

An interactive, web-based application for population pharmacokinetic model diagnostics: PopPK FitCheckR

Ruilian Zhang

`ruilian.zhang@merck.com`

Merck & Co., Inc., Rahway, NJ, USA (United States)

Bela Patel

Merck & Co., Inc., Rahway, NJ, USA (United States)

Brittany Walker

Merck & Co., Inc., Rahway, NJ, USA (United States)

Research Article

Keywords: population pharmacokinetics, NONMEM, model evaluation, R Shiny, goodness-of-fit

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11134016/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

**An interactive, web-based application for population pharmacokinetic model
diagnostics: PopPK FitCheckR**

PopPK FitCheckR for population pharmacokinetic model diagnostics

Ruilian Zhang¹, Bela Patel¹, Brittany Walker¹

1. Merck & Co., Inc., Rahway, NJ, USA

Corresponding author:

Ruilian Zhang

Merck & Co., Inc., Rahway, NJ, USA

Tel +1 9177056254

E-mail: ruilian.zhang@merck.com

Abstract

PopPK FitCheckR is a web-based application that generates interactive, customizable diagnostic plots and facilitates the model diagnostic process of population pharmacokinetic (popPK) modeling. With the ability to read a variety of nonlinear mixed effects modeling (NONMEM) output files, the application automates four types of goodness-of-fit (GOF) plots in both static and interactive formats, including observed values versus population predictions, observed values versus individual predictions, conditional weighted residuals (CWRES) versus predictions, and conditional weighted residuals versus time. It also generates flexible correlation plots, including ETA plots that display the individual random effects estimated for each subject, with different plot types and regression methods available. The application supports interactive filtering, grouping, color stratification, log scale transformation, user-defined axes, and downloading results. Additionally, an interactive variable name mapping interface with optional variable type selection allows flexibility in input files, enabling the use of diverse naming conventions and data structures without modification of model code. By combining flexible input handling with interactive filtering, stratification, and configurable diagnostics, PopPK FitCheckR facilitates popPK model evaluation and supports more efficient and consistent diagnostic review.

Keywords

population pharmacokinetics, NONMEM, model evaluation, R Shiny, goodness-of-fit

Introduction

Population pharmacokinetic (popPK) modeling is an essential component of modern drug development, enabling characterization of drug exposure, interindividual variability, and the influence of covariates within target populations. These models are routinely applied to inform dose selection, optimize clinical study design, support special population assessments, and contribute to regulatory submissions. Given their central role in decision-making, robust model evaluation is critical to ensure that the final model adequately describes the observed data and provides a reliable basis for simulation and inference. The diagnostics considered standard for continuous-data pharmacometric models, including observed versus population and individual predictions, conditional weighted residuals versus predictions and time, and graphical assessment of random effects, have been formalized through community consensus [1]. In practice, however,

diagnostic assessment of popPK models can be time-consuming and technically demanding. While standard GOF plots can be generated with existing tools, customizing them—such as filtering, stratifying, or otherwise subsetting the data—often requires additional coding effort. In addition, diagnostic workflows may vary across users and projects, resulting in inconsistent plot formatting and reduced efficiency.

Established tools for diagnostic evaluation of popPK models in the NONMEM ecosystem are centered primarily around PsN and Xpose, with newer options such as ggPMX, and are often supplemented by custom R-based workflows. NONMEM remains the standard platform for nonlinear mixed effects (NLME) modeling in pharmacometrics and is widely used for development of popPK models [2]. While NONMEM includes a built-in \$SCATTERPLOT option for generating basic diagnostic plots, graphical diagnostics in pharmacometric model development are commonly constructed using external tools that extend NONMEM's functionality. Building on NONMEM outputs, PsN (Perl-speaks-NONMEM) has become a well-established toolkit for model automation and simulation-based diagnostics, supporting tasks such as visual predictive checks (VPCs), bootstrapping, and covariate model building [3]. Xpose is similarly well recognized as a diagnostic and visualization framework, enabling generation of GOF plots, residual diagnostics, and other graphical summaries to facilitate model evaluation [4]. More recently, ggPMX has been developed as an open-source R package that standardizes diagnostic plotting for NLME models, providing a predefined set of publication-quality diagnostics aligned with community recommendations [5].

PsN is highly effective for automation and simulation workflows, but it does not itself render graphics. As a command-line toolkit, it produces the numerical output underlying diagnostics, and can auto-generate R scripts that rely on R and Xpose to draw the corresponding figures. Producing clear, standardized diagnostic plots therefore depends on complementary tools and familiarity with R. Xpose provides substantial flexibility for visualization, but its use typically depends on familiarity with R and may require additional coding and customization to generate consistent, review-ready outputs across models and projects, and exploring a different data subset or stratification typically requires editing and re-running the underlying script. Similarly, ggPMX provides a standardized and reproducible plotting workflow, but it operates as an R package requiring users to write code to configure and generate diagnostics, and its outputs are static. More

recently, Certara.ModelResults offers a Shiny-based graphical interface for generating GOF plots and diagnostic tables from NONMEM and NLME output [6]. However, its use requires users to first construct an Xpose database object within an R session before launching the interface, and it operates within the broader Certara commercial software ecosystem, which may limit accessibility in organizations that do not use Certara products. It is also worth noting that Monolix (Lixoft SAS, a Simulations Plus company) provides an alternative NLME modeling platform with diagnostic visualization tools integrated directly into its graphical interface, reducing reliance on external tooling within its own ecosystem [7]. However, Monolix operates independently from NONMEM and produces platform-specific outputs that are not interchangeable with NONMEM table files; its built-in diagnostics are therefore not accessible to users working within the NONMEM ecosystem. In addition, Monolix is proprietary software and generally requires a paid license. Consequently, many diagnostic workflows remain fragmented, involving multiple software components, manual intervention, and varying levels of programming expertise.

PopPK FitCheckR was developed to address these practical gaps by offering a more direct and automated approach to routine model diagnostics. In contrast to more general-purpose tools, PopPK FitCheckR emphasizes standardized and accessible generation of GOF plots and correlation plots, reducing the need for extensive scripting or post-processing. Its main advantage is therefore in improving efficiency, interactivity, and consistency of model diagnostic review, particularly for users or teams seeking a streamlined workflow for NONMEM-based popPK model assessment. PopPK FitCheckR has been in routine use by pharmacometricians and clinical pharmacologists within our organization for approximately three years, supporting model evaluation across multiple development programs, and its current feature set reflects iterative refinement in response to user feedback gathered over that period.

R Shiny is an R-based framework for building interactive web applications that enables analytical workflows and visualizations to be delivered through a browser-accessible graphical user interface (GUI) without requiring extensive web development [8]. It is particularly well suited for scientific applications because it allows R-based analyses to be translated into user-friendly interfaces that support a smooth end-to-end workflow from data input to analysis and results visualization. Importantly, its interactive nature lets users adjust and refine visualizations on the fly, offering greater flexibility than a static, standardized plotting script. Built using R Shiny, PopPK FitCheckR

leverages these capabilities to automate routine diagnostic plotting while preserving flexibility, interactivity, and ease of use.

The following features make PopPK FitCheckR a useful platform that benefits the day-to-day workflow of pharmacometricians and modelers: interactive data filtering that updates both static and interactive GOF plots in real time; customizable interactive GOF plots with stratification, color grouping, and flexible axis control; flexible data column mapping; user-defined correlation plot collections; and simple exporting options with informative file names.

This article starts with the introduction of PopPK FitCheckR workflow, followed by the description of each tab in the application, and the use cases of leveraging this application to facilitate model diagnostics.

Methods

Application workflow

The PopPK FitCheckR application provides a structured workflow for model diagnostics plots, shown in **Fig. 1**. Upon launching the app, users are directed to an Instructions tab detailing data requirements and available features (**Fig. 2a**). For GOF analysis, users first upload a NONMEM output table file (sdtab). Once the data are loaded, an optional filtering step becomes available to subset desired data slices for visualization. This interactive filtering can be applied at any time after the data are uploaded, allowing users to see the plots update in real time as the data are sliced. Column name standardization is then performed to map actual column names to the required identifiers (PRED, IPRED, DV, CWRES, TIME, ID) (**Fig. 2b**), with an optional variable type assignment step, in which each column is designated as continuous or categorical. Following standardization, two plot outputs are accessible: a static GOF plot panel displaying observed vs. predicted and CWRES diagnostic plots, and an interactive GOF plot supporting stratification, color grouping, log transformation, axis customization, and individual data point identification (**Fig. 2c**). Both types of plots can be downloaded as .png files. For correlation analysis, users upload a separate NONMEM table file (e.g., patab or cotab) and specify the X and Y axes of interest, after which scatter or box plots with optional regression overlays are rendered and available for download.

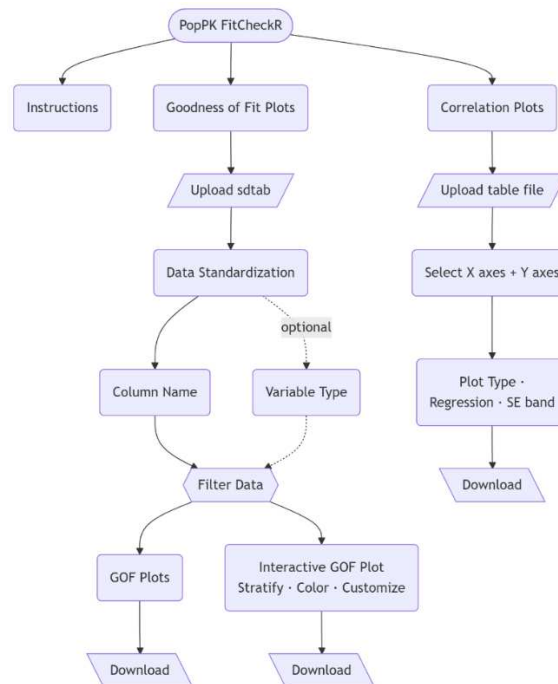


Fig. 1 PopPK FitCheckR application workflow. The diagram illustrates the end-to-end diagnostic workflow, beginning with data upload, proceeding through column name standardization and optional variable type assignment, and ending with generation of static GOF plots, interactive GOF plots, and correlation plots, all of which are downloadable as .png files



Fig. 2 Representative screenshots of the PopPK FitCheckR application interface. **(a)** The Instructions tab provides an overview of the application workflow and describes data input requirements, available analysis features, and key behavioral defaults including file format handling, automatic column-name matching, and variable type auto-detection. **(b)** The Column Name sub-tab within Data Standardization, illustrating automatic pre-population of column mappings for the six required standardized variable names (PRED, IPRED, DV, CWRES, TIME, ID) using Jaro-Winkler string similarity matching against the uploaded NONMEM output file. **(c)** The Interactive GOF Plot sub-tab demonstrating simultaneous use of categorical filters (MDV = 0; DVID = 1) and a continuous time-window filter (TIME: 0–168 hours), with faceting by DVID and color stratification by dose group, producing a subset Observed vs. Predicted plot

Next, the function of each tab and sub-tab will be individually introduced.

Instructions tab

The Instructions tab serves as the landing page of the app, orienting users before any data is uploaded. It outlines the overall workflow, describes the purpose and usage of each sub-tab, and documents data input requirements and key behavioral defaults such as file format handling, automatic column mapping, and variable type detection. Users are encouraged to review this tab before proceeding to the analysis tabs.

Goodness of Fit (GOF) Plots tab

The Goodness of Fit (GOF) Plots tab is the primary analysis tab for evaluating NONMEM model fit diagnostics. Users begin by uploading a NONMEM output table file (e.g., sdtab) via the Browse button in the left panel. The tab contains four sub-tabs: Data Standardization — Column Name, Data Standardization — Variable Type, GOF Plots, Interactive GOF Plot.

Data Standardization — Column Name sub-tab

This tab maps columns from the uploaded file to the six standardized variable names required for plotting: DV (observed), PRED (population predicted), IPRED (individual predicted), CWRES (conditional weighted residuals), TIME, and ID (subject identifier). Column mappings are automatically pre-populated using the Jaro-Winkler string similarity algorithm, a character-matching metric that quantifies similarity between two strings based on the proportion of matching

characters and their positional proximity, while giving additional weight to shared prefixes [9]. This makes it particularly well suited for short identifier names that share common prefixes or differ only by minor typographic variations.

Matching is performed against a Jaro-Winkler similarity threshold of 0.8, a cutoff commonly applied in record linkage and fuzzy string matching of short identifiers to match candidate pairs where the names are nearly identical in spelling and character order [10]. At this threshold, minor variations that frequently occur in NONMEM output column naming (such as TIME versus TIME72) are correctly recognized as matches, while semantically distinct columns that share only a few characters (such as DV and MDV) are appropriately kept separate. When more than one column exceeds the similarity threshold for a given identifier, an exact name match takes precedence; otherwise the column with the highest similarity is selected as the default. Columns below the threshold are left unmatched and need to be assigned manually by selecting from a drop-down menu. All mappings in the Column Name tab should be reviewed before proceeding to the GOF Plots and the Interactive GOF Plot sub-tab.

Data Standardization — Variable Type sub-tab

This is an optional sub-tab that allows users to override the automatically assigned variable type (continuous or categorical) for any column in the uploaded dataset. This tab is populated by all columns in the dataset upon upload.

The default assignment considers both column data type and how many distinct values the column contains relative to the dataset size. A column is classified as continuous only if it is numeric and its number of unique values exceeds 10% of the total number of rows, with a minimum of 10 unique values required. Columns that do not meet both criteria are classified as categorical by default. For example, a numerically coded dose group column containing five distinct values across 500 rows would be classified as categorical, because its five unique values (1% of rows) fall well below the threshold. This behavior is intentional: such a variable represents a discrete group rather than a truly continuous measurement, and treating it as categorical ensures that the appropriate multi-value filter selector is displayed rather than a numeric range input.

The variable type classification affects how filter controls are rendered in the left panel's Filter Data section. Continuous variables produce numeric range inputs, while categorical variables

produce a multi-value selector. Users can override any auto-assignment in this sub-tab if the default classification does not match the intended use of a variable.

Interactive data filtering

An interactive Filter Data function is provided in the left panel, available at any point after uploading data. Users can select one or more columns to subset the data before or during plotting. Filter controls adapt to the variable type: continuous columns show start and end value inputs and are pre-populated with the min/max value observed in the data, and categorical columns show a multi-select dropdown pre-populated with all unique values. Filters are applied reactively and affect all plots in both GOF Plots and Interactive GOF Plot tabs.

GOF Plots sub-tab

The purpose of this tab is to get an initial quick read on the data. This tab displays four static GOF plots arranged in a 2×2 grid: Observed vs. Predicted, Observed vs. Individual Predicted, CWRES vs. Predicted, and CWRES vs. Time. Each plot includes a locally estimated scatterplot smoothing (LOESS) [11] smoothing line in red and a reference line in black (identity line for observed vs. predicted plots; horizontal reference line at $y = 0$ for CWRES plots). Aspect ratio is fixed at 1 for observed vs. predicted plots, while for CWRES plots the axes ratio can be automatically adjusted to the browser size. The full four-panel figure can be downloaded as a .png file. The date and time are automatically attached to the downloaded file name, which allows for clear arrangements and traceability for multiple downloads.

Interactive GOF Plot sub-tab

This tab provides a single interactive plot view implemented using the Plotly package [12], enabling dynamic exploration of diagnostic results. The plot type can be selected using the dropdown menu provided. Faceting by any user-selected column(s) is supported, where the plot is split into separate panels grouped by the values of the selected variable(s), with panel labels displaying both the variable name and value. Multiple columns may be selected simultaneously, producing a multi-way stratification. When a continuous variable is selected for faceting, it is automatically binned into up to five equal-width intervals to avoid an unmanageable number of panels; categorical variables are faceted directly by their observed levels. Whether a variable is

treated as continuous or categorical is determined by the user's selection in the Variable Type sub-tab if one has been made; otherwise, the default unique values rule mentioned above applies.

Color stratification is available as an optional overlay, allowing data points to be colored by any column in the dataset. The same variable type logic governs the color scale: variables classified as categorical are treated as discrete factors and mapped to a categorical color palette, while variables classified as continuous are mapped to a continuous color gradient. When no color variable is selected, all points are rendered in a uniform blue.

The plot is rendered as an interactive Plotly visualization. Hovering over any data point displays a tooltip showing the subject ID, time, and the X and Y axis values for that observation. Axis scale can be toggled between linear and logarithmic: both axes jointly for the observed-versus-predicted plots, and the x-axis only for the CWRES plots, whose signed residuals retain a linear y-axis. When log scale is selected, observations with zero, negative, or missing values on the affected axes are automatically excluded. The number and percentage excluded are reported both in a persistent on-screen warning and beneath the plot title, so that the exclusion remains visible in downloaded figures. When stratification is applied, panels may be given independent axis ranges: both axes for the observed-versus-predicted plots, and the x-axis only for the CWRES plots, so that residual magnitudes remain comparable across panels. Custom numeric start and end values can be specified for both the X and Y axes to override the default data-driven range; these inputs are automatically hidden when free scaling is enabled. The X-axis title, Y-axis title, and plot title can each be overridden with user-specified text; if left blank, they default to the variable name or GOF plot type label. The finalized plot can be downloaded as a .png file.

Correlation Plots tab

The Correlation Plots tab enables exploratory analysis of relationships between any two variables in the uploaded dataset, for example, relationships between model parameters and covariates. Users upload a separate NONMEM output table file containing the variables of interest (e.g., patab for empirical Bayes estimates of random effects, cotab for covariates) and select one or more variables for the X axes and Y axes using multi-select dropdown inputs. All pairwise combinations of the selected X and Y variables are generated and plotted simultaneously in a grid layout with two columns. For datasets containing multiple records per subject (e.g., standard NONMEM output tables that include one row per observation rather than one row per individual),

a “Deduplicate by ID” option is available to retain only the first record per subject prior to plotting; this is recommended when the variables of interest are time-invariant subject-level parameters such as empirical Bayes estimates or baseline covariates.

Two plot types are available. The default line plot renders a scatter plot with a regression line overlay, where the regression method can be chosen between ordinary linear regression (lm) and locally estimated scatterplot smoothing (LOESS) [11]. When the linear model method is selected, the fitted equation and coefficient of determination (R^2) are annotated directly on each panel using the ggpmisc package [13]. A 95% confidence band around the regression line can be optionally toggled on or off. The alternative box plot selection displays the Y-axis variable as a continuous distribution grouped by the factor levels of the X-axis variable, with individual observations overlaid as jittered points. This is particularly useful for examining the relationship between categorical covariates (e.g., sex, race) and continuous model parameters (e.g., ETAs, clearance).

The complete set of pairwise plots can be downloaded as a single .png file. As with the GOF tab, the download filename is automatically appended with a date-time stamp to support traceability across multiple analysis iterations.

Results

Case study 1: GOF diagnostics for an investigational monoclonal antibody PK model

To demonstrate the utility of PopPK FitCheckR, the app was applied to NONMEM output from a population pharmacokinetic (PK) model for drug A, an investigational monoclonal antibody developed as an anticoagulant therapy. The analysis dataset comprised 97 subjects enrolled across two study parts in a Phase I single ascending dose and multiple ascending dose clinical trial, with dose levels of 8, 20, 25, 40, 50, 75, and 120 mg. The PK model consisted of a two-compartment pharmacokinetic model with first-order absorption (parameterized as clearance [CL], central volume [V2], peripheral volume [V3], inter-compartmental clearance [Q], absorption rate constant [KA], and relative bioavailability [F1]). In total, 1,257 pharmacokinetic observations were available for model evaluation.

GOF diagnostics

The NONMEM standard diagnostic table (**Supplementary File 1**) was uploaded to the GOF Plots tab. **Fig. 3** shows the four-panel GOF output generated by the app for the PK component of the model. The Observed vs. Predicted plot shows that population-level predictions generally track the observed concentrations, while the Observed vs. Individual Predicted plot shows tighter agreement following individual post-hoc estimation. The CWRES vs. Predicted and CWRES vs. Time plots show residuals distributed symmetrically about zero with no obvious systematic trends.

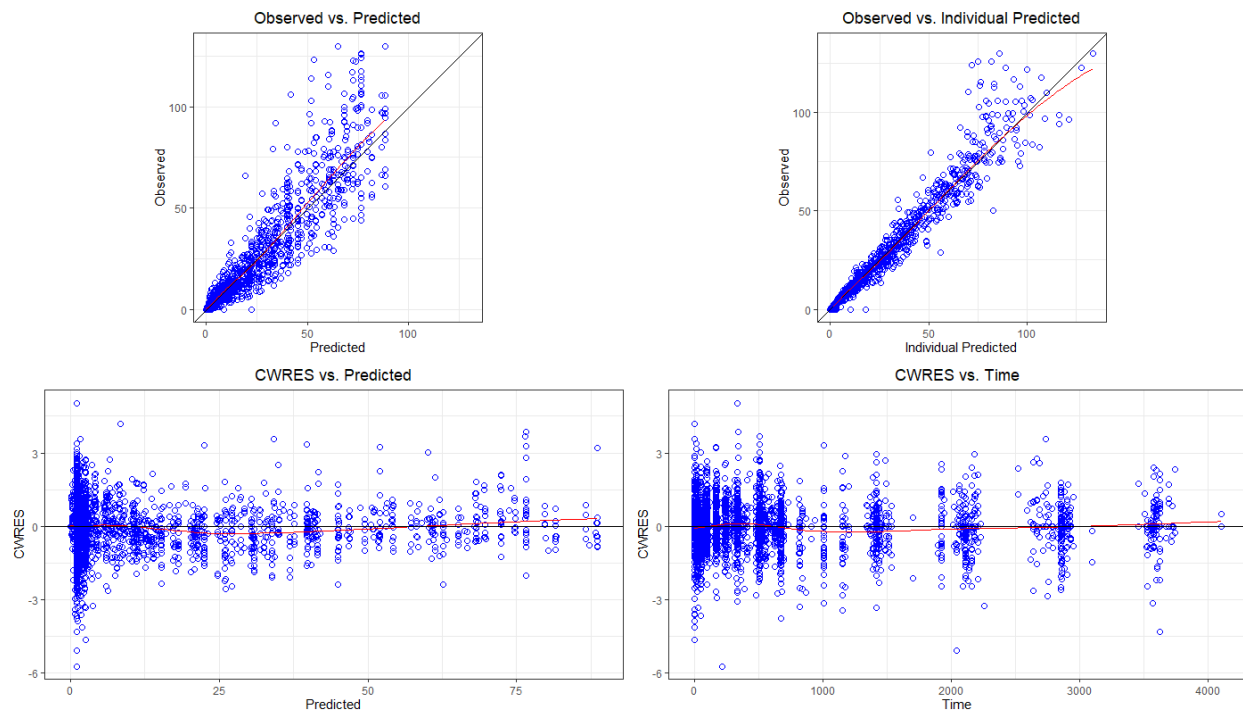


Fig. 3 Four-panel goodness-of-fit (GOF) diagnostic plots generated by PopPK FitCheckR for the PK component of drug A popPK model. Top left: Observed vs. Predicted (population predictions); top right: Observed vs. Individual Predicted (individual post-hoc predictions); bottom left: Conditional Weighted Residuals (CWRES) vs. Predicted; bottom right: CWRES vs. Time. Each panel includes a LOESS smoothing line (red) and a reference line (black; identity line for observed vs. predicted plots, $y = 0$ for CWRES plots)

Stratified exploration

Using the Interactive GOF Plot tab, the Observed vs. Predicted plot was stratified by dose group via the Stratification dropdown. **Fig. 4** displays the resulting faceted plot with seven dose-group panels (placebo subjects excluded from PK observations). The faceting shows comparable

agreement across dose levels, with the LOESS smoothing line closely tracking the line of identity within each panel. This visual inspection suggests that there are no apparent dose-dependent trends in the predicted–observed relationship.

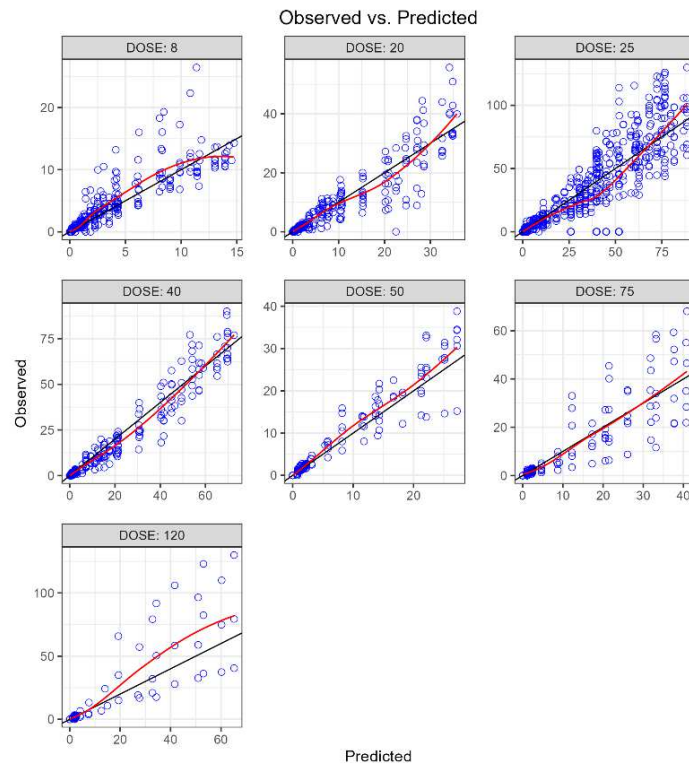


Fig. 4 Faceted Observed vs. Predicted plot stratified by dose group (placebo excluded), generated using the Interactive GOF Plot tab. Seven panels correspond to the 8, 20, 25, 40, 50, 75, and 120 mg dose levels. Each panel displays the scatter of observed concentrations against predicted values with a LOESS smoothing line (red) and the line of identity (black)

ETA–ETA correlation analysis

To assess independence between individual random effects for distribution and absorption parameters, pairwise ETA–ETA correlation plots were generated for the healthy volunteer (HV) subpopulation using a NONMEM parameter output table file restricted to subjects without end-stage renal disease (ESRD) (**Supplementary File 2**), retaining one record per subject via the "Deduplicate by ID" option. Four PK ETAs estimated in this subpopulation were examined: ETA1 (CL) and ETA3 (Q), representing the disposition parameters, and ETA5 (KA) and ETA6 (F1),

representing the absorption parameters. ESRD-specific ETAs (ETA7–ETA10) were excluded as they are only estimable in the ESRD subpopulation and would appear as zero for all HV subjects, precluding meaningful correlation assessment. As shown in **Fig. 5**, all four pairwise plots displayed near-flat LOESS regression lines with wide 95% confidence bands encompassing zero across the full range of ETA values. These plots did not reveal systematic correlation between disposition and absorption random effects, suggesting no strong dependence between the inter-individual variability in CL and Q and that in KA and F1 in the healthy volunteer population.

This case study illustrates how PopPK FitCheckR enables rapid, standardized model evaluation across multiple diagnostic dimensions, from standard GOF assessment to dose-stratified diagnostics and covariate screening.

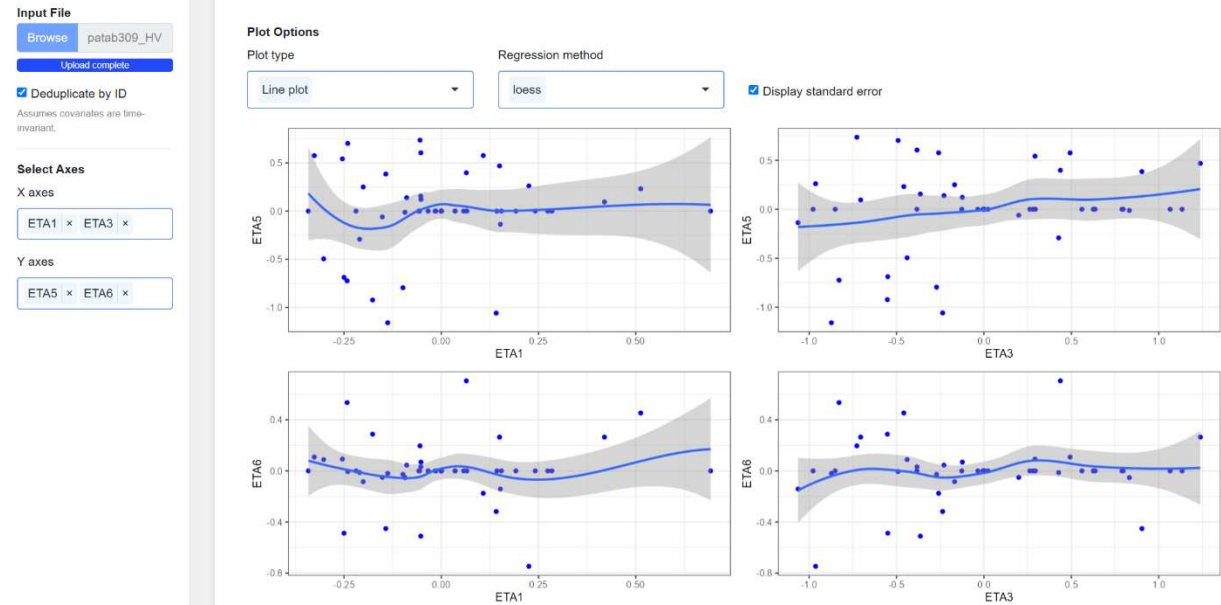


Fig. 5 Pairwise correlation plots of empirical Bayes estimates of inter-individual variability for disposition parameters (ETA1 and ETA3, corresponding to CL and Q) versus absorption parameters (ETA5 and ETA6, corresponding to KA and F1) for the healthy volunteer subpopulation of the drug A population pharmacokinetic model. One record per subject was retained prior to plotting. Scatter plots are shown with LOESS regression lines and 95% confidence bands

Case study 2: covariate–parameter correlation analysis for a general-purpose remifentanyl PK model

To further demonstrate the app's covariate–parameter correlation capabilities, PopPK FitCheckR was applied to NONMEM output from the Eleveld *et al.* general-purpose population pharmacokinetic model for remifentanyl [14]. The model was developed by pooling data from three published studies [15–17], comprising 2,634 remifentanyl arterial concentration observations from 131 individuals. The study population spanned a wide demographic range, including neonates (5 days old) to elderly adults (85 years), with body weights from 2.5 to 106 kg, reflecting the model's general-purpose design. The final model is a three-compartment pharmacokinetic model (parameterized as central volume [V1], peripheral volumes [V2, V3], elimination clearance [CL], and inter-compartmental clearances [Q2, Q3]) incorporating allometric scaling based on fat-free mass, age-dependent maturation of clearance, aging effects on volumes and clearances, and sex-based corrections. Inter-individual variability was estimated on all six structural parameters (V1, V2, V3, CL, Q2, Q3).

Covariate–parameter correlations

The individual-level parameter and covariate table file (**Supplementary File 3**) was uploaded to the Correlation Plots tab. This file contained one row per subject with demographic covariates (AGE, WGT, HGT, BMI, M1F2 [sex]), model-derived covariates (BSIZ [allometric body-size scalar]), study identifier (STDY), empirical Bayes estimates of inter-individual variability (ETA1–ETA6), and individual post-hoc pharmacokinetic parameter estimates (V1, V2, V3, CL, Q2, Q3).

To assess whether residual covariate effects remained after model estimation, the six ETA values (ETA1 through ETA6, corresponding to V1, V2, V3, CL, Q2, and Q3, respectively) were selected as Y-axis variables, and the continuous demographic covariates AGE and BMI were selected as X-axis variables. **Fig. 6** presents the resulting 12-panel scatter plot matrix with LOESS regression lines and 95% confidence bands. The LOESS trends are flat across all covariate–ETA combinations, with confidence bands encompassing zero, consistent with the allometric and covariate structure having accounted for the influence of age and BMI. This is consistent with the model's design intent as a general-purpose model spanning pediatric to elderly populations.

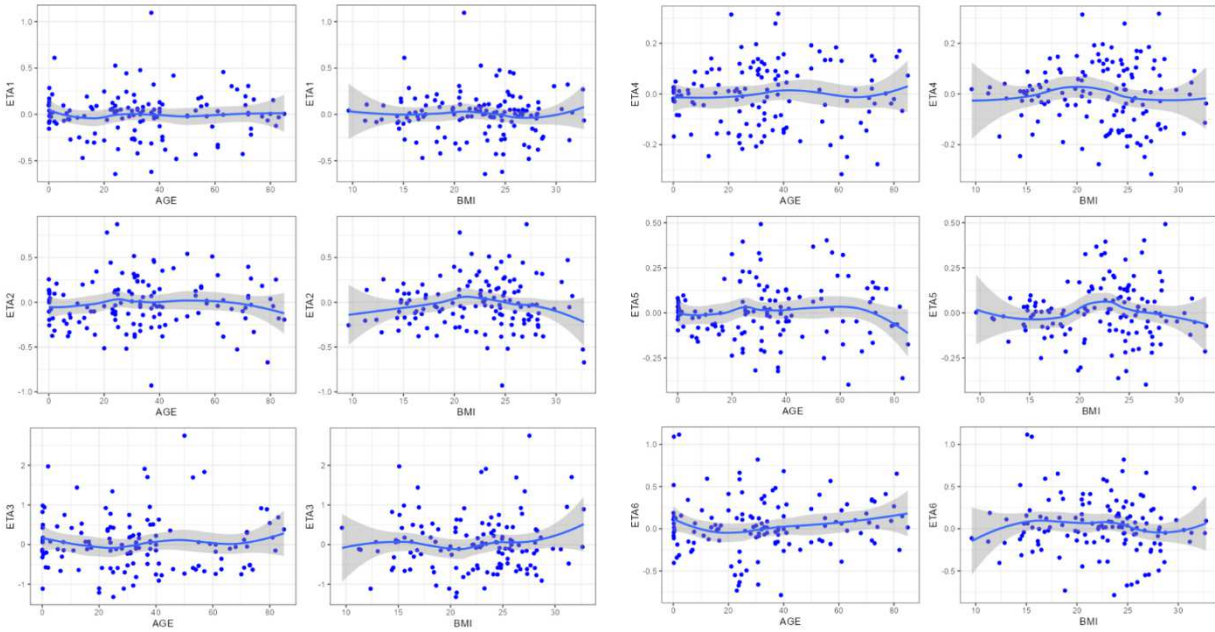


Fig. 6 Correlation plots of empirical Bayes estimates of inter-individual variability (ETA1–ETA6, corresponding to V1, V2, V3, CL, Q2, and Q3) versus continuous demographic covariates (AGE, BMI) for the remifentanyl population pharmacokinetic model. Scatter plots are shown with LOESS regression lines and 95% confidence bands

To further explore categorical covariate effects, box plots of ETA1 through ETA6 were generated by sex (M1F2) and by study (STDY). **Fig. 7** shows that the ETA distributions are centered near zero for both sexes and across all three component studies, with no apparent systematic differences by sex or study.

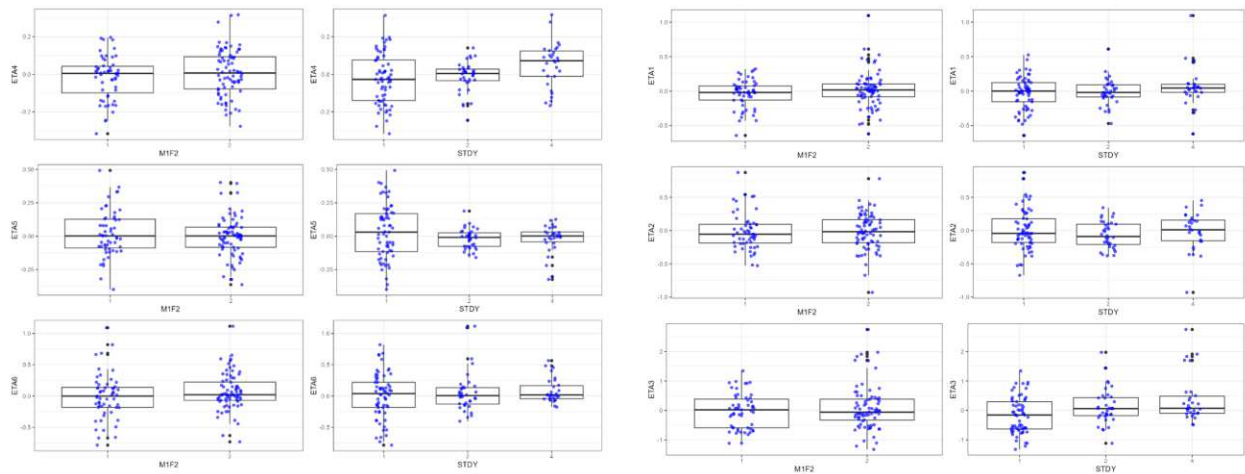


Fig. 7 Box plots of empirical Bayes estimates of inter-individual variability (ETA1–ETA6) grouped by sex (M1F2) and study (STDY) for the remifentanyl population pharmacokinetic model. Individual observations are overlaid as points

This case study illustrates how the Correlation Plots tab in PopPK FitCheckR facilitates efficient multi-dimensional covariate screening. The ability to simultaneously visualize all pairwise combinations of ETAs and covariates in a single panel, with the option to toggle between scatter plots and box plots, enables rapid identification of residual covariate effects that may warrant further model refinement.

Discussion

The PopPK FitCheckR application addresses a practical gap in the population pharmacokinetic modeling workflow: the absence of a lightweight, interactive tool for generating standard goodness-of-fit diagnostics directly from NONMEM output tables. While established tools such as PsN and Xpose remain essential for advanced simulation-based diagnostics and flexible exploratory visualization, their application to routine GOF assessment often requires familiarity with R programming, project-specific scripting, and manual configuration [3, 4]. PopPK FitCheckR was designed to complement these tools by providing a browser-based interface that may reduce the amount of coding typically required to produce review-ready diagnostic plots from NONMEM output.

One practical advantage of the application is its ease of use. By accepting NONMEM output table files directly and employing automated column-name matching via Jaro-Winkler string similarity [9, 10], the tool minimizes the need for users to manually parse, reformat, or pre-process model outputs prior to visualization. This is particularly useful in environments where pharmacometricians with varying levels of programming experience collaborate on shared models, or where rapid turnaround of diagnostic outputs is needed during iterative model development.

The interactive plotting functionality extends the utility of the tool beyond static GOF figures. Users can stratify, filter, and color-code plots by any variable present in the output table, facilitating exploratory diagnostics such as examining residual behavior across dose groups, study populations, or covariate strata. The integration of logarithmic axis transformations and faceted layouts within

the same interface further reduces the number of workflow steps typically involved in evaluating model adequacy across subgroups.

The correlation plots module provides additional value for random effects diagnostics. By allowing users to select arbitrary pairwise combinations of ETA parameters and covariates, the application supports visual screening of potential covariate–parameter relationships, a standard step in covariate model building [18]. The availability of both linear regression and locally estimated scatterplot smoothing (LOESS) [11] overlays enables users to assess both linear and nonlinear trends without configuring separate analyses.

From a workflow integration perspective, PopPK FitCheckR promotes consistency. Because the same set of diagnostic plots is generated through a standardized interface, outputs are visually uniform across models, analysts, and projects. This consistency is valuable in regulatory submissions and internal review processes, where reviewers benefit from predictable figure formats and labeling conventions [19]. The downloadable outputs can be readily incorporated into technical reports, internal review documents, or presentation materials.

Several limitations should be acknowledged. The current implementation expects input files generated with the ONEHEADER option in the NONMEM \$TABLE record (though this is a common practice among modelers). The application is designed for standard GOF diagnostics and pairwise correlation plots and does not support simulation-based methods such as VPCs, including prediction-corrected VPCs [20]. Additionally, the current version does not support automated batch processing of multiple model output files simultaneously, which may limit efficiency in large-scale model comparison exercises. Users with highly customized plotting requirements may still prefer scripted workflows in R or Xpose for full control over aesthetic and layout choices. Compared to Certara.ModelResults [6], which provides R code generation and report rendering from its interface, PopPK FitCheckR does not currently generate underlying R scripts or formatted reports. These capabilities may be important for users who require end-to-end traceability from diagnostic exploration to final reporting within a single tool.

Future development may include support for additional diagnostic plot types (e.g., individual concentration–time profiles, quantile–quantile plots for residuals), visualization of simulation-based diagnostics such as VPCs from simulation output generated externally by tools such as PsN, batch mode operation for multi-model comparisons, and integration with cloud-based NONMEM

execution environments. Extension to pharmacodynamic and exposure–response models is also under consideration.

Conclusions

In conclusion, PopPK FitCheckR provides a practical, accessible, and standardized solution for routine goodness-of-fit assessment of population pharmacokinetic models developed in NONMEM. By reducing the programming effort and time typically required to generate diagnostic plots, PopPK FitCheckR allows pharmacometricians to focus more on model interpretation and less on plot generation, supports more efficient model development cycles, and promotes consistent graphical outputs that can be shared across teams and projects.

Application access and data handling

PopPK FitCheckR is hosted on shinyapps.io and can be accessed via <https://ruilian-zhang.shinyapps.io/PopPK-FitCheckR/>. The app code is published at https://github.com/MSDLLCpapers/PopPK_FitCheckR under the terms of GNU General Public License Version 3 [21].

Uploaded files are held in a temporary, session-scoped location managed by Shiny for the duration of the session, read into memory for analysis, and removed when the session ends. No data are written to a database or other persistent application storage, or retained between sessions.

The application hosted on shinyapps.io is intended for demonstration and evaluation with de-identified or non-confidential data. For confidential, proprietary, or subject-level clinical data, users should download the source code provided in the GitHub repository above and run the application locally or deploy it within their own environment, in which case no data leave the user's network.

Data availability

No datasets were generated or analysed during the application development.

Acknowledgements

The authors would like to acknowledge Brian Maas and Ziheng Hu for the guidance on case study design, Dan Li and Lihong Du for providing the NONMEM output datasets used in case study 1,

Franco Szeto, Maja Skataric, Oluwaseun Francis Egbelowo, and Ziheng Hu for reviewing the manuscript, and colleagues in the Quantitative Pharmacology and Pharmacometrics group at Merck & Co., Inc. who provided feedback that guided the development of the application's features.

References

1. Nguyen THT, Mouksassi MS, Holford N, Al-Huniti N, Freedman I, Hooker AC et al (2017) Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol* 6:87–109.
<https://doi.org/10.1002/psp4.12161>
2. Beal SL, Sheiner LB, Boeckmann AJ, Bauer RJ (2024) NONMEM 7.x users guides. ICON plc, Gaithersburg
3. Lindbom L, Pihlgren P, Jonsson EN (2005) PsN-Toolkit—a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed* 79:241–257.
<https://doi.org/10.1016/j.cmpb.2005.04.005>
4. Jonsson EN, Karlsson MO (1999) Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* 58:51–64. [https://doi.org/10.1016/S0169-2607\(98\)00067-4](https://doi.org/10.1016/S0169-2607(98)00067-4)
5. Gassem A, Bieth B, Baltcheva I, Dumortier T, Bartels C, Bhattacharya S et al (2025) ggPMX: 'ggplot2' based tool to facilitate diagnostic plots for NLME models. R package version 1.3.2. <https://doi.org/10.32614/CRAN.package.ggPMX>
6. Craig J, Talley M (2025) Certara.ModelResults: generate diagnostics for pharmacometric models using 'shiny'. R package version 3.0.1.
<https://doi.org/10.32614/CRAN.package.Certara.ModelResults>
7. Lixoft (2023) Monolix version 2023R1. Lixoft SAS, a Simulations Plus company, Antony. <https://lixoft.com/products/monolix/>. Accessed 15 September 2026

- 476 8. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2024) shiny: web application
477 framework for R. R package version 1.8.1.1.
478 <https://doi.org/10.32614/CRAN.package.shiny>
- 479 9. Winkler WE (1990) String comparator metrics and enhanced decision rules in the Fellegi-
480 Sunter model of record linkage. In: Proceedings of the section on survey research
481 methods. American Statistical Association, Alexandria, pp 354–359
- 482 10. Cohen W, Ravikumar P, Fienberg S (2003) A comparison of string distance metrics for
483 name-matching tasks. In: Proceedings of the IJCAI-2003 workshop on information
484 integration on the web, Acapulco, pp 73–78
- 485 11. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J
486 Am Stat Assoc 74:829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- 487 12. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M et al (2024) plotly:
488 create interactive web graphics via 'plotly.js'. R package version 4.10.4.
489 <https://doi.org/10.32614/CRAN.package.plotly>
- 490 13. Aphalo PJ (2026) ggpmisc: miscellaneous extensions to 'ggplot2'.
491 <https://doi.org/10.32614/CRAN.package.ggpmisc>. Accessed 15 September 2026
- 492 14. Eleveld DJ, Proost JH, Vereecke H, Absalom AR, Olofsen E, Vuyk J et al (2017) An
493 allometric model of remifentanyl pharmacokinetics and pharmacodynamics.
494 Anesthesiology 126:1005–1018. <https://doi.org/10.1097/ALN.0000000000001634>
- 495 15. Minto CF, Schnider TW, Egan TD, Youngs E, Lemmens HJM, Gambus PL et al (1997)
496 Influence of age and gender on the pharmacokinetics and pharmacodynamics of
497 remifentanyl. I. Model development. Anesthesiology 86:10–23.
498 <https://doi.org/10.1097/00000542-199701000-00004>
- 499 16. Ross AK, Davis PJ, Dear GL, Ginsberg B, McGowan FX, Stiller RD et al (2001)
500 Pharmacokinetics of remifentanyl in anesthetized pediatric patients undergoing elective
501 surgery or diagnostic procedures. Anesth Analg 93:1393–1401.
502 <https://doi.org/10.1097/00000539-200112000-00008>

- 503 17. Mertens MJ, Engbers FHM, Burm AGL, Vuyk J (2003) Predictive performance of
504 computer-controlled infusion of remifentanyl during propofol/remifentanyl anaesthesia. *Br*
505 *J Anaesth* 90:132–141. <https://doi.org/10.1093/bja/aeg046>
- 506 18. Gastonguay MR (2004) A full model estimation approach for covariate effects: inference
507 based on clinical importance and estimation precision. *AAPS J* 6(S1):W4354
- 508 19. U.S. Food and Drug Administration (2022) Population pharmacokinetics: guidance for
509 industry. FDA, Silver Spring. [https://www.fda.gov/regulatory-information/search-fda-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/population-pharmacokinetics)
510 [guidance-documents/population-pharmacokinetics](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/population-pharmacokinetics). Accessed 15 September 2026
- 511 20. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO (2011) Prediction-corrected visual
512 predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J* 13:143–151.
513 <https://doi.org/10.1208/s12248-011-9255-z>
- 514 21. Free Software Foundation (2007) GNU general public license version 3.
515 <https://www.gnu.org/licenses/gpl-3.0.en.html>. Accessed 15 September 2026

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S](#)
- [S](#)
- [S](#)