Citation: CPT Pharmacometrics Syst. Pharmacol. (2014) 3, e143; doi:10.1038/psp.2014.41

www.nature.com/psp

ORIGINAL ARTICLE

Simultaneous Exposure–Response Modeling of ACR20, ACR50, and ACR70 Improvement Scores in Rheumatoid Arthritis Patients Treated With Certolizumab Pegol

BD Lacroix[1,2], MO Karlsson[1] and LE Friberg[1]

The Markovian approach has been proposed to model American College of Rheumatology's (ACR) response (ACR20, ACR50, or ACR70) reported in rheumatoid arthritis clinical trials to account for the dependency of the scores over time. However, dichotomizing the composite ACR assessment discards much information. Here, we propose a new approach for modeling together the three thresholds: a continuous-time Markov exposure–response model was developed, based on data from five placebo-controlled certolizumab pegol clinical trials. This approach allows adequate prediction of individual ACR20/50/70 time-response, even for non-periodic observations. An exposure–response was established over a large range of licensed and unlicensed doses including phase II dose-ranging data. Simulations from the model (50–400 mg every other week) illustrated the range and sustainability of response (ACR20: 56–68%, ACR50: 27–42%, ACR70: 11–22% at week 24) with maximum clinical effect achieved at the recommended maintenance dose of 200 mg every other week.
CPT Pharmacometrics Syst. Pharmacol. (2014) **3**, e143; doi:10.1038/psp.2014.41; published online 29 October 2014

The American College of Rheumatology's (ACR) core dataset[1] has caused a major improvement in the standardization of clinical trials for rheumatoid arthritis (RA). Response criteria have been defined as 20%, 50%, and 70% improvement from baseline and named ACR20, ACR50, and ACR70. ACR20[2] has been accepted as the efficacy benchmark in RA clinical trials,[3] showing greater discriminant capacity over ACR50 and ACR70 to distinguish active treatment from placebo control.[4] However, 20% improvement does not represent optimal clinical progress and is not a meaningful clinical response for rheumatologists.[5] In clinical trials conducted since 1997, similar discriminant capacities were demonstrated for ACR20 and ACR50 to distinguish active versus control treatments, which seems to attest the superiority of newer therapeutic approaches and drugs, achieving higher level of ACR response.[6] ACR50 and ACR70 are more desirable targets for patients and provide useful information in addition to ACR20.[6,7] Dichotomizing the composite ACR assessment into binary variables such as ACR20, ACR50, and ACR70 discards information.[8,9] Classification as "ACR20 responder" ignores whether the subject actually reached 50 or 70% improvement over baseline. Thus, simultaneous modeling of the three available endpoints ACR20, ACR50, and ACR70 is a more effective use of the data.[10] ACR-N[11–13] or hybrid-ACR[14] analysis would represent a further step toward use of a continuous measure based on ACR assessment.

Successive ACR assessments collected repeatedly over time for each individual are not independent.[15] Not accounting for this property in the implementation of ordered logistic regression, the standard method to model ordered categorical data[16] is expected to lead to over-prediction of the number of transitions between the different grades.[17] The inclusion of Markov elements in the logistic regression[18] lead to better characterization of the transitions between response and non-response for the binary score ACR20.[15] Nevertheless, the complexity of the model increases dramatically with the number of categories and constant influence of preceding score(s) is implicitly assumed, whatever the observation schedule. The latent variable (LV) approach, that relates the observed scores to an underlying continuous measure of the disease activity,[19] was proposed to model the three ACR scores simultaneously.[20,21] This approach essentially focuses on describing the average probability of the population of subjects and was shown to over-predict the number of transitions between scores when applied to stand-alone ACR20 modeling in its simple implementation.[15] It was recently extended to accommodate extra-correlation between longitudinal dichotomous data[22] but remains complex to implement and is not integrated in standard software packages. The continuous-time Markov model was first proposed in the PKPD literature to model the tablet position in the gastrointestinal tract.[23] It was applied to characterize various types of ordered categorical data, such as pain in animal models[24] and side effects in clinical trials: grades of proteinuria in cancer patients treated with anti-angiogenic drug[17] or extrapyramidal side effects in schizophrenic patients treated with antipsychotic drugs.[25] This approach addresses the dependence between successive observations with the influence of the previous score decreasing with increasing time between observations. It is more sparing than inclusion of Markov elements in proportional odds models when more than two categories are considered and allows for accurate simulations for various observation schedules.

In this article, we present the development of a continuous-time type mixed-effect Markov model characterizing the ACR20, ACR50, and ACR70 responses as a function of certolizumab

[1]Department of Pharmaceutical Biosciences, Pharmacometrics Group, Uppsala University, Uppsala, Sweden; [2]UCB Pharma, Braine-l'Alleud, Belgium. Correspondence: BD Lacroix (brigitte.lacroix@ucb.com)
Received 11 April 2014; accepted 7 August 2014; published online 29 October 2014. doi:10.1038/psp.2014.41

pegol exposure from five phase II and phase III clinical trials including both licensed and unlicensed doses. Certolizumab pegol is PEGylated Fc free anti-TNF (Fc = crystallizable fragment) licensed for treatment of moderate to severely active RA in several countries, including United States and Europe, at the dose of 400 mg initially and at weeks 2 and 4, followed by 200 mg every 2 week (Q2W); for maintenance dosing, 400 mg every 4 weeks (Q4W) can also be considered.

RESULTS

Data

The patient population and study design of the five double-blind placebo-controlled clinical trials[26–29] have been described previously.[15] Data from 2,380 subjects with RA were analyzed (24,519 observations); 1,747 subjects received certolizumab pegol at doses ranging from 50 to 800 mg, and 633 subjects received placebo. Treatment was administered subcutaneously Q2W or Q4W for 12 weeks up to 1 year, as monotherapy or in combination with methotrexate (patients were on stable dose at inclusion), with a loading dose for the two pivotal studies (**Table 1**). The predicted plasma concentrations at the time of ACR assessment (Cp) ranged from 0.4 to 174 µg/ml. Typical steady-state average concentration for the 200 mg Q2W and 400 mg Q4W label treatments was 23.2 µg/ml.

Structural ACR model

At each visit, subjects were classified as ACR20 non-responder (ACR-NR, score = 0), ACR20 but not ACR50 responder (ACR20-50, score = 1), ACR50 but not ACR70 responder (ACR50-70, score = 2), or ACR70 responder (score = 3). The probabilities of the scores, modeled as compartment amounts, were defined by four ordinary differential equations, the sum of the probabilities of the four responder states remaining equal to 1 at any time point (**Figure 1**). At the time of observation, the actual score is known, i.e., the probability of observing this score is equal to 1. The probabilities of the other responder states and the integration time were reset to 0 (see example dataset in **Supplementary Material**). Between two observations, the probability "amount" distributes with time between the different states, with a rate determined by the transition parameters. At each time, the total probability "amount" in the different states/compartments is 1 as the system is closed (**Figure 2**).

$$\begin{aligned}
\frac{d\Pr(0)}{dt} &= K_{01}\cdot\Pr(1) - K_{10}\cdot\Pr(0)\\
\frac{d\Pr(1)}{dt} &= K_{10}\cdot\Pr(0) + K_{12}\cdot\Pr(2) - (K_{01}+K_{21})\cdot\Pr(1)\\
\frac{d\Pr(2)}{dt} &= K_{21}\cdot\Pr(1) + K_{23}\cdot\Pr(3) - (K_{12}+K_{32})\cdot\Pr(2)\\
\frac{d\Pr(3)}{dt} &= K_{32}\cdot\Pr(2) - K_{23}\cdot\Pr(3)
\end{aligned} \qquad (1)$$

with $K_p = \mathrm{TVK}_p\cdot\exp(\eta_{p,i})$ and $\mathrm{TVK}_p = f_p\,(\mathrm{K}_{p,0}, \mathrm{g}_p(t), \mathrm{h}_p(q_{ij}))$

where Pr(0) to Pr(3) are the probability of the scores from ACR20-NR to ACR70, the K_p are the transfer rate parameters between the compartments where the first digit of the rate parameter represents the current score and the second digit the preceding score (e.g., K_{21} can be read as the transfer rate to score 2 given current score of 1). $K_{p,0}$ reflects the baseline value of the *p*th transfer rate parameter at the time of the first dose of study medication, $g_p(t)$ and $h_p(q_{ij})$ the time and drug effects on the transfer rate, t the continuous time and q_{ij} the exposure measurement for the *j*th observation of *i*th subject. The predicted concentration at the time of observation, Cp, was selected as marker for drug exposure based on the objective function value (OFV).

The $\eta_{p,i}$, representing the random effect around parameter K_p for subject i, was assumed to have a normal probability distribution of mean zero and variance ω^2; its empirical Bayes predictions allowed for subject-specific predictions of the corresponding transfer rate.

Final model parameters are summarized in **Table 2**. Whenever possible, the model was reduced by combining effect parameters on different transfer rates. All patients were considered as ACR20-NR at baseline (time = 0) and the score could improve naturally after this time point. Later in the study, scores tended to worsen with time in placebo treated patients: transfer rate to lower scores were greater than transfer rate to higher scores, i.e., an ACR20 responder subject receiving placebo was more likely to revert back to ACR20-NR than to become ACR50 responder.

A positive exposure–response relationship was identified. *E*max-models with positive E_{max} value (K_{10}) and linear functions with positive slope (K_{21} and K_{32}, common slope) described the effect of the drug concentration on upward transfer rate parameters. *E*max-models with negative estimates of E_{max}, described the drug effect on downward rate parameters (K_{01}, K_{12}, $K_{23,}$ common E_{max}). K_p were restricted to be positive by setting the lower boundaries of the initial parameter estimates.

$$K_p = \left(K_{p,0}\cdot\left(1+\mathrm{slt}_p\cdot\mathrm{time}\right)\cdot\left(1+E_{\max\,p}\cdot\frac{\mathrm{Cp}}{\mathrm{EC}_{50}+\mathrm{Cp}}\right)\right)\cdot\exp(\eta_p)$$

for K_{10}, K_{01}, K_{12}, and K_{23}

$$K_p = \left(K_{p,0}\cdot\left(1+\mathrm{slt}_p\cdot\mathrm{time}\right)\cdot\left(1+\mathrm{sld}_p\cdot\mathrm{Cp}\right)\right)\cdot\exp(\eta_p)$$

for K_{21} and K_{32}

where slt_p and sld_p represent the slopes of the linear relationship to continuous time and concentration, respectively, $E_{\max\,p}$ is the maximum change in the pth transfer constant value as a function of exposure measure and EC_{50} the concentration to reach half of the maximum effect related to Cp, common to the relevant transfer rate parameters.

Incorporation of a full variance-covariance matrix characterizing the correlations between the transition rate parameters markedly improved the fit of the model to the data (decrease in OFV of 338 units) and the simulation performance of the model. Transfer rate parameters to higher ACR were highly correlated (K_{10}, K_{21}: 62%, K_{10}, K_{32}: 60%, K_{21}, K_{32}: 82%). Transfer rate parameters to lower ACR were less correlated, except 79% for K_{01}, K_{12}. However, inclusion of these 15 additional parameters dramatically increased the run time, and resulted in failure of the covariance step.

Upward transfer rates were decreased (lower ACR response) with increased baseline swollen joint count and concomitant corticosteroid treatment. Downward transfer rates increased with increased age, resulting in decreased ACR response for elderly subjects. Upward transfer rates

Table 1 Study design and available data

Study	Duration (weeks)	Active doses (mg)	Control	Dosing schedule[a]	Formulation	Pharmacodynamic assessment (weeks)	No. of subjects[b] Total/active	Dropout frequency (%) Active/placebo
1 (phase II)	8	50, 100, 200, 400, 600, 800	Placebo	Q4W	Liquid pH5.5	1, 2, 4, 5, 6, 8, 9, 10, 12	320/239	11/37
2[c] (phase III)	24	400	Placebo	Q4W	Lyophilized solution	1, 2, 4, 8, 12, 16, 20, 24	219/111	32/73
3[c] (phase III)	24	400	Placebo+ MTX[a]	Q4W	Lyophilized solution	1, 2, 4, 8, 12, 16, 20, 24	243/124	21/45
4[c] (phase III)	52	Load[d]+200, 400	Placebo + MTX	Q2W	Lyophilized solution	1, 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52	979/780	32/78[e]
5[c] (phase III)	24	Load[d]+200, 400	Placebo+MTX	Q2W	Liquid pH 4.7	1, 2, 4, 6, 8, 12, 14, 16, 20, 24	619/493	28/87[e]

[a]MTX, methotrexate; Q2W, every other week; Q4W, every 4 weeks. [b]All subjects who received at least one study treatment dose and provided at least one certolizumabpegol concentration were included in the analysis. [c]Studies 2, 4, and 5 are the FAST4WARD,[26] RAPID 1 (ref. 28), and RAPID 2 (ref. 29) studies, respectively. Study 3 was also reported.[27] [d]Load = 400 mg at weeks 0, 2, and 4. [e]Includes the dropout defined by study design for subjects who were assessed as ACR20 non-responders at weeks 12 and 14.

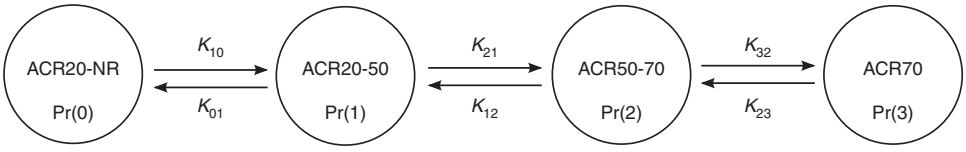


Figure 1 Transition model for the four ACR scores based on the clinical responses ACR20, ACR50, and ACR70.

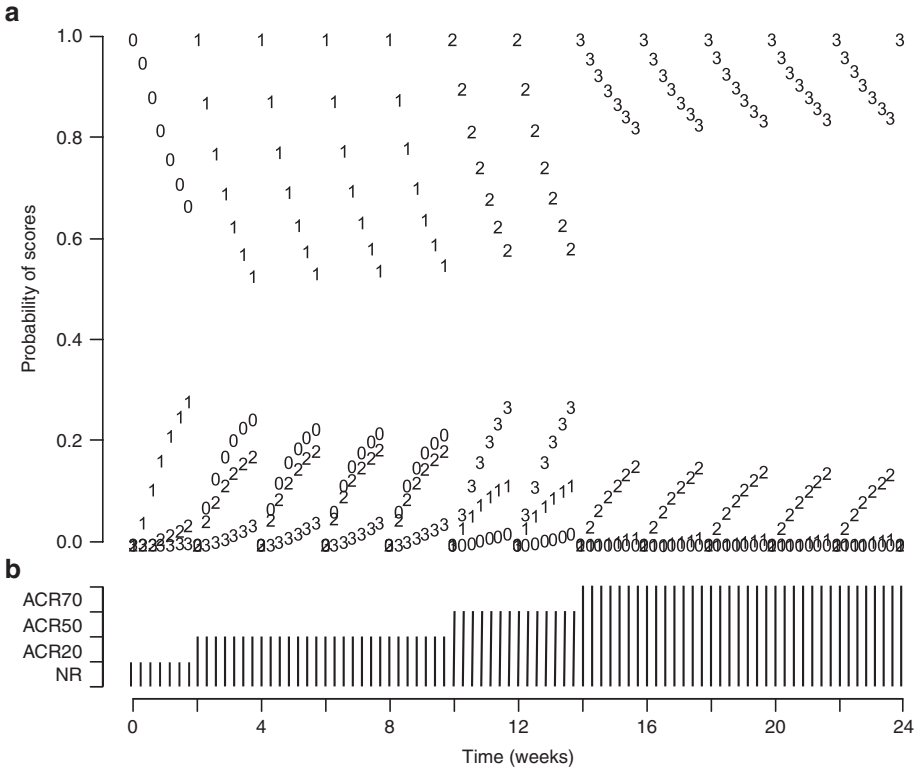


Figure 2 Individual profile, observation every other day. (**a**) Evolution of the probability of the four modeled scores with time for a subject reaching ACR70 response. 0, 1, 2, and 3 symbols represent the probability of the modeled scores ACR20-NR, ACR20-50, ACR50-70, and ACR70 scores, respectively; (**b**) corresponding observed ACR grades from non-responder to ACR70.

were higher for subjects originated from Central and Southern America than for North Americans and Western, Eastern, and Northern Europeans.

The concomitant use of methotrexate was not found to affect the probability of ACR response. This is not surprising, as patients had to be on a stable methotrexate dosage

upon enrollment in the relevant studies. The ACR assesses improvement from baseline, thus the methotrexate effect should already be included in the baseline assessment.

Dropout model

The subjects' dropout status (yes/no) was assessed at each visit. The dropout probability was modeled using a logistic model. No random effect around the dropout probability was estimated because subjects could only drop out once during the trial. The increase in dropout with time from the start of the trial was described by a Hill function on the logit scale where the time to reach half of the maximum dropout was estimated to approximately 3 weeks. In addition, ACR70 and ACR50 responders had the same, lower probability of dropping out than ACR20 responders, who had lower probability of dropping out than ACR20 non-responders:

$$\begin{aligned} \text{PACR} &= \varphi_{\text{ACR20}} \cdot \text{ACR20} + \theta_{\text{ACR50}} \cdot \text{ACR50} + \theta_{\text{ACR70}} \cdot \text{ACR70} \\ \text{Logit} &= \varphi + \frac{E_{\max T} \cdot \text{time}^{\gamma}}{T_{50}^{\gamma} + \text{time}^{\gamma}} + \text{PACR} \\ \text{Pr}_{\text{dropout}} &= \frac{\exp(\text{Logit})}{1 + \exp(\text{Logit})} \end{aligned} \quad (3)$$

where φ reflects the baseline (time = 0) logit score of the dropout, $E_{\max T}$ the maximum change in the logit related to time, T50 the time to reach half of this maximum change. The parameters θ_{ACR20}, θ_{ACR50}, and θ_{ACR70} reflects the effects of previous ACR outcome (ACR20 = 1 if the patient was an ACR20-50 and 0 otherwise, similarly for ACR50-70 and ACR70).

Model evaluation

The model simulated well the observed ACR levels (**Figure 3a**) and the observed proportions of transitions between the ACR levels (**Figure 3**, drug data in panel b and placebo data in panel c) for the whole dataset. For a few of the schedules, e.g., 200 mg Q2W, some small misfit is apparent when the ACR clinical scores by treatment were predicted (**Figure 3d**). Inter-study or inter-treatment variability was not included, that may impact the fit of individual studies.

Simulations

Simulations were performed to illustrate the exposure-response relationship. Typical PK exposure was derived from the population PK model (71.5 kg subject who did not develop anti-drug antibody) and the ACR scores simulated from the continuous Markov model. For example, at week 12, the rate of response ranged from 53.7 to 67.3% for ACR20, from 22.4 to 35.0% for ACR50 and from 6.5 to 14.4% for doses ranging from 50 mg Q2W to 400 mg Q2W (**Figure 4**). This range of doses extending outside the label dosages was chosen for illustration purpose, as a plateau in the response is rapidly reached and the dose of 400 mg Q2W is not markedly better in term of efficacy than the 200 mg Q2W dose. Subjects that dropped out were considered as non-responders in the computation of these statistics, corresponding to what was done in the traditional statistical analysis of the performed clinical trials (worst case scenario). At week 24, the ACR20 outcome from the two pivotal trials RAPID 1 (ref. 28) and RAPID 2 (ref. 29) were somewhat lower than simulated from the model (for 200 mg Q2W: 58.8 and 57.3% observed ACR20 response vs. 65.2% simulated, 37.1 and 37.7% observed ACR50 response vs. 37.7% simulated, 21.4 and 15.9% ACR70 response vs. 16.7% simulated). An explanation could be that compared to the simulations that were based on a typical subject's PK, these studies included subjects with lower exposure due to higher body weight and/or formation of anti-drug antibodies.

To exemplify the potential use of the model for clinical purpose, some other outcomes were derived from the simulations (**Figure 4**). At long term, the probability of responder subjects at week 12 still being responder at week 48 increased with

Table 2 Final model parameter estimates

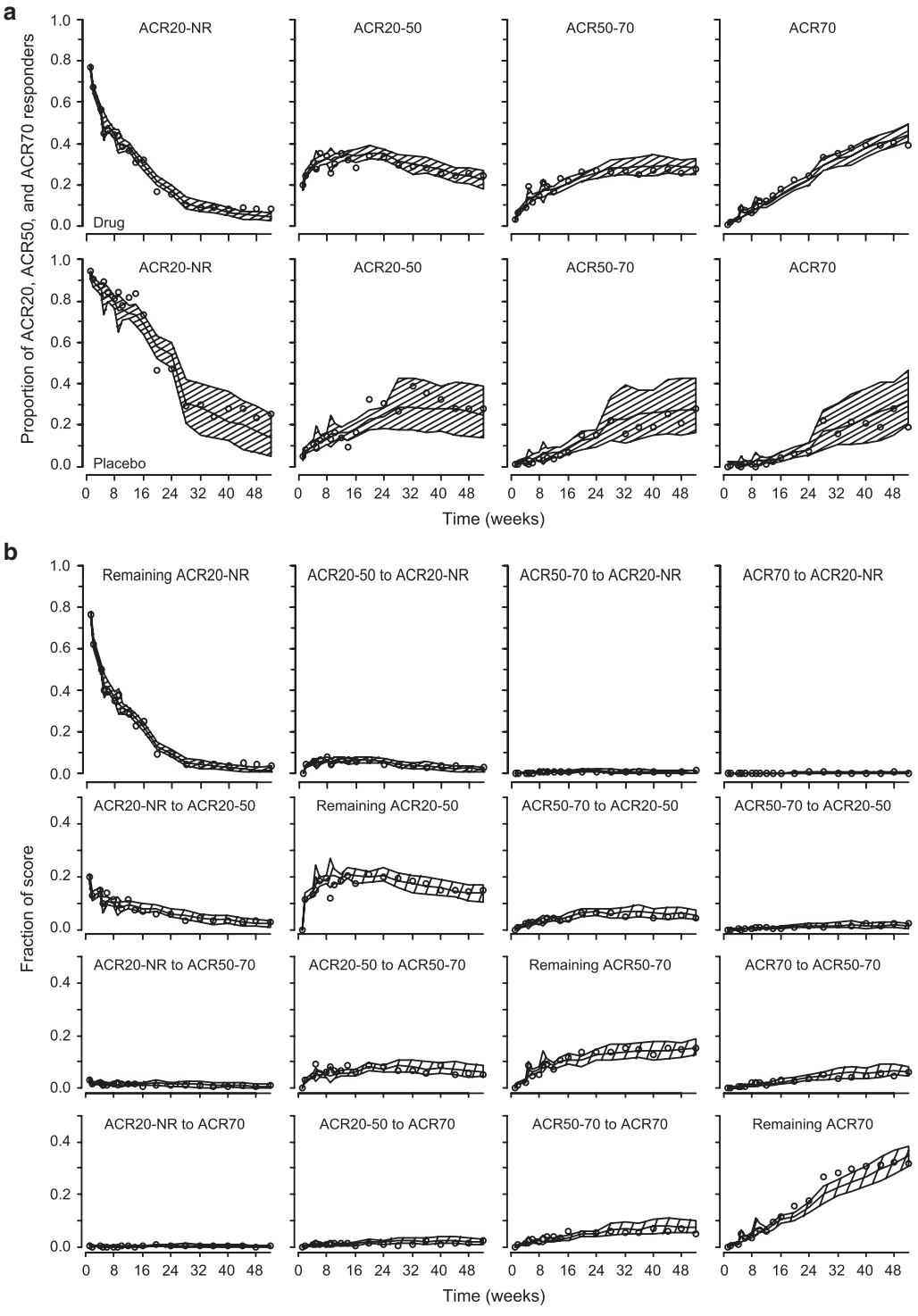
Parameter[a]	Estimate	Bootstrap median	Bootstrap SD
Transfer rate constants			
→ $K_{10,0}$ (day^{-1})	0.00677	0.00677	0.00073
→ $K_{21,0}$ (day^{-1})	0.0168	0.0166	0.0029
→ $K_{32,0}$ (day^{-1})	0.00778	0.00775	0.00156
← $K_{01,0}$ (day^{-1})	0.0859	0.0860	0.3301
← $K_{12,0}$ (day^{-1})	0.1800	0.1800	0.0592
← $K_{23,0}$ (day^{-1})	0.1330	0.1330	0.0504
Time effects on K_p			
→ Slope(t) on K_{10} (day^{-1})	0.0103	0.0103	0.0019
→ Slope(t) on K_{21}/K_{32} (day^{-1})	0.00256	0.00256	0.00355
← Slope(t) on $K_{01}/K_{12}/K_{23}$ (day^{-1})	–0.00159	–0.00157	0.00039
Exposure effects on K_p			
→ E_{max}(Cp) on K_{10} (day^{-1})	4.21	4.25	0.52
← E_{max}(Cp) on $K_{01}/K_{12}/K_{23}$ (day^{-1})	–0.697	–0.697	0.041
↔ EC_{50}(Cp) on $K_{10}/K_{01}/K_{12}/K_{23}$ (μg/ml)	3.04	3.04	0.43
→ E_{max}(slope) on K_{21}/K_{32} (day^{-1})	0.0103	0.0103	0.0031
Dropout model			
Intercept	–6.47	–6.45	0.30
From ACR20 response	–1.49	–1.47	0.13
From ACR50 response	–2.60	–2.61	0.33
From ACR70 response	–2.65	–2.63	0.28
Time effect on dropout			
$E_{max T}$	4.08	4.04	0.32
T_{50} (days)	23.1	23.8	3.7
Gamma	1.58	1.60	0.21
Covariate effects			
→ Steroids on $K_{10}/K_{21}/K_{32}$	–0.208	–0.208	0.037
→ BS28 on $K_{10}/K_{21}/K_{32}$	–0.00628	–0.01003	0.00687
← AGE on $K_{01}/K_{12}/K_{23}$	0.0139	0.0138	0.0037
→ North America, Western Europe	0.397	0.396	0.090
→ Central and South America	1.89	1.89	0.06
→ Northern Europe	–0.271	–0.269	0.075
Inter-individual variability[b]			
→ K_{10}	1.13	1.13	0.08
← K_{01}	2.19	2.19	0.26
→ K_{21}	0.897	1.03	0.111
← K_{12}	1.41	1.41	0.209
→ K_{32}	1.24	1.53	0.39
← K_{23}	2.39	2.39	0.26

[a]The arrows help in visualizing upward and downward parameters. [b]All estimates of the full correlation block are provided as initial estimate in the NONMEM control stream provided in **Supplementary Material**.

the dose: 72.1–77.3% for ACR20, 71.4–78.6% for ACR50 and 56.9–68.8% for ACR70 for treatment from 50 to 400 mg Q2W. At week 24, the likelihood of remaining responder was high, between 77 and 94%, for the three ACR scores whatever the dose, characterizing a sustained response. On the other hand, non-responders at W4 had between 41 and 48% chance to achieve ACR20 at week 12, 12–17% to achieve ACR50 and 2–5% to achieve ACR70 depending on the dose.

DISCUSSION

This pooled analysis established the relationships between exposure to the anti-TNFα drug certolizumab pegol and the ACR20, ACR50, and ACR70 responses in five phase II and III clinical trials, including both licensed and unlicensed doses. The model development was based on the database that was previously used to develop stand-alone models for ACR20



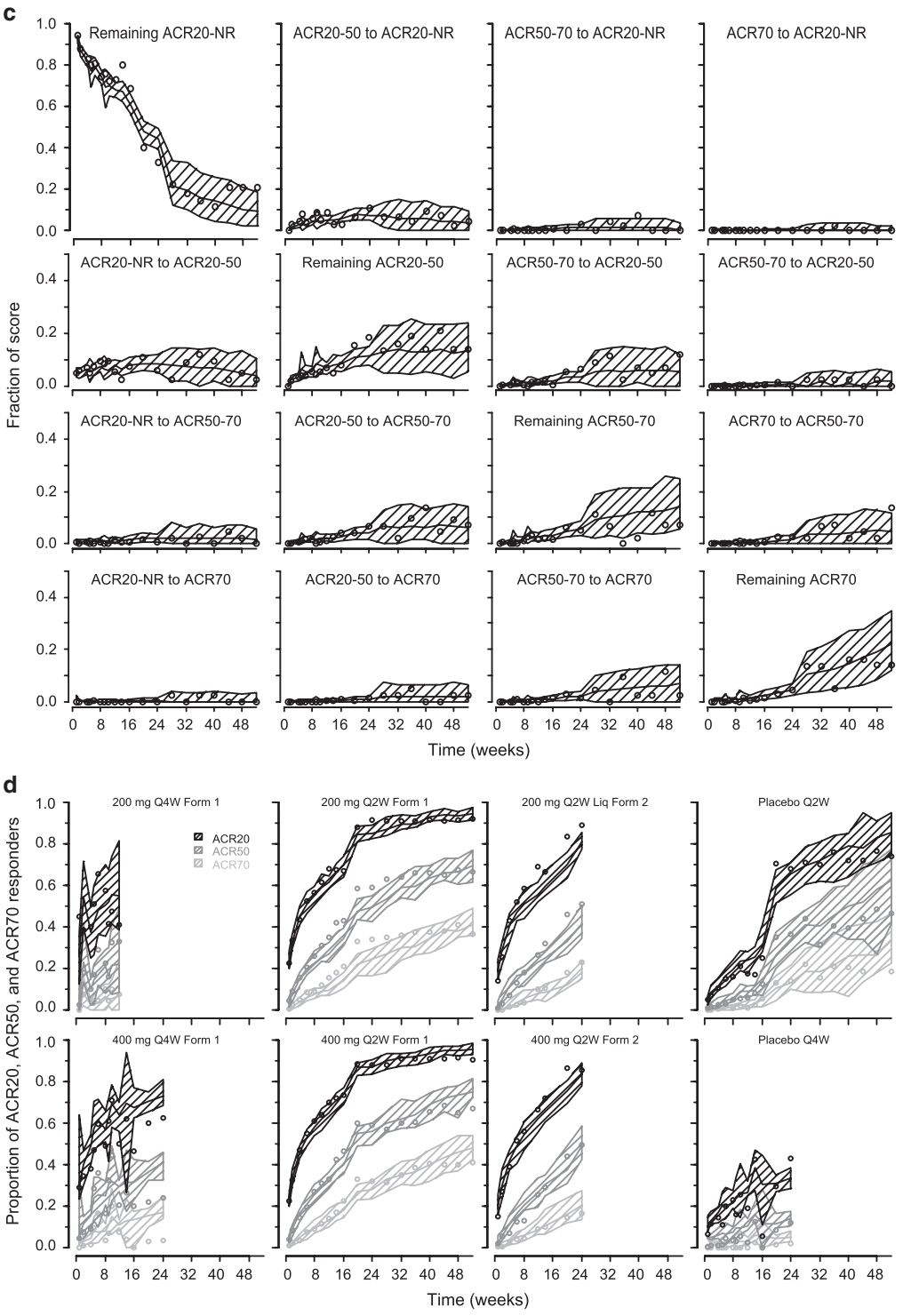


Figure 3 Visual predictive checks of the final model. Circles represent the observations and shaded areas the 95% confidence interval from the final model. Q2W, every other week; Q4W, every 4 weeks. *Load = 400 mg at weeks 0, 2, and 4. (**a**) Modeled scores ACR20-NR, ACR20-50, ACR50-70, ACR70, drug (top panel) and placebo (bottom panel). (**b**) Transitions between modeled scores, drug data. (**c**) Transitions between modeled scores, placebo data. (**d**) ACR clinical scores by treatment (black: ACR20, middle gray: ACR50 and light gray: ACR70).

(ref. 15) and ACR50 (ref. 10) responses; however, this new approach has the advantage of gaining more information from the same amount of data, as various levels of response are characterized simultaneously. The model adequately predicts both the proportion of response for the three ACR thresholds and the transition between the four statuses of response.

The exposure–response relationship was established over a large range of doses, including doses from the dose-ranging

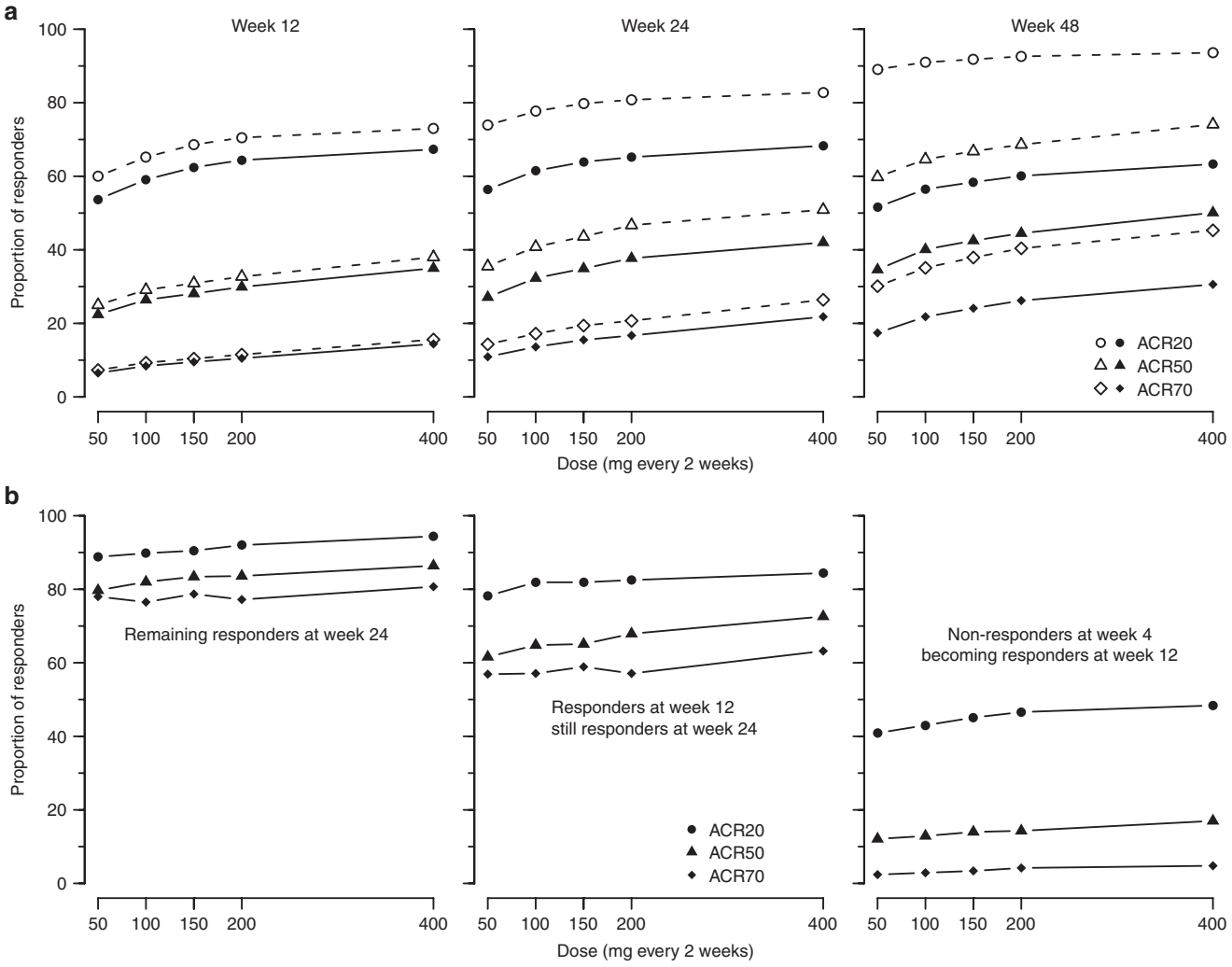


Figure 4 Simulation from the models, 50–400 mg Q2W (subject simulated as dropouts were considered as non-responders when not stated otherwise). (**a**) Proportion of responders at week 12, 24, and 48 (open symbols: only subjects simulated to not drop out were taken into account). (**b**) Illustration of sustained response at W24 and prediction of long-term response in various settings.

phase II study that are outside of the range of recommended doses for maintenance certolizumab pegol treatment in the RA indication, 200 mg Q2W and 400 mg Q4W. A plateau in the probability of becoming a responder and achieving a higher ACR response is reached, with the 200 mg and 400 mg Q2W doses providing similar level of response. The dropout is also affected, as subjects with higher level of ACR response are less likely to drop out. The model characterized a sustained response to treatment, a factor that is recognized as key in evaluating the treatment response and disease status in patients with RA,[7] the aim being to prevent joint damage and achieve sustained low disease activity or remission.[30] As such, the new EULAR/ACR recommendations request to include time to onset of response and sustainability of response when evaluating treatment response in clinical trials.[31,32]

Depending on the audience, each of the three ACR levels may provide the most relevant information. Regulatory authorities may be interested in the ACR20 and ACR50 responses to discriminate between drugs, rheumatologists may consider higher thresholds as a more relevant target to be achieved in clinical trials and might rather consider the time to reach ACR50, while patients will appreciate sustained ACR70 response to treatment. The model can be utilized to provide clinically interesting information through population simulations, with a large population, or through trial simulations, with a limited number of subjects and a specific design, according to the question of interest. For example, this could be the prediction of the long term response or the evaluation of the chance for an ACR20 non-responder at an early time point to achieve an ACR20, ACR50, or ACR70 response at a later time.

Accounting for the dropout and simultaneous modeling of the three ACR thresholds enables accurate trial simulations with specific study design. For example, in two of the studies used for the present analysis, a study-design dropout was defined based on the ACR20 response at weeks 12 and 14, which was an issue for developing a stand-alone ACR50 model.[10] It also allows for applying different post-processing rules to compute the response rate, e.g., to consider dropped out subjects as non-responders.

Table 3 Covariates

Category	Covariate	Continuous covariates: median (min–max) Categorical covariates: % subjects by category
Patient-related covariates	Age	53 (18–83) years
	Body weight	72 (41–164) kg
	Gender	Females (81%)
		Males (19%)
	Geographical regions	Eastern Europe (31%)
		Northern Europe and Baltic states (25%)
		Western Europe (20%)
		North America (15%)
		Central and South America (6%)
		Asia/Oceania (3%)
Disease-related covariates	Disease duration at entrance in the study	6 (0.3–47) years
	Baseline reduced tender joint count	17 (0–28)
	Baseline reduced swollen joint count	14 (0–28)
	Baseline C-reactive protein level	15 (0.2–273) mg/l
	Physician's assessment of disease activity	65 (14–100) mm
	Patient's assessments of disease activity	64 (6–100) mm
	Patient's assessments of physical function	1.6 (0–3)
	Patient's assessments of arthritis pain	65 (0–100) mm
Concomitant medications	Methotrexate dose at baseline	10 (0–30) mg/week
	Use of corticosteroids	39% : 46.3% on placebo; 37.6% on drug treatment
	Use of non-steroidal anti-inflammatory drugs	60% ; 62.9% on placebo; 59.2% on drug
	Use of analgesics	16%: 16.9% on placebo; 15.6% on drug
Previous medications	Number of previous disease-modifying anti-rheumatic drugs (DMARDs) other than methotrexate	1 (0–9)

The dependency between the successive ACR assessments was accounted for by mimicking a compartment model where the probability flows between the compartments are estimated. Any transitions between the four levels of ACR response are allowed, but the model is parameterized such that only transitions between neighboring states are estimated, making it more parsimonious than a proportional odds model with Markov elements. For nonconsecutive observed scores (e.g., ACR20 to ACR70), it is assumed that the intermediate unobserved score (ACR50 in this example) must have been reached at an intermediate time. The influence of previous observations decreases with increasing time between observations, i.e., if observations are made with large time interval, the influence of the previous observation will be small compared if two observations are made close in time. The half-life for transfer between the compartments were found to be dependents on drug concentration and time in the study. In addition, the inter-individual variability allowed for different patients to have different transfer rates.

The latent variable (LV) approach has also been proposed to model the ACR20, ACR50 and ACR70 scores simultaneously.[20] The concept was, however, different, aiming at describing the average probability of ACR20/50/70 response in the population rather than an accurate description of an individual's evolution of the response with time. It had the appeal to be parsimonious while flexibility can be added by fitting separate intercepts for the different thresholds. The continuous Markov model enables larger flexibility by allowing different covariate and drug effects for each transition rate, but this increases further the complexity of the model. We tried, however, to keep the model as simple as possible by using the same effects on the different transfer rates whenever possible.

Similar to models for ordered categorical data and the ACR20 Markov model, the parameters of the continuous Markov model may be difficult to interpret directly, and their clinical relevance is most easily interpreted through simulations from the model. The latent variable model may be considered as a step further for a more mechanistic interpretation, with parameters such as E_{max} that can be interpreted directly. Run time may be another limitation of the continuous Markov model—coded as differential equations—particularly with large datasets and increased number of categories.

In conclusion, an exposure–response model that simultaneously characterized the temporal course of the ACR20, ACR50, and ACR70 assessments in patients suffering from RA enrolled in phase II and phase III clinical trials, treated with certolizumab pegol and placebo, was successfully developed. The model predicted that increased certolizumab pegol exposure resulted in an increased probability of attaining higher level of ACR response, with the majority of the clinical effect being attained at the exposure achieved with the 200 mg Q2W label maintenance dose. This approach is applicable to model any ordered categorical score where consecutive observations are dependent.

METHODS

The present analysis was performed retrospectively on pooled data from five clinical trials, not following the protocol analysis and hence the results cannot be directly compared with the published results of the respective studies.

Data

Clinical score: ACR20, ACR50, and ACR70 responses are defined as a decrease of 20, 50, or 70%, respectively, in both tender and swollen joint counts, as well as the same magnitude of improvement in at least three of the five other measures of the ACR core dataset: the patient's and physician's global assessments of disease activity, the patient's assessment of pain and physical function, and an acute phase reactant (C-reactive protein level or erythrocyte sedimentation

rate). The C-reactive protein level was used to compute the ACR scores in the present analysis.

Study population and design: Subjects with active RA diagnosed at adult age since at least 6 months and inadequate response to methotrexate monotherapy were enrolled in the five clinical trials (**Table 1**). Study protocols were approved by the medical ethics committees or institutional review boards of the participating centers, and all subjects provided written informed consent prior to initiation of study procedures.[33]

In studies 4 and 5, subjects assessed ACR20 non-responders at both week 12 and week 14 withdrew at week 16 with the option of entering an open-label extension study (data not included in the analysis). This design feature was reproduced in the model evaluation.

Covariate data: Covariates tested during model development are summarized in **Table 3**.

Certolizumab pegol exposure measures: Certolizumab pegol plasma concentrations were determined using a validated enzyme-linked immunosorbent assay (ELISA) method. Individual exposure measurements (individual certolizumab pegol concentration at the time of clinical observation Cp and average certolizumab pegol concentration over the last dosing interval Cavg) were derived from a previously developed one-compartment population model using the empirical Bayes *post hoc* estimates from NONMEM. These exposure measures were set to 0 for placebo data.

Software and data analysis

The model was developed in NONMEM 7.1.2 (ref. 34) (ICON Development Solution, Hanover, MD) and finalized in NONMEM 7.3.0. launched with PsN (Uppsala University, Uppala, Sweden).[35] The Laplacian method was used to approximate the marginal likelihood; it uses a second-order expansion around the empirical Bayes predictions of the inter-individual random effects. R version 2.11 and 2.15 (ref. 36) (R Foundation for Statistical Computing, Vienna, Austria) were used for data management and graphical outputs.

Model development

Model structure: See Results section.

Model building and selection: During the model building, efforts were made to keep the model as simple and parsimonious as possible. Modeling was initiated based on placebo data. Various functions of increasing complexity, such as linear, E_{max}, Hill, second order and quadratic polynomial were tested to describe the influence of time since start of the study on the transfer rates between the four scores and on the dropout probability. After addition of data from certolizumab pegol-treated subjects, drug-effect models were evaluated. The effect of drug on the transition rates were modeled as functions of exposure, using a treatment flag, the dose, Cp or Cavg as markers for certolizumab pegol exposure. The exposure relationships for each of the transfer rate were explored using linear, E_{max} and Hill equations. Effects were tested separately on each transition rate and the parameters were shared between rate constants if found to be similar and this did not lead to a significant increase in OFV. Equations retained in the final model are presented in the results section.

Inter-individual variability was assessed for the six transfer rate parameters.

Covariate model: Covariates were included in the model based on their statistical significance. The potential covariate effects were added on all transition rate constants (see example in the NONMEM control file). Continuous covariates were introduced as

$$K_p = K_{p,0} \cdot \exp\left(\theta_{cov} \cdot (\mathrm{CONT} - \mathrm{CONTmed})\right) \quad (4)$$

where θ_{cov} is the estimated covariate effect, CONT the continuous covariate value and CONTmed the median of the covariate in the dataset.

Co-medications were evaluated as binary covariates (use/no-use). Categorical covariates were introduced as

$$K_p = K_{p,0} \cdot \begin{pmatrix} 1 + \theta_{cov,CAT1} \cdot \mathrm{CAT1} + \theta_{cov,CAT2} \cdot \\ \mathrm{CAT2} + \theta_{cov,CAT3} \cdot \mathrm{CAT3.} + \ldots \end{pmatrix} \quad (5)$$

where the $\theta_{cov,CATx}$ are the estimated covariate effects, CATx = 1 for the corresponding covariate, CATx = 0 otherwise.

A manual stepwise approach was used, beginning with a univariate testing (α = 0.05). Significant covariates were included in both forward inclusion (α = 0.05) and backward elimination (α = 0.001) steps, which continued until only significant covariates remained in the model.

Model evaluation: Model selection was based on changes in the NONMEM OFV evaluated using a likelihood ratio test (nested models, *P = 0.*05) and simulations from the models (visual predictive checks) to evaluate the model's capacity to capture the observed frequency of scores and transitions between the scores over time. For that purpose, 250 datasets were simulated from the model of interest. Additional rows at planned visits were added in case of early dropout in the original dataset in order to reproduce the theoretical observation scheme and allow subjects to drop out according to model simulations. The frequency of ACR outcomes and transitions between the four ACR scores were plotted as a function of time and compared with the observed data. Residuals were not available for evaluating the goodness of fit because the model estimated the probability of scores and not a prediction of the scores themselves.

A bootstrap with 100 samples, stratified on study and treatment, was performed to evaluate the precision of estimation of the parameter estimates. For each model parameter, the median and the standard deviation of the parameters' distributions from the bootstrap runs were computed and reported.

Simulations: For performing simulations, a probability amount of 1 was allocated in the compartment corresponding to the simulated score (subjects are set to non-responder at baseline, see model file and an example simulation dataset in **Supplementary Material**).

Acknowledgments. The authors would like to acknowledge Emilie Hénin (Université Claude Bernard Lyon 1—Hospices Civils de Lyon), who provided valuable scientific advice and Laetitia Laurent (RealDolmen), who compiled the database. The research leading to these results has received support

from DDMoRe, which is an Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115156, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution. The DDMoRe project is also supported by financial contribution from Academic and SME partners. This work does not necessarily represent the view of all DDMoRe partners. B.D.L. was funded by UCB Pharma.

Conflict of Interest/Disclosure. The clinical trials on which the manuscript is based were sponsored by UCB Pharma. Brigitte D. Lacroix is full-time employee of UCB Pharma. Mats O. Karlsson has received research funding from UCB Pharma, but not related to this work. L.E.F. declared no conflict of interest. Deputy Editor in Chief L.E.F. was not involved in the review or decision process for this paper.

Author Contributions. B.D.L., M.O.K., and L.E.F. wrote the manuscript, designed the research, performed the research, and analyzed the data.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

✓ Two approaches have been proposed to model the ACR improvement criteria used in rheumatoid arthritis clinical trials. The latent variable approach focuses on describing the average probability of response of the population of subjects. The Markovian approach considers the dependency of the assessments with time, but its complexity increases dramatically with the number of categories and constant influence of preceding scores is implicitly assumed.

WHAT QUESTION DID THIS STUDY ADDRESS?

✓ The study investigated if an exposure–response relationship can be established that simultaneously characterizes the three levels of ACR response in patients enrolled in phase II and phase III clinical trials.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

✓ Simultaneous analysis of the three levels of ACR response in clinical trials can be performed to leverage the knowledge and gain more information based on the same amount of data.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

✓ The model can be utilized in drug development to characterize and simulate expected ACR assessments under various settings and assumptions, and to derive clinically meaningful outcomes.

1. Felson, D.T. *et al.* The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum.* **36**, 729–740 (1993).
2. Felson, D.T. *et al.* American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum.* **38**, 727–735 (1995).
3. US Food and Drug Administration. *Guidance for Industry. Clinical Development Programs for Drugs, Device and Biological Products for the Treatment of Rheumatoid Arthritis (RA)* (Center for Drug Evaluation and Research, Rockville, MD, 1999).
4. Felson, D.T., Anderson, J.J., Lange, M.L., Wells, G. & LaValley, M.P. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? *Arthritis Rheum.* **41**, 1564–1570 (1998).
5. Pincus, T. & Stein, C.M. ACR 20: clinical or statistical significance? *Arthritis Rheum.* **42**, 1572–1576 (1999).
6. Chung, C.P., Thompson, J.L., Koch, G.G., Amara, I., Strand, V. & Pincus, T. Are American College of Rheumatology 50% response criteria superior to 20% criteria in distinguishing active aggressive treatment in rheumatoid arthritis clinical trials reported since 1997? A meta-analysis of discriminant capacities. *Ann. Rheum. Dis.* **65**, 1602–1607 (2006).
7. Dougados, M. *et al.* Evaluation of different methods used to assess disease activity in rheumatoid arthritis: analyses of abatacept clinical trial data. *Ann. Rheum. Dis.* **68**, 484–489 (2009).
8. Fedorov, V., Mannino, F. & Zhang, R. Consequences of dichotomization. *Pharm. Stat.* **8**, 50–61 (2009).
9. Schmitz, S., Adams, R., Walsh, C.D., Barry, M. & FitzGerald, O. A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. *Ann. Rheum. Dis.* **71**, 225–230 (2012).
10. Lacroix, B.D., Lovern, M.R., Stockis, A., Sargentini-Maier, M.L., Karlsson, M.O. & Friberg, L.E. Exposure-response modeling of the ACR50 score in rheumatoid arthritis patients treated with certolizumab pegol. *PAGE* **18** (2009) <http://www.page-meeting.org/?abstract=1585>.
11. Siegel, J.N. & Zhen, B.G. Use of the American College of Rheumatology N (ACR-N) index of improvement in rheumatoid arthritis: argument in favor. *Arthritis Rheum.* **52**, 1637–1641 (2005).
12. Boers, M. Use of the American College of Rheumatology N (ACR-N) index of improvement in rheumatoid arthritis: argument in opposition. *Arthritis Rheum.* **52**, 1642–1645 (2005).
13. Hu, C., Xu, Z., Zhang, Y., Rahman, M.U., Davis, H.M. & Zhou, H. Population approach for exposure-response modeling of golimumab in patients with rheumatoid arthritis. *J. Clin. Pharmacol.* **51**, 639–648 (2011).
14. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. *Arthritis Rheum.* **57**, 193–202 (2007).
15. Lacroix, B.D., Lovern, M.R., Stockis, A., Sargentini-Maier, M.L., Karlsson, M.O. & Friberg, L.E. A pharmacodynamic Markov mixed-effects model for determining the effect of exposure to certolizumab pegol on the ACR20 score in patients with rheumatoid arthritis. *Clin. Pharmacol. Ther.* **86**, 387–395 (2009).
16. Sheiner, L.B. A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin. Pharmacol. Ther.* **56**, 309–322 (1994).
17. Keizer, R.J. *et al.* A model of hypertension and proteinuria in cancer patients treated with the anti-angiogenic drug E7080. *J. Pharmacokinet. Pharmacodyn.* **37**, 347–363 (2010).
18. Zingmark, P.H., Kågedal, M. & Karlsson, M.O. Modelling a spontaneously reported side effect by use of a Markov mixed-effects model. *J. Pharmacokinet. Pharmacodyn.* **32**, 261–281 (2005).
19. Hutmacher, M.M., Krishnaswami, S. & Kowalski, K.G. Exposure-response modeling using latent variables for the efficacy of a JAK3 inhibitor administered to rheumatoid arthritis patients. *J. Pharmacokinet. Pharmacodyn.* **35**, 139–157 (2008).
20. Hu, C., Xu, Z., Rahman, M.U., Davis, H.M. & Zhou, H. A latent variable approach for modeling categorical endpoints among patients with rheumatoid arthritis treated with golimumab plus methotrexate. *J. Pharmacokinet. Pharmacodyn.* **37**, 309–321 (2010).
21. Hu, C., Xu, Z., Mendelsohn, A.M. & Zhou, H. Latent variable indirect response modeling of categorical endpoints representing change from baseline. *J. Pharmacokinet. Pharmacodyn.* **40**, 81–91 (2013).
22. Hutmacher, M.M. & French, J.L. Extending the latent variable model for extra correlated longitudinal dichotomous responses. *J. Pharmacokinet. Pharmacodyn.* **38**, 833–859 (2011).
23. Bergstrand, M., Söderlind, E., Weitschies, W. & Karlsson, M.O. Mechanistic modeling of a magnetic marker monitoring study linking gastrointestinal tablet transit, *in vivo* drug release, and pharmacokinetics. *Clin. Pharmacol. Ther.* **86**, 77–83 (2009).
24. Snelder, N., Diack, C., Benson, N., Ploeger, B. & van der Graaf, P. A proposal for implementation of the Markov property into a continuous time transition state model in NONMEM. *PAGE* **18** (2009). <http://www.page-meeting.org/?abstract=1536>.
25. Pilla Reddy, V., Petersson, K.J., Suleiman, A.A., Vermeulen, A., Proost, J.H. & Friberg, L.E. Pharmacokinetic-pharmacodynamic modeling of severity levels of extrapyramidal side effects with markov elements. *CPT. Pharmacometrics Syst. Pharmacol.* **1**, e1 (2012).
26. Fleishmann, R. *et al.* Efficacy and safety of certolizumab pegol monotherapy every 4 weeks in patients with rheumatoid arthritis failing previous disease-modifying antirheumatic therapy: the FAST4WARD study. *Ann. Rheum. Dis.* **68**, 805–811 (2009).

27. Keystone, E., Mason, D. & Fleischmann, R.M. Certolizumab pegol 400 mg every 4 weeks as monotherapy rapidly reduces disease activity in active rheumatoid arthritis. *Arthritis Rheum.* **56 (suppl 9)**, S156 (2007).
28. Keystone, E. *et al.* Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum.* **58**, 3319–3329 (2008).
29. Smolen, J. *et al.* Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A randomised controlled trial. *Ann. Rheum. Dis.* **68**, 797–804 (2009).
30. Aletaha, D., Eberl, G., Nell, V.P., Machold, K.P. & Smolen, J.S. Attitudes to early rheumatoid arthritis: changing patterns. Results of a survey. *Ann. Rheum. Dis.* **63**, 1269–1275 (2004).
31. Aletaha, D. *et al.* Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Ann. Rheum. Dis.* **67**, 1360–1364 (2008).
32. Karonitsch, T. *et al.* Methods of deriving EULAR/ACR recommendations on reporting disease activity in clinical trials of patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **67**, 1365–1373 (2008).
33. World Medical Association. World Medical Association Declaration of Helsinki : ethical principles for medical research involving human subjects. *JAMA.* **310**, 2191–2194 (2013).
34. Beal, S.L., Sheiner, L.B., Boeckmann, A. & Bauer, R.J. *NONMEM User Guides.* (Icon Development Solutions, Ellicott City, MD, 2009).
35. Lindbom, L., Pihlgren, P., Jonsson, E.N. & Jonsson, N. PsN-Toolkit–a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput. Methods Programs Biomed.* **79**, 241–257 (2005).
36. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2008). <http://www.R-project.org>.

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://www.nature.com/psp)