initial parameter estimate selection. An equivalent form of its underlying IDR model was also used earlier on other clinical endpoints for psoriasis [7, 8]. Hu [2] substantiated the mechanistic rendition of latent variable IDR models by deriving the latent variable representations for common scenarios that ordered categorical endpoints may be applicable, and summarized related theoretical characteristics and practical implementations.

It is not uncommon that two clinical endpoints are considered of similar importance, such as DAS28 and ACR responses in RA trials, where each has been used as the primary endpoint. In addition, multiple clinical endpoints may measure similar disease status and share the use of some similar information, e.g., both DAS28 (CRP) and ACR20/50/70 use tender and swollen joint counts, as well as CRP. While a panel of clinical endpoints are more informative than a single endpoint in clinical interpretation, they are usually modeled separately in practice [9]. The natural question is whether the same conclusions could be reached by considering these endpoints (1) individually; (2) separately yet together, or (3) jointly. This could be answered to some extent using individual modeling; however the joint modeling approach directly addresses the consistency among the three methods through shared parameters. In addition, joint modeling facilitates the simulation of correlated endpoints, thus the correct evaluation of joint decision criteria based on both endpoints. The correlation between the endpoints should indicate the level of similarity between the disease components measured by the endpoints. Indeed, a strong correlation between DAS28 and ACR response has been shown (e.g., [10]).

More importantly, when multiple clinical endpoints measure similar disease status, as with DAS28 and ACR20/50/70, the joint model framework provides the advantage of allowing the exploration of similar fixed effect and between-subject variability (BSV) parameters, guided by mechanistic rationale. Furthermore, joint modeling allows the remaining unexplained residual correlations to be estimated at the within-subject random effect level. For continuous endpoints, this can be easily handled by assuming a multivariate normal distribution for the residuals. For categorical endpoints, this can be accommodated by the latent variable framework, albeit the implementation is complex [11]. In the case of a continuous and a categorical endpoint, Hu et al. [12] provided a conditional approach to allow for the implementation of the joint model in NONMEM [13], a widely used software application for E–R modeling. Specifically, the residual correlation is modeled at the latent variable level with a bivariate normal distribution, and a negative correlation

coefficient indicates that, given the structural and BSV model parameters, the chance of achieving ACR response criteria increases as DAS28 decreases. While estimating the residual correlation may not affect individual endpoint predictions, it improves the prediction of joint probability distribution of the endpoints, more specifically the assessment of the proportion of subjects achieving desirable responses simultaneously in more than one endpoint. This has been well established in statistical literature [11].

The objective of this manuscript was to explore the level of improvement that joint modeling may gain in contrast to separate modeling of a continuous and a categorical endpoint measuring the same disease status. Both approaches were applied to model the DAS28 score and ACR20, ACR50, ACR70 response, using data from two previously published phase III clinical trials for the treatment of RA for which we published a latent variable IDR model of the ACR20, ACR50, ACR70 response previously [6]. The results indicated that improved fit of ACR data could be achieved by leveraging subject-specific random effects whose estimation were made possible only with DAS28 data.

Methods

Data and information used for E–R modeling

The DAS28 score and ACR20, ACR50, ACR70 response were available in data from two phase III, parallel, placebo-controlled clinical trials of intravenously administered mAb X, Study 1 [14] and Study 2 [15], in patients with active RA despite prior use of methotrexate (MTX) therapy. Data used were the same as in the E–R modeling of ACR20, ACR50, ACR70 response described previously [6], with the additional inclusion of DAS28 scores. Briefly, Study 1 investigated the mAb X dose regimen of 2 mg/kg given at Weeks 0, 4, and every 8 weeks thereafter, briefly written hereafter as the q8 weekly regimen. Study 2 studied the mAb X dose regimens of 2 and 4 mg/kg given every 12 weeks. Both trials had MTX as placebo control arms, and subjects on the placebo arms were switched to the active arms of mAb X + MTX at Week 16, at which time they were eligible for early escape. The numbers of subjects in the E–R modeling dataset were 395, 197, 129, 126, and 129 respectively for the following treatment arms: mAb X 2 mg/kg + MTX q8 weeks (Study 1), placebo 1 (MTX, Study 1), mAb X 2 mg/kg + MTX q12 weeks (Study 2), mAb X 4 mg/kg + MTX q12 weeks (Study 2), and placebo 2 (MTX, Study 2).

Both studies had data through Week 48 available for analysis. At least ten serum concentration measurements of mAb X per patient were scheduled over the study period.

DAS28, ACR20, ACR50, and ACR70 responses were evaluated at the time of each treatment administration in both studies. Study 1 had an additional evaluation at Week 2.

A population PK analysis using a two-compartment structural model implemented in NONMEM was performed using data from patients available for E–R modeling and additional data from other studies. The model described the data adequately and the details of the PK study data and analysis are described elsewhere. Results were consistent with a previous confirmatory analysis [16]. Empirical Bayesian parameter estimates were then used in a sequential modeling approach for the E–R modeling discussed below.

The final dataset contained 8474 DAS28 score and 7535 ACR response data points from 976 patients. More detailed description on study designs can be found in Hu et al. [6].

E–R model

The E–R model is specified by the DAS28 score and ACR response components along with the residual correlation model.

DAS28 score component

DAS28 scores were modeled by adopting a semi-mechanistic approach applied in earlier E–R analyses [7] as

$$DAS28(t) = b - f_{DAS28,p}(t) - f_{DAS28,d}(t) + \varepsilon \quad (1)$$

where $DAS28(t)$ is the observed DAS28 score, b is baseline, $f_{DAS28,p}(t)$ is placebo effect, $f_{DAS28,d}(t)$ is drug effect, and $\varepsilon \sim N(0, \sigma^2)$ represents the within-subject variability. The placebo effect was modeled empirically as

$$f_{DAS28,p}(t) = b\, F_{p,DAS28}[1 - \exp(-r_{DAS28}t)] \quad (2)$$

where $0 \leq F_{p,DAS28} \leq 1$ is the fraction of maximum placebo effect and r_{DAS28} is the rate of onset. The drug effect was modeled with

$$f_{DAS28,d}(t) = b(1 - F_{p,DAS28})E_{max}[1 - R_{DAS28}(t)] \quad (3)$$

where $0 \leq E_{max} \leq 1$ represents fraction of maximum drug effect and, following a previous approach [7, 8], the drug effect was assumed to be driven by a latent variable $R_{DAS28}(t)$ governed by:

$$\frac{d\,R_{DAS28}(t)}{dt} = k_{in,DAS28}\left(1 - \frac{C_p}{IC_{50,DAS28} + C_p}\right) - k_{out,DAS28}\,R_{DAS28}(t) \quad (4)$$

It was further assumed that at baseline $R_{DAS28}(0) = 1$, yielding $k_{in,\,DAS28} = k_{out,\,DAS28}$.

The BSV on the parameters constrained to be between 0 and 1, i.e., E_{max} and $F_{p,DAS28}$, were modeled with normal distributions after a logit transformation. The BSV on other parameters were modeled with lognormal distributions. Correlation between BSV was assessed with the corresponding multivariate normal distributions..

ACR response component

As in previously implementations [6], the three endpoints ACR20, ACR50, and ACR70 were combined into one ordered categorical endpoint ACR having 4 possible outcomes: ACR = 1, if achieving ACR70; ACR = 2, if achieving ACR50 but not ACR70; ACR = 3, if achieving ACR20 but not ACR50; and ACR = 4, if not achieving ACR20. To facilitate residual correlation modeling [12], the following mixed-effect probit regression was used instead:

$$\Phi^{-1}[prob(ACR \leq k)] = \alpha_k + f_{p,ACR}(t) + f_{d,ACR}(t) + \eta \quad (5)$$

where Φ is the normal cumulative distribution function, k = 1, 2, or 3, α_k are monotonically increasing in k, and $\eta \sim N(0, \omega^2)$ represents inter-subject variability. As in previous applications [2, 6, 12], α_k were re-parameterized as (α_2, d_1, d_3) with $d_1, d_3 > 0$ such that $\alpha_1 = \alpha_2 - d_1$ and $\alpha_3 = \alpha_2 + d_3$, in order to stabilize parameter estimation. The placebo effect was modeled with a previously used exponential function [4, 9]:

$$f_{p,ACR}(t) = -F_{p,ACR} \exp(-r_{p,ACR}t) \quad (6)$$

In the latent variable level representation, this has the same form as Eq. 2 as shown by Hu [2]. It is more flexible than that used in Hu et al. [6] and was chosen here to allow exploration of similarities between the endpoints. For the same reason, it is assumed that the drug effect is driven by a latent variable $R_{ACR}(t)$, governed by the same form as Eq. 4;
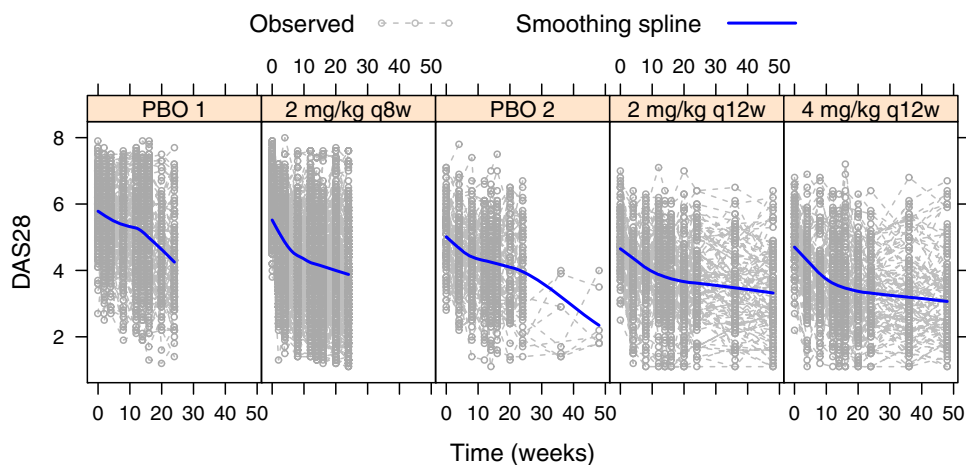
$$\frac{d\,R_{ACR}(t)}{dt} = k_{in,ACR}\left(1 - \frac{C_p}{IC_{50,ACR} + C_p}\right) - k_{out,ACR}\,R_{ACR}(t) \quad (7)$$

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

$$f_d(t) = f_{d,ACR}(t) = DE_{ACR}[1 - R_{ACR}(t)] \quad (8)$$

where DE_{ACR} is a parameter to be estimated that determines the magnitude of drug effect. We have previously

Fig. 1 Observed 28-joint disease activity (DAS28) scores by treatment group overlaid with smoothing spline

Results

Figure 1 shows the observed DAS28 time course by treatment group. High variability was apparent, both between and within subjects. In addition, baseline DAS28 in Study 1 appeared to be notably larger than that in Study 2.

Initial DAS28 model

Equations 1–4 were fitted to the DAS28 data. Early modeling explorations using the Laplace option could not reliably estimate any more BSV terms than those on b and IC_{50}, and led to a sizable estimate ($\cong 5$, or >70 %) for BSV on IC_{50}. While parameter estimate appeared reasonable, standard error estimation appeared unstable. Since the adequacy of the Laplace approximation degrades as the magnitude of BSV increases, the IMP estimation option in NOMEM was used for key model runs. An additional parameter b_s was used to account for the baseline difference between Study 1 and Study 2, and BSV terms were included on b, p_{DAS28}, $IC_{50,DAS28}$ and E_{max}, with a full variance–covariance matrix accounting for their correlations. Attempting to reduce the BSV terms or the correlation parameters or to include additional BSV terms resulted in either notably worsening or lack of sufficient improvement in the fit. Table 1 shows the parameter estimates. Estimation precision was reasonable, with standard errors (SE) generally a magnitude lower than the estimates. SE is presented instead of the often used relative standard errors in order to provide appropriate comparison of estimation precision among different models. Figure 2 shows the VPC results, where high variability of the observed data is consistent with Fig. 1. Overall, the model reasonably described the observed data trends.

shown that this representation of drug effect is equivalent to that of a change-from-baseline [6], and that DE_{ACR} may be interpreted as the baseline of the latent variable [2]. The mechanistic interpretation of the latent variable representation of Eqs. 5–8 is the same as that of Eqs. 1–4 [2], which allows for easy exploration of shared parameters between the DAS28 and ACR models.

DAS28-ACR joint model and residual correlation

Equations 1–8 were first fitted to DAS28 and ACR data separately, and then simultaneously with shared parameters explored. The joint model allows the potential residual correlation between ACR and DAS28 responses to be modeled using a bivariate normal distribution of $\varepsilon = (\varepsilon_{DAS28}, \varepsilon_{ACR})$ with a correlation parameter ρ. Further details on the joint model likelihood specification and NONMEM implementation code were given in Hu et al. [12].

Model estimation and evaluation

The sequential PK/PD modeling approach was used by first fixing the previously obtained individual empirical Bayesian estimates of PK parameters. PK/PD model parameter estimation was implemented in NONMEM, using the LAPLACE method for early exploration and the importance sampling (IMP) method for key model runs. The NONMEM objective function value (OFV) approximates −2 times loglikelihood. A change in NONMEM objective function value of 10.83, corresponding to a nominal *p* value of 0.001, was used as a criterion of including an additional parameter. Visual predictive check (VPC) was used for model evaluation by simulating 500 replicates of the dataset and comparing simulated and model-predicted DAS28 score and ACR response frequencies over the treatment period.

Table 1 DAS28 exposure–response model parameter estimates

Fixed and residual effect estimate (SE)	b	p_{DAS28}	$IC_{50,DAS28}$ (µg/mL)	E_{max}	r_{DAS28} (1/day)	$k_{out,DAS28}$ (1/day)	b_s	ρ^a	σ_{DAS28}
Initial separate model	4.91 (0.0436)	0.161 (0.0104)	0.132 (0.0681)	0.23 (0.0145)	9.54×10^{-3} (7.46×10^{-4})	0.16 (0.03)	0.961 (0.0578)	NA	0.632 (5.75×10^{-3})
Final joint model	4.93 (0.0428)	0.15 (9.72×10^{-3})	1.24 (0.0742)	0.221 (0.0296)	1.01×10^{-2} (7.89×10^{-4})	0.125 (0.0154)	0.932 (0.0566)	−0.655 (0.0103)	0.624 (5.48×10^{-3})

Variance–covariance matrix of between-subject random effect estimate (SE)

	b	p_{DAS28}	$IC_{50,DAS28}$	E_{max}
Initial separate model				
b	0.0177 (1.26×10^{-3})			
p_{DAS28}	0.0386 (9.84×10^{-3})	1.95 (0.203)		
$IC_{50,DAS28}$	0.0605 (0.0294)	−0.838 (0.564)	1.6 (0.652)	
E_{max}	−0.0506 (0.0114)	0.173 (0.156)	−0.0596 (0.354)	0.91 (0.117)
Final joint model				
b	0.0201 (1.24×10^{-3})			
p_{DAS28}	0.0248 (0.0101)	2.03 (0.241)		
$IC_{50,DAS28}$	0.0887 (0.0403)	−0.89 (1.15)	3.29 (2.25)	
E_{max}	−0.0197 (0.0102)	0.101 (0.148)	0.193 (0.531)	1.12 (0.237)

Between-subject random effect parameters are for the appropriately transformed scale: logit-normal for E_{max} and lognormal for the rest of the parameters

DAS28 28-joint disease activity score, *SE* standard error, *b* baseline, p_{DAS28} fraction of maximum placebo effect, $IC_{50,DAS28}$ potency, E_{max} fraction of maximum drug effect, r_{DAS28} rate of placebo effect onset, $k_{out,DAS28}$ disease amelioration rate for DAS28, b_s baseline DAS28 difference between Study 1 and Study 2, ω_b^2 variance of between-subject variability for baseline DAS28, σ_{DAS28}^2 variance of within-subject variability, ρ residual correlation between ACR and DAS28 score, *NA* not available

[a] Shared between ACR and DAS28 models

Fig. 2 Visual predictive check of the 28-joint disease activity (DAS28) score for the initial model. The 5th, 50th and 95th percentiles of observed DAS28 scores are overlaid with the 90 % prediction intervals (PI) of their model predictions at planned observation times by treatment. *PBO* placebo
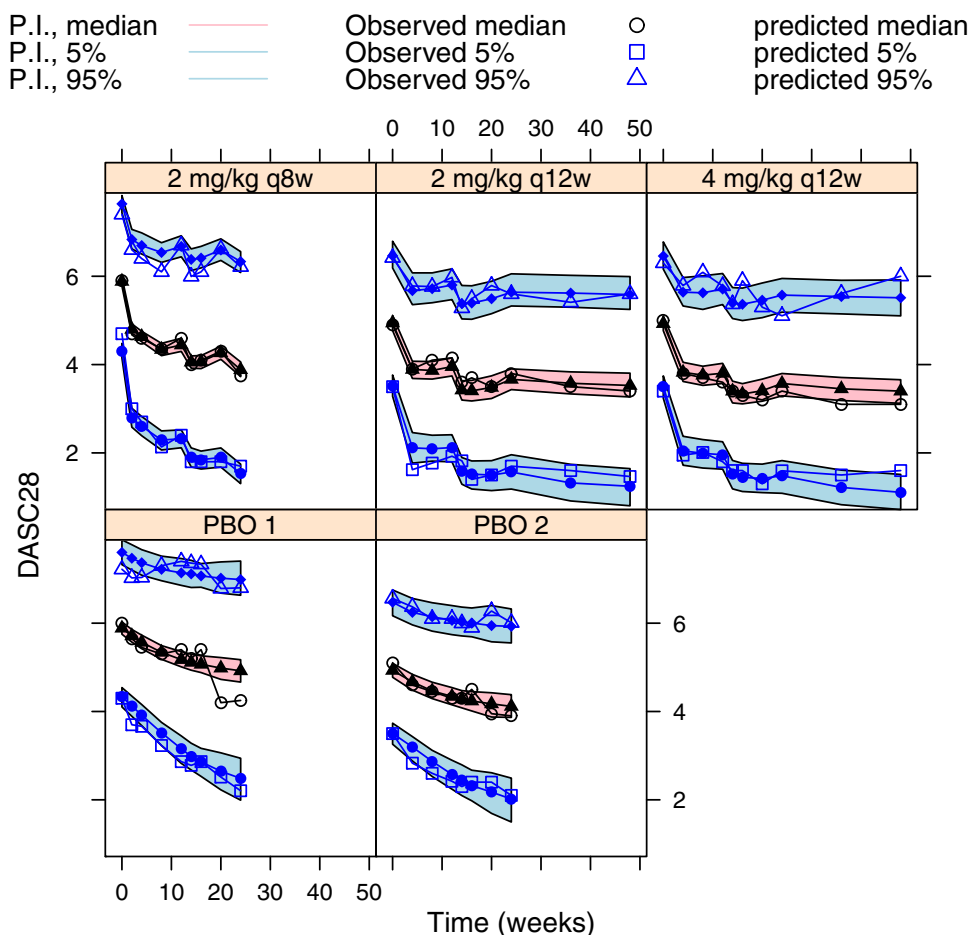
Initial ACR model

The model described by Eqs. 5–8 was applied to the ACR response data. Table 2 shows the parameter estimates. Estimation precision was reasonable. IC_{50} and K_{out} estimates were similar to those obtained previously [6]. From a theoretical perspective, DE estimate from logistic regression could be expected to be larger than from probit regression by approximately a factor of 1.8 ($\cong \sqrt{\pi^2/3}$), due to the fact that variances for the standard logistic and the standard normal distributions are $\pi^2/3$ and 1 respectively. Taking account of this difference, differences between the DE estimates did not appear to be unexpected. Other parameter estimates were different in part due to the earlier use of a reduced placebo model. Estimation precision was reasonable, with SEs at generally a magnitude lower than the estimates. VPC results are shown in Fig. 3, which are similar to those in the previous analyses [6] as could be expected [2]. As previously noted, the high observed placebo responses at Weeks 20 and 24 may be due to early escape [6].

Joint E–R model

The base scenario of fitting Eqs. 1–8 simultaneously with no shared parameters between the DAS28 and ACR model component is equivalent to estimating the DAS28 and ACR models separately. Indeed, the sum of the OFVs of the DAS model and ACR models was nearly identical to the OFV of the simultaneously fitted model with no shared parameters. From Tables 1 and 2, the rate parameters, namely r_{DAS28} and k_{out}, appeared similar between the endpoints, along with IC_{50}. Indeed, this was substantiated by the joint model evaluations showing insignificant OFV change for sharing each parameter in the joint model.

This result motivated the question of whether further similarities between the endpoints could be found. It could be hypothesized that, based on binding, a single latent variable could govern both endpoints through IDR models. The similarity of the K_{out} parameters suggested the possibility of using a single IDR model instead of two. If so, the latent variable may not be separately identifiable from DAS28. The question however is in what sense the placebo

Table 2 ACR exposure–response model parameter estimates

Parameter estimate (SE)	α_2	d_1	d_3	$r_{p,ACR}$ (1/day)	k_{out} (1/day)	IC_{50} (µg/mL)	DE_{ACR}	$F_{p,ACR}$	Lm	ω^2
Initial separate model	−1.47 (0.121)	1.06 (0.0321)	1.29 (0.0266)	0.00916 (1.17×10^{-3})	0.111 (0.0209)	0.168 (0.043)	1.43 (0.0948)	1.93 (0.0961)	NA	1.69 (0.111)
Final joint model	−4.69 (0.0826)	1.21 (0.0356)	1.43 (0.0297)	NA	NA	NA	NA	NA	0.125 (2.79×10^{-3})	0.38 (0.0397)

ACR American College of Rheumatology disease severity index, *DAS28* 28-joint disease activity score, *SE* standard error, α_2, d_1, d_3 intercepts parameters, $r_{p,ACR}$ rate of placebo effect onset, $k_{out,ACR}$ disease amelioration rate, $IC_{50,ACR}$ potency, DE_{ACR} drug effect, $F_{p,ACR}$ fraction of maximum placebo effect onset, *Lm* maximum latent variable effect, ω^2 variance of between-subject variability, *NA* not applicable

and drug effect could be considered similar for both endpoints, particularly because DAS28 and ACR endpoints have different scales. This rationale motivated the use of DAS28 change-from-baseline as ACR endpoints are defined as change-from-baseline and the latent variable is defined only up to a constant [2, 4]. Following the notation and motivation provided by Hu [2], this led to the following ACR model in place of Eqs. 5–8:

$$\Phi^{-1}[prob(ACR \leq k)] = \alpha_k - M(t) + \eta \qquad (9)$$

where $M(t) = L_m[f_{DAS28,p}(t) + f_{DAS28,d}(t)]/b$ is the scaled change-from-baseline DAS28 score model prediction given in Eq. 1, with BSV terms as given in the initial DAS28 score model.

Fitting Eqs. 1–4 and 9 simultaneously to the DAS28 and ACR response data resulted in a NONMEM objective function decrease of over 2000, indicating a significant improvement of the fit, despite using four fewer parameters than the base scenario. Including the residual correlation between DAS28 and ACR responses further reduced the NONMEM objective function by over 1900. This was considered as the final model, with parameter estimates of the DAS28 and ACR response components given in Tables 1 and 2, respectively.

Table 1 shows that for DAS28, the joint model parameter estimates and associated SEs were generally similar to those obtained with the initial model using only DAS28 data. Table 2 shows that for ACR, the joint model used no additional parameters for the placebo and drug effects other than the scaling parameter L_m. Estimates of the intercept parameter α_2 between the initial and joint models are not directly comparable, as the average intercept value is determined up to a constant with the latent variable [2]. Estimates of d_1 and d_3, the intercept differences, were similar between the initial and joint models. The estimate of ω^2 was smaller in the joint model, due to the fact that the treatment effect predictor $M(t)$ in Eq. 9 contains BSV components whereas Eqs. 6–8 do not. This contribution of BSV components of the DAS28 model is the main explanation for the improved fit of the joint model in light of the NONMEM objective function decrease of over 1900 mentioned above. Coupled with a high absolute value of correlation parameter estimate (0.655) shown in Table 1, this confirms that the two endpoints measure the same component of the disease.

VPC results of the joint model for DAS28 was visually indistinguishable from Fig. 2, and thus is not shown. This is consistent with the similarity between the parameter estimates of the joint and the separate models. VPC results of the joint model for the ACR response are shown in Fig. 4. It is noted that the VPC results of the joint model before incorporating the residual correlation component were visually indistinguishable from Fig. 4 and are not shown. The results appeared largely similar to Fig. 3;

Fig. 3 Median model
predictions at planned
observation times and 90 %
prediction intervals (PI), in
overlay with observed
American College
Rheumatology (ACR) response
frequencies, for the initial ACR
model. ACR20/50/70,
20 %/50 %/70 % improvement
in the American College of
Rheumatology criteria
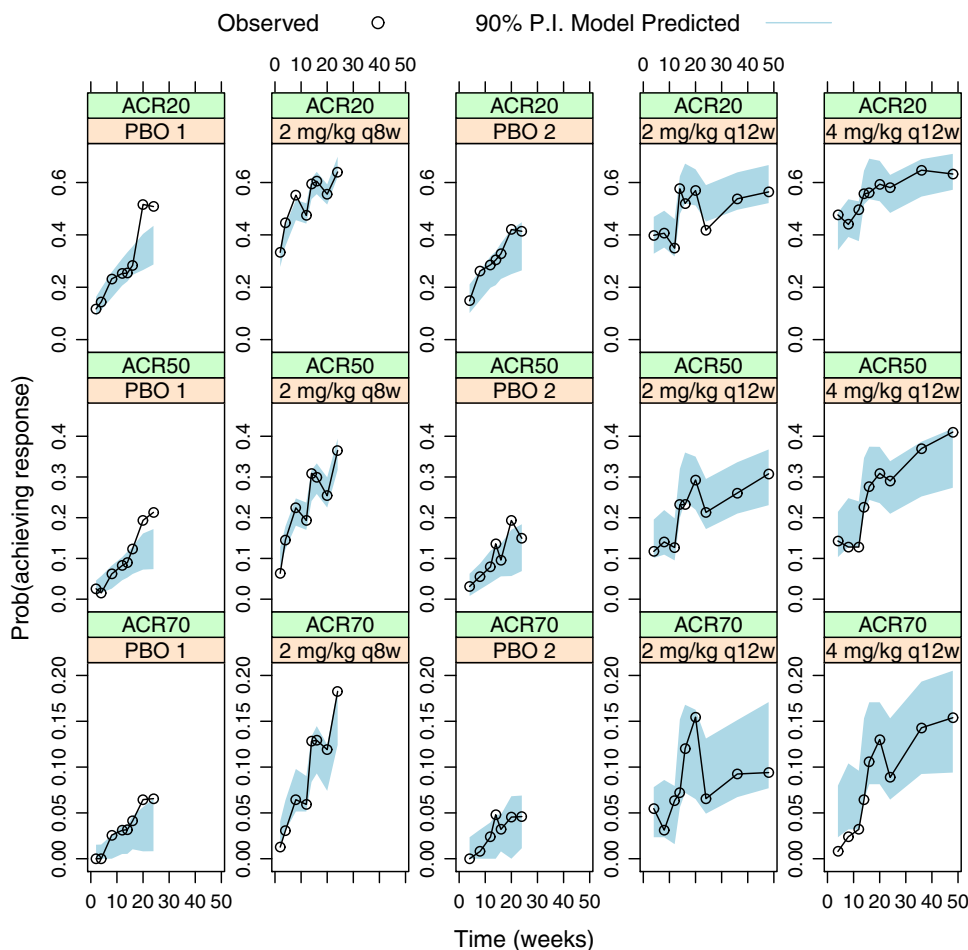
where differences occurred, it may appear difficult to determine whether the joint model or the individual ACR model prediction better represent reality. However, it is reasonable to expect that the ACR70 response rates for the placebo arms to be small in early treatment periods, and the joint model prediction in Fig. 4 is better than the separate ACR model prediction in Fig. 3. This may be attributed to the more reasonable partitioning of BSV onto additional model parameters under the joint model instead of only at the intercept level under the separate ACR model. It can be seen that under Eq. 5, the expected ACR response rate at time of (or near) 0 is given by [2, 17]

$$E\{\text{prob}[ACR \leq k]\} = \Phi(\alpha_k / \sqrt{1 + \omega^2}) \qquad (10)$$

and that the larger predicted ACR70 response rate in the separate model was caused by the larger ω^2 estimate. In contrast, the joint model had part of this BSV component partitioned onto other parameters, namely through BSV terms on b, p_{DAS28}, $IC_{50,DAS28}$ and E_{max}, none from which contributed to the BSV of ACR responses at time near 0 since $M(0) = 0$, which led to the smaller overall BSV predictions near time 0. This therefore suggested that the latent variable joint model may partition the BSV more

appropriately by allowing it to vary over time, as opposed to remaining constant under the separate model.

In order to understand the source of improvement in the joint model fitting, it may be tempting to examine the OFV changes in the DAS component and ACR component. However this would be exceedingly difficult, because the two components are not independent without conditioning on the BSV terms. Nevertheless there are two relevant observations in comparing the separate and the joint models: (1) in the DAS component, changes of parameter estimates and associated SE were minor, and the VPC results were virtually unchanged; (2) in the ACR component, the original BSV variance ω^2 was markedly reduced in the joint model. These suggested that the improvement in OFV mostly came from the ACR component, through the presence of the additional BSV terms in the latent variable.

Discussion

Our analysis supported and utilized the relatedness between DAS28 and ACR responses that are designed to measure the same disease component through shared

Fig. 4 Median model predictions at planned observation times and 90 % prediction intervals (PI), in overlay with observed American College Rheumatology (ACR) response frequencies, for the final joint model. ACR20/50/70, 20 %/50 %/70 % improvement in the American College of Rheumatology criteria
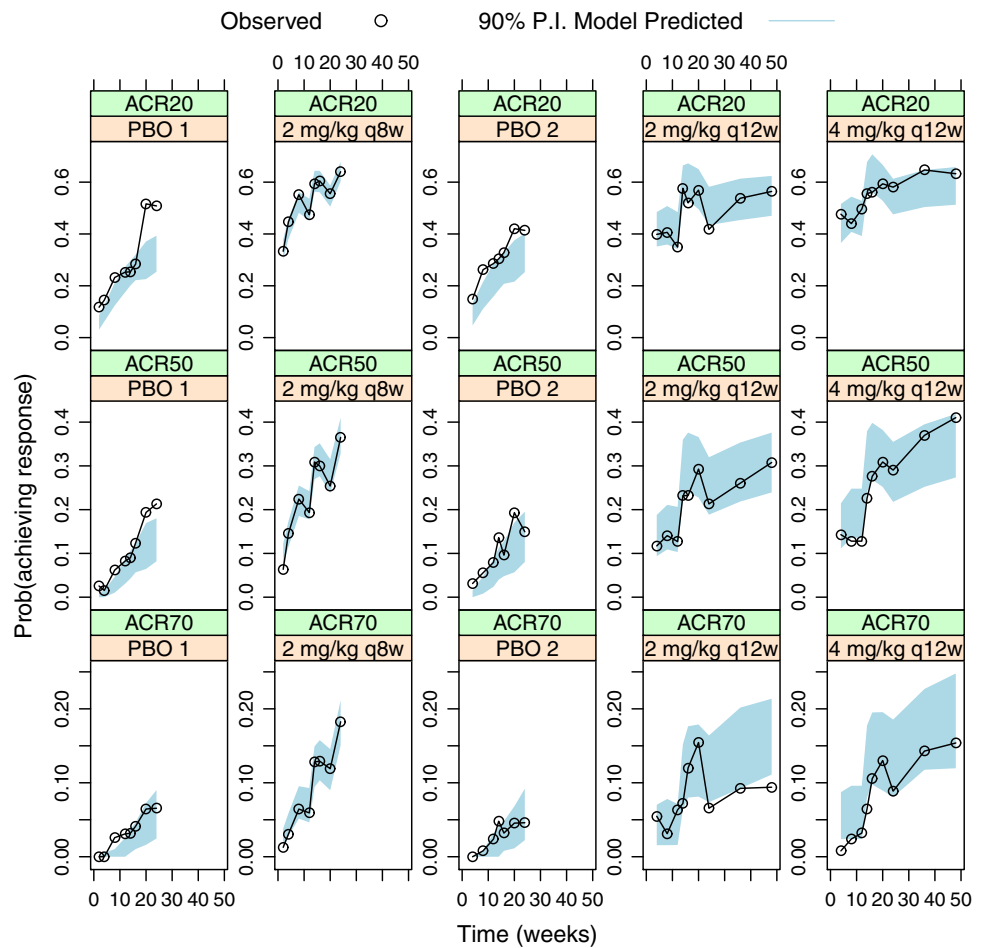
structural model component along with the related BSVs and the residual correlation. A common view on the modeling of multiple endpoint is that, while joint modeling may improve the overall fit as measured by the likelihood or predictions of correlated responses, it would not affect the descriptions or predictions of the individual endpoints [11]. On the other hand, the joint model for DAS28 and ACR responses developed here improved the characterization of the ACR response, in a manner unrelated to the residual correlation component. This is due to the fact that the number of random effects used to describe ACR response actually increased by three under the joint model, namely through the BSV terms under the latent variable M(t) in Eq. 9. The estimation of these additional random effects could not be reliably supported by ACR data alone [6], and was made possible only with the added DAS28 data under the joint model framework. This feature may hold more generally when the endpoints include both the continuous and the categorical types, where the continuous endpoint data may support random effects not estimable from the categorical endpoint data alone, and thus resulting in better description of the categorical endpoints. This

demonstrates more clearly the improvement achievable by the joint modeling approach than the previous applications [11, 12]. It is noted that the joint approach requires considerably more effort and computational time. On the other hand, categorical endpoint modeling alone typically cannot support the estimation of more than one BSV term. It is noted that the magnitude of BSV at the intercept level determines the response probability together with underlying exposure, as can be seen in Eq. 10 (see more details in Hutmacher and French [17]). The question thus arises on how this would affect model predictions, especially the associated variability. The joint model can provide valuable insights into this important question in clinical development.

A study effect term was used to account for baseline DAS28 differences between the two studies. Estimation results of the remaining parameters were similar to those in an earlier model without the study effect on baseline. However, the VPC of this earlier model showed systematic differences between the model predicted and observed data trends, which could easily lead to doubt regarding the predictive ability of the model. The study effect has

descriptive but not predictive value, and the inclusion of such covariates should be exercised with caution. As subjects are randomized only within the studies and not to them, the study effect on baseline served the purpose of adjusting for imbalances between studies. Inclusion of study effect on other model components could easily lead to interpretation difficulties and is thus generally not advisable.

An often encountered practical difficulty with IDR modeling is to explain the nature of the underlying E–R relationship to unfamiliar audiences. It is especially difficult to illustrate how it influences the time course of the response, since the E–R relationship can only be plotted under a theoretical steady-state infusion, which may be of little help to understand the actual response in clinical applications where dosing periods are often at their half-lives or longer. However, IDR models have been used increasingly in E–R modeling of clinical endpoints and have certain advantages compared with other approaches such as the Markov transition approach or direct correlations between the primary efficacy endpoint and simple exposure metrics using area under drug concentration–time curve (AUC) or trough drug concentration (Ctrough) [2]. In practical applications, placebo effects have been modeled in different forms under the IDR models and may be broadly classified as 1-pathway and 2-pathway approaches [2, 18]. In part due to the lack of mechanistic understanding, practical choices were perhaps often made according to the relative ease of implementation in the particular application, and may differ even when multiple endpoints were obtained from the same trials [9]. Maintaining consistency of mechanism interpretations of the model forms for the different endpoints is important to allow joint modeling to fully achieve its benefits.

References

1. Felson DT, Anderson JJ, Boers M et al (1995) American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 38(6):727–735
2. Hu C (2014) Exposure-response modeling of clinical end points using latent variable indirect response models. CPT: Pharmacomet Syst Pharmacol 3:e117. doi:10.1038/psp.2014.15
3. Dayneka NL, Garg V, Jusko WJ (1993) Comparison of four basic models of indirect pharmacodynamic responses. J Pharmacokinet Biopharm 21(4):457–478
4. 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
5. Hu C, Xu Z, Rahman MU, Davis HM, Zhou H (2010) A latent variable approach for modeling categorical endpoints among patients with rheumatoid arthritis treated with golimumab plus methotrexate. J Pharmacokinet Pharmacodyn 37(4):309–321
6. 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
7. 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
8. 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
9. Ma L, Zhao L, Xu Y, Yim S, Doddapaneni S, Sahajwalla CG, Wang Y, Ji P (2014) Clinical endpoint sensitivity in rheumatoid arthritis: modeling and simulation. J Pharmacokinet Pharmacodyn 41(5):537–543. doi:10.1007/s10928-014-9385-x
10. Ringold S, Chon Y, Singer NG (2009) Associations between the American College of Rheumatology pediatric response measures and the continuous measures of disease activity used in adult rheumatoid arthritis. Arthritis Rheum 60(12):3776–3783
11. 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
12. 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 41:335–349
13. Beal SL, Sheiner LB, Boeckmann A, Bauer RJ (2009) NONMEM user's guides (1989-2009). Icon Development Solutions, Ellicott City
14. Weinblatt ME, Bingham CO, Mendelsohn AM, Kim L, Mach M, Lu J, Baker D, Westhovens R (2012) Intravenous golimumab is effective in patients with active rheumatoid arthritis despite methotrexate therapy with responses as early as week 2: results of the phase 3, randomised, multicentre, double-blind, placebo-controlled GO-FURTHER trial. Ann Rheum Dis 72:381–389
15. Kremer J, Ritchlin C, Mendelsohn A, Baker D, Kim L, Xu Z, Han J, Taylor P (2010) Golimumab, a new human anti-TNFalpha antibody, administered intravenously in patients with active rheumatoid arthritis: 48-Week efficacy and safety results of a phase 3, randomized, double-blind, placebo-controlled study. Arthritis Rheum. doi:10.1002/art.27348
16. Hu C, Zhou H (2008) An improved approach for confirmatory phase III population pharmacokinetic analysis. J Clin Pharmacol 48:812–822
17. Hutmacher MM, French JL (2011) Extending the latent variable model for extra correlated longitudinal dichotomous responses. J Pharmacokinet Pharmacodyn 38:833–859
18. Hutmacher MM, Mukherjee D, Kowalski KG, Jordan DC (2005) Collapsing mechanistic models: an application to dose selection for proof of concept of a selective irreversible antagonist. J Pharmacokinet Pharmacodynam 32:501–520