Blinatumomab, a Bispecific T-cell Engager (BiTE®) for CD-19 Targeted Cancer Immunotherapy: Clinical Pharmacology and Its Implications

Min Zhu,1* Benjamin Wu,1 Christian Brandl,2 Jessica Johnson,1 Andreas Wolf,2 Andrew Chow,1 Sameer Doshi1

1Amgen Inc., Thousand Oaks, CA, USA; 2Amgen Research (Munich), Munich, Germany

*Corresponding Author: Min Zhu, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA; Tel: 805-447-2236; Fax: 805-376-1871; Email: min.zhu@amgen.com
Population PK Methods

This population pharmacokinetic analysis was performed using a nonlinear mixed-effects modeling approach by extended least squares regression using NONMEM 7.2 (ICON Development Solutions, Ellicott City, MD). This approach estimated the typical value of model parameters, their variances, and residual noise in blinatumomab serum concentrations. Studies with various designs were combined and evaluated simultaneously in this meta-analysis.

Outlier detection was based primarily on visual examination of individual and pooled (per study) blinatumomab serum concentration vs time profiles and the absolute values of the conditional weighted residuals (> 6).

Based on review of blinatumomab concentration vs time profiles, a one-compartment pharmacokinetic model was tested to characterize blinatumomab pharmacokinetics. The model was then selected as the base model to describe the pharmacokinetics of blinatumomab in the patients. The model was parameterized in terms of systemic clearance from central compartment (CL) and volume of distribution for the central compartment (V).

Between-patient variability (inter-individual variability, IIV) and within-patient variability (inter-occasion variability, IOV) in model parameters were evaluated assuming a log-normal distribution. The magnitude of IIV and IOV in the pharmacokinetic parameters was expressed as a coefficient of variation (CV). IOV was evaluated and included in the model if there was more than a 20% reduction in IIV and residual variability. Residual variability was evaluated using an additive error model after natural logarithmic transformation of the measured concentrations and model predictions. In addition, the possibility of including additional residual variability for studies with expected higher pharmacokinetic variability (multicenter late development studies with sparse samples vs single center early development studies with intensive sampling) was also investigated.
As blinatumomab concentrations below the quantification limit (BQL) were available in the dataset (14.2% of the available concentrations were BQL), the impact of blinatumomab serum values BQL on the pharmacokinetic estimates was evaluated by considering the lower limit of quantitation (LLOQ) to be 100 pg/mL for Studies MT103-104 and MT103-202 and 50 pg/mL for Studies MT103-206 and MT103-211. These data were treated as censored data, and the likelihood for BQL observations was maximized with respect to the model parameters using the M3 method [1].

To identify the best structural and statistical model, a series of models was evaluated that was perceived to potentially describe the observed data. In selection of the preferred model, models that converged successfully, had a successful estimation of the standard errors, produced reasonable parameter estimates, and had low inter- and intra-individual variability were preferred over others. In addition, for each model, the change in the minimum value of the objective function (MVOF) and the improvement in the fit obtained was assessed. For hierarchical models, this change represented a statistic, which was assumed to be asymptotically distributed like \( \chi^2 \), with degrees of freedom (df) equal to the number of parameters added to or deleted from the model. For example, for a single comparison, a change in the MVOF (\( \Delta \text{MVOF} \)) of \( \geq 10.8 \) is required to reach statistical significance (\( \alpha = 0.001 \)) for the addition of 1 fixed effect.

The baseline covariates included in the analysis were demographic factors (age, body weight, BSA, sex), renal function (CrCL), liver function (alanine transaminase [ALT], aspartate transaminase [AST], albumin, total bilirubin), and disease status (performance status, lactate dehydrogenase [LDH], hemoglobin, CD19\(^*\) B-cell counts, CD3\(^*\) T-cell counts, and the ratio of CD19\(^*\) B to CD3\(^*\) T-cell counts [BTCR]). Race was not tested as a covariate because more than 90% of patients were white.
Covariate screening was based on visual graphical inspection and stepwise linear regression of the relationships between the individual post hoc Bayesian model parameters and the covariates. Only covariates that were statistically significant ($P < 0.01$) and had coefficient of determination $r^2 > 0.1$ with model parameters in the screening analysis were considered potentially as clinically relevant and were further tested. The inclusion of 1 covariate in the population model was considered if a decrease in the MVOF of greater than 7.88 ($\chi^2$ test, df: 1, $P < 0.005$) was obtained during the univariate analysis.

After the univariate analysis in NONMEM, the covariates with a statistically significant effect on pharmacokinetic parameters were incorporated into the population model all at one time in order to get the intermediate population pharmacokinetic model. At this stage, the exploratory covariate analysis was repeated and additional significant covariates (if any) were included in the population model. A full model was determined when no additional covariates could improve the model. Then, the relative contribution of each covariate to the goodness-of-fit was reevaluated by deleting it from the full model. If the exclusion of a fixed effect resulted in an increase in MVOF less than 10.83 (df = 1, $P < 0.001$), the covariate was dropped out of the model [2]. With this methodology, only covariates showing significant contributions were conserved in the population pharmacokinetic model.

In addition, the improvement in the model fit after incorporating the covariates was assessed by the reduction in the IIV and residual variability, the reduction of the standard errors, and the examination of diagnostic plots as previously described.

A non-parametric bootstrap analysis was performed [3, 4] as an internal model evaluation technique for the final model. Replication of the original dataset (a bootstrap sample) was obtained by random draws of individual data (with replacement) from the dataset. The population pharmacokinetic model was re-fitted to each new dataset and this process was
repeated 1000 times with different random draws. The stability of the model was evaluated by visual inspection of the distribution of the model parameter estimates obtained from the newly replicated datasets and was compared with that obtained from the fit of the original dataset. The parameter estimates of the model were compared to the mean and 95% confidence interval (CI) of the non-parametric bootstrap replicates that had been successfully minimized. If the parameter estimates were within the 95% CI obtained from the bootstrap analysis, the model was considered acceptable. Bootstrap runs with unsuccessful minimization were evaluated and were excluded from further analysis, if deemed appropriate.

Model evaluation also included prediction corrected visual predictive checks (pcVPC). The pcVPC differs from traditional visual predictive check in that the dependent variable has been subjected to a prediction correction before the statistics are calculated. Prediction correction aims to correct for the differences within a bin coming from different independent variables in the model and hence, clearly diagnose model mis-specifications in both fixed and random effects. The pcVPC is presented with 3 solid red lines which represent the 5th, 50th, and 95th percentiles of the observed blinatumomab plasma concentrations values, while the associated shaded area constitutes the 95% CI for the corresponding model-based predicted percentiles computed for each bin across time and replicates [5].

**Population PK Results**

The dataset consisted of a total of 3015 serum samples from 322 patients who received blinatumomab as a clV infusion over 4 weeks at doses ranging from 0.5 to 90 µg/m²/day or as a fixed dose of 9 or 28 µg/day. Blinatumomab serum concentrations were BQL in 428 samples.

**Model Selection**
An open one-compartment pharmacokinetic model parameterized in terms of CL and V was suitable to describe the time course of blinatumomab concentrations after cIV administrations of various doses in patients with hematological malignancies, including patients with r/r ALL. Due to the sparse sampling scheme implemented in Studies MT103-206 and MT103-211, there was limited information to uniquely identify IIV for both V and CL, thus IIV was estimated for CL only. Residual variability was estimated to be higher than 50%, probably because there is random IIV in the residual error, which might potentially be due to the short half-life of blinatumomab, the logistics associated with the administration schedule, as well as the different sampling schemes across the analyzed studies. The inclusion of IIV on the residual variability improved the model significantly (ΔMVOF: −469.367). After inclusion of IIV on the residual, it was evident that residual variability was still higher in Study MT103-211 compared to other studies. To account for this difference, the residual variability was separated for Study MT103-211 versus the other studies (ΔMVOF: −38.312; df: 1; P < 0.0001).

Additionally, after partitioning residual variability, the IIV in CL was still high with a 60% CV. To explore why there was such a large variability in CL, a mixture model was tested to identify any subpopulations with different CL. Implementation of the mixture model identified a subpopulation (10% of patients) that had faster CL, and the IIV in CL was lowered to 46% (ΔMVOF: −25.572). Thus, a model parameterized by V, a bimodal distribution of CL (mixture model), IIV in CL and residual variability, and separate estimates of residual variability for Study MT103-211 versus other studies was identified as the reference model for covariate exploration.

Graphical analysis between the random effects of CL and the covariates evaluated suggested a potential effect of CrCL and age on CL. Univariate inclusion of the CrCL effect on CL was statistically significant (ΔMVOF: −40.888, df: 1; P < 0.0001) and the IIV in CL was reduced by 8.9%. Notably, after the inclusion of CrCL effect on CL, the age effect on CL was no longer
relevant, probably because of the correlation between age and CrCL. In addition, using the MDRD equation instead of Cockcroft–Gault equation to quantify renal function did not substantially improve the fit. Consequently, the model did not include additional covariate effects other than CrCL on CL. Attempts to include IOV were also performed, however after inclusion of IOV, IIV on CL was not reduced, and residual variability only diminished by 11.7%. Given that the number of data points per occasion was limited (73.6% of the occasions had two or less than two data points), IOV was not included in the population pharmacokinetic model.

The effect of blinatumomab serum concentrations BQL was explored by analyzing the data using the M3 method, which resulted in changes in fixed effects of less than 4.5%, except the CL estimate for subpopulation 2, which was estimated to be 46% higher, and a slight increase in IIV of CL (3%), the residual variability (10%) and its associated IIV variability (7%). Furthermore, the goodness-of-fit plots did not show any significant improvement when the M3 method was used. Given the limited pharmacokinetic information added by the BQL concentrations, the number of BQL data, and the short half-life of blinatumomab, it was deemed adequate to exclude the BQL observations from the final analysis dataset.

Thus, the model that included a mixture model of CL, adjusted by CrCL, IIV in CL and residual variability, and had separate estimates of residual variability for Study MT103-211 was considered the final model.

**Model Evaluation**

Figure 1 shows the goodness-of-fit plots for the final model. The observed versus predicted plots (upper panels) show random normal scatter around the identity line indicating the absence of systematic bias and the adequacy of the model to describe the data. In addition, conditional weighted residuals (middle panels) show random normal scatter around zero with no specific pattern suggesting model mis-specification. Notably, the distribution of conditional weighted
residuals versus time remains fairly constant, which indicates the absence of time-dependent pharmacokinetics. Also, there were no absolute values of the conditional weighted residuals higher than 6, and therefore, none of the data points in the analysis dataset were identified as outliers. The NPDE mean was 0.007 (95% CI, –0.031 to 0.045) and SD was 0.970 (95% CI, 0.943 to 0.998), which suggested the absence of bias in the reference model and show a trend to overpredict the variability given that the 95%CI of the NPDE SD is slightly lower than 1. The estimated parameters for the final model and the corresponding 95%CI from the non-parametric bootstrap are presented in Table 1 in the body of the manuscript. Both fixed and random effects parameters were estimated with good precision (RSE < 18%).

Additionally, the results of the pcVPC performed are depicted in Figure 2 and show that the model developed is appropriate to describe the time course of blinatumomab and its variability in patients with hematological malignancies.

References

Supplementary Figure 1. Goodness of fit plots for the final population PK model
Supplementary Figure 2. Prediction corrected visual predictive check for the final population PK model.