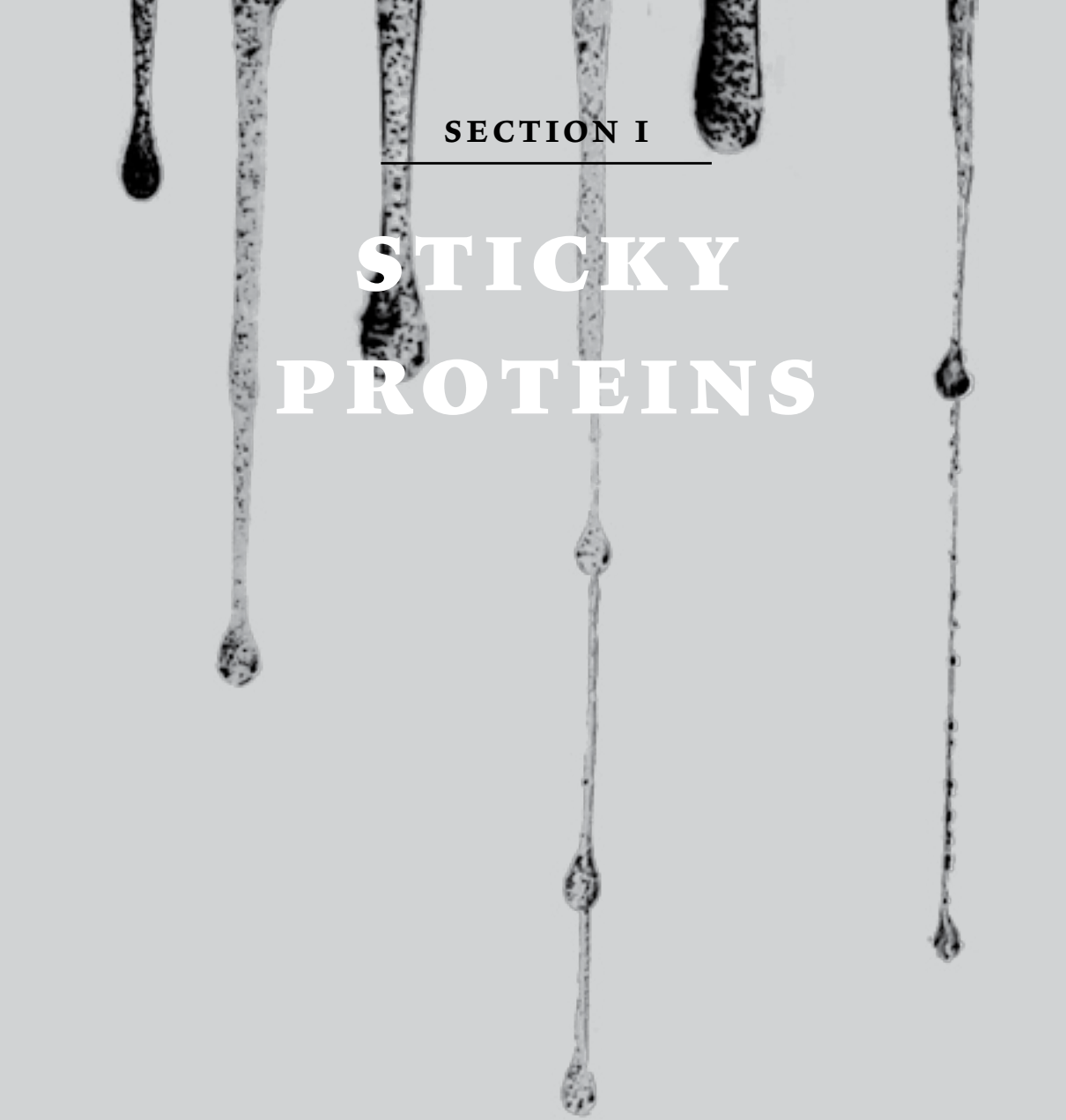

SECTION I

STICKY PROTEINS

CHAPTER I

Use of population approach non-linear mixed effects models in the evaluation of biosimilarity of monoclonal antibodies

J.A.A. Reijers, T. van Donge, F.M.L. Schepers, J. Burggraaf, J. Stevens

Population pharmacokinetic analyses (PPK) have been used to establish bioequivalence for small molecules and some biopharmaceuticals. We investigated whether PPK could also be useful in biosimilarity testing for monoclonal antibodies (MABs).

Data from a biosimilarity trial with two trastuzumab products were used to build population pharmacokinetic models. First, a combined model was developed and similarity between test and reference product was evaluated by performing a covariate analysis with trastuzumab drug product (test or reference) on all model parameters. Next, two separate models were developed, one for each drug product. The model structure and parameters were compared and evaluated for differences.

Drug product could not be identified as statistically significant covariate on any parameter in the combined model and the addition of drug product as covariate did not improve the model fit. A similar structural model described both the test and reference data best. Only minor differences were found between the estimated parameters from these separate models.

PPK can also be used to support a biosimilarity claim for a MAB. However, in contrast to the standard non-compartmental analysis, there is less experience with a PPK approach. Here, we describe two methods of how PPK can be incorporated in biosimilarity testing for complex therapeutics.

During the past decades, many biopharmaceuticals have been marketed, mostly for use in the field of oncology and rheumatology. Although efficacious, high costs often limit the availability of these therapies or greatly burden the health care system. For example, treatment of a rheumatological US patient with biopharmaceuticals costs on average $20,000 to $30,000 annually,[1] and a single cycle of bevacizumab or other monoclonal antibody (MAB) can cost $2,000 or more.[2] In 2014, the top 20 in global medicinal product sales contained 10 biopharmaceuticals, generating revenues between 4.4 and 11.8 billion dollar each.[3] Because of the growing number of patent expirations for the original biopharmaceuticals, it is expected that research of biosimilars will increase.

A first requirement for registration of the novel compound is to establish pharmacokinetic 'biosimilarity'. Although the terminology slightly differs between the regulatory agencies, all agree on the basic concept of biosimilarity, which is that the novel ('test') compound should be highly similar to its originator

('reference') in terms of quality, efficacy, and safety, and that any remaining difference should be clinically insignificant.[4–6]

Notwithstanding specific criteria for biopharmaceuticals, often parts of guidelines for establishing bioequivalence – not biosimilarity – between chemically derived substances ('small molecules') are applied. These guidelines require that similarity should be demonstrated for key pharmacokinetic parameters, most commonly area under the concentration-time curve (AUC) and maximum concentration (C_{max}), based on predefined acceptance limits at the highest dose level used. According to an evaluation by the World Health Organisation, studies proving biosimilarity generally use the 80–125% equivalence range due to lack of specific acceptance criteria for biopharmaceuticals.[4]

Although it is widely recognised that a non-compartmental analysis (NCA) is less appropriate when dealing with complex pharmacokinetics, it is still the most commonly used analytical method for demonstrating biosimilarity. Mentré's group has extensively studied the use of population pharmacokinetic techniques in bioequivalence testing and found that it yielded similar results, with the modelling approach leading to a better understanding of the underlying biological system, and the NCA being a relatively easy approach that does not require modelling and whose results can be used in a statistical analysis. The same was found for two biopharmaceuticals, somatropin and epoetin-α.[7–9]

We investigated whether a population pharmacokinetic analysis (PPK) could also be useful in bioequivalence testing for monoclonal antibodies (MABs), which display complex elimination mechanisms, including non-linear routes, and have a plasma half-life of one to multiple weeks. Two approaches in modelling pharmacokinetic (PK) data were studied. First, we developed a combined model built on all available data for both the test and reference product, and tested whether adding product (test/reference) as a covariate would improve the model, indicating non-similarity. Second, we developed separate models, one for test and one for reference product. This approach does not assume similarity as a starting point and allows comparison of the model structures and parameters. For this exercise, we chose the humanised MAB trastuzumab, which targets the HER2-receptor.

METHODS

Study population and treatment

Data were gathered in a phase I randomised, single dose, parallel group bioequivalence study, preceded by a placebo-controlled dose escalation part.[10] In this study, 110 male volunteers, aged 18–45 years, who were deemed healthy after a full medical screening, received trastuzumab in 250 mL 0.9% NaCl as an intravenous infusion over 90 min. Two trastuzumab products were administered: the biosimilar product (test, T), codenamed FTMB (Synthon BV, Nijmegen, The Netherlands), and the EU-licensed product (reference, R), marketed as Herceptin®.

Studied dose levels of the test product in the dose escalation part were 0.5 mg/kg ($n = 6$), 1.5 mg/kg ($n = 6$), 3 mg/kg ($n = 6$). The bioequivalence part consisted of 92 participants, who randomly received test ($n = 46$) or reference ($n = 46$) product at 6 mg/kg.

Based on the trastuzumab content of the used test and reference product vials, the actual dose levels were determined to be 0.49, 1.48, 2.96, 5.92 mg/kg for T, and 6.44 mg/kg for R.

Bioanalyses

Trastuzumab was quantitated in serum samples collected pre-dose and at 45 min, 1.5, 2, 3, 4, 5, 6, 8, and 24 h, and at 2, 4, 8, 14, 21, 28, 35, 42, 49, and 63 days after start of administration. A detailed description of the assay is given by Wisman *et al.*[10] The lower limit of quantification (LLOQ) was 0.060 µg/mL. All pre-dose trastuzumab concentrations <LLOQ were set to zero prior to analysis. Post-dose concentrations below LLOQ were not included. A serum sample for the quantification of serum HER2 extracellular domain (ECD) was collected prior to administration. This assay had a LLOQ of 0.50 ng/mL.

In the original clinical study protocol, the sample at day 63 was not collected for PK analysis and hence not included in the previously reported NCA results.[10] However, as this sample provided valuable insight in the non-linear clearance of trastuzumab, it was included in the analyses reported in this chapter.

PPK

General modelling approach

Population pharmacokinetic analysis (PPK) followed a stepwise approach. First, a general model for trastuzumab, hereafter referred to as 'combined model',

was developed based on all available PK data for both test and reference product, including dose levels of the dose escalation part (0.49, 1.48, and 2.96 mg/kg). To investigate potential bias in the PK models due to analysing test and reference products simultaneously, PK models were also developed for the test (model T) and reference product (model R) separately and included only data from subjects who were exposed to 6 mg/kg test or reference product. These are hereafter referred to as 'separate models'. The separate models were developed in parallel in order to maintain a structurally similar model for the test and reference product. Consequently, the model was only adopted if the corresponding model in the other treatment arm was preferred over its parent as well.

Model development

Model development was performed using NONLINEAR Mixed Effects Modelling (NONMEM 7.2.0, Icon Development Solutions, Ellicott City, Maryland, USA) and closely followed the FDA and EMA guidelines for PPK.[11,12] Models were built under ADVAN 13 with tolerance (TOL) 9 and the first-order conditional method with interaction (FOCE-I) was used for parameter estimation. NONMEM reports an objective function value (OFV), which is the -2 · log likelihood. Model hypothesis testing used the likelihood ratio test under the assumption that the difference in OFV is χ^2-distributed with degrees of freedom being determined by the number of additional parameters in the more complex model. Hence, with a decrease in OFV of ≥7.88 points ($p < 0.005$), the model is preferred over its parent model. Also, model performance was evaluated by means of goodness-of-fit plots, using the software package R (v3.2.2, R foundation for statistical computing, Vienna, Austria, 2015).

Several structural models with 2 or 3 compartments, including combinations of linear and non-linear clearance, were fitted to the data to determine the best structural model. Log-normal distribution of the between-subject variability (η) was assumed and several residual error (ε) structures were tested (proportional, additive and combined).

Potential covariate correlations, defined as a significant Pearson's product-moment correlation coefficient ($r^2 > 0.5$ with $p < 0.01$), were tested in the model development, in linear and exponential manners, and incorporated based on improvement in model performance. Explored covariates included lean body weight (LBW),[13] body weight, body surface area (BSA),[14] height, BMI, age, HER2-ECD concentration, dose, and product.

Model evaluation and predictive performance

To evaluate the robustness and predictive performance of the developed model, a visual predictive check (VPC) with 500 simulations was performed.[15] Prediction intervals of 95% were obtained by simulating the model results from the original data. Model evaluation was performed by calculating the coefficient of variation to derive the uncertainty in the parameter estimates of the model, which was considered acceptable if lower than 50%. Also, shrinkage, as defined by Karlsson and Savić,[16] was calculated to exclude model misspecification; shrinkage less than 30% was considered acceptable.

Individual pharmacokinetic profiles

Individual pharmacokinetic profiles were simulated in R using the individual participant's model parameter estimates. Integration was performed from the start of administration until the time point when the concentration in the central compartment dropped below 0.01 µg/mL. For the simulations, the following integration intervals were used: 1 s from administration until 24 h, 1 min until day 80, and 1 day thereafter. The concentrations were stored at original sampling times and at every 5 min. Trastuzumab concentration at the start of administration was assumed to be 0 µg/mL.

Comparison to NCA

For comparison to a standard NCA, AUCs were derived using model *simulated* (predicted) individual concentrations at the original sampling times. AUC from administration (time 0) to the time of the last concentration ≥LLOQ (AUC_l) was calculated using the linear trapezoidal method. AUC extrapolated to infinity (AUC_∞) based on the apparent terminal elimination rate constant was calculated as well.

Biosimilarity statistics were performed on AUC_∞ and AUC_l of all participants who were exposed to 6 mg/kg, comparing T to R in an unpaired t-test, using the software package R. AUCs were natural log (ln)-transformed prior to statistical analysis. The estimated difference in means and the corresponding 90% confidence interval (CI) were back-transformed to obtain the relative geometric mean ratio (GMR) of T over R (T/R). These results were then compared to those calculated in a standard NCA.

TABLE 1.1 Demographics

PARAMETER	TEST 0.5 mg/kg ($n=6$)	TEST 1.5 mg/kg ($n=6$)	TEST 3.0 mg/kg ($n=6$)	TEST 6.0 mg/kg ($n=46$)	REFERENCE 6.0 mg/kg ($n=46$)
Age (year)	26.9 (8.9)	33.0 (9.1)	23.4 (2.3)	26.0 (7.3)	24.1 (5.8)
Height (cm)	183 (12.0)	176 (6.5)	184 (3.3)	184 (7.5)	182 (6.2)
Weight (kg)	73.1 (12.6)	73.0 (8.7)	72.0 (7.5)	79.5 (11.2)	77.1 (10.2)
BMI ($kg\,m^{-2}$)	21.7 (3.3)	23.5 (2.6)	21.2 (2.1)	23.4 (2.5)	23.2 (2.7)
LBW (kg)	59.4 (8.4)	57.5 (5.1)	59.1 (3.8)	62.6 (6.6)	61.0 (5.6)
BSA (m^2)	1.93 (0.21)	1.89 (0.13)	1.92 (0.10)	2.01 (0.17)	1.97 (0.15)
ECD ($\mu g\,L^{-1}$)	12.7 (1.8)	11.8 (2.1)	11.4 (1.5)	11.3 (1.8)	11.8 (1.8)

Mean (standard deviation) demographics per treatment arm.
BSA: body surface area | ECD: HER2 extracellular domain | LBW: lean body weight

To correct for the difference between actual (5.92 mg/kg and 6.44 mg/kg) and labelled dose (6 mg/kg), a linear normalisation to 6 mg/kg was applied to the individual AUCs in the NCA. In the PPK, individual profiles were simulated with the actual and labelled dose. Both corrected and uncorrected AUCs were calculated and statistically compared.

RESULTS

Population

Pharmacokinetic data were gathered from 110 healthy male volunteers, whose demographics are presented in *table 1.1*. In total, 1,247 serum trastuzumab concentrations were available for the test product (T), of which 143 were <LLOQ (64 pre-dose). In the 6 mg/kg test group, 60/906 observations were <LLOQ (46 pre-dose) and in the reference product (Herceptin®), 51/912 observations (44 pre-dose).

Model development

First step: combined model

Initial exploration of the data suggested that a 2 or 3 compartment model would describe the data best. Based on the observed non-linear kinetics, Michaelis-Menten kinetics was incorporated, described in terms of maximum rate of elimination (V_{max}), and the concentration where $\frac{1}{2} \cdot V_{max}$ is reached (K_m). Addition of a linear elimination pathway, defined by elimination rate constant (k_e), significantly improved the model fit for both the 2 and 3 compartments models.

Adding the third compartment accounted for a delayed clearance effect. The 3-compartment model, parameterised in terms of a central (V_1) and peripheral volumes (V_2 and V_3) of distribution, and inter-compartmental clearances (Q_1 and Q_2), resulted in a significant improvement compared to the 2-compartment model. This was confirmed by an improved goodness-of-fit, especially for the lower doses of trastuzumab. Thus, the 3-compartment model was considered superior over the 2-compartmental model (*figure 1.1*). A combined residual error structure (ε) proved best fit for purpose.

After identification of the structural model, individual estimates of random effects for between-subject variability were identified for the parameters V_1, K_m, and k_e, with final coefficient of variation values of 14.8, 35.9, and 17.2%, respectively. The residual coefficient of variation of the best model was 14.98%. An omega block was required to correct for the parameter correlation between K_m and k_e in the model.

Significant correlations were found between V_1 and lean body weight (LBW), body weight, body surface area (BSA), height, and BMI, with correlation coefficients of 0.61, 0.55, 0.60, 0.54, and 0.28, respectively. Linear regression analysis of LBW *vs.* BSA resulted in a coefficient of 1, and for LBW *vs.* body weight in 0.96. Furthermore, significant correlation coefficients were observed between BMI and k_e (0.60), between serum concentrations HER2-ECD and k_e (0.29), and between serum concentrations HER2-ECD and K_m (0.18).

Implementing LBW as a linear covariate on V_1 (*equation 1.S1*), significantly improved the objection function value (OFV) and was added to the model. Incorporating other weight-related covariates (body

weight, height and BMI) separately in the model did not result in a significant improvement compared to LBW; accordingly, they were not implemented in the model. Covariate analyses identified BMI as the one most significantly correlated to k_e. Incorporating this covariate linearly on k_e (*equation 1.s2*), further improved the model, and BMI was thus added to the model. Including HER2-ECD as a covariate did not improve the model fit. Interestingly, the model favours lean body weight as a size descriptor to scale trastuzumab dose compared to body weight, which is used clinically in dose calculation.

Adding trastuzumab drug product (test or reference) as a covariate to the model did not explain any relevant variability. A maximum decrease in OFV of only 5.80 points ($p > 0.01$) was observed when treatment was added as a covariate on K_m. Thus, drug product as covariate did not significantly improve model fit. All PK parameter estimates obtained with the best fit of the models are listed in *table 1.2*.

Additionally, the rates of the linear and non-linear elimination pathway *vs.* trastuzumab concentration were calculated. At low trastuzumab concentrations (<10 µg/mL), total elimination was almost independent of serum drug concentration, *i.e.* the non-linear elimination exceeded the linear elimination. At high concentrations, this pathway became saturated and the influence of non-linear elimination seemed negligible (*figure 1.s2*).

Also, a more complex mechanistic model approach was applied to characterise the distribution and clearance of trastuzumab: the target-mediated drug disposition (TMDD) model.[17,18] In addition to receptor and drug-receptor complex quantification, such models are able to provide information on binding affinity of the drug to the receptor. Fitting a TMDD model to our data proved difficult due to over-parameterization. A simplified approximation TMDD model approach with a dissociation constant K_d[19] still resulted in an incorrect fit and instability of the model, and the TMDD model approach was abandoned.

Second step: separate models

Model development of the separate models, including only data from participants who were exposed to 6 mg/kg, followed a similar approach as the combined model to ensure the structural similarity. For both trastuzumab products, a third compartment could be identified, as well as a linear and a non-linear route of elimination, described by Michaelis-Menten kinetics.

For the separate models, individual estimates of random effects for the between-subject variability were identified for the parameters V_1, V_{max}, and k_e, with

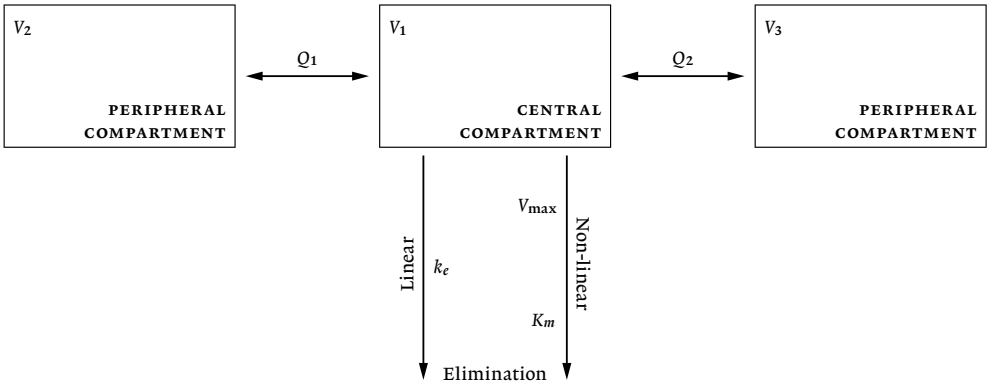

FIGURE 1.1 Schematic representation of the structural PK model with a parallel linear and non-linear elimination pathway | Linear elimination is described by an elimination rate constant (k_e) and non-linear elimination is calculated as

$$\frac{V_{max} \cdot C}{K_m + C},$$

where V_{max} is the maximum rate of elimination, K_m the concentration which produces half of the V_{max}, and C the concentration. V_1, V_2, and V_3 are the distribution volumes; Q_1 and Q_2 are the inter-compartmental clearances to the peripheral compartments.

TABLE 1.2 Population pharmacokinetic parameter estimates from the full covariate model for trastuzumab

PARAMETER[a]	COMBINED MODEL		SEPARATE MODEL T		SEPARATE MODEL R	
Fixed effects, estimate (CI)						
V_1 (L)	3.28	(3.185–3.367)	3.59	(3.418–3.752)	3.13	(3.028–3.232)
V_2 (L)	1.89	(1.325–2.457)	6.82	(-5.572–19.21)	44	(28.18–59.77)
V_3 (L)	1.96	(1.736–2.179)	2.15	(1.858–2.443)	2.09	(1.929–2.244)
Q_1 (L h^{-1})	2.91	(2.02–3.79)· 10^{-3}	2.82	(1.081–4.566)	3.92	(3.58–4.25)· 10^{-3}
Q_2 (L h^{-1})	4.34	(3.66–5.01)· 10^{-2}	3.75	(2.787–4.706)	4.67	(4.12–5.21)· 10^{-2}
V_{max} (µg h^{-1})	178	(162.3–193.1)	172	(138.6–205.7)	127	(111–143.4)
K_m (µg L^{-1})	937	(759.6–1115)	995	(674.6–1316)	1440	(1189–1699)
k_e (h^{-1})	2.20	(2.02–2.38)· 10^{-3}	1.95	(1.33–2.57)· 10^{-3}	1.76	(1.62–1.9)· 10^{-3}
Random effects, estimate (CV)						
Between-subject variability						
$\omega^2 V_1$	0.0217	(14.8%)	0.0270	(16.5%)	0.0122	(11.1%)
$\omega^2 V_{max}$	—		0.0163	(12.8%)	0.0347	(18.8%)
$\omega^2 K_m$	0.121	(35.9%)	—		—	
$\omega^2 k_e$	0.0292	(17.2%)	0.0355	(19.0%)	0.0286	(17.0%)
Residual error						
σ^2 proportional	0.0222		0.0207		0.0198	
σ^2 additive	1520		3090		790	

a. Explanation of parameters is given in *figure 1.1*.
CI: confidence interval | CV: coefficient of variation | σ^2: residual variance | ω^2: between-subject variance

final coefficient of variation values in model T of 16.5, 12.8, and 19%, respectively. The residual coefficient of variation of the best model was 14.5%. In model R, the final coefficients of variation were 11.1, 18.8, and 17%, with a residual coefficient of variation of 14.1%.

As with the combined model, the best model fit with the greatest reduction in OFV was obtained by incorporating LBW as linear covariate on V_1 and BMI on k_e for both separate models.

Model evaluation and predictive performance

Combined model

Goodness-of-fit plots of the combined model (*figure 1.2*) showed that all predictions lay around the line of unity. There was one outlier in the reference group, where one subject had a very low mid-infusion concentration of 0.088 µg/mL. Virtually all conditional weighted residuals with interaction (CWRESI) lay randomly scattered around zero without apparent bias.

The variability of the parameters V_1, K_m, and k_e on the η density histograms (*figure 1.S4*) seemed normally distributed around zero with acceptable coefficient of variation values, indicating correct description of the between-subject variability. Furthermore, no significant shrinkage was observed for parameters for which between-subject variability was identified (<8.04%).

The VPC proved good predictive performance (*figure 1.3*) of the combined model. For the doses >1.48 mg/kg, no signs of bias were apparent and most observations lay within the 95% prediction interval. Only for the lowest dose administered (0.49 mg/kg), a slightly higher prediction of the population mean was observed, especially in the lower concentration range. However, even for this dose group, most of the observations were within the 95% prediction interval.

Separate models

The goodness-of-fit plots of the separate models (*figures 1.S1 & 1.S2*) showed that predictions lay around the line of unity and that the CWRESI were observed near the central line. No bias or trend in the model prediction could be determined. The shrinkage observed for the parameters for which between-subject variability was identified (V_1, V_{max}, k_e) was not significant (<17.80% for model T, <15.50% for model R). Additionally, the variability on the η density plots (*figures 1.S5 & 1.S6*) seemed normally distributed around zero.

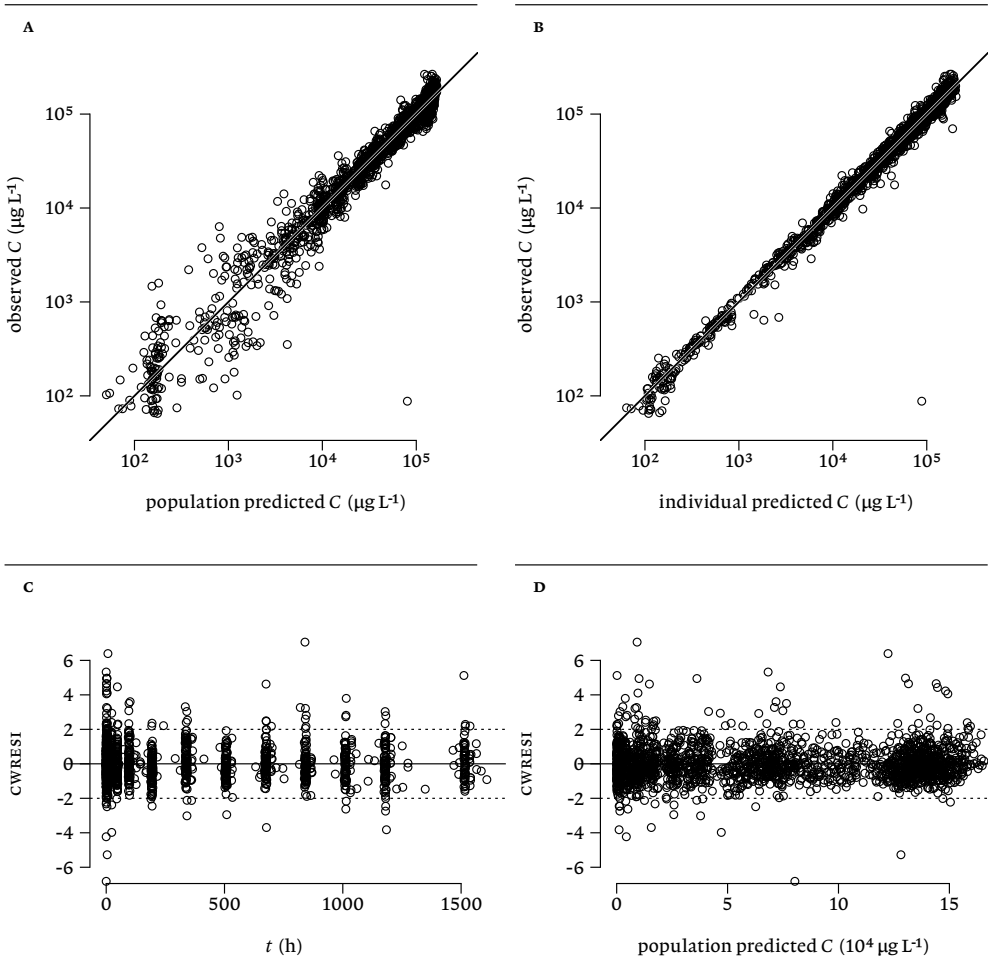

FIGURE 1.2 Goodness-of-fit plots combined model
Observed *vs.* population predicted concentrations (A), observed *vs.* individual predicted concentrations (B), conditional weighted residuals with interaction (CWRESI) *vs.* times (C), and conditional weighted residuals *vs.* population predictions (D) of the combined model.

The population PK parameter estimates from the full covariate model are presented in *table 1.2*. When comparing parameter estimates, most parameter distributions overlapped. The parameter estimates for V_2 differed between model T and model R, but were in the same order of magnitude. However, Q_1 and Q_2 for model T were higher compared to model R. In contrast to the combined model, where between-subject variability was identified for V_1, K_m, and k_e, in the separate models, these were found for V_1, V_{max}, and k_e.

Comparison to NCA

The geometric mean (GM) AUC_1 obtained from the standard NCA was 1,301 µg day mL^{-1} for the test (T) and 1,588 µg day mL^{-1} for the reference (R) product. The AUC_1 remained virtually unchanged when the same calculations were repeated with simulated concentrations, regardless of whether the combined model or the separate models were used (*table 1.3*). Similar results were obtained with regard to AUC_∞. The GM ratio (GMR) T/R with all AUC methods was 81.66-82.54% with the

lower limit (LL) of the associated 90% confidence interval (CI) below the predefined equivalence boundary of 80% (*table 1.3*). Applying a linear correction to the NCA results caused the difference T-R in AUC_t and AUC_∞ to decrease (GMR T/R 89.11-89.55%, LL 90% CI >84.66%). Further reductions were achieved when an equal dosage of 6 mg/kg was *simulated* for both trastuzumab products, which affected the AUCs in the reference product arm more profoundly, and increased the GMR with approximately 2 %-point (*table 1.4*).

Using the entire simulated profile, as opposed to only the simulated concentrations at the original sampling times, generally resulted in a small decrease of 1–2% compared with the NCA for both AUC_t and AUC_∞, with the exception of the AUC_∞ calculated on profiles derived with model R, where an average increase of 1.7% was observed (*tables 1.S1 & 1.S2*). Conversely, with the combined model, lower AUCs were obtained compared with the separate models for T only.

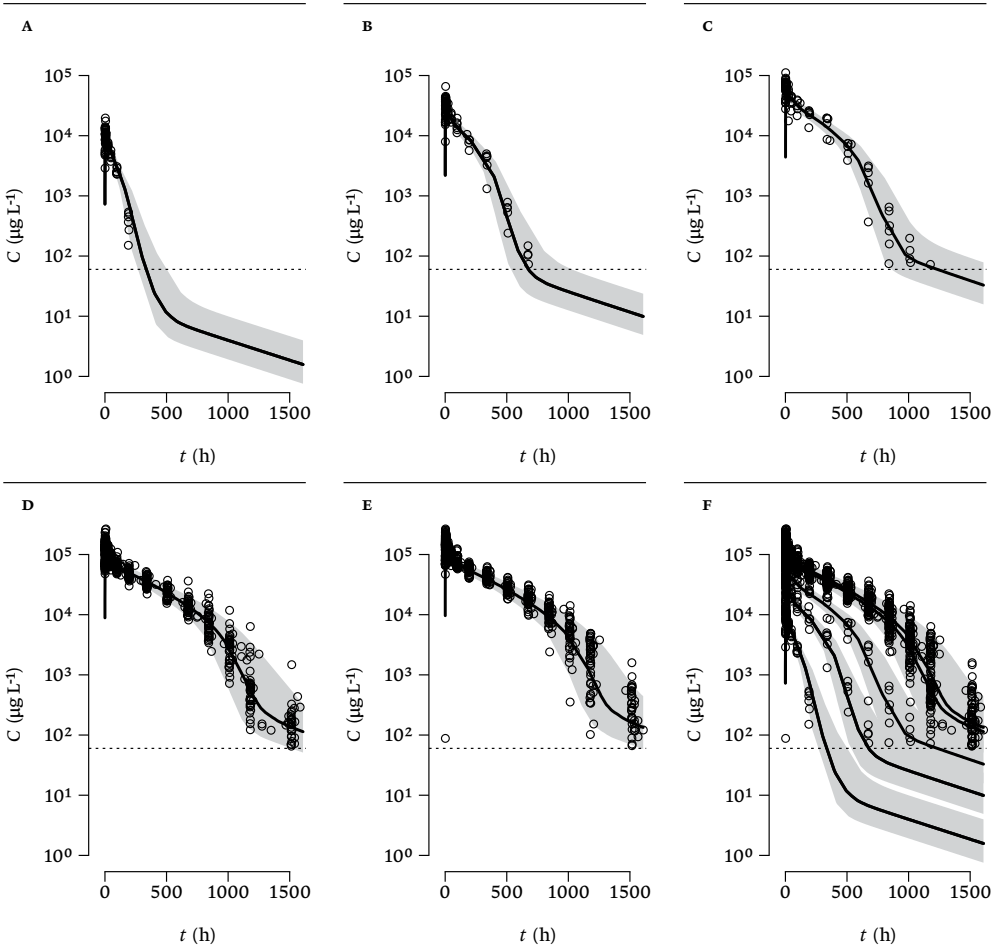

FIGURE 1.3 Visual predictive check (VPC)

Visual predictive check (VPC) of the best combined model, conditioned per dose of test product (0.49 [A], 1.48 [B], 2.96 [C], 5.92 mg/kg [D]) or reference product (6.44 mg/kg, E). In F all doses are displayed. The circles indicate the observations for the different trastuzumab doses administered, the lines are the typical predicted concentrations by the model for each dose, and the grey area is the 95% prediction interval. The dashed line is the assay's lower limit of quantification (LLOQ) for trastuzumab (0.060 µg/mL).

TABLE 1.3 AUC comparison actual dose

	GM		GMR (%)	
	T	R	T/R	
AUC_I				
NCA	1301	1588	81.91	(77.82–86.22)
Separate models	1300	1588	81.86	(78.08–85.82)
Combined model	1296	1588	81.59	(77.88–85.47)
AUC_∞				
NCA	1311	1593	82.32	(78.17–86.69)
Separate models	1313	1591	82.54	(78.70–86.57)
Combined model	1300	1592	81.66	(77.93–85.56)

Geometric mean (GM, μg day mL^{-1}) and GM ratio (GMR) with the 90% confidence interval for the actual dose (5.92 mg/kg for T; 6.44 mg/kg for R) as derived by different methods per treatment arm.

AUC: area under the concentration-time curve | AUC_I: AUC from administration (time 0) to the time of the last concentration ⩾LLOQ (lower limit of quantification) | AUC_∞: AUC_I extrapolated to infinity NCA: non-compartmental analysis | R: reference | T: test

TABLE 1.4 AUC comparison after dose correction

	GM		GMR (%)	
	T	R	T/R	
AUC_I				
NCA	1318	1479	89.11	(84.66–93.79)
Separate models	1323	1455	90.93	(86.72–95.35)
Combined model	1319	1443	91.41	(87.25–95.76)
AUC_∞				
NCA	1329	1484	89.55	(85.03–94.30)
Separate models	1337	1457	91.74	(87.46–96.24)
Combined model	1324	1446	91.54	(87.37–95.92)

Geometric mean (GM, μg day mL^{-1}) and GM ratio (GMR) with the 90% confidence interval for the labelled dose (6 mg/kg) as derived by different methods per treatment arm. For the NCA-results, a linear dose correction was applied; in the models, the labelled dose was used to simulate the individual profiles (see main body).

AUC: area under the concentration-time curve | AUC_I: AUC from administration (time 0) to the time of the last concentration ⩾LLOQ (lower limit of quantification) | AUC_∞: AUC_I extrapolated to infinity NCA: non-compartmental analysis | R: reference | T: test

DISCUSSION

As long as generical products are being developed, controversy and scepticism regarding the claims of therapeutical equality have followed marketed bioequivalent products. Recently, Bate *et al.*[20] advocated that for the more complex pharmaceuticals, two allegedly bioequivalent drug products may not be interchangeable, which could have adverse consequences. MABs are certainly among the most complex therapeutics and establishing similarity to the reference product can thus be challenging.

For demonstrating pharmacokinetic biosimilarity in a human population, a NCA is virtually always performed and its results (AUC and C_{max}) compared statistically, even though it is widely recognised that the NCA is less suitable for drugs with complex non-linear kinetics, as is the case for MABs.

Population approach pharmacokinetic (PK) modelling and simulation techniques have been successfully applied to quantitatively describe the PK of MABs in humans.[21–26] Such an approach has been used in bioequivalence studies, also for biopharmaceuticals,[9] where it was found to give indistinguishable results on the standard NCA-parameters (AUC and C_{max}), as was the case in our analysis. However, as was also argued by Dubois *et al.*,[9] a PK model can provide valuable insight in the biological systems underlying the PK properties. Although the standard NCA-derived parameters, such as C_{max}, AUC_∞, terminal half-life, *etc.*, may seem similar, the two drug products could behave quite differently in terms of PK, a feature that goes undetected in a NCA.[27] Furthermore, similar plasma concentrations do not invariably mean similar concentrations at the site of action.

Here, we describe two methods of incorporating PK modelling in biosimilarity research. The first approach is developing a model on all available data from both test and reference product(s) and carefully examining possible bias in one of the treatment groups. Testing for (statistically) significant differences between drug products can be done for all the model parameters via covariate analysis. Covariate testing follows a well-established statistical distribution that can be used for statistical inference.[28,29] If no significant correlations can be identified between the drug products and if attempts to incorporate treatment as covariate in the model fail to improve it, the biosimilarity claim is supported.

The second method entails the development of different models, one for each test and reference product(s), which in contrast to a combined model does not assume similarity between test and reference product as a starting point. This method allows comparison of the model structure, that should be

identical for biosimilar products, and of model parameters for both test and reference product.

Comparing different PK models inevitably reveals minor differences for which the clinical significance needs to be discussed. For example, in model T the optimal inter-compartmental clearances (Q_1 and Q_2) were estimated to be a factor 10^2–10^3 higher than the corresponding parameters in the other models, while the striking dissimilarity did not seem to affect the descriptive properties of the overall profiles. However, as the (fictive) second and third compartment were not sampled, this finding merely reflects a mathematical solution to a rather complex problem and not necessarily a true (*e.g.* physiological or pharmacological) difference. Additionally, the higher dose administered for the reference product could have allowed a better characterisation of the terminal portion of the PK profile (elimination parameters), which also affects the estimation of remaining parameters, such as Q_1 and Q_2.

This represents an important limitation of the second method, which may be of particular relevance when modelling PK data from two different populations separately. Unfortunately, pharmacokinetic biosimilarity of biopharmaceuticals is regularly investigated in trials of parallel design, because of the long half-life and the potential of anti-drug antibodies development, which could influence the pharmacokinetics.[30] Theoretically, all MABs share common pharmacokinetic properties, *e.g.* small central volume of distribution, no renal excretion due to large molecular size, metabolism into amino-acids and peptides, both specific (non-linear) and non-specific (linear) cellular uptake and degradation elimination mechanisms.[31–35] Thus, the remaining variability is probably determined by patient characteristics. When comparing the model parameters of the separate models, one of the most prominent differences is the population estimate for V_1, which is unlikely caused by a difference between test and reference product.

The combined model equally well described the data, without bias in either the test or reference group. Adding trastuzumab drug product as covariate to the model could not explain any residual variability, which strongly supports the biosimilarity claim, but also indicates that the difference in AUCs must be attributed to population characteristics.

From a regulatory perspective, another limitation of the second method is the lack of proper statistical inference testing on the model parameters. One might consider overlapping confidence intervals for parameter estimates indications for biosimilarity, but many parameters are related, so that – for example – a low inter-compartmental clearance may be 'compensated' in the model by a low volume of distribution. An extension of 'bioequivalence statistics' has been applied to model parameters by Wilkens *et al.*,[36] although their method suffers from the aforementioned limitations as well.

Notwithstanding the limitations of PPK, it has several benefits over a NCA. Importantly, a PPK is not concerned with differences in administered doses. Although the EMA allows a dose correction in the bioequivalence guideline (for chemically derived products) if the difference exceeds 5%, the NCA assumes linearity in its correction, which is not appropriate for MABs, that display non-linear pharmacokinetics. Other benefits of PPK include the possibility to identify and thus correct for certain covariates, and the relative robustness of a PPK against protocol deviations, with regard to timing of sample collection, missing samples, duration of intravenous administration, and incomplete administration.[8,37]

Simulations with model R revealed that the two allowed extremes for protein content per batch (effective doses 5.28 mg/kg and 7.2 mg/kg) would result in a 90% CI for the GMR for AUC_1 of 146.39–147.22% in a crossover design ($n = 46$). If such batch-to-batch variations are not considered relevant, then the consequences on the standard biosimilarity parameters may also be argued to be irrelevant.

With a PK model, multiple scenarios can be simulated within these extremes, which can be used to build the case that the test product achieves therapeutical drug concentrations, similar to the reference product, when administered according to a certain dosing regimen. This approach also circumvents some of the aforementioned limitations of direct comparison of two or more models. If a biomarker or pharmacological effect can be measured in the biosimilarity trial and incorporated in a pharmacokinetic-pharmacodynamic model (pharmacodynamic model), a relevant clinical target may be simulated and lend further support to a biosimilarity claim.

The NCA will most likely remain a gold standard in biosimilarity research, even for the complex MABs. Nonetheless, the model approach can serve as an elegant add-on. Questions that need to be addressed before a PPK can fully substitute the NCA in demonstrating biosimilarity relate to selection of the most

meaningful PK or pharmacodynamic parameter from the model, and the minimal population size to detect with sufficient statistical power relevant (model) differences.

Previously, the benefits of modelling and simulation have been proposed for proof of biosimilarity, to which this chapter adds similar benefits for MABs.

REFERENCES

1 Bonafede M, Joseph GJ, Princic N, Harrison DJ. Annual acquisition and administration cost of biologic response modifiers per patient with rheumatoid arthritis, psoriasis, psoriatic arthritis, or ankylosing spondylitis. *J Med Econ* 2013; **16**: 1120–8.

2 Francis SM, Heyliger A, Miyares MA, Viera M. Potential cost savings associated with dose rounding antineoplastic monoclonal agents. *J Oncol Pharm Pract* 2015; **21**: 280–4.

3 IMS health. IMS Health Top 20 Global Products 2014. 2015. http://www.imshealth.com/files/web/Corporate/News/Top-Line Market Data/2014/Top_20_Global_Products_2014.pdf (accessed April 1, 2016).

4 WHO Expert Committee on Biological Standardization. Guidelines on evaluation of similar biotherapeutic products (SBPs). Genève, 2009.

5 US Food and Drug Administration. Scientific considerations in demonstrating biosimilarity to a reference product. Silver Spring: US Department of Health and Human Services, 2015.

6 Committee for Medicinal Products for Human Use (CHMP). Guideline on similar biological medicinal products. London: European Medicines Agency, 2015.

7 Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat Med* 2005; **24**: 1509–24.

8 Dubois A, Gsteiger S, Pigeolet E, Mentré F. Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses: evaluation of estimates of sample means and type I error for different designs. *Pharm Res* 2010; **27**: 92–104.

9 Dubois A, Gsteiger S, Balser S, *et al.* Pharmacokinetic similarity of biologics: analysis using nonlinear mixed-effects modeling. *Clin Pharmacol Ther* 2012; **91**: 234–42.

10 Wisman LAB, De Cock EPM, Reijers JAA, *et al.* A phase I dose-escalation and bioequivalence study of a trastuzumab biosimilar in healthy male volunteers. *Clin Drug Investig* 2014; **34**: 887–94.

11 US Food and Drug Administration. Population pharmacokinetics. Silver Spring: US Department of Health and Human Services, 1999.

12 Committee for Medicinal Products for Human Use (CHMP). Guideline on reporting the results of population pharmacokinetic analyses. London: European Medicines Agency, 2008.

13 Janmahasatian S, Duffull SB, Ash S, Ward LC, Byrne NM, Green B. Quantification of lean bodyweight. *Clin Pharmacokinet* 2005; **44**: 1051–65.

14 Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med* 1987; **317**: 1098.

15 Post TM, Freijer JI, Ploeger BA, Danhof M. Extensions to the visual predictive check to facilitate model performance evaluation. *J Pharmacokinet Pharmacodyn* 2008; **35**: 185–202.

16 Karlsson MO, Savić RM. Diagnosing model diagnostics. *Clin Pharmacol Ther* 2007; **82**: 17–20.

17 Luu KT, Bergqvist S, Chen E, Hu-Lowe D, Kraynov E. A model-based approach to predicting the human pharmacokinetics of a monoclonal antibody exhibiting target-mediated drug disposition. *J Pharmacol Exp Ther* 2012; **341**: 702–8.

18 Mager DE, Jusko WJ. General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J Pharmacokinet Pharmacodyn* 2001; **28**: 507–32.

19 Gibiansky L, Gibiansky E, Kakkar T, Ma P. Approximations of the target-mediated drug disposition model and identifiability of model parameters. *J Pharmacokinet Pharmacodyn* 2008; **35**: 573–91.

20 Bate R, Mathur A, Lever HM, *et al.* Generics substitution, bioequivalence standards, and international oversight: complex issues facing the FDA. *Trends Pharmacol Sci* 2016; **37**: 184–91.

21 Bruno R, Washington CB, Lu J-F, Lieberman G, Banken L, Klein P. Population pharmacokinetics of trastuzumab in patients with HER2+ metastatic breast cancer. *Cancer Chemother Pharmacol* 2005; **56**: 361–9.

22 Charoin J-E, Jacqmin P, Banken L, Lennon S, Jorga K. Population pharmacokinetic analysis of trastuzumab (Herceptin) following long-term administration using different regimens. *PAGE* 2004; **13**: Abstr 489.

23 Fukushima Y, Charoin J-E, Brewster M, Jonsson EN. Population pharmacokinetic analysis of trastuzumab (Herceptin®) based on data from three different dosing regimens. *PAGE* 2007; **16**: Abstr 1121.

24 Kloft C, Graefe E-U, Tanswell P, *et al.* Population pharmacokinetics of sibrotuzumab, a novel therapeutic monoclonal antibody, in cancer patients. *Invest New Drugs* 2004; **22**: 39–52.

25 Kuester K, Kovar A, Lüpfert C, Brockhaus B, Kloft C. Refinement of the population pharmacokinetic model for the monoclonal antibody matuzumab: external model evaluation and simulations. *Clin Pharmacokinet* 2009; **48**: 477–87.

26 Van Hasselt JGC, Boekhout AH, Beijnen JH, Schellens JHM, Huitema ADR. Population pharmacokinetic-pharmacodynamic analysis of trastuzumab-associated cardiotoxicity. *Clin Pharmacol Ther* 2011; **90**: 126–32.

27 Rescigno A, Powers J, Herderick EE. Bioequivalent or nonbioequivalent? *Pharmacol Res* 2001; **43**: 543–7.

28 Khandelwal A, Harling K, Jonsson EN, Hooker AC, Karlsson MO. A fast method for testing covariates in population PK/PD models. *AAPS J* 2011; **13**: 464–72.

29 Meibohm B, Derendorf H. Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling. *Int J Clin Pharmacol Ther* 1997; **35**: 401–13.
30 Committee for Medicinal Products for Human Use (CHMP). Guideline on similar biological medicinal products containing monoclonal antibodies: non-clinical and clinical issues. London: European Medicines Agency, 2012.
31 Wang W, Wang EQ, Balthasar JP. Monoclonal antibody pharmacokinetics and pharmacodynamics. *Clin Pharmacol Ther* 2008; **84**: 548–58.
32 Tabrizi MA, Tseng CL, Roskos LK. Elimination mechanisms of therapeutic monoclonal antibodies. *Drug Discov Today* 2006; **11**: 81–8.
33 Ternant D, Bejan-Angoulvant T, Passot C, Mulleman D, Paintaud G. Clinical pharmacokinetics and pharmacodynamics of monoclonal antibodies approved to treat rheumatoid arthritis. *Clin Pharmacokinet* 2015; **54**: 1107–23.
34 Dirks NL, Meibohm B. Population pharmacokinetics of therapeutic monoclonal antibodies. *Clin Pharmacokinet* 2010; **49**: 633–59.
35 Keizer RJ, Huitema ADR, Schellens JHM, Beijnen JH. Clinical pharmacokinetics of therapeutic monoclonal antibodies. *Clin Pharmacokinet* 2010; **49**: 493–507.
36 Wilkins JJ, Gautier A, Lowe PJ. Bioequivalence, bootstrapping and case-deletion diagnostics in a biologic: a model-based analysis of the effect of formulation differences in a monoclonal antibody. *PAGE* 2008; **17**: Abstr 1284.
37 Charles B. Population pharmacokinetics: an overview. *Aust Prescr* 2014; **37**: 210–3.

SUPPLEMENT

$$V_{1,i} = \theta_p \frac{x_i}{\tilde{x}} e^{\eta_i} \qquad \textbf{(1.S1)}$$

$$k_{e,i} = \theta_p \frac{x_i}{\tilde{x}} e^{\eta_i} \qquad \textbf{(1.S2)}$$

Here, i is the i^{th} individual, V_1 the volume of distribution, k_e the elimination rate constant, θ_p the population parameter estimate, x the covariate (respectively LBW and BMI), $\tilde{x}$ the median of x, and η the inter-individual variability.

TABLE 1.S1 AUC comparison actual dose

	TEST						REFERENCE					
	AUC		%-change[a]		%-change[b]		AUC		%-change[a]		%-change[b]	
AUC_1												
NCA	1318	(220)					1602	(220)				
Separate models												
Actual time	1315	(211)	100.0	(5.0)			1599	(190)	100.1	(4.5)		
Continuous time	1292	(206)	98.3	(4.9)			1571	(188)	98.4	(4.5)		
Combined model												
Actual time	1310	(204)	99.8	(5.4)	99.7	(1.1)	1599	(193)	100.1	(4.9)	100.0	(1.5)
Continuous time	1287	(199)	98.0	(5.2)	99.7	(1.1)	1571	(191)	98.4	(5.0)	100.0	(1.5)
AUC_∞												
NCA	1328	(225)					1607	(224)				
Separate models												
Actual time	1329	(216)	100.3	(5.2)			1602	(192)	100.0	(4.5)		
Continuous time	1312	(208)	99.0	(5.0)			1630	(201)	101.7	(4.7)		
Combined model												
Actual time	1315	(206)	99.3	(5.7)	99.1	(3.3)	1604	(195)	100.1	(4.9)	100.1	(1.5)
Continuous time	1298	(200)	98.1	(5.3)	99.1	(1.1)	1577	(193)	98.4	(5.0)	96.8	(1.2)

Mean (standard deviation) AUCs (μg day mL^{-1}) and mean (standard deviation) percentage change for the actual dose (test 5.92 mg/kg; reference 6.44 mg/kg), as derived by different methods per treatment arm. AUCs of combined and separate models are compared to the NCA result (a); AUCs of combined model are compared to the AUCs of the separate model (b, actual time compared to actual time, continuous time to continuous time). For comparison to a standard NCA, AUC_1 was calculated using model simulated (predicted) individual concentrations at the original sampling times ('actual time'). Extrapolation to infinity (AUC_∞) was based on the apparent terminal elimination rate constant. Additionally, the AUCs were derived by integration of the simulated concentration-time profiles ('continuous time'); AUC_1 from the administration time to the last concentration used in the NCA, AUC_∞ until infinity.
AUC: area under the concentration-time curve|NCA: non-compartmental analysis

TABLE 1.S2 AUC comparison after dose correction

	TEST						REFERENCE					
	AUC		%-change[a]		%-change[b]		AUC		%-change[a]		%-change[b]	
AUC_t												
NCA	1335	(223)					1493	(205)				
Separate models												
Actual time	1338	(215)	100.5	(5.1)			1465	(176)	98.4	(4.5)		
Continuous time	1315	(210)	98.7	(5.0)			1439	(174)	96.7	(4.4)		
Combined model												
Actual time	1334	(208)	100.2	(5.4)	99.7	(1.1)	1453	(176)	97.7	(4.8)	99.2	(1.2)
Continuous time	1310	(202)	98.5	(5.3)	99.7	(1.2)	1427	(174)	95.9	(4.8)	99.2	(1.3)
AUC_∞												
NCA	1346	(228)					1497	(209)				
Separate models												
Actual time	1353	(220)	100.7	(5.2)			1467	(177)	98.3	(4.5)		
Continuous time	1335	(212)	99.5	(5.0)			1493	(185)	100.0	(4.7)		
Combined model												
Actual time	1339	(210)	99.8	(5.8)	99.1	(3.4)	1456	(177)	97.6	(4.8)	99.2	(1.2)
Continuous time	1322	(204)	98.6	(5.3)	99.1	(1.2)	1432	(175)	96.0	(4.8)	96.0	(1.0)

Mean (standard deviation) AUCs (μg day mL^{-1}) and mean (standard deviation) percentage change for the labelled dose (6 mg/kg), as derived by different methods per treatment arm. For the NCA-results, a linear dose correction was applied; in the models, the labelled dose was used to simulate the individual profiles (see main body). AUCs of combined and separate models are compared to the NCA result (a); AUCs of combined model are compared to the AUCs of the separate model (b, actual time compared to actual time, continuous time to continuous time). For comparison to a standard NCA, AUC_t was calculated using model simulated (predicted) individual concentrations at the original sampling times ('actual time'). Extrapolation to infinity (AUC_∞) was based on the apparent terminal elimination rate constant. Additionally, the AUCs were derived by integration of the simulated concentration-time profiles ('continuous time'); AUC_t from the administration time to the last concentration used in the NCA, AUC_∞ until infinity.

AUC: area under the concentration-time curve | NCA: non-compartmental analysis

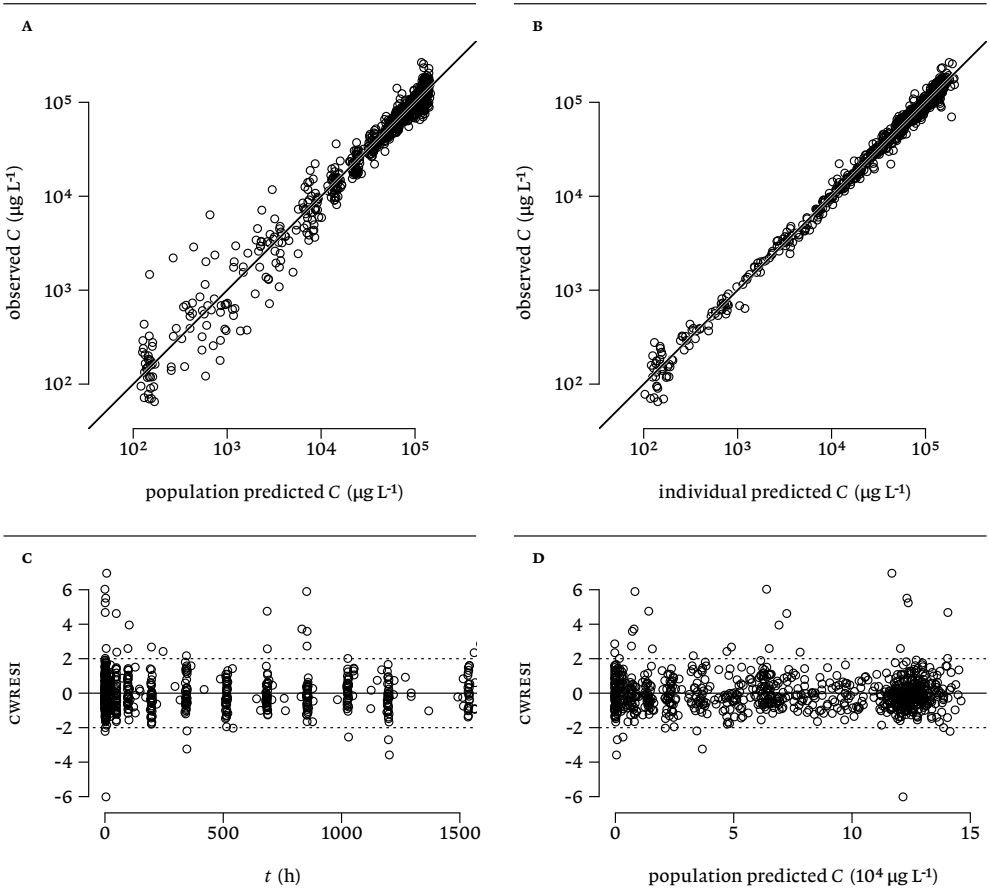

FIGURE 1.S1 Goodness-of-fit plots model T

Observed *vs.* population predicted concentrations (A), observed *vs.* individual predicted concentrations (B), conditional weighted residuals with interaction (CWRESI) *vs.* times (C), and conditional weighted residuals *vs.* population predictions (D) of the separate model for the test product.

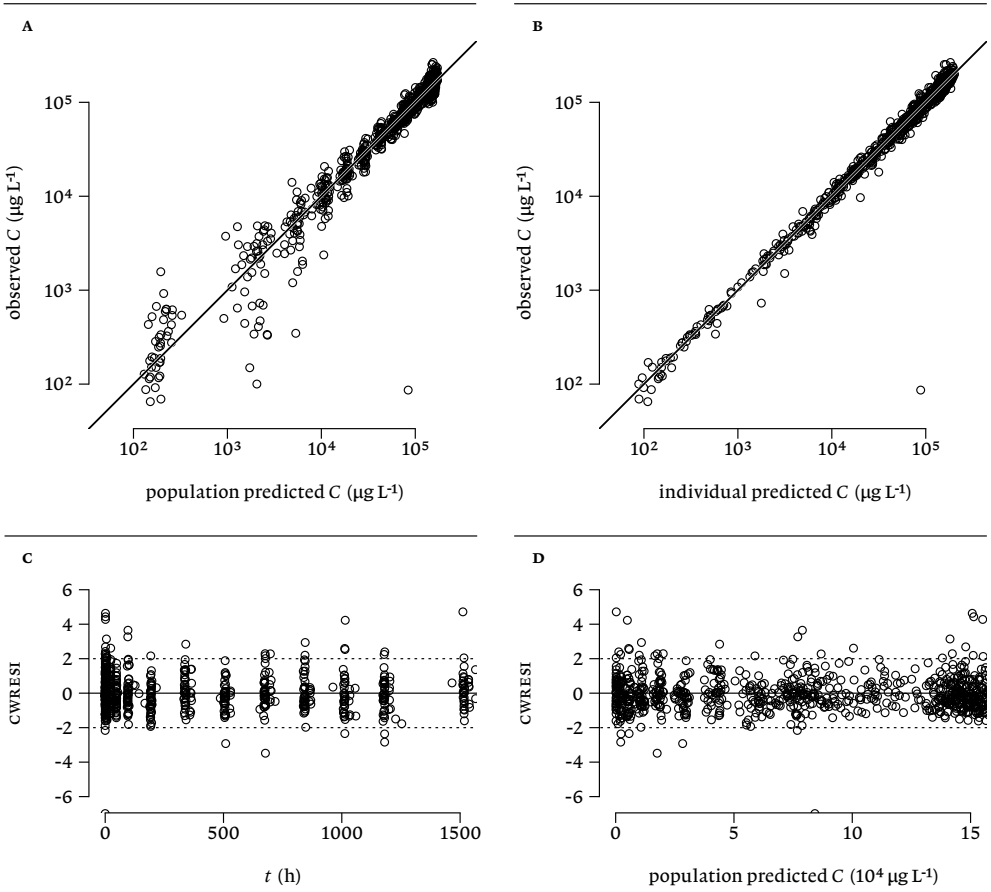

FIGURE 1.S2 Goodness-of-fit plots model R

Observed *vs.* population predicted concentrations (A), observed *vs.* individual predicted concentrations (B), conditional weighted residuals with interaction (CWRESI) *vs.* times (C), and conditional weighted residuals *vs.* population predictions (D) of the separate model for the reference product.

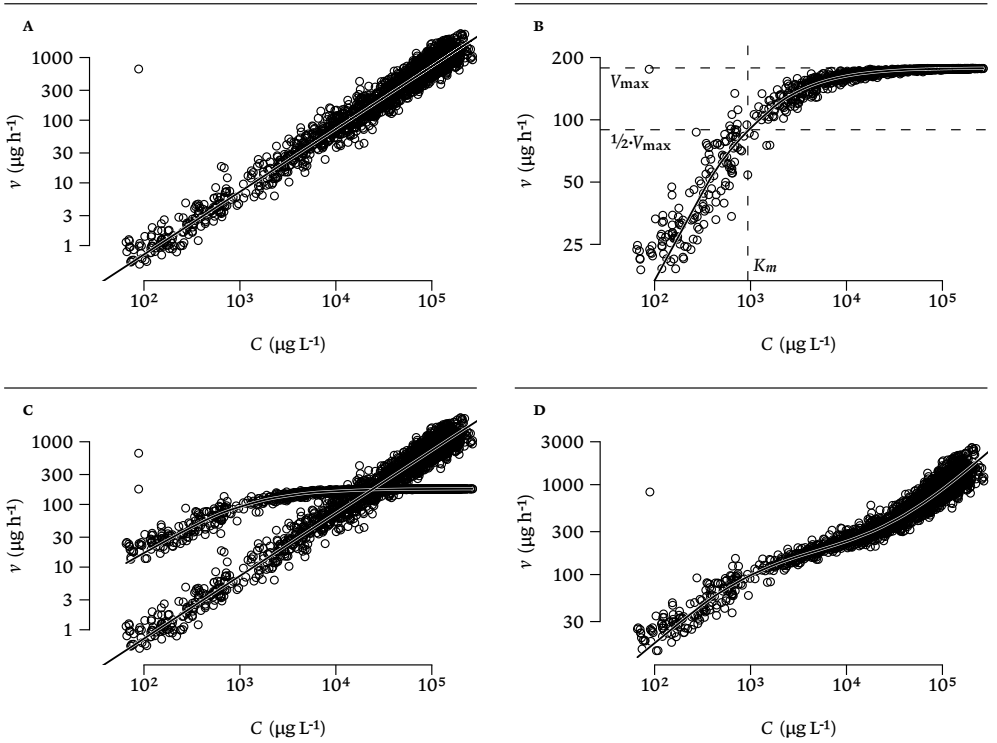

FIGURE 1.S3 Linear and non-linear clearance combined model

The linear clearance (A), the non-linear clearance (B), the combined linear and non-linear clearances (C), and the total clearance (D) of the combined model.

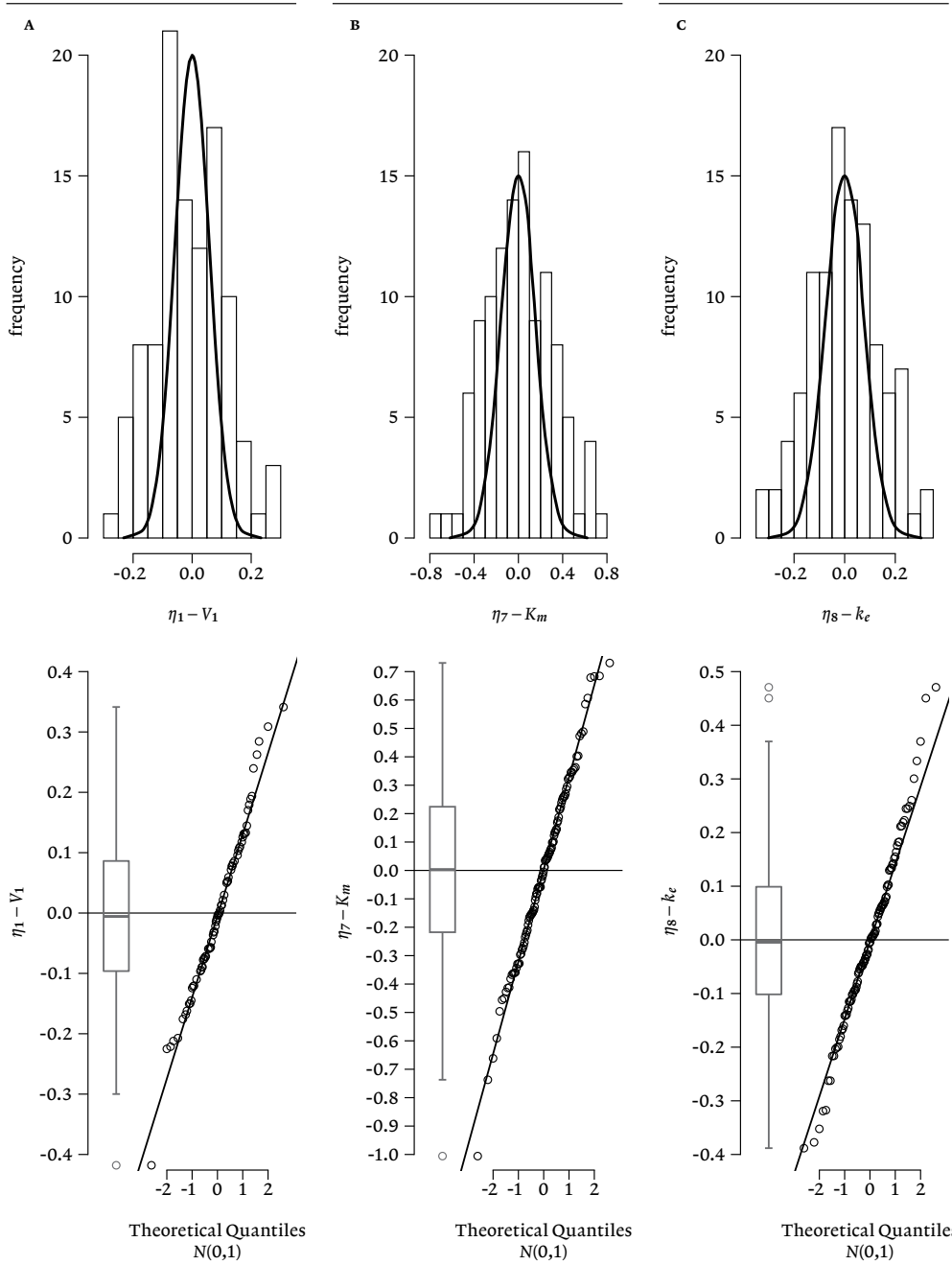

FIGURE 1.S4 η density combined model
Distribution of ηs in the combined model for V_1 (η_1, A), K_m (η_7, B), and k_e (η_8, C).

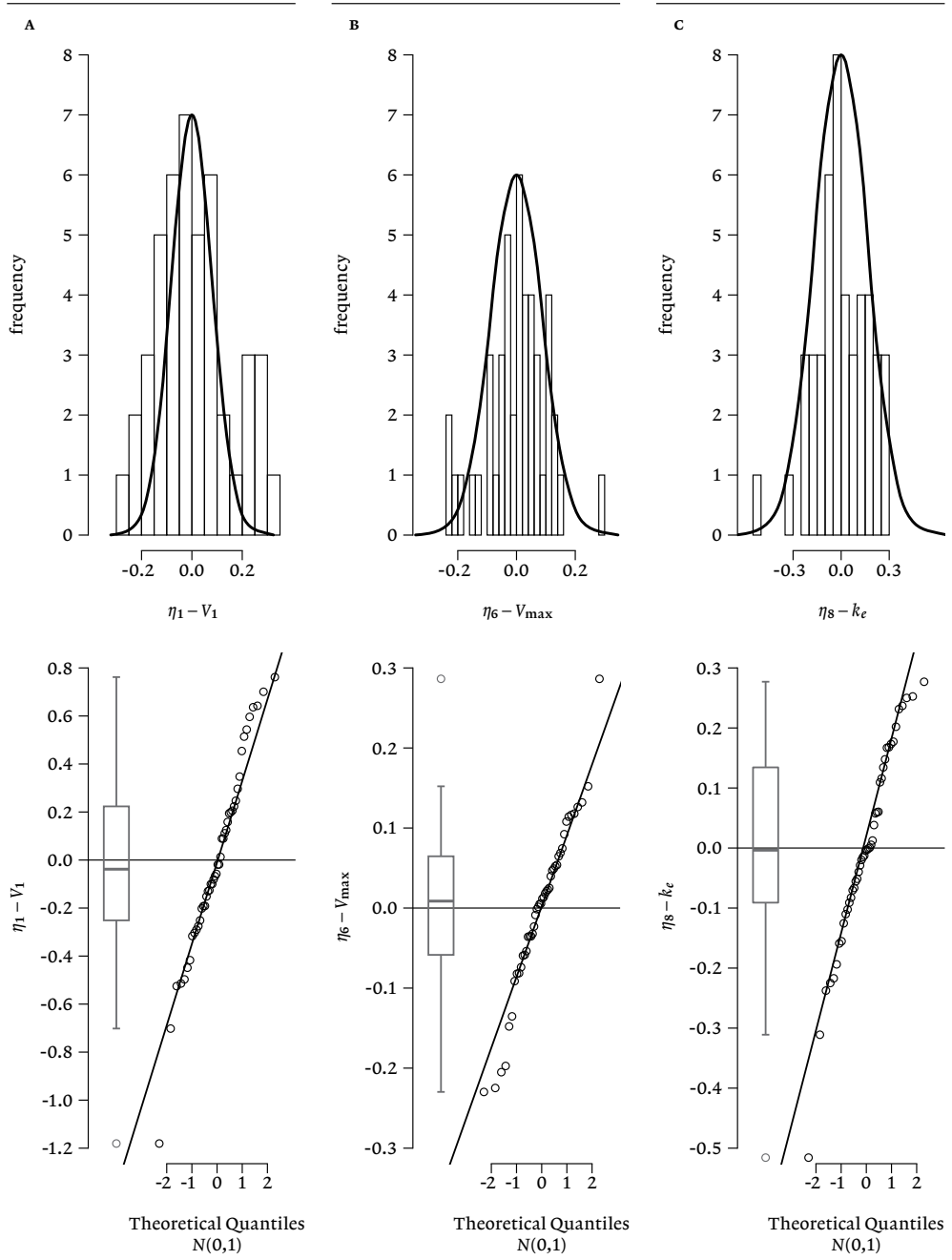

FIGURE 1.S5 η density model T
Distribution of ηs in model T for V_1(η_1, A), V_{max}(η_6, B), and k_e(η_8, C).

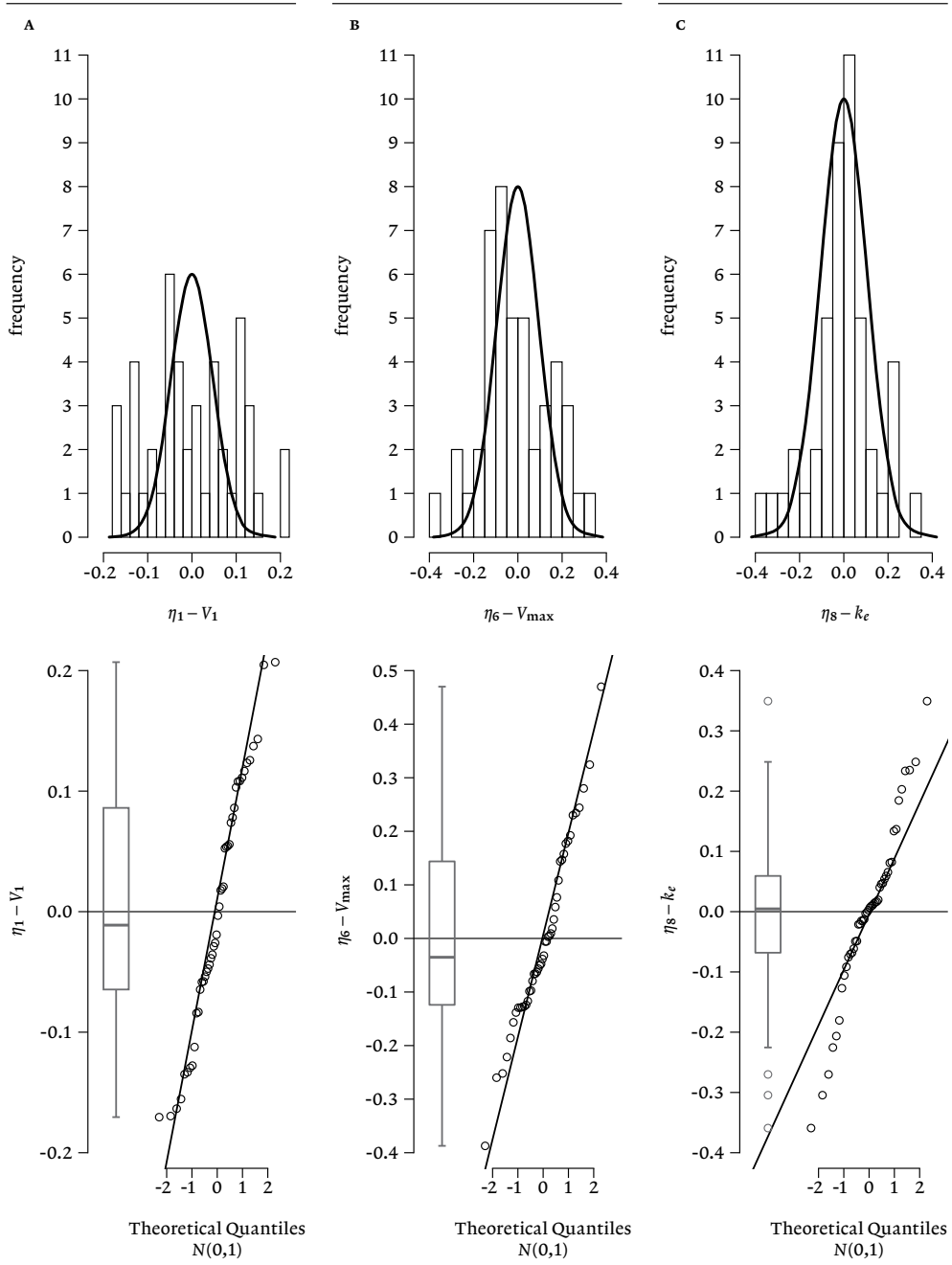

FIGURE 1.S6 η density model R
Distribution of ηs in model R for $V_1(\eta_1, \text{A})$, $V_{max}(\eta_6, \text{B})$, and $k_e(\eta_8, \text{C})$.