British Journal of Clinical Pharmacology

Br J Clin Pharmacol (2016) **82** 1568–1579 1568

PHARMACOKINETIC DYNAMIC RELATIONSHIPS

Comparative assessment of clinical response in patients with rheumatoid arthritis between PF-05280586, a proposed rituximab biosimilar, and rituximab

Correspondence Donghua Yin, Pfizer Inc, 10777 Science Center Drive, CB1, San Diego, California 92121, USA. Tel.: +1 858 526 4942; E-mail: donghua.yin@pfizer.com

Received 19 January 2016; **Revised** 8 July 2016; **Accepted** 8 August 2016

Jason H. Williams[1], Matthew M. Hutmacher[2], Matthew L. Zierhut[1], Jean-Claude Becker[3], Barry Gumbiner[1], George Spencer-Green[4], Lisa A. Melia[1], Kai-Hsin Liao[1], Matthew Suster[1], Donghua Yin[1], Ruifeng Li[4] and Xu Meng[1]

[1]Pfizer Inc., San Diego, California, USA, [2]Ann Arbor Pharmacometrics Group, Ann Arbor, Michigan, USA, [3]Pfizer Inc., New York, New York, USA, and [4]Pfizer Inc., Cambridge, Massachusetts, USA

Keywords Biosimilar, pharmacodynamics, pharmacokinetics

AIMS

To evaluate potential differences between PF-05280586 and rituximab sourced from the European Union (rituximab-EU) and USA (rituximab-US) in clinical response (Disease Activity Score in 28 Joints [DAS28] and American College of Rheumatology [ACR] criteria), as part of the overall biosimilarity assessment of PF-05280586.

METHODS

A randomised, double-blind, pharmacokinetic similarity trial was conducted in patients with active rheumatoid arthritis refractory to anti-tumour necrosis factor therapy on a background of methotrexate. Patients were treated with 1000 mg of PF-05280586, rituximab-EU or rituximab-US on days 1 and 15 and followed over 24 weeks for pharmacokinetic, clinical response and safety assessments. Key secondary end points were the areas under effect curves for DAS28 and ACR responses. Mean differences in areas under effect curves were compared against respective reference ranges established by observed rituximab-EU and rituximab-US responses using longitudinal nonlinear mixed effects models.

RESULTS

The analysis included 214 patients. Demographics were similar across groups with exceptions in some baseline disease characteristics. Baseline imbalances and group-to-group variation were accounted for by covariate effects in each model. Predictions from the DAS28 and ACR models tracked the central tendency and distribution of observations well. No point estimates of mean differences were outside the reference range for DAS28 or ACR scores. The probabilities that the predicted differences between PF-05280586 *vs.* rituximab-EU or rituximab-US lie outside the reference ranges were low.

CONCLUSIONS

No clinically meaningful differences were detected in DAS28 or ACR response between PF-05280586 and rituximab-EU or rituximab-US as the differences were within the pre-specified reference ranges. TRIAL REGISTRATION Number: NCT01526057.

DOI:10.1111/bcp.13094

© 2016 Pfizer Inc. British Journal of Clinical Pharmacology published by John Wiley & Sons Ltd on behalf of British Pharmacological Society.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT

- Pharmacodynamics and clinical response data collected during early clinical pharmacology studies can add to the totality of the evidence, reduce residual uncertainty, in support of the overall demonstration of biosimilarity, as described in regulatory guidance. However, no report is available to share the knowledge on and experience with this topic.

WHAT THIS STUDY ADDS

- This analysis introduces an objective approach to define 'meaningful effect' and quantifies a reference range in which to perform comparative assessments.
- Model-based approaches may help reduce the number of studies necessary to demonstrate biosimilarity to bring affordable versions of biologics to patients who otherwise have no or limited access.

Tables of Links

TARGETS	
Other protein targets [2]	
CD20	TNF-alpha

LIGANDS
Rituximab

These Tables lists key protein targets and ligands in this article that are hyperlinked to corresponding entries in http://www.guidetopharmacology.org, the common portal for data from the IUPHAR/BPS Guide to PHARMACOLOGY [1], and are permanently archived in the Concise Guide to PHARMACOLOGY 2015/16 [2].

Introduction

Therapies based on biologics are increasingly important in treating many diseases due to their success in addressing previously unmet medical needs. In particular, biologics have greatly improved the management of rheumatoid arthritis (RA), demonstrating efficacy and safety in alleviating symptoms, inhibiting bone erosion and preventing loss of function [3]. However, access to biologic medicines can be limited, particularly in resource-constrained countries [4]. Biosimilars, biologic products that are highly similar to a reference product in terms of quality, biological activity, safety and efficacy, are expected to be an essential component in reducing health care costs and enhancing patient access to these important, often lifesaving medications [4]. In the USA, the Biologics Price Competition and Innovation Act of 2009 established an abbreviated pathway for US Food and Drug Administration (FDA) licensure of products that are biosimilar to a licensed reference product [5]. Similar regulatory pathways for biosimilar applications have been established elsewhere [6, 7]. A key component of the abbreviated pathway is extrapolation of clinical data to one or more additional indications [6, 8].

The role of clinical pharmacology studies in demonstration of biosimilarity is emphasised in the FDA's guidance for industry [9]. These studies provide data that describe the degree of similarity in drug exposure, also referred to as pharmacokinetic (PK) similarity, between the proposed biosimilar and reference product(s). In addition, clinical pharmacology studies often include pharmacodynamic (PD) end points and pharmacometric analyses to assess whether clinically meaningful differences exist between the proposed biosimilar and reference product(s). Clinical pharmacology data can add to the totality of the evidence and reduce residual uncertainty, thus guiding the need for, and design of, subsequent clinical testing. Furthermore, clinical pharmacology data can provide scientific justification supporting extrapolation of clinical data to one or more additional indications.

Rituximab is a chimeric anti-CD20 monoclonal antibody that selectively targets and depletes B cells and has demonstrated significant clinical benefit in patients with RA and non-Hodgkin's lymphoma [10, 11]. PF-05280586 has the same primary amino acid sequence as rituximab, with similar physicochemical and *in vitro* functional properties, and is under development as a potential biosimilar to rituximab [12, 13]. The PK similarity of PF-05280586 to rituximab sourced from the European Union (rituximab-EU) and US (rituximab-US), as well as PK similarity of rituximab-EU to rituximab-US, was demonstrated in a multicentre, multinational, randomised double-blind, controlled trial in patients with active RA on a background of methotrexate who had an inadequate response to one or more tumour necrosis factor antagonist therapies [12]. Use of reference products sourced from different regions (i.e. EU and US) is part of standard PK similarity study design for not only meeting the reference-specific PK similarity requirement but also for providing scientific justification for use of a single reference product in subsequent trials [8].

This trial was designed to demonstrate PK similarity, yet clinical response end points were also collected during the 24-week study period. The study was therefore not powered for standard statistical evaluation of efficacy. Using a population PK/PD (PopPK/PD) modelling approach that was planned prospectively, analysis of clinical end points was conducted to assess any potential clinically meaningful difference between the proposed biosimilar and a reference product. The approach took advantage of the multiple repeated measurements for each clinical end point and variability observed between the two reference products using the

assumption that differences in clinical responses between the two reference products would not be clinically meaningful if PK similarity was established. The key aspect of this approach was to utilise data from the two reference arms for constructing a reference range of 'no clinically meaningful difference' for comparative assessments of PF-05280586 to the reference products. We present this PopPK/PD modelling analysis as a case study for utilizing clinical response data from a clinical pharmacology study to add to the overall demonstration of biosimilarity.

Methods

This study is registered at ClinicalTrials.gov (NCT01526057) and was conducted in compliance with the Declaration of Helsinki and with all International Conference on Harmonisation Good Clinical Practice guidelines. In addition, all local regulatory requirements were followed, in particular, those affording greater protection to the safety of trial participants. The final protocol, amendments and informed consent documentation were reviewed and approved by Institutional Review Boards and/or Independent Ethics Committees at each participating centre. A signed and dated informed consent was required from each patient before any screening procedures were conducted.

Study design

The study was a randomised, double-blind, controlled trial in patients with active RA on a background of methotrexate who had inadequate responses to one or more tumour necrosis factor antagonist therapies [12]. The primary objective was to demonstrate PK similarity of PF-05280586, rituximab-EU and rituximab-US. The secondary objectives included those described herein, which were to use PopPK/PD modelling approaches to integrate PK and PD data for the purpose of detecting potential differences in PK/PD profiles. Other secondary objectives included evaluation of overall safety, tolerability and immunogenicity. The results of this study, except for those from the PopPK/PD modelling, have been presented elsewhere [12]. Full details of the study design have been described [12]. Briefly, eligible participants were adults (aged ≥18 years) with confirmed diagnosis of RA based on 2010 American College of Rheumatology (ACR)/European League Against Rheumatism classification criteria [14]. Patients were required to: meet class I, II or III of the ACR 1991 revised criteria for Global Functional Status in Rheumatoid Arthritis [15]; have RA seropositivity, as documented by a screening assessment for rheumatoid factor; and/or anticyclic citrullinated peptide antibodies; active disease, as defined by the following: at least six tender/painful joints (of 68 assessed) and six or more swollen joints (of 66 assessed) at screening and baseline, high-sensitivity C-reactive protein (CRP) greater than the upper limit of normal, or Patient's Global Assessment of arthritis score ≥50 at screening, and Disease Activity Score in 28 joints (DAS28)–CRP >3.2 at screening; be on a stable dose of oral or parenteral methotrexate 10–25 mg per week (or as low as 7.5 mg per week, in the case of prior poor tolerance) for at least 3 months and receiving the stable dose for at least 4 weeks prior to first dose of study drug; and have inadequate response, in the opinion of the investigator, to one or more approved TNF-antagonist therapies, defined as failure to achieve adequate clinical response during prior TNF-antagonist therapy, relapse following clinical response to TNF-antagonist therapy or an adverse event resulting in discontinuation of TNF antagonist. A sample size calculation based on standard bioequivalence determined that up to 210 subjects were to be enrolled to ensure at least 195 subjects (65 subjects in each treatment arm) completed the study procedures per protocol. The PopPK/PD analysis population was defined as all randomised subjects who receive full doses of the assigned study treatment and had at least one protocol-specified measurement for drug concentration and the PD response of interest collected after receiving study treatment, as well as the respective baseline values. Patients received 1000 mg of PF-05280586, rituximab-EU or rituximab-US by intravenous infusion on days 1 and 15. All subjects were to receive premedication with 100 mg intravenous methylprednisolone (or its equivalent) prior to each infusion.

Clinical response

Efficacy following rituximab therapy in patients with RA included improvements in measures of clinical response such as DAS28 and categorical American College of Rheumatology (ACR) response [16–18]. DAS28 is a function of four components: tender joint counts 28 (TJ28) and swollen joint counts 28 (SJ28), CRP and patient's global assessment of arthritis (PGA). The ACR responder criteria are a function of a core set of seven components: TJ68, SJ66, CRP, PGA, Health Assessment Questionnaire Disability Index, patient's assessment of arthritis pain, and physician's global assessment of arthritis. A patient was considered a responder on the ACR20 assessment if >20% improvement in tender and swollen joint counts was observed with >20% improvement in three of the five other component variables; ACR50 and ACR70 were defined similarly, with 50% and 70% thresholds, respectively. Components of DAS28 and ACR responder status end points were collected at predose (baseline) and at 2, 4, 8, 12, 16, 20 and 24 weeks after start of treatment.

An enzyme-linked immunosorbent assay for determining concentrations of rituximab and PF-05280586 in human serum samples was developed and validated by QPS, LLC (Newark, DE, USA; see Supplemental Methods). Actual time of collection for all PK and clinical response measurements was used for the analysis.

Model development and implementation

Longitudinal nonlinear mixed-effects models were fit to the PK, DAS28 and ACR response data. The models were implemented in NONMEM (Icon Development Solutions, Ellicott City, MD, USA), version 7.1.2 or higher, using first-order conditional estimation for continuous end points and Laplace estimation for categorical end points. The models were similar to those described for characterization of the time-course of both continuous [19] and categorical [20–23] measures of clinical response in patients with RA. As previously described, these models provide a semi-mechanistic framework to describe the effect functions for both placebo and drug, each of which can include parameters related to the effect onset

and maximum. This is more intuitive for continuous end points. For categorical end points, an underlying unobservable 'latent variable' (LV) is mapped into binary or ordered categorical response through the probability mass between thresholds. In this way, the threshold parameter is a numerical value on the unobservable continuous scale, which determines positivity on the binary scale. The LV is often assumed to be a function of an indirect response mechanism (e.g. reduction of inflammation) and any acting placebo effects (e.g. a patient's perception of their disease) which may be aggregated into a single clinical response end point (e.g. ACR response rate). Constructing the LV as a function of the indirect response thus yields a pharmacologically interpretable type of model and therefore may be useful for assessing similarity of response between two drug products. In addition, for a more efficient use of ACR response data, simultaneous modelling of ACR20, ACR50 and ACR70 end points were implemented, assuming that ACRm, where m = 20, 50 or 70, was achieved when an m% improvement form baseline was observed [19].

Covariate effects were modelled focusing on components of baseline disease activity appropriate for each clinical response end point, which are centred on their median observed values, and product specific parameters applied to three key model components and were estimated simultaneously. Detailed mathematical descriptions of structures for DAS28 change from baseline (DAS28cfb) and ACR responder models, and procedures for their suitability evaluations and applications for comparative assessments are provided in Supplemental Methods.

Reference range for comparative assessments

The rationale for using the reference range to define 'clinically meaningful effect' relied on the assumption that differences in clinical responses between the two reference products would not be clinically meaningful if PK similarity was established. This assumption is based on the following: two rituximab reference products, while sourced from different regions, share the same prescribing practice for the indications approved in the regions, consistent with the biosimilarity definition based on the principle of no clinically meaningful differences between two biologic products under consideration; and confirmation that equivalent exposures were achieved in the two reference products ensuring that any difference in clinical response between the two are not due to differing exposure.

The 90% confidence interval (CI) of the model-predicted difference between area under effect curve (AUEC) means of rituximab-EU and rituximab-US treatments defined the reference range for a given clinical response end point. The model predictions of AUEC differences between means for PF-05280586 and rituximab-US and between PF-05280586 and rituximab-EU were compared with the reference range. Predictions within the reference range implied that no meaningful difference was observed between PF-05280586 and the reference.

Comparative assessments

Monte Carlo techniques (1000 simulation replicates of 10 000 individuals per arm) were used to derive the 90% CIs used for the comparative assessments. A smoothed parametric bootstrap procedure was used to generate the Monte Carlo replicates in which vectors of parameters were drawn for each simulated replicate from a multivariate normal distribution, with mean vector equal to the population model estimates and uncertainty incorporated via the estimated covariance matrix of these population estimates. Additionally, a covariate vector for each individual within the replicate was sampled from the empirical distribution of the study using a bootstrap technique with replacement of patients during resampling. The bootstrap technique was not stratified by treatment to ensure covariate distributions were balanced on average across the groups over the simulated replicates. From these simulated data, population AUEC means were computed across the individual and the covariate distribution per replicate and per arm to derive the prediction for the replicate. The predictions were not dependent upon the individual covariate effects, because these were integrated from the population mean prediction. Thus, differences between treatments were based on fixed effects of treatment differences and their uncertainties. The 90% CI was calculated using the 5th and 95th percentile from the distribution of the AUEC differences between the test and reference products.

Case–control analysis

A *posthoc* case–control analysis was performed as an orthogonal approach to verifying the impact of the imbalance in baseline patient disease characteristics among groups on the comparative assessments. Multivariate and propensity score matching was implemented using the Matching package in R Software (R Foundation for Statistical Computing, Vienna, Austria) [24]. Variables selected for inclusion in the propensity score model were based on those that showed an imbalance in the original dataset while also considering observed correlations between different variables. Matching was performed without replacement of subjects using a calliper of one-quarter the standard deviation of the propensity score. For a given propensity score model, matching was performed first to balance the two reference groups (rituximab-EU vs. rituximab-US) using 1:1 matching. The two matched reference groups were then compared for how well balance was achieved. If adequate, matching was then performed using the same propensity score model to balance the reference group (rituximab-EU/US) to the test group (PF-05280586) using 2:1 matching. The matched groups were then compared for how well balance was achieved. This process was repeated for various models, which included either a single or multiple variables. Performance was assessed at each step graphically and statistically. Graphical comparison of clinical response measures (DAS28cfb and ACR20 responder rate) between the three balanced treatment groups was performed to show the effects of matching.

Results

Baseline characteristics

A total of 214 patients were included in the PopPK/PD modelling analysis population (Table 1). Demographics were similar across groups; however, some imbalances with respect

Table 1

Patient demographics and baseline disease characteristics for the PK/PD modelling analysis population

	PF-05280586(*n* = 71)	Rituximab-EU(*n* = 72)	Rituximab-US(*n* = 71)
Demographics			
Mean age ± SD, years	54.8 ± 11.7	55.7 ± 10.2	53.8 ± 11.8
Male, *n* (%)	13 (18.3)	16 (22.2)	19 (26.8)
Mean body weight ± SD, kg	86.2 ± 22.0	82.6 ± 19.8	80.4 ± 21.6
Mean body surface area ± SD, m^2	1.9 ± 0.2	1.9 ± 0.2	1.9 ± 0.3
Mean disease duration since first diagnosis ± SD, years	12.7 ± 8.4	11.8 ± 8.3	10.6 ± 8.1
Disease characteristics, mean ± SD			
DAS28-CRP	5.64 ± 0.85	5.80 ± 0.96	6.2 ± 0.89
HAQ-DI	1.67 ± 0.56	1.61 ± 0.53	1.74 ± 0.62
Swollen joint count 28	11.4 ± 5.0	13.0 ± 6.6	14.0 ± 6.0
Tender joint count 28	14.3 ± 6.5	15.1 ± 6.8	18.0 ± 6.5
Swollen joint count 66	15.4 ± 8.8	17.9 ± 10.6	18.9 ± 8.4
Tender joint count 68	22.7 ± 12.6	23.7 ± 13.2	29.7 ± 15.0
CRP, mg l^{-1}	12.4 ± 14.9	14.7 ± 17.6	18.2 ± 25.1
PhGA	64.6 ± 15.3	66.1 ± 15.5	70.1 ± 15.6
Patient's assessment of arthritis pain	65.6 ± 17.8	66.1 ± 21.0	72.1 ± 18.5
PGA	67.4 ± 16.8	67.7 ± 20.9	74.8 ± 16.0

CRP, C-reactive protein; DAS28-CRP, Disease Activity Score in 28 joints based on C-reactive protein; HAQ-DI, Health Assessment Questionnaire Disability Index; PD, pharmacodynamics; PGA, patient's global assessment of arthritis; PhGA, physician's global assessment of arthritis; PK, pharmacokinetics; SD, standard deviation.

to baseline disease characteristics and duration of RA were noted. Specifically, patients in the rituximab-US arm had the highest baseline DAS28 with a CRP component, as well as the highest mean values of the individual components that make up DAS28 and ACR assessments. The most noticeable differences were in CRP; PGA and physician's global assessment of arthritis based on visual analogue scale measurements; SJ28 and SJ66; and TJ28 and TJ68. Duration of RA was also shortest in the rituximab-US group. Taken together, these data indicate that patients randomised to the rituximab-US group entered the study with more severe yet shorter duration of disease than patients in the rituximab-EU or PF-05280586 groups.

Population PK/PD models

The DAS28 dataset included 214 baseline and 1382 postbaseline observations. Model development occurred in stages proceeding from modelling DAS28 in normal scale, DAS28cfb scale, inclusion of exposure effects and lastly, inclusion of covariate effects (see Supplemental Results for additional information). Additional details, including the incorporation of between and within-subject variability, are included in Supplemental Methods.

The final DAS28cfb model incorporated parameters related to onset (kp_i) and maximum magnitude ($PMAX_i$) of treatment effects to characterise the exponential decline in disease activity following a single course of treatment. A maximum drug effect model dependent on product concentration (C_{ij}) included parameters representing maximum effect ($Emax_i$) and the concentration at which half the maximal effect is predicted (exp (θ_{EC50})). Subject covariates and drug exposure were predictive of clinical response (Supplemental Table S1). The final DAS28cfb model predictions tracked the central tendency and distribution of DAS28cfb observations (Figure 1). In addition, inspection of plots produced for model validation indicated that the DAS28cfb model was appropriate for comparative assessment of clinical response (data not shown).

Concentration effects were included in the DAS28cfb model using individual predicted concentrations from the final PopPK model. A total of 2673 observations were included in the PopPK dataset and the percentage of observations below the LLOQ was low (1.5%). Other than those observations excluded that were <LLOQ, two other PK observations were excluded due to a missing sample time or missing data value. The final two-compartment PopPK model included covariates baseline body surface area and sex on clearance and central volume, similar to previously described in patients with RA [25]. The model adequately characterised the time course of drug exposure for all three products as assessed by a visual predictive check (Supplemental Figure S1) and individual prediction plots (not shown).

The ACR response dataset included 1402 observations. The effect of product treatment on probability of response over time is represented by a function consisting of parameters related to maximum effect ($PMAX_i$) and onset of effect

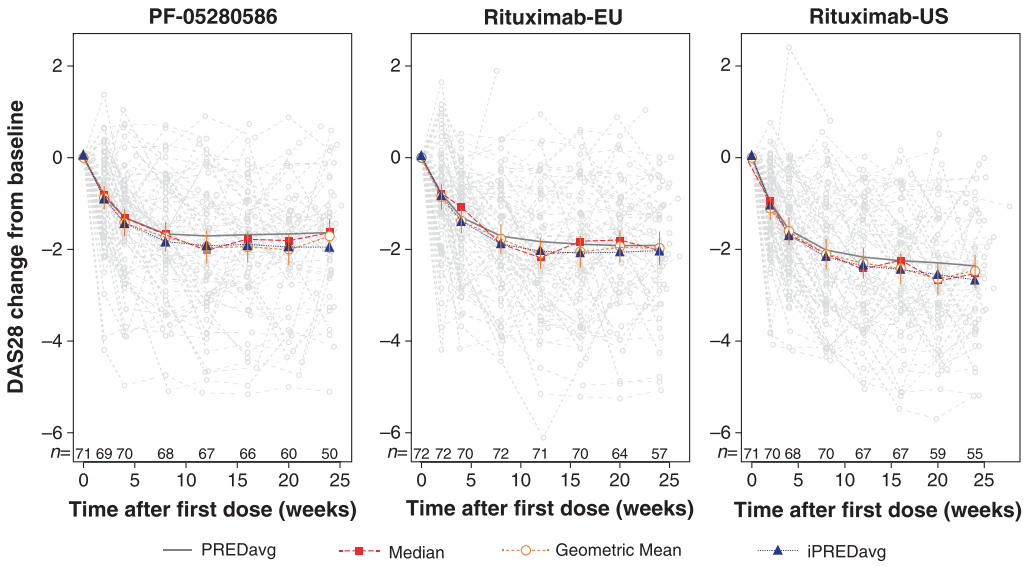

Figure 1

Observed and predicted DAS28 change from baseline *vs.* time, by treatment. The light grey dashed lines and symbols represent individual patient response profiles, and solid lines represent the population predicted mean (PREDavg; black), observed median (red), geometric mean (orange), and individual predicted mean (iPREDavg; blue). The number of patients contributing to each time point is listed along the bottom of the x axis. DAS28, Disease Activity Score in 28 joints

(kp_i), similar to the DAS28cfb model. Unlike the DAS28cfb model, rituximab exposure effects could not be supported in the ACR model.

Covariate effects were predictive of ACR response and lead to some interesting findings. For example, patients with higher numbers of tender joints had a lower probability of being a responder, whereas those with a higher number of swollen joints had a higher probability of response. The population mean predictions tracked the observed data for ACR20, ACR50 and ACR70 in general (Figure 2). In certain cases, the individual predictions captured the trend of the data better than population predicted (e.g. ACR20 for rituximab-US). This may be in part due to the drop-out characteristics since the average individual predicted (Figure 2) was only computed over the subjects who remained in the study, and therefore this method or prediction would have the same dropout characteristics as the observed data. This finding, along with inspection of plots produced for model validation (data not shown), indicated the ACR responder rate model was considered accurate for comparative assessment of clinical response.

Comparative reference range

The comparative reference range was established from the nonlinear mixed-effects PopPK/PD model. This range was defined as the 90% CI of the standardised difference between covariate adjusted mean predictions of an integrated clinical response (AUEC) of the two reference arms in this trial. The comparative reference range was −0.05 to 0.57 for DAS28cfb and −0.16 to 0.023 for ACR20 which were approximately similar and somewhat larger, respectively, than the range of clinical responses observed in historic trials [16, 26–28]. The asymmetry (i.e. reference range not centred on zero) was a result of the use of fixed effects in the models to describe clinical response profiles for the two reference products, rather than pooling the two reference arms. The resulting reference range was considered adequate for comparative assessments based on the objective of the analysis. As an alternative approach, which emphasised the absolute differences between the covariate corrected means of the AUEC, an absolute comparative reference range (or boundary) was also established as the 90th percentile of the absolute difference between the covariate corrected means of the AUEC of the two reference arms. This boundary was 0.50 for DAS28cfb and 0.14, 0.13 and 0.099 for ACR20, ACR50 and ACR70, respectively.

Comparative assessments

Comparative assessment for the standardised DAS28cfb AUEC mean difference and absolute mean difference between PF-05280586 and rituximab-EU (0.076 and 0.083, respectively) and between PF-05280586 and rituximab-US (0.32 and 0.33, respectively) showed that the differences were within the comparative reference range and absolute comparative reference range (Figure 3). Similarly, the ACR20 AUEC differences between PF-05280586 and rituximab- EU (−0.016 and 0.012, respectively) and between PF-05280586 and rituximab-US (−0.084 and 0.077, respectively; Figure 4) were within the reference range and absolute reference range. This was also confirmed for ACR50 and ACR70 using the absolute mean differences and their reference ranges, respectively (Supplemental Figure S2).

Case–control analysis

Prior to matching, several subcomponents of DAS28 and ACR response rate were statistically different between reference groups as well as between the pooled reference group (rituximab-EU/US) and PF-05280586 group (Supplemental Table S2). A propensity score model, which included two

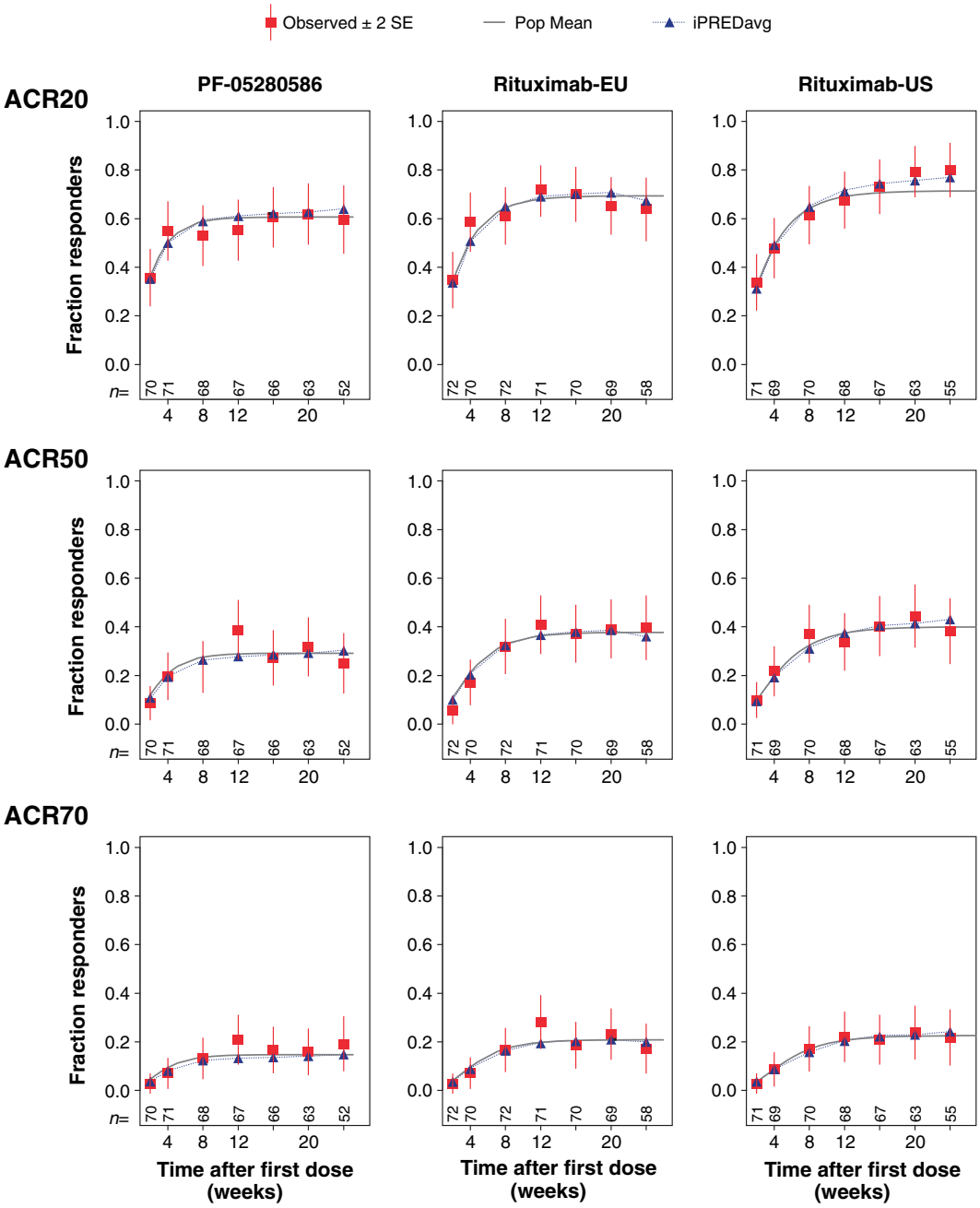

Figure 2

Observed and predicted ACR20, ACR50 and ACR70 responder rates *vs.* time, by treatment. Red symbols and error bars represent the observed proportion and ±2 standard errors (SE), respectively, at each time point achieving the defined ACR response criteria. Solid black lines represent the population-predicted proportion (Pop Mean) and blue lines are mean individual predicted proportion (iPREDavg). The number of patients contributing to each time point is listed along the bottom of the x axis. ACR20/50/70, American College of Rheumatology improvement in rheumatoid arthritis of 20%/50%/70%

subcomponents, SJ66 and patient's assessment of pain, was sufficient to balance all disease characteristics (including DAS28; Supplemental Table S2). The resulting subset was ~70% of the original dataset. The resulting mean DAS28cfb and ACR20 profiles in this subset of patients showed lack of an effect due to imbalance in baseline patient disease characteristics among all treatment groups compared with those in the original dataset (Figure 5).

Discussion

We describe a PopPK/PD approach for comparative assessments of clinical responses between PF-05280586, a proposed biosimilar to rituximab, and two rituximab products (references) in patients with RA. This approach was designed to take advantage of multiple repeated measurements for each clinical end point and variability observed between the two

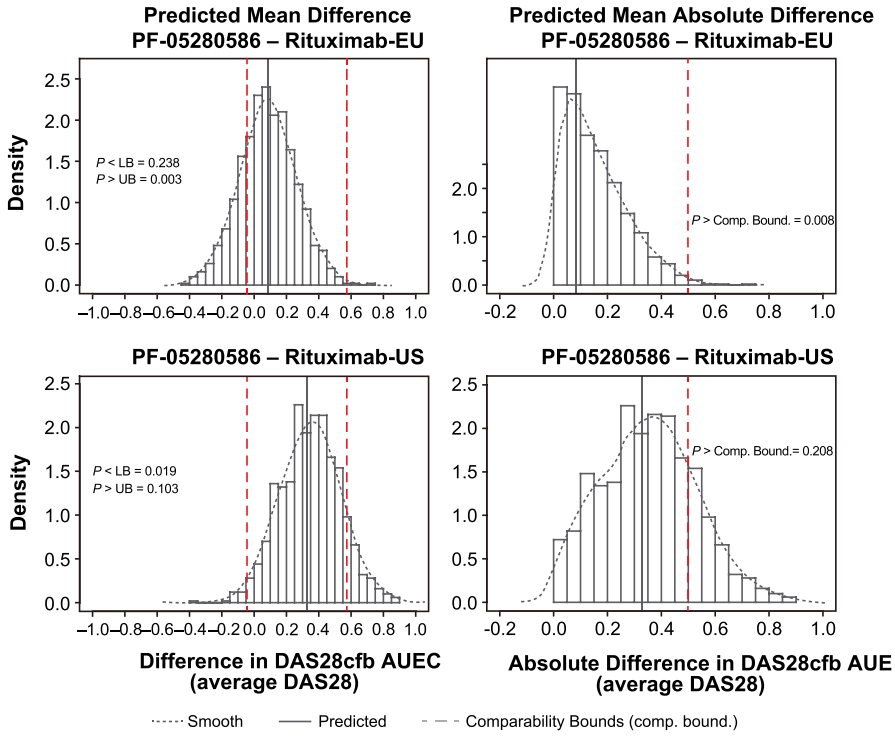

Figure 3

Comparative assessment between PF-05280586 and two rituximab comparators for DAS28cfb normalised AUEC. The reference range established from the two reference treatment groups for the signed difference approach (left panels) is defined by the 90% confidence interval (dashed red vertical lines) of the distribution of time-average normalised mean AUEC differences between rituximab-EU *vs.* rituximab-US. The reference range established from the two reference treatment groups for the absolute difference approach (right panels) is defined by the 90th percentile (dashed red vertical lines) of the distribution of time-average normalised mean AUEC differences between rituximab-EU *vs.* rituximab-US. Predictions within the comparative reference range implied that no meaningful difference was observed between PF-05280586 and the reference. The point estimate (solid black vertical line) and distribution (histogram) of mean differences between the test and a single reference product is shown from this study for DAS28cfb. AUEC, area under the effect curve; DAS28cfb, Disease Activity Score in 28 joints change from baseline

reference products used in the PK study. A procedure for determining the criteria as to whether a meaningful difference between PF-05280586 and reference products was observed in the study was prospectively defined, providing a framework for the comparative assessments. Based on these criteria, any mean difference within the reference range would not be considered meaningful. Establishing the reference range relied on the assumption that both reference products had equivalent potency and that, given PK similarity was established, any difference in clinical responses observed between the two references in a single trial are due to study design attributes (not product-related), such as between-subject variability or imperfect randomization in baseline disease characteristics among treatment groups. Thus, emphasis is placed on the distribution of differences between references that may be expected between two similar products in a PK trial setting with ~70 subjects per arm. In comparison to historical data extracted from several randomised clinical trials, reference ranges established in this trial are similar in size to those observed across trials for DAS28 but somewhat larger than those for ACR20 [16, 26–28].

The traditional approach – using ACR clinical response rate at the end of a specified period as the primary end point in randomised clinical trials to distinguish active treatment from placebo – may not be optimal for the biosimilarity assessment between two active treatments especially when the clinical dose is located on the plateau of the dose response curve [22, 29]. This limitation was a motivating factor for utilizing the repeated clinical response measurements for the model-based comparative assessments, along with assessment of another clinical response endpoint (i.e. DAS28), in an attempt to improve detection of a potential difference and add to the totality of evidence for demonstration of biosimilarity. Integrating all PK/PD (clinical response) measurements over time using the model-based approach allowed examination of onset of effect and magnitude of response along with precise estimation of AUEC, a parameter of interest for clinical comparability assessments consistent with the FDA draft guidance on clinical pharmacology data to support demonstration of biosimilarity [9].

In this study, an underlying imbalance in the rituximab-EU and rituximab-US groups was evident from baseline patient disease characteristics. To adequately characterise response profiles for each group, separate fixed-effects were estimated for each reference group, in addition to inclusion of covariates to account for the imbalance. The aim was to evaluate whether treatment differences could be explained

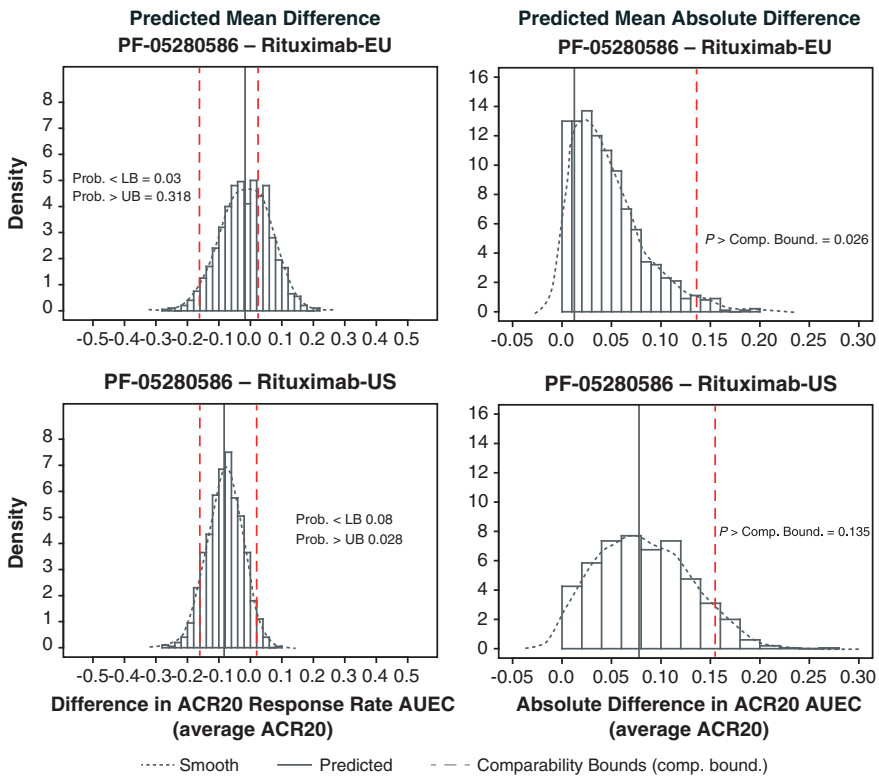

Figure 4

Comparative assessment between PF-05280586 and two rituximab comparators for ACR20 responder rates normalised AUEC. The reference range established from the two reference treatment groups for the signed difference approach (left panels) is defined by the 90% confidence interval (dashed red vertical lines) of the distribution of time-average normalised mean AUEC differences between rituximab-EU *vs.* rituximab-US. The reference range established from the two reference treatment groups for the absolute difference approach (right panels) is defined by the 90th percentile (single vertical dashed red line) of the distribution of time-average normalised mean AUEC differences between rituximab-EU *vs.* rituximab-US. Predictions within the comparative reference range implied that no meaningful difference was observed between PF-05280586 and the reference. The point estimate (solid black vertical line) and distribution (histogram) of mean differences between the test and a single reference product is shown from this study for ACR20. ACR20, American College of Rheumatology improvement in rheumatoid arthritis of 20%; AUEC, area under the effect curve

or predicted by the covariates through the imbalance between groups. Therefore, the simultaneous estimation approach was preferable to forward selection of covariates which may introduce bias in the estimated values [30, 31]. Incorporation of covariates in the models assuming a pooled rituximab-EU/US population (data not shown) did not account for all the imbalance, indicating unmeasured covariates also may have contributed to the underlying imbalance. In addition to the pooling approach, an attempt to incorporate group-specific random effects to provide correlation within the groups was conducted. This confirmed that it was not possible to estimate a precise between-group variability term with the small number of groups to characterise the remaining difference between the two reference products after inclusion of covariates in the model. As may be expected, comparative assessments between two reference products conducted with fixed effects resulted in an asymmetric comparison. That is, the distribution of mean estimated differences between two reference products is shifted, resulting in a nonzero mean. Details of this particular finding, while insightful, may not be easily understood without evaluation of simulation of various hypothetical scenarios. An alternative perspective, which is included in our analysis, was to evaluate each assessment using absolute difference. Since the emphasis of similarity is on the magnitude of difference, the absolute difference is consistent with the objectives of the given analysis and offered an intuitive interpretation of the analysis results. Lastly, imbalances in baseline disease characteristics, which tend to occur in PK trial settings, can be addressed to a great extent in the mixed-effect modelling framework through a typical covariate analysis. In this regard, the models developed in this analysis included the individual components of baseline disease activity and elucidated potential confounding factors. In particular, the ACR model indicated that patients with a higher swollen joint count had a higher probability of being an ACR responder, while those with a higher tender joint count had a lower probability. As others have described, joint tenderness might indicate chronification of the pain reaction rather than ongoing inflammation and could therefore be a confounding factor [32]. That is, a disconnection between the number of swollen and tender joints may appear in RA patients who have

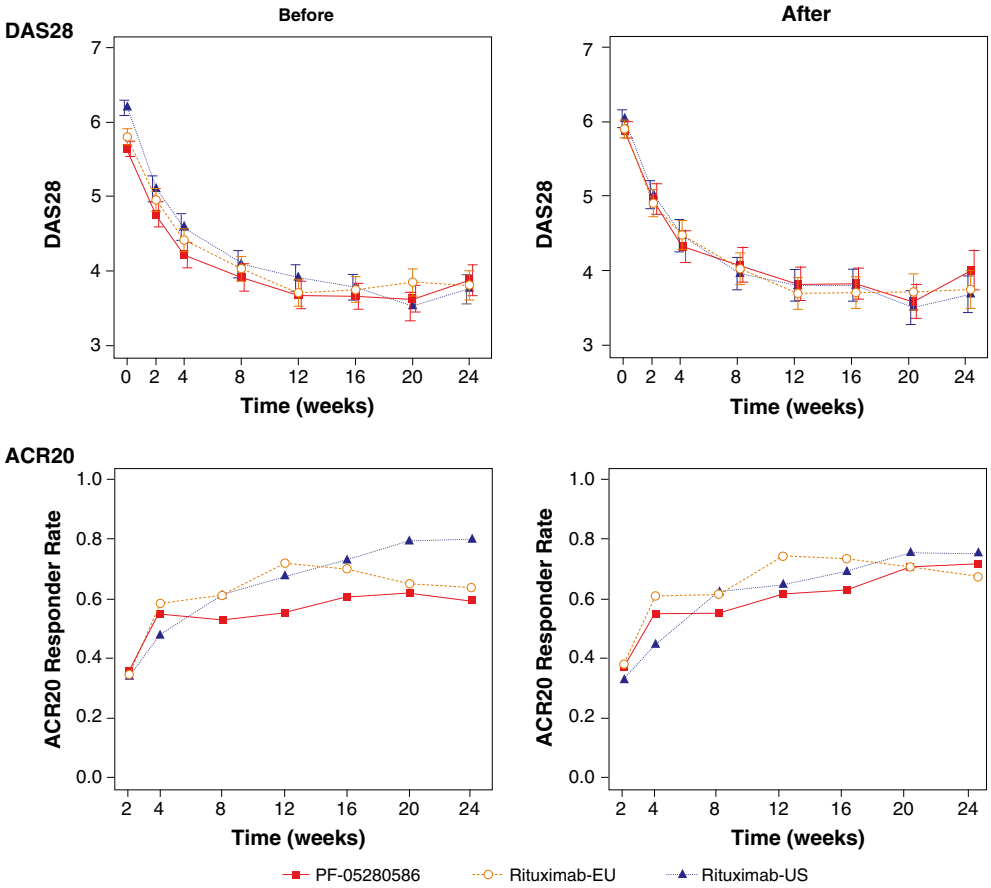

Figure 5

Clinical response profiles for three treatment groups before (left panels) and after (right panels) matching, based on swollen joint counts and patient's assessment of pain. Points represent mean (DAS28cfb, top) or proportion (ACR20, bottom). ACR20, American College of Rheumatology improvement in rheumatoid arthritis of 20%; DAS28cfb, Disease Activity Score in 28 joints change from baseline

developed widespread pain syndromes or other hyperalgesic conditions [32] and consequently could result in somewhat different populations with respect to their potential to respond to treatment.

Still, as found in this analysis, there may be unaccounted imbalances in the data presumably due largely to the randomization with small sample sizes and no stratification based on disease activity. In verifying this, an orthogonal approach using case–control analysis to match patients based on one or more variables may be useful, as described for an exposure–response analysis [33]. This approach is often used to account for imbalance in nonrandomised studies and typically with much larger sample sizes. Applied here, this approach helped identify two important measures of disease activity, swollen joint count and patient's assessment of pain, which, when used for matching, resulted in balanced baseline disease activity among treatment groups. The resulting balanced groups all had similar clinical response profiles.

In conclusion, no point estimates of mean differences were found outside the reference ranges for either DAS28 or ACR20 scores. Furthermore, the estimated probabilities that the predicted differences for the given comparisons, PF-05280586 *vs.* rituximab-EU or PF-05280586 *vs.* rituximab-US, lie outside the boundaries of the respective reference ranges of absolute differences are low. Our analysis provided a case study where a model-based PK/PD approach was utilised for biosimilarity assessment using clinical response data from the PK similarly study comparing a proposed biosimilar against two reference products. The results may therefore be used as a PK/PD demonstration of no clinically meaningful differences in support of overall biosimilarity determination. This case study offers an example for additional applications of clinical pharmacology studies and approaches aimed for efficient biosimilar development that are essential for improving patient access to the important, often lifesaving, biologic therapies.

Competing Interests

All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf (http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare the following financial relationships with organizations that might have an

interest in the submitted work in the previous 3 years: BG, LAM, DY, RL and XM are employees of Pfizer Inc., JHW, MLZ, and GSG were employees of Pfizer Inc at the time of the study, MMH and J-CB were paid consultants of Pfizer Inc.

We acknowledge the investigators and patients who contributed to this study. We thank Jian Du, Dian Xu and Stuart Pearce for preparation of NONMEM datasets. We thank Michael Amantea for helpful advice. This study was funded by Pfizer Inc. Editorial assistance was provided by Christina McManus, PhD, of Engage Scientific Solutions and funded by Pfizer.

Contributors

JHW and XM designed the research. J-CB, BG, LAM, GSG, DY, RL, JHW, MMH, MLZ, K-HL and XM performed the research. MMH, JHW, MLZ, K-HL and XM analysed the data. JHW and XM wrote the manuscript.

References

1 Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SP, *et al*. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucl Acids Res 2016; 44: D1054–D1068.

2 Alexander SP, Kelly E, Marrion N, Peters JA, Benson HE, Faccenda E, *et al*. The Concise Guide to PHARMACOLOGY 2015/16: Overview. Br J Pharmacol 2015; 172: 5729–43.

3 Pope J, Combe B. Unmet needs in the treatment of rheumatoid arthritis. Sci Res 2013; Available at http://dx.doi.org/10.4236/ojra.2013.32011 (last accessed 6 June 2016).

4 McCamish M, Woollett G. The state of the art in the development of biosimilars. Clin Pharmacol Ther 2012; 91: 405–17.

5 US Food and Drug Administration. Title VII: improving access to innovative medical therapies. Biologic price competition and innovation (BPCIA) provisions of the Patient Protection and Affordable Care Act (PPACA). Silver Springs, MD: US Department of Health and Human Services, 2010. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/UCM216146.pdf (last accessed 6 June 2016).

6 European Medicines Agency. Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. London: European Medicines Agency, 2013. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/01/WC500180219.pdf (last accessed 6 June 2016).

7 World Health Organization. Guidelines on evaluation of similar biotherapeutic products (SBPs). Geneva: World Health Organization, 2009. Available at http://www.who.int/biologicals/areas/biological_therapeutics/BIOTHERAPEUTICS_FOR_WEB_22APRIL2010.pdf (last accessed 6 June 2016).

8 US Food and Drug Administration. Scientific considerations in demonstrating biosimilarity to a reference product: guidance for industry. Silver Springs, MD: US Department of Health and Human Services, 2015. Available at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm291128.pdf (last accessed 6 June 2016).

9 US Food and Drug Administration. Clinical pharmacology data to support a demonstration of biosimilarity to a reference product: guidance for industry. Silver Springs, MD: U.S. Department of Health and Human Services, 2014. Available at http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm397017.pdf (last accessed 6 June 2016).

10 F. Hoffmann-La Roche Ltd. MabThera (rituximab) summary of product characteristics. Basel, Switzerland: European Medicines Agency, 2009. Available at http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000165/WC500025821.pdf (last accessed 6 June 2016).

11 Genentech Inc. Rituxan® (rituximab) prescribing information. South San Francisco, CA, 1997. Available at http://www.gene.com/download/pdf/rituxan_prescribing.pdf (last accessed 6 June 2016).

12 Cohen S, Emery P, Greenwald M, Yin D, Becker JC, Melia LA, *et al*. A phase I pharmacokinetics trial comparing PF-05280586 (a potential biosimilar) and rituximab in patients with active rheumatoid arthritis. Br J Clin Pharmacol 2016 82: 129–38

13 Karnik S, Thompson M, DeGruttola H, Ferrari S, Gu L, Aggarwal P, *et al*. Characterization and comparison of PF-05280586 – a proposed rituximab biosimilar to the licensed product. Proceedings of the American Association of Pharmaceutical Scientists-2013 National Biotechnology Conference; 20–22 May 2013; San Diego, CA.

14 Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, *et al*. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010; 69: 1580–8 [correction of 2010;69:1892].

15 Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992; 35: 498–502.

16 Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, *et al*. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. Arthritis Rheum 2006; 54: 2793–806.

17 Sanz I, Anolik JH, Looney RJ. B cell depletion therapy in autoimmune diseases. Front Biosci 2007; 12: 2546–67.

18 Sibilia J, Gottenberg JE, Mariette X. Rituximab: a new therapeutic alternative in rheumatoid arthritis. Joint Bone Spine 2008; 75: 526–32.

19 Levi M, Grange S, Frey N. Exposure-response relationship of tocilizumab, an anti-IL-6 receptor monoclonal antibody, in a large population of patients with rheumatoid arthritis. J Clin Pharmacol 2012; 53: 151–9.

20 Hu C, Xu Z, Rahman MU, Davis HM, Zhou H. A latent variable approach for modeling categorical endpoints among patients with rheumatoid arthritis treated with golimumab plus methotrexate. J Pharmacokinet Pharmacodyn 2010; 37: 309–21.

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

22 Lacroix BD, Karlsson MO, Friberg LE. Simultaneous exposure-response modeling of ACR20, ACR50, and ACR70 improvement scores in rheumatoid arthritis patients treated with certolizumab pegol. CPT Pharmacometrics Syst Pharmacol 2014; 3: e143.

23 Lacroix BD, Lovern MR, Stockis A, Sargentini-Maier ML, Karlsson MO, Friberg LE. 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 2009; 86: 387–95.

24 Sekhon J. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. J Stat Softw 2011; 42: 1–52.

25 Ng CM, Bruno R, Combs D, Davies B. Population pharmacokinetics of rituximab (anti-CD20 monoclonal antibody) in rheumatoid arthritis patients during a phase II clinical trial. J Clin Pharmacol 2005; 45: 792–801.

26 Edwards JC, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, Close DR, *et al*. Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N Engl J Med 2004; 350: 2572–81.

27 Emery P, Deodhar A, Rigby WF, Isaacs JD, Combe B, Racewicz AJ, *et al*. Efficacy and safety of different doses and retreatment of rituximab: a randomised, placebo-controlled trial in patients who are biological naive with active rheumatoid arthritis and an inadequate response to methotrexate (Study Evaluating Rituximab's Efficacy in MTX iNadequate rEsponders (SERENE)). Ann Rheum Dis 2010; 69: 1629–35.

28 Emery P, Fleischmann R, Filipowicz-Sosnowska A, Schechtman J, Szczepanski L, Kavanaugh A, *et al*. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment: results of a phase IIB randomized, double-blind, placebo-controlled, dose-ranging trial. Arthritis Rheum 2006; 54: 1390–400.

29 Mandema JW, Salinger DH, Baumgartner SW, Gibbs MA. A dose–response meta-analysis for quantifying relative efficacy of biologics in rheumatoid arthritis. Clin Pharmacol Ther 2011; 90: 828–35.

30 Gastonguay M. Full covariate models as an alternative to methods relying on statistical significance for inferences about covariate effects: a review of methodology and 42 case studies. Population Approach Group in Europe (PAGE), 2011. Available at http://www.page-meeting.org/?abstract=2229 (last accessed 6 June 2016).

31 Harrell FE. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis, 2nd edn. New York: Springer International Publishing, 2010.

32 Kristensen LE, Bliddal H, Christensen R, Karlsson JA, Gulfe A, Saxne T, *et al*. Is swollen to tender joint count ratio a new and useful clinical marker for biologic drug response in rheumatoid arthritis? Results from a Swedish cohort. Arthritis Care Res (Hoboken) 2014; 66: 173–9.

33 Yang J, Zhao H, Garnett C, Rahman A, Gobburu JV, Pierce W, *et al*. The combination of exposure-response and case–control analyses in regulatory decision making. J Clin Pharmacol 2013; 53: 160–6.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

http://onlinelibrary.wiley.com/doi/10.1111/bcp.13094/suppinfo.

Table S1 Parameter estimates for Disease Activity Score in 28 Joints change from baseline and American College of Rheumatology responder rate models.
Table S2 Patient demographics and baseline disease characteristics for the pharmacokinetic/pharmacodynamic modelling analysis population before and after propensity score matching.
Figure S1 Visual predictive check of the rituximab pharmacokinetic model.
Figure S2 Comparative assessment between PF-05280586 and two rituximab comparators for ACR50 and ACR70 responder rates normalised area under effect curves.